ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

FACULTY OF COMPUTER AND MATHEMATICAL SCIENCES

DEPARTMENT OF COMPUTER SCIENCE

CHARACTERIZATION OF AMHARIC VOWEL SOUND UNITS

BY: Tessfu Geteye Fantaye

A Thesis Submitted to the School of Graduate Studies of Addis Ababa University in Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer Science

November, 2011

Addis Ababa University

School of Graduate Studies

Faculty of Computer and Mathematical Sciences

Department of Computer Science

CHARACTERIZATION OF AMHARIC VOWEL SOUND UNITS

BY

TESSFU GETEYE

APPROVED BY

EXAMINING BOARD:

1. Sebsibe H/Mariam, Advisor (PhD) ____________________________
2. ____________________________ ____________________________
3. ____________________________ ____________________________
4. ____________________________ ____________________________

Acknowledgements

First of all, I am truly thankful to my greatest father GOD who has blessed my work from the beginning up to the ending. He supports me for every success in my life.

I would like to take this opportunity to thank all the people who have helped me to complete this thesis work. I would like to thank sincerely my advisor Dr. Sebsibe H/Mariam for his invaluable guidance throughout this study. He will always be remembered by me as a person with a professional attitude to research. I would like to thank all my colleagues at the Department of Computer Science for their support during information and ideas exchanging, data collection, testing, preparation of the document and for commenting and proofreading the document.

I would like to express my heartily thanks to Beza Tamiru for all her appreciation, supporting and help while handling this thesis work. I am always praying for her heavenly and earthly blessing. I am always with you in every situation throughout my life.

I would like to express my special thanks to my families, who had put a lot to make this work possible, father, mother, my only brother and sisters for their continues support and encouragements throughout my study.

Dedication

To all people who have played a role in shaping me into the person I am today.

TABLE OF CONTENTS

LIST OF TABLES v

LIST OF FIGURES vi

LIST OF APPENDICES vii

LIST OF ACRONYMS viii

ABSTRACT ix

Chapter One 1

INTRODUCTION 1

1.1 Background 1

1.2 Statement of the problem 6

1.3 Motivation 6

1.4 Objectives 7

1.4.1 General objective 7

1.4.2 Specific objectives 7

1.5 Scope 7

1.6 Methodology 7

1.6.1 Data collection 8

1.6.2 Feature Extraction 8

1.6.3 Tools 8

1.6.4 Testing and Analysis 9

1.7 Application of results 9

1.8 Organization of the thesis 10

Chapter Two 11

LITERATURE REVIEW 11

2.1 Introduction 11

2.2 Human Speech Production 11
2.3 Speech Unit Characterization Techniques 17
2.3.1 Formant based Characterization 17
2.3.2 Duration based Characterization 20
2.3.3 Fractal Dimension (FD) 23
2.4 Acoustic-Phonetic Labeling (Segmentation) 26
Chapter Three 29
RELATED WORKS 29
3.1 Characterization of vowel speech units for English Language 29
3.1.1 Control method used in a study of the vowels 29
3.1.2 Acoustic characteristics of American English vowels 30
3.2 Characterization of vowel speech units for Bangla Language 32
3.2.1 Acoustic classification of Bangla vowels 32
3.2.2 Acoustic space of Bangla vowels 33
3.2.3 Bangla vowel characterization based on Analysis by Synthesis 34
3.3 Characterization of vowel speech units for Amharic Language 35
Chapter Four 36
VOWELS 36
4.1 Introduction 36
4.2 Vowel Articulation 36
4.2.1 Vowel Height 38
4.2.2 Vowel Backness 38
4.2.3 Vowel Roundedness 39
4.3 Characteristics of Vowels 40
4.3.1 Vowel Formants 40

4.3.2 Vowel length 41

4.3.3 Vowel duration 42

4.4 Types of Vowels 43

Chapter Five 44

THE AMHARIC LANGUAGE PHONOLOGY 44

5.1 Introduction 44

5.2 Phonology of the Amharic words 47

5.2.1 Vowels 47

5.2.2 Consonants 49

5.2.3 Amharic word Transcription 52

5.3 Phonological rules of Amharic Phonemes 52

5.3.1 Vowel sequences 52

5.3.2 Gemination 54

5.3.3 Consonant clustering 56

5.3.4 Stress 59

5.3.5 Phoneme length 59

Chapter Six 60

CONTEXT SPECIFICATION AND ARCHITECTURE OF AMHARIC VOWEL CHARACTERIZATION 60

6.1 Introduction 60

6.2 Context specification for vowel characterization 60

6.2.1 Corpus preparation 60

6.2.1.1 Text corpus preparation 60

6.2.1.2 Speech corpus preparation 64

6.2.2 Listening test 65

6.3 Architecture of Amharic vowel sound units characterization 66

6.3.1 Labeling / Segmentation 68
6.3.2 Feature extraction/ Acoustic Measurement 68
6.3.3 Aggregate feature extractors 71
6.3.4 Duration and fundamental frequency extractor 72
6.3.5 Acoustic vowel spaces Analysis 72
6.3.6 Duration effect on formant frequency values analyzer 73
6.3.7 Formant frequency variability analyzer 73
Chapter Seven 75
EXPERIMENTATION AND DISCUSSION 75
7.1 Introduction 75
7.2 Acoustic measurement result and analysis 75
7.2.1 Vowel duration and fundamental frequency 76
7.2.2 Formant Frequency 77
7.2.2.1 Formant Frequency Ranges 82
7.2.2.2 Effect of Duration on Formant frequency values 84
7.2.2.3 Acoustic Vowel Space 86
7.2.2.4 Formant frequency variability 87
Chapter Eight 92
CONCLUSION AND FUTURE WORKS 92
9. REFERENCES 95
APPENDIXES 101

LIST OF TABLES

Table 4.1: Japanese vowels (mainly adopted from [48]) *42*

Table 5.1: Amharic alphabets with their seven orders *46*

Table 5.2: Categories of Amharic consonants (adopted from [54]) *51*

Table 6.1: The possible combination of contexts to make CVC syllable structure data *63*

Table 6.2: The collected text corpus for the research *64*

Table 7.1: The aggregate features of the seven Amharic vowels *75*

Table 7.2: The average duration of the seven Amharic vowels at normal and fast speaking rates *76*

Table 7.3: The average fundamental frequency values of the seven Amharic vowels..................... *77*

Table 7.4: Aggregate features of the seven Amharic vowels for each context at normal rate extracted from 77 contexts... *79*

Table 7.5: F1 values of vowel /ኤ/ [ie] in both short and long duration... *85*

Table 7.6: F2 values of vowel /ኤ/ [ie] in both short and long duration... *85*

Table 7.7: F3 values of vowel /ኤ/ [ie] in both short and long duration... *86*

Table 7.8: The number of utterances and their percentage value which makes formant frequency variation among similar CVC context utterances of vowel /ኢ/ [ii]... *88*

Table 7.9: The number of instances and their percentage values for the seven Amharic vowels extracted from 11 contexts.... *89*

Table 7.10: Contexts which varies the formant frequency values of the seven Amharic vowels..... *90*

LIST OF FIGURES

Figure 2.1: Epiglottis and Vocal folds structure in larynx (mainly adopted from [13]) 12

Figure 2.2: General flow chart of source filter speech production system 14

Figure 2.3: Block diagram of the human speech production system [15] 14

Figure 2.4: Block diagram of a basic voicing source model [15] 16

Figure 5.1: The seven Amharic vowels with their features (mainly adopted from [54]) 48

Figure 5.2: The word /በራ/- [b e r r a] contain geminated consonant (/ር/ [r]) 55

Figure 5.3: The word /በራ/- [b e r a] contain nongeminated consonant (/ር/ [r]) 56

Figure 5.4: Medial consonant clusters of a word /ድንግል/ [d ix n g ix l] 57

Figure 5.5: Final consonant clusters of a word /ሊለብስ/ [l ii l e b s] 58

Figure 5.6: Final consonant clusters of a word /መንግስት/ [m e n g ix s t] 58

Figure 6.1: General Architecture of Amharic vowel sound units characterization 67

Figure 6.2: Duration measurement of vowel /ኤ/ [ie] in /ጼፕ/ [xx ie p] context 69

Figure 6.3: Steady state time measurement of vowel /ኤ/ [ie] in /ጼፕ/ [xx ie p] context 69

Figure 6.4: Formant frequency extraction of vowel /ኢ/ [ii] in /ጢዝ/ [tx ii z] context using speech analyzer tool 70

Figure 7.1: The average duration of the seven Amharic vowels 76

Figure 7.2: F1 vs. F2 space for 770 individual data points or utterances 78

Figure 7.3: F1 vs. F2 space of the seven vowels extracted from 77 utterances 82

Figure 7.4: Formant frequency ranges of the seven Amharic vowels 83

Figure 7.5: Acoustic vowel space of the seven Amharic vowels 87

LIST OF APPENDICES

Appendix A: Amharic Phonetic List, IPA Equivalence and its ASCII Transliteration ---------- 101

Appendix B: Contexts used in the thesis with their target vowel and transcription-------------- 102

LIST OF ACRONYMS

AMF	Average Magnitude Function
ASCII	American Standard Code for Information Interchange
CV	Consonant - Vowel
CVC	Consonant - Vowel - Consonant
DFT	Discrete Fourier Transform
EEG	Electroencephalogram
FD	Fractal Dimension
FFT	Fast Fourier Transform
F0	Fundamental Frequency (pitch)
F1	First Formant Frequency
F2	Second Formant Frequency
F3	Third Formant Frequency
F4	Fourth Formant Frequency
IFFT	Inverse Fast Fourier Transform
IPA	International Phonetic Alphabet
KHZ	Kilo Hertz
LP	Linear Prediction
LPC	Linear Predicative Coding
NLP	Natural Language Processing
POA	Place of Articulation
TTS	Text to Speech
VC	Vowel - Consonant
Vs.	Verses

ABSTRACT

Vowel sound unit characterization is a process of studying the acoustic characteristics or behaviors of vowel speech units within different contexts. It is used to describe the phonetic and acoustic features of the respective speech units and avoid inappropriate parameter selection in concatenative and formant based Text to Speech (TTS) systems.

This study focused on the acoustic characteristics of Amharic vowel sound units. The characteristics of vowels are studied in Consonant-Vowel-Consonant (CVC) contexts using formant and duration based characterization approaches. In the study, context specification and text corpus preparation was done and listening test was conducted on the recorded speech signals. Then the speech signal is labeled to extract the vowel segment features such as first formant frequency (F1), second formant frequency (F2), third formant frequency (F3), fundamental frequency (F0) and duration. Finally, aggregate feature extractors were conducted on the sample vowel features in order to acquire the acoustic characteristics of the vowel sound units.

At run time, the effect of speaking rate on duration, the influence of duration on formant frequency values and formant frequency variability occurred due to loudness/ intensity and adjacent contexts were described.

The characterization shows that the duration of vowels is affected by the speaking rates. This implies that vowel duration is short within contexts that are recorded at fast speaking rate while the vowel duration is long within contexts that are recorded at normal speaking rate. Similarly, the formant frequency values of vowels are affected by the duration. So, F1 is inversely proportional to duration while F2 and F3 are directly proportional to duration.

The characterization was evaluated by visualizing acoustic vowel space and through comparing and contrasting with International Phonetic Alphabet (IPA) maps of vowels. The evaluation shows that acoustic vowel space is dependent on duration, context and loudness/ intensity. However, these factors have not a big influence on the acoustic space rather they make a little shift of vowel position on the space.

Keywords: Vowel, **Vowel Characterization**, **Formant**, **Acoustic vowel space**, **Articulatory vowel space**

Chapter One

INTRODUCTION

1.1 Background

Natural Language Processing (NLP) is a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications [1]. The foundation of NLP lies in a number of disciplines, like computer and information sciences, linguistics, mathematics, electrical and electronic engineering, artificial intelligence and robotics, psychology, etc. NLP deals with the interaction between computer and human (natural) languages [2].

The most explanatory method for presenting what actually happens within NLP system is by looking at the different levels of linguistic analysis. There are seven levels of linguistic analysis. These are: phonology, morphology, lexical, syntactic, semantic, discourse and pragmatic. In order to understand natural languages, it is important to be able to distinguish among the seven interdependent levels that people use to extract meaning from text or spoken languages [3, 4].

The first linguistic analysis level is phonetic or phonological level that deals with pronunciation, which deals with the interpretation of speech sound units within and across words. The second level, morphological level focuses on the smallest meaning bearing of a word appeared as affix of a root. The root and the affixes are collectively called morphemes. Since the meaning of each morpheme remains the same across words, humans can break down an unknown word into its constituent morphemes in order to understand its meaning. The lexical level deals with lexical meaning of words and parts of speech analysis. To do lexical interpretation of the words, lexical analysis needs the lexicon. The fourth level, syntactic level, deals with the structural relationships between words in a sentence. It requires both a grammar and a parser, the output of which is a representation of the sentence that reveals the structural dependency relationships between the words. This structural dependency can be represented using a parse tree. The semantic level focuses on the meaning of words, phrases and sentences. It requires knowledge of the meaning of the component words which is called lexical semantics and how components combine to form larger meanings which is known as compositional semantics. These five linguistic analyses are called lower levels [1].

The discourse level deals with the properties of the text as a whole that convey meaning by making connections between component sentences. It imposes meaning and structure on individual sentences (utterances) that go well beyond the compositional meaning of sentences in isolation. For instance, the meaning of a text is entirely different when we reorder the sentences and a sentence in isolation is frequently ambiguous, whereas a sentence in a naturalistic discourse context is rarely ambiguous. The last level in linguistic analysis is the pragmatic level. It deals with the knowledge that comes from the outside world, that means, from outside the contents of the document. It also is concerned with powerful use of language in situations and utilizes context over and above the context for understanding. These two levels are also called higher level of linguistic analysis [3, 4].

Phonological level of linguistic analysis deals with the interpretation of speech sound units within and across words. It is the systematic use of sound to encode meaning in any spoken human language, or the field of linguistics studying. In simple language, phonological analysis is concerned with the function, behavior and organization of speech sound units as linguistic items. It is also important in order to distinguish phone and phonemes in the language. In phonological analysis, there are three types of rules: phonetic, phonemic and prosodic.

The first phonological analysis rule is phonetic rule used to describe about speech sound units within words; how these speech sound units are produced; and the corresponding acoustic result of the speech articulation and generally deal with phones in words. The phonetic nature of phones is affected by speaking rate and loudness (intensity) of the speakers and contexts in which phones are located. The second phonological analysis rule is phonemic rules. They are applied for variations of pronunciation when words are spoken together. This variation comes due to the speech sound units of the phonemes used in the words. The third phonological rule is prosodic rule. This focuses on fluctuation in stress and intonation across a sentence. This fluctuation comes due to the health, mood and emphasis problems of the speaker of the sentence [1].

In the phonetic rule of phonological analysis, there is a concept of phonetics which is the study of the actual speech sound units of the language. It is concerned with the physical properties of speech sound units or phones, their physiological production, acoustic properties, auditory perception, and neurophysiologic status. Languages contain different phonemes which are an ideal speech sound unit with a complete set of articulatory gestures and the basic theoretical unit for describing how speech conveys linguistic meaning [5]. The three branches of phonetics are articulatory phonetics, acoustic phonetics and auditory phonetics. Articulatory phonetics focuses on the manner in which the speech sound units are produced by the articulators of the vocal system. Acoustic phonetics deals with the sound units of speech through the analysis of the speech waveform and its spectrum. One can get the properties of sound units through the analysis of waveform and spectrum frequencies, amplitude and harmonic structure. Auditory phonetics studies about the perceptual response to speech sound units as reflected in listener trials like speech perception, speech categorization and recognition of speech sound units [5].

The acoustic phonetics is a part of phonetics which deals with acoustics aspects of speech sound units. It investigates properties like the mean squared amplitude of a waveform, its duration, its fundamental frequency, formant frequency or other properties of its frequency spectrum, and the relationship of these properties to other branches of phonetics like articulatory and auditory phonetics and to abstract linguistic concepts like phones, phrases, or utterances [6]. An acoustic signal is formed when the vocal organs move resulting in a pattern of disturbance to the air-molecules that is propagated outwards in all directions eventually reaching the ear of the listener when the speech is produced. Acoustic signal is the physical representation of the speech [6].

Speech is the vocalized form of human communication. It is based upon the syntactic combination of lexical and names that are drawn from very large (usually >10,000 different words) vocabularies. Each spoken word is created out of the phonetic combination of a limited set of vowel and consonant speech sound units. These vocabularies, the syntax which structures spoken words and the respective set of speech sound units, differ in creating the existence of many thousands of different types of mutually unintelligible human languages but the acoustic characteristics of speech sound units such as vowels and consonants is not elaborated for many languages [7].

Vowel speech sound unit characterization is a means of studying the detail characteristics or behaviors of the speech sound units. There are two characterization methods of vowel sound units. These are: articulatory characterization and acoustic characterization. Articulatory characterization of vowels has been done based on the position of the tongue during the vowel articulation. According to this characterization, vowels are grouped into front, middle and back due to tongue movement horizontally and high, middle and low due to tongue movement vertically. This characterization is sometimes called articulatory vowel space [6].

The second characterization ways of vowels is acoustic characterization. Acoustic characterization of vowel speech sound units is a part of acoustic phonetics rule in NLP that deals with the extraction of distinctive features of vowels of a language at phonological level of linguistic analysis. This characterization is conducted based on the source - filtered speech production system. In source - filtered system, the source and filters are two independent components. Source is the vibration of the vocal folds in the larynx whereas filter represents the resonances of the vocal tract which changes over time [5].

The technique used to achieve acoustic characterization of vowel sound units is formant based characterization approach. Formant based characterization can represent the tongue position by the formant frequency values of vowels. Formant frequency is the frequency produced or formed from the resonance of the vocal tract (filter) during the variation of the vocal tract shape or size. The first two formant frequencies are used to accomplish vowel characterization because F1 corresponds to the height of the vowels such as high, middle and low whereas F2 corresponds to the frontness of the vowels like front, middle and back [6].

Acoustic characterization of vowel speech sound units is important and relevant in order to understand the phonetic nature of the language; for instance, to develop acoustic vowel space and to solve the problems that exhibit in teaching and learning process of the language phonology, to conduct appropriate unit and parameter selection in different speech synthesis and recognition systems, to reduce the problem in man-machine interaction in the area of speech systems and to improve the quality of speech in speech synthesis system in terms of naturalness and intelligibility by providing important features of the speech units.

The reason that motivated to study acoustic characterization of vowel speech units is the advancement of speech technologies in the world. These speech technologies are becoming popular in avoiding the problems related to speech. But, still there is a performance problem on the speech systems due to lack of detail phonological linguistic analysis study. It is also difficult to select the best speech sound units or appropriate parameters in the development of speech systems like speech synthesis and speech recognition because phonemes speech units do not have uniform characteristic in the languages. So, to overcome these problems studying the acoustic characteristics of vowel speech units should be the first stage for different languages.

Amharic is the language for country-wide communication and was used for a long period as the principal language for literature and the medium of instruction in primary and junior schools of Ethiopia (while higher education is carried out in English). Amharic words use consonantal roots with vowel variation expressing difference in interpretation. In modern written Amharic, each syllable pattern comes in seven different forms (called orders), reflecting the seven vowels [8]. The first order is the basic form; the other orders are derived from it by more or less regular modifications indicating the different vowels. There are 33 basic forms, giving 231(7 * 33) Consonant-Vowel (CV) syllable patterns (syllographs), or "fidels". The thirty-three core orthographic symbols, each of which has seven different shapes to represent the seven vowels (e, u, ii, a, ie, ix, o). Each consonant and the seven vowels in combination represent CV syllables of the Amharic language [9].

Generally, the script of Amharic language is nearly phonetic in nature. So, it is important to deal with the phonology of vowels and consonants. This provides us a better insight to explore ideas that are directly related to phonology. There are limited numbers of researches done on Amharic language. So, acoustic characterization of vowel sound units for Amharic language is one of the research areas to be conducted in order to study the acoustic characteristics of Amharic vowel sound units in different contexts with the intent of improving various speech technology researches.

1.2 Statement of the problem

Acoustic characterization of vowels property is important to increase the quality of speech systems. Different researches have been conducted on acoustic characterization of vowel speech sound units on various languages such as English, Spanish, and Bangla in order to characterize the vowels of the language and achieve better quality of speech systems. The quality of Amharic speech systems such as text to speech systems, speech to text systems, and speech enhancement had not been reached the maximum achievable level. This is mainly due to lack of knowledge on specific acoustic characteristics of the vowel speech sound units of the Amharic language.

1.3 Motivation

Currently, technology is growing in a very fast rate in our world. Technological products are serving people in the day to day activities. Speech systems are technological products that enable computer and human being to dialog one with the other efficiently and effectively. These systems have enabled people to easily get information in different working areas, like in telecommunication and they are very important to humans who have some kind of disabilities such as visual or auditory in order to disseminate information from or to others.

To get the aforementioned benefits, there are many speech systems developed for Amharic language by following various techniques but the quality had not reached to the level that could be used to solve the real world problem. This is primarily due to the absence of acoustic characterization of the phonemes of the language. From phonemes of Amharic language, vowels are embedded in each consonant and have greater effect on the speech systems due to variation in speech units in different conditions and contexts. So, it is important and relevant issue to deal with the acoustic characteristics of vowel phonemes of the language.

1.4 Objectives

1.4.1 General objective

The main objective of the research is to obtain acoustic characterization of Amharic vowel sound units.

1.4.2 Specific objectives

The specific objectives of the research work are:

- To characterize Amharic vowels based on their contextual and spectral properties
- To determine context selection used for acoustic characterization of Amharic vowels
- To build text and speech corpora used for acoustic characterization of Amharic vowels
- To build the general steps and procedures used for acoustic characterization of Amharic vowels
- To conduct detail analysis about acoustic characteristics of Amharic vowels

1.5 Scope

The main target of this research work is to characterize Amharic vowel sound units. Vowels are chosen as they are most frequently occurring sound units in the language and addressing characterization of vowels will contribute a lot to the improvement of the quality of speech systems. In this study, the scope is limited to characterize all the seven Amharic vowels within some or reduced contexts. The contexts considered during this characterization are stops, fricatives and liquids. From these contexts, the contexts which have less coarticulation effect on vowels and less segmentation challenges are selected for this research.

1.6 Methodology

The techniques or methods that we follow to achieve the objectives of the research are formant based characterization [10] and duration based characterization at different vowel contexts. These methods are used to characterize the vowel sound units by considering and extracting basic parameters or features of the speech units from the waveform and spectrogram description during speech production.

The set of steps that we follow in this thesis to accomplish its objectives are:-

1.6.1 Data collection

In data collection, the contexts are chosen by considering and assessing the coarticulation effect on the vowels and the level of segmentation challenge as well as the distribution of consonants with their manner and place of articulation.

Hence, based on the assessment, the contexts which have less coarticulation influence on the vowels and less segmentation challenge when segmenting the vowels are: stops (/ት/ [t], /ግ/ [g], /ጥ/ [tx], /ፕ/ [p] and /ጵ/ [px]), fricatives (/ፍ/ [f], /ስ/ [s], /ዝ/ [z] and /ጽ/ [xx]) and liquids (/ር/ [r]) constricted in labial, velars and alveolar place of articulations.

Therefore, the text corpus contains all the seven vowels and the possible CVC combination of the above contexts such as [t] v [z], [t] v [r], [f] v [g], [f] v [r], [p] v [xx], [tx] v [z], [tx] v [f], [s] v [p], [xx] v [p] and [xx] v [g], where v stands for the vowel that represents all Amharic vowel sound units.

Once the text corpus preparation is finished, recording is conducted to get the corresponding speech corpus. Speech corpus contains all the speech files of vowels and CVC syllable structure contexts.

1.6.2 Feature Extraction

Once the labeled speech corpus is made ready from the vowels and CVC syllable structure acoustic data, we have extracted the acoustic parameters from the wave file. Praat and speech analyzer, speech analysis tools are used to extract formant frequencies and other relevant information from the speech files. Formant frequencies are the most important features to characterize the vowel sound units. In speech analyzer we have adopted, peak picking formant extraction mechanism, together with the first order LPC (Linear Predictive Coding) tracking filters.

1.6.3 Tools

During feature extraction, we used praat and speech analyzer, speech analysis tools. Praat is very important tool in three phases. Segmentation of the vowels from the given CVC syllable structure contexts; after segmentation, it is used to extract duration and steady state time features of vowels and it is also used to draw the acoustic vowel spaces whereas speech analyzer is used to extract the formant frequency and other acoustic information from the speech files. The formants are the most important features to characterize the vowel sound units. Speech analyzer is also used during recording the speech units from the speakers. On the other hand, Microsoft Office Excel 2007 is used to conduct the aggregate feature extractions from individual sample vowel features in order to acquire the acoustic characteristics of Amharic vowel sound units.

1.6.4 Testing and Analysis

Perceptual or listening test has been used to distinguish the common contexts intended by a speaker and perceived through the listeners. The descriptive statistics such as average and standard deviation are used to obtain appropriate acoustic characteristics of vowels.

1.7 Application of results

There are many speech systems such as speech recognition and speech synthesis. These speech systems have various application areas but their performance is affected by different conditions. In this research, we focus on characterization of Amharic vowel sound units.

Some of the applications of our research are:

- To improve quality of researches on speech recognition and synthesis
- To study the nature of Amharic phonemes
- To understand the phonological nature of the language to researchers who perform research work on Amharic speech systems.

1.8 Organization of the thesis

The thesis is organized into Eight Chapters including the current Chapter.

The Second Chapter focuses on human speech production system, the different techniques or approaches used in characterization of speech sound units. Chapter Three focuses on related works which are done for other languages. Chapter Four focuses on vowel articulation, the general characteristics of vowels and types of vowels.

Chapter Five brings up a discussion on the Amharic language phonological structure. It mainly focuses on consonants, vowels, Amharic word transcription and the phonological rules of Amharic phonemes. The Sixth Chapter targets context specification for vowel characterization and the architecture of vowel sound unit characterization. It includes data or corpus preparation, listening test, labeling/segmentation of speech files, feature extraction and analysis of the extracted features.

Chapter Seven contains the discussions on experimentation and its results. It includes acoustic characteristics of vowels such as duration, fundamental frequency, the first three formant frequency ranges, and effect of duration on formant frequency values, acoustic vowel spaces and formant frequency variability. Chapter Eight concludes the research by putting conclusion and some future works to be done.

Chapter Two

LITERATURE REVIEW

2.1 Introduction

Speech characterization is a means of studying the detail characteristics of speech units. Phoneme speech unit characterization is a speech characterization that investigates all the phonemic, phonetic and phonological properties of phonemes in a language. Speech unit characterization is conducted based on the analysis of speech waveform and spectrogram.

In this chapter, we will have a brief look on literature review. The literature review section contains the explanation about human speech production process, the different speech unit characterization techniques and the acoustic- phonetic labeling or segmentation.

2.2 Human Speech Production

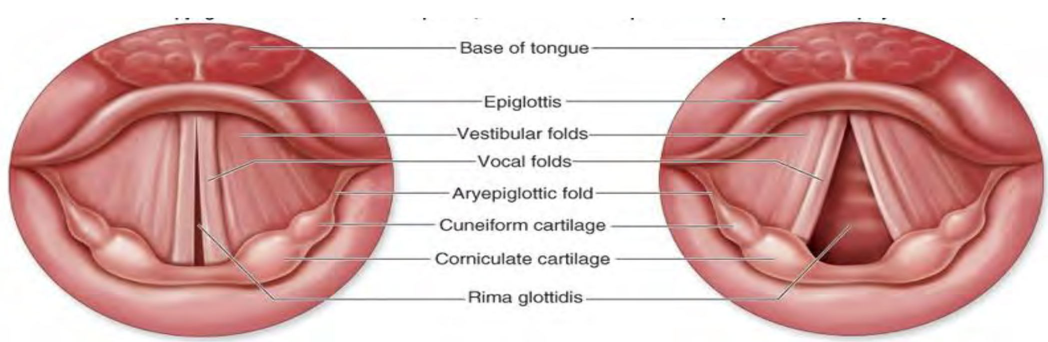
Speech is a natural form of communication for human beings. Computers with the ability to understand speech and speak with human beings are expected to contribute to the development of more natural man-machine interfaces [11]. Speech is produced by the speech production systems. Speech production is the process by which spoken words are selected to be produced, have their phonetics formulated and then finally is articulated by the motor system in the vocal apparatus. Human being uses speech in everyday life almost unconsciously, without understanding the mechanisms of its production. This discussion will help to clarify which brain section processes speech formation; vocal organs involved and the set of steps in speech production.

The mechanism of human speech production is composed of four processes: language processing, in which the content of an utterance is converted into phonemic symbol sequences in the brain‟s language center called cerebral cortex; generation of motor commands to the vocal organs in the brain‟s motor center; articulatory movement for the production of speech by the vocal organs based on these motor commands; and the speech generation is the emission of air from the lungs through the vocal tract to produce the desired speech [11].

Vocal organs are parts in the speech production system responsible for the generation of sound units. These include lungs, trachea(windpipe), larynx (voice box), pharyngeal cavity (throat), oral and nasal cavity, lips, tongs, teeth, hard and soft palates, velum, jaw, glottis and pharyngeal and oral cavities, and the nasal cavity. The region where air passes from the larynx to the tip of mouth is called vocal tract and the nasal cavity also referred as nasal tract.

Articulatory motion and speech generation processes involve four processes: initiation, phonation, oro-nasal and articulation [12]. The initiation phase is the moment when the air is expelled from the lungs based on the command that comes from the motor control. It is the base of all other articulatory movements and speech generation processes such as phonation, oro-nasal and articulation.

Phonation is a phase that occurs at the larynx. The larynx has two horizontal folds of tissue in the passage of air; they are called vocal folds shown in Figure 2.1. The gap between these folds is called glottis. The glottis can be nearly closed, as in Figure 2.1 (a). Then, nearly less amount of air can pass. The glottis can also have a narrow opening which can make the vocal folds vibrate producing the "voiced sounds". Finally, it can be wide open as in Figure 2.1 (b), as in normal breathing and, thus, the vibration of the vocal folds is reduced, producing the "voiceless sounds".



(a) (b)

Figure 2.1: Epiglottis and Vocal folds structure in larynx (mainly adopted from [13])

Voiced and unvoiced sound units produced by the vibration of the vocal folds at the phonation phase are not sufficiently loud to be heard as speech, nor can the various timbres of different vowel sound units be produced without the vocal tract, only a buzzing sound would be heard. The sound must be filtered and pass through the vocal tract. The vocal tract can vary the resonance value to adjust the speech quality based on the shape of vocal organs like mouth, lips and jaw. After filtering the sound by vocal tract, the desired speech will be differeciate and produced as oral and nasal speech unit. This phase is called oro-nasal phase.

The articulation phase takes place in the mouth and it is the process through which we can differentiate most speech sound units. In the mouth we can distinguish between the oral cavity, which acts as a resonator, and the articulators, which can be active or passive: upper and lower lips, upper and lower teeth, tongue (tip, blade, front, back) and roof of the mouth (alveolar ridge, palate and velum). So, speech sound units are distinguished from one another in terms of the place where and the manner how they are articulated.

Generally, after the motor control gives a command to the vocal organs, humans produce speech in the following manner:

1. Air pressure from the lungs creates a steady flow of air through the trachea; larynx and pharynx (back of the throat).
2. The vocal folds in the larynx vibrate, creating fluctuations in air pressure that are known as sound pulses or the vocal folds remain open producing turbulence.
3. Resonances in the vocal tract modify these waves according to the position and shape of the lips, jaw, tongue, soft palate, and other speech organs, creating formant regions and thus different qualities of sonorant or voiced sound.
4. Mouth and nose openings radiate the sound waves into the environment.

According to [10] speech production may be a nonlinear process that sparked great interest in the area of nonlinear analysis of speech giving rise to many studies. This research is based on the natural hypothesis that nonlinear processes occur in speech production, due to turbulent air flow produced in the vocal tract; nonlinear neuro muscular processes that should occur at the level of vocal folds and of the larynx.

2.2.1 Source - Filter Speech Production System

The main idea in source filter speech production system is the decomposition of the speech signal as a speech source passed through a linear time-varying filter. This filter can be derived from models of speech production based on the theory of acoustics that states the source represents the air flow at the vocal folds and the filter represents the resonances of the vocal tract, which change over time. The two independent components of source filter speech production model are source and filter. Figure 2.2 shows the interaction of input (source), process and output (speech). Input is the larynx buzzes come due to turbulence noise or vocal fold vibration and output will be unvoiced or voiced sound units.

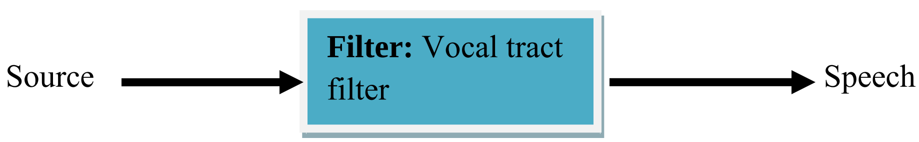


Figure 2.2: General flow chart of source filter speech production system

Such a source-filter model is further illustrated in Figure 2.3 [14]. The filter (i.e. a set of resonators) is excited by a source, which can be either a vocal fold vibration for voicing, or a noise occurs due to constriction somewhere in the vocal tract. The sound wave is created in the vocal tract, and then radiates through the lips [15].

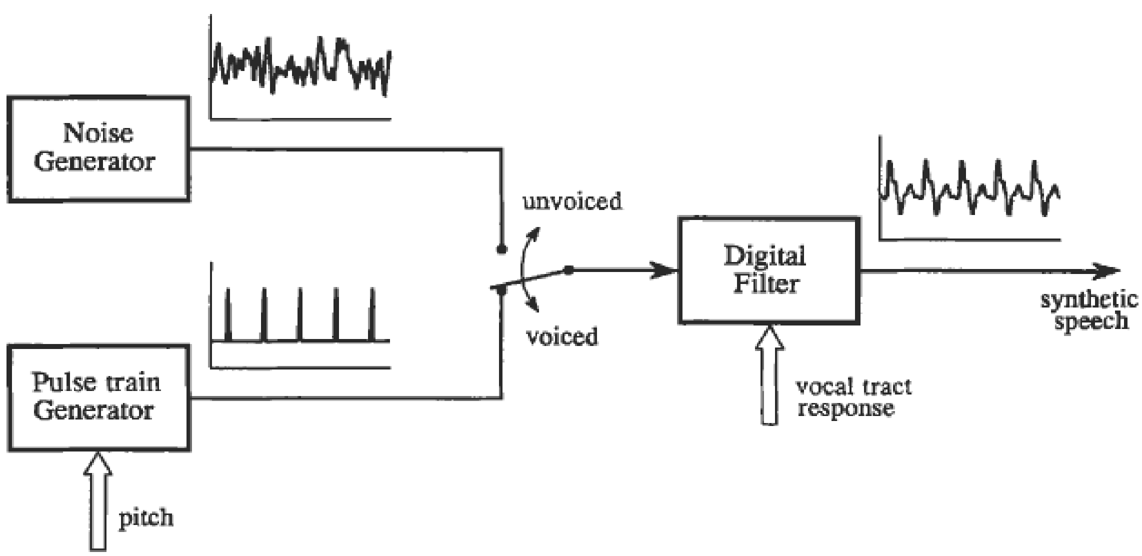


Figure 2.3: Block diagram of the human speech production system [15]

In this model, there is no interaction between the source and the filter other than the fact that the filter imposes its resonant characteristics on the source [15]. Hence, the individual acoustic properties of the source and the filter can be separately simulated. The vocal tract filter can be modeled as an acoustic tube with a varying cross-sectional area formed by the pharynx, the oral cavity, the nasal cavity and the lips.

In source-filter speech production system, there are two excitation sources that are needed for understanding the properties of speech units: a source producing a quasi-periodic wave (the voicing source) with the vocal fold vibration for voiced sound generation and a noise generator (the friction source) with turbulence noise for voiceless sound generation [15].

During the voiced sound unit generation, the vocal folds vibrate in a pseudo periodic pattern that produces a series of air pulses called glottal pulses. The rate at which the vocal folds vibrate is what determines the pitch of the sound produced or the speaker. This rate of cycling (opening and closing) of the vocal folds in the larynx during phonation of voiced sound units is called the fundamental frequency. The fundamental frequency also contributes more than any other single factor to the perception of pitch (the semi-musical rising and falling of voice tones) in speech. These air pulses that are created by the vibrations finally pass along the rest of the vocal tract where some frequencies resonate. It is generally known that women and children have higher pitched voices than men. It is therefore important to include the pitch period in the analysis and synthesis of speech if the final output is expected to accurately represent the original input signal. Vowels are an example of voiced sound units. Vowels often have high average energy levels and very distinct resonant or formant frequencies. In Figure 2.4, voiced sound units are represented by the pulse train generator, with the pitch (i.e., the fundamental frequency of the waveform) being an adjustable parameter.

Unvoiced or voiceless sound units generally had less energy and higher frequencies than voiced sounds. The production of unvoiced sound involves air being forced through the vocal tract in a turbulent flow. During this process, the vocal folds do not vibrate; instead, they stay open until the sound is produced. Pitch is not an important attribute of unvoiced speech since there is no vibration of the vocal folds and no glottal pulses.

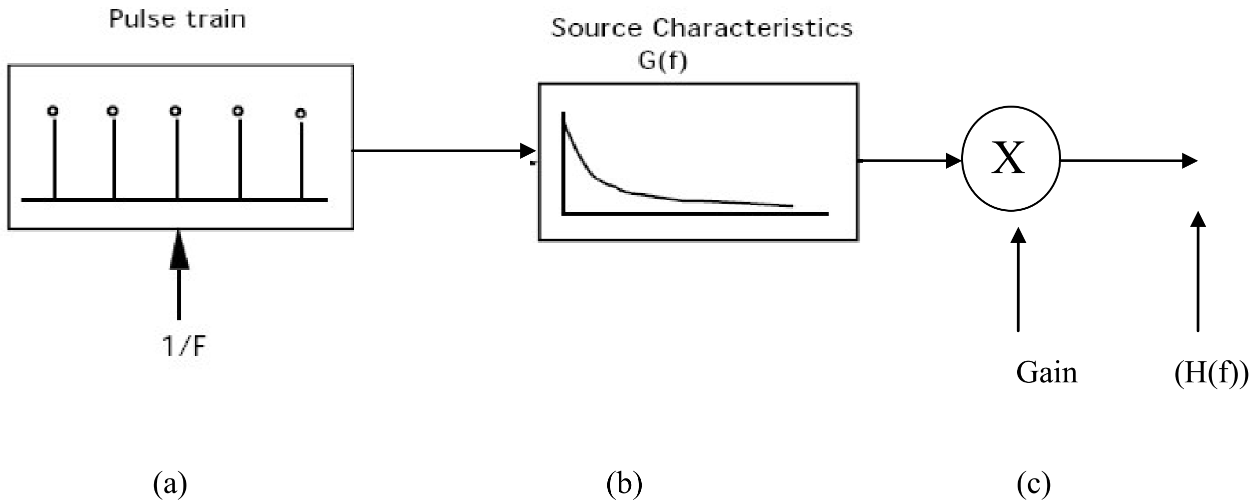


Figure 2.4: Block diagram of a basic voicing source model [15]

As shown in Figure 2.4, this model is composed of an impulse train generator that produces pulses at the rate of one per fundamental period. As indicated in Figure 2.4 (a), this impulse excitation simulates the generation of acoustic energy at the instant of the opening of the vocal folds. This signal impulse excitation has frequency spectrum function G(f) (Glottal function) that approximates the glottal wave form. A gain control device allows the adjustment of the voicing amplitude as the source pass through the linear filter whose frequency response is H (f).

In contrast, voiceless sound units originate as random noises. The vocal folds will be in non vibrating mode and are held open. This occurs when the air flow is nearly blocked by the tongue, lips, and/or teeth, resulting in air turbulence near the constriction. This phenomenon is due to a pressure drop across a constriction formed in the vocal tract, where the flow of air becomes turbulent [14, 15].

Since the glottal wave is periodic, consisting of fundamental frequency (F0) and a number of harmonics (integral multiples of F0), it can be analyzed as a sum of sine waves. The resonances of the vocal tract (above the glottis) are excited by the glottal energy. Usually, the vocal tract is assumed as a straight tube of uniform cross-sectional area, closed at the glottal end, open at the lips. Depending on the shape of the acoustic tube (mainly influenced by tongue position), a sound wave traveling through it will be reflected in a certain way so that interferences will generate resonances at certain frequencies. These resonances are called formants and the frequency of these peaks of energy is called formant frequency. By changing the relative position of the tongue and lips, the formant frequencies can be changed in both frequency and amplitude [14, 16].

As a conclusion, speech unit characterization determines the spectral properties of consonants and vowels for efficient synthesis as well as recognition of consonants and vowels. It depends on the speech production system to analyze isolated words which contain the required phonemes. This leads to the extraction of spectral parameters from the speech production system in order to characterize speech units. Due to this reason, we focus on the speech production system.

2.3 Speech Unit Characterization Techniques

Speech characterization is the means of studying the detail characteristics of speech units in the languages. There are different characterization techniques of speech units. The most common techniques are: **Formant** based**, Duration** based and **Fractal dimension** based characterization techniques.

2.3.1 Formant based Characterization

According to the acoustic theory of speech production [17], all speech sound units can be analyzed as a product of a "source" and a "filter." In vowels, the sound source is the vibration of the vocal folds and the filter is the system of resonators or cavities whose shapes depend on the positions of the various articulators. The source provides the raw material, which is modified by the vocal tract filter so as to produce the desired vowel quality. The vibrations at the glottis are responsible for the fundamental frequency (F0) which represents "the first harmonic and the distance between the following harmonics" [18] and the filter is responsible for the formation of acoustic "formants," or spectral energy peaks.

Gunnar Fant in [17] and Benade in [19] stated that formants are spectral peaks of the sound spectrum envelope in frequency domain representation of speech signal. Formants in the sound of the human voice are particularly important because they are essential components in the intelligibility of speech units. For example, the identification of the vowel speech units can be attributed to the differences in their first three formant frequencies. Producing different vowel speech amounts to retuning these formants within a general range of frequencies.

The process of articulation determines the frequencies of the vowel formants. In the vowel spectrum F1, F2, and F3 represent the first, second, and third well-known spectral maxima. By positioning articulators such as the tongue, the jaw, and the lips, the speaker controls the formant frequencies of each vowel. Sundberg in [20] has identified portions of the vocal anatomy which he associates with these formant frequencies. The jaw opening, which constricts the vocal tract toward the glottal end and expands it toward the lip end, is the deciding factor for the first formant. This formant frequency rises as the jaw is opened wider. The first formant (F1) is affected by the size of constriction, used as a cue for manner of articulation and unrelated to the place of articulation. The second formant is most sensitive to the shape of the body of the tongue and the forward movement of the tongue tends to increase the second formant frequency (F2), and the third formant is most sensitive to the tip of the tongue. However, both the second and third formants F2 and F3 are affected by the place of articulations. Additionally, other articulator that affects the formant is lip rounding that is sometimes invoked to lower formant frequencies [18]. Most often the first two formants, F_1 and F_2, are enough to identify the vowels. These two formants determine the quality of vowels in terms of the high /low and front/back dimensions (based on the position of the tongue). Thus the first formant (F_1) corresponds to the vowel height or vowel openness. It has a lower frequency for a high (closed) vowel and a higher frequency for a low (open) vowel; for instance, English vowels [i] and [u] have similar low first formants, whereas [a] has higher first formant. [a] is a low or open vowel, so its F1 value is higher than that of [i] and [u], which are high vowels. And the second formant (F_2) corresponds to vowel frontness. It has a higher frequency for a front vowel and a lower frequency for a back vowel. For example: front vowel [i] has a much higher F2 frequency than [u] and [a] vowels. But F1 and F2 representation of vowel space is true when the vowels are least influenced by its surrounding phonetic context and is considered to be either a point in the time course of the vowel or else a section of time during which the vowel position remains stable [21].

The conventional method of depicting the F1 and F2 does not adequately represent the multi-dimensional nature of vowel quality. Syed Skhter in [18] showed that the third formant influenced listeners judgments of vowel quality and more recent experiments have determined that the higher formants have a combined influence on vowel perception.

More recent studies have examined global spectral features suggesting that the [F3 - F2] difference is a more accurate way of identifying vowel frontedness than [F2], [F2-F0] and [F2-F1]. Syed, Lutfar and Farruk in [18] have shown that the separation between back and front vowels is more closely linked to the [F3 - F2] difference than the [F2 -F1] difference. It is important to recognize, however, that F3 and F4 vary more than F1 and F2 as a result of speaker characteristics whereas they are relatively stable across vowel categories in contrast to F1 and F2, which vary greatly as a result of vowel quality. The higher formants are therefore less effective carriers of phonetic information than the lower formants [18].

The factors affecting the measurement of vowels formant frequencies are: (1) linguistic factors such as dialectal and sociolectal differences and (2) non-linguistic factors such as physical anatomy, age, gender, and emotional state of the speaker. So when we measure the formant frequencies of vowels we must consider the above conditions.

Formant based characterization is easy to compute the formant parameters and it is also preferable method to characterize the acoustic vowel space. However, this approach has difficulty to identify the vowels in acoustic vowel space which overlap to each other.

2.3.1.1 Formant Extraction Techniques

Formant features can be interpreted as adaptive non-uniform samples of the signal spectrum that are located in the resonance frequencies of the vocal tract. The number and the position of these frequencies along the frequency axis might differ depending on the phonemes and the position of the window along the phoneme (i.e. beginning or ending part of a phoneme). Along with the formants (the resonance frequency), we might use the bandwidth and amplitude of the spectrum in that particular frequency to encode the properties of the speech and use them in different speech applications such as speech characterization, synthesis, recognition, enhancement, noise reduction and hearing aid adaptive filters.

Formants, the resonance of the vocal tract, are the key parameters in the characterization of the vowel speech units. The frequency at which they occur is said to be the formant frequency. These formants are extracted from the speech signal by using different methods. But, in most cases the vocal tract system H (z) is modeled using an all pole system shown in equation 2.1:

$$H(z) = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}}.$$ ----------------- Equation (2.1)

Where, G is the gain factor which is called the air volume and a_k are the Linear Prediction (LP) coefficients, z is the resonance frequency and p is the order of the linear prediction. In Linear Prediction, we assume a signal *y*[*n*] to be "predicted" from a linear combination of its past p values and the current input *x*[*n*] as:

$$y[n] = \sum_{k=1}^{p} a_k y[n-k] + Gx[n]$$ --------------Equation (2.2)

The Linear Predictors are used to separate speaker dependent information such as vocal-tract length, pitch, formant frequencies from speech. Formant frequency information is extracted by using different methods such as spectral peak picking method, root extraction method, analysis by synthesis method and formant tracking using context-dependent phonetic information [22, 23].

2.3.2 Duration based Characterization

Duration based characterization is the vowel sound unit characterization approach which is used to realize the phonemic vowel length with the physical measurement using speech analysis tools. In the sound systems of many languages, vowels can be characterized by their contrastive use of vowel quality and quantity. Vowel quality refers to the relative phonological resonance or timbre of a sound, whereas vowel quantity refers to the phonologically distinctive length of a vowel relative to one or more vowels of similar quality in the language. Vowel quantity is used in phonetics and poetics for the length of a vowel, it is usually indicated in phonetic transcription by a length mark colon [:] after a vowel, vowels so marked have in general greater duration than the same vowels with no such mark. Vowels so marked are described as long and unmarked vowels are short, a distinction known as vowel length [24].

A distinction in vowel quantity (vowel length) is generally realized acoustically by the duration of a vowel, with a phonologically long vowel having a duration which extends over more time than a phonologically short vowel. However, the measurable duration of vowels also depends on at least two other factors: (1) Vowel height, in terms of the position of the tongue; (2) Environment, in terms of preceding and following sound units. The duration of vowels is shortened before some consonants and lengthened before others. In many varieties, short vowels may have greater duration than long vowels. To avoid the confusion, some phoneticians consider it preferable to treat the length mark as a mark of quantity (vowel length) rather than duration, and refer to „heavy" (long) and „light" (short) vowels.

On the other hand, if a vowel has longer duration, there is time for the organs that form it to move into their target positions and remain there briefly before moving to the next target. Such a vowel is described as tense. If the vowel is too short, the organs have to leave the target as soon as they reach it, and in extreme cases may not reach the target at all. Such a vowel is described as lax. Lax vowels are pronounced with muscles of jaw in relaxed manner whereas tense vowels are produced and pronounced with muscular effort and have relatively long duration than lax vowels.

Vowel duration can be measured from the onset of the vowel to the formation of the constriction of the following consonant or to offset of the vowel. The onset of a vowel is a place where either a release of the constriction of the preceding consonant or the beginning of the vowel constriction. At this point, the acoustic signal wave amplitude and complexity begin to increase.

The offset of a vowel can be determined by observing the position of the vowel in the word or phrase. If the vowel is in final position, the offset of the vowel is where detectable acoustic energy ceases whereas if the vowel is in nonfinal position, the offset of the vowel will be the onset of the following consonant. This is the point at which wave amplitude and complexity stops decreasing. Detecting the precise locations of the release and the constriction of consonants depends on the kind of consonant itself [25].

The factors that are affecting the measurement of vowel duration are natural characteristics of vowels and contexts [25, 26, 27].

The first factor which affect the vowel duration is the natural properties of vowels. These characteristics of vowels are phonemic vowel length (phone identity), vowel height, and tone of vowels and gemination of vowels. The first characteristic is phonemic vowel length which is the perceived length of vowel when it is produced. Here vowel length and vowel duration are directly proportional to each other. This indicates that if the vowel is phonemically long or short, the duration of the vowel is also long or short. The second characteristic is vowel height; it is the height of the vowel measured based on the position and height of the tongue. It has its own influence on the vowel duration. For instance, English low vowels have longer duration than high vowels. The third one is tone of the vowel, it can affect the duration. For example: vowels which have contour tone (diphthongs and triphthongs vowels) have longer duration than those vowels which have simple tone (monophthongs vowels). The last one is gemination of vowels, if there is a gemination of vowels; the duration is longer than those single vowels.

The second factor which affect the vowel duration is the contexts around the vowels. The contexts which have an influence on duration are: preceding and following phone identities, domain position of vowel in words or phrases, stress, syallable structure and consonant clustering. The preceding and following phone identities may be voiced stops, voiceless stops, glottalized stops, voiced fricatives, voiceless affricatives, glottalized fricatives, voiced affricatives, voiceless affricatives, glottalized affricatives, nasals, liquids and semivowels. All these phones have their own influence on the adjacent vowel duration. For instance: English vowels which are before voiced consonants are longer that those which are before voiceless consonants. The other contextual factor is domain position with in words and phrases. If the vowels located in different position in a word or phrase, the duration is varies. For example: In English language, vowels which are located in final position is longer than in medial position. The final factor that has an influence on vowel duration is stress. It also affect the duration of vowels. For example: English vowels which found in stressed context are longer than in unstressed context.

In order to investigate one of these factors, the other factors must be controlled. For instance: If domain position within phrasing is the experimental factor, hold all the other factors constant, meaning all test vowels have the same vowel length, vowel height, stress level, tone, and the same following and preceding consonants or contexts.

2.3.3 Fractal Dimension (FD)

A fractal is a rough or fragmented geometric shape that can be subdivided into parts, each of which is (at least approximately) a reduced-size copy of the whole. Fractals are generally self-similar and independent of the scale. There are many mathematical structures that are fractals; e.g. Sierpinski triangle, Koch snowflake, Peano curve, Mandelbrot set, and Lorenz attractor [28]. Fractals also describe many real-world objects, such as clouds, mountains, turbulence, and coastlines that do not correspond to simple geometric shapes. The language of fractals has renewed interest in science, computation and mathematics etc. It is notable that fractals provide a framework for characterization and modeling of irregular, seemingly complicated structure in nature. For random fractals like noise, music, mountain and cloud, the characterization and modeling of them has been investigated from aspect of the fractional Brownian motion. As an example of fluctuated time sequential data, it is extensively accepted that the human electroencephalogram (EEG) possesses a self-affinity property with fractal dimension corresponding to a chaotic behavior of the neurons. Due to the invention of the name of fractal, Mandelbrot‟s fractal geometry has revolutionized the application of non-Euclidean geometry concepts to the natural sciences.

B. Mandelbrot in [28] coined the term fractal to identify a family of shapes that describe irregular and fragmented patterns in nature. Differentiating irregular shapes from the pure geometric shapes measured by the Euclidean geometry is the role of fractal geometry. Fractal geometry is the geometry of the irregular shapes found in nature [28]. It is based on the idea of self-similar forms. To be self similar, a shape must be able to be divided into parts that are smaller copies which are more or less similar to the whole. Because of the smaller similar divisions of fractals, they appear similar at all magnifications. However, while all fractals are self-similar, not all self-similar forms are fractals. (For example, a straight Euclidean line and a tessellation are self-similar, but are not fractals because they do not appear similar at all magnifications). Many times, fractals are defined by recursive formulas. Fractals often have a finite boundary that determines the area that it can take up, but the perimeter of the fractal continuously grows and is infinite and, in general, fractals are characterized by infinite details, infinite length, self-similarity, fractal dimensions, and the absence of smoothness or derivative using the concept of fractal geometry [28].

Fractal dimension is one of the fractal object characterization technique which refers to a non integer or fractional dimension of a geometric objects. Fractal dimension (FD) analysis is frequently used in biomedical signal processing, including EEG analysis [29] and in speech sound characterization for improving the computation and application of the speech systems such as speech recognition and speech synthesis. Basically, the applications of FD include two types of approaches, time domain and phase space domain. The first approaches estimate the FD directly in the time domain or waveform domain, where the waveform is considered as a geometric figure or a fractal object.

The second approaches estimate the FD of an attractor in state space domain. The phase space representation of a nonlinear, autonomous, dissipative system can contain one or more attractors with generally fractional dimension. This attractor dimension is invariant, even under different initial conditions. This explains why the FD of attractors has been used widely for system characterization. However, estimating the FD of these attractors involves a large computational burden. An embedding system has to be constructed from the time-domain signal, based on the method of delays [29] and the attractor of this system has to be untangled before estimating its FD. At present, the algorithms developed to assess the FD of the attractor are very slow, due to a considerable requirement for preprocessing. The most popular method for doing this is the algorithm from Grassbergerand Proccacia [29], which estimates the correlation dimension or FD of the attractor.

However, the time domain FD approach is better than phase space FD approach to calculate the FD value of waveform or signal. But their computational requirements are expensive. Calculating the FD of waveforms is useful for transient detection occurred during speech production, with the additional advantage of fast computation. It consists of estimating the dimension of a time-varying signal (waveform) directly in the time domain, which allows significant savings in program run-time.

Fractal characterization of speech waveforms based on the time or waveform domain approach was first reported by Pickover and Khorasani. Remarkably, they found the self affinity and fractal dimension FD=1.66 for human speech. Later such fractal properties in music have been investigated by Voss in detail, anlayising a statistical scaling property of the variance and also the exponent of spectral density. He concluded that the relation between 1/f noise and the musical melody with the fluctuation is concerned with a similarity of different melodies. The fluctuation component of a time sequence data may be well specified in terms of fractal dimension relate with self- affinity property. Verify recent fractal property of vowels has been studied on the basis of the scaling property of the variance, the capacity, the correlation dimension and the lyapunov analysis. Therein the local and the global properties have been characterized in terms of the fractal dimensions. It has been concluded; however, that there exists certain difficulty to evaluate the fractal dimension in the high precision from the scaling property of the power spectrum since the vocal sounds involves strong fluctuation in the power spectrum profile.

As point out on speech production, under source filter speech production system section the vowels are produced due to the vibration of vocal fold but consonants are produced due to turbulence noise. During consonant production dynamics of speech airflow might create small or large degrees of turbulence or noise during production of speech sound units by the human vocal-tract system. This turbulence effect makes speech production non linear system. The geometry of speech turbulence and non linearity of speech production as reflected in the fragmentation of the time signal is quantified by using fractal models. Waveform of speech sound produced by the turbulences effect is very complex than the vibration of the vocal fold. Nonlinear dynamic methods are used to describe the complexity of natural speech. Fractal dimension is a geometric invariant measure used for characterizing nonlinear systems. Based on the above point"s fractal dimension characterization is appropriate to the consonants phonemes. In addition to the above, the reasons why FD is appropriate for consonants are due to:

- Mostly turbulence occurs in speech with consonants during production.
- The vowels waveform is not fractal object [30].
- Consonant waveforms may be indistinguishable in time or frequency domain easily.

2.4 Acoustic-Phonetic Labeling (Segmentation)

To measure the formant frequency and duration of the speech unit there should be segmented or labeled speech unit from the specific context. Therefore we must focus on the segmentation section. Acoustic-phonetic labeling involves segmenting the speech signal in terms of phonetic characteristics such as stop occlusion, frication, nasality, vowels, semivowels and glottalisation [31]. The process of labeling the acoustic speech called segmentation. Segmenting the speech signal into phone-sized units is somewhat of an artificial task, since the gestures used to produce successive speech sound units overlap to a great degree. Nevertheless, there are often salient acoustic landmarks that correspond straightforwardly to recognizable articulatory events [32]. For instance, onsets and releases of oral consonantal constrictions for the production of stops, fricatives, and affricates often coincide with abrupt spectral changes.

The output of segmentation is phone. A segment is defined by an initial boundary at the left side of the segment and a final boundary at the right. The phonetic or phonemic labeling of speech signal is normally performed manually by phonetic or speech communication experts. However, due to human variability of visual and acoustic perceptual capabilities and to the difficulty in finding a clear common labeling strategy, the manual labeling or segmenting procedure is implicitly incoherent. Other drawback of manual intervention in labeling speech signals is that it is extremely time consuming. Even if it has a disadvantage the performance of manual segmentation is highly accurate and gives high performance than the automatic segmentation or labeling [32].

The segmentation process has the following procedures. 1) The segmentation environment, the audio and visual facilities available for segmenting the speech material which are spectrogram, waveform and formants; 2) the chosen or target phonetic alphabet for segmenting; 3) The common adopted segmentation strategy or criteria‟s and practical suggestion.

2.4.1 Segmentation Criteria

The segmentation criteria that we present in this section are all based on the more general strategy of finding constriction onsets and releases of the speech units. It is based on spectral characteristics most easily seen in spectrograms and waveform of the speech. The most dominant segmenting environment used for segmenting phonemes is spectrogram. However, Waveforms can also be useful for segmentation since they show dips and rises in amplitude, which often correspond to the onsets of constrictions and their release. In addition, voicing is often easier to see in waveforms than in spectrograms. However, amplitude dips can sometimes be gradual on waveform displays, particularly when constrictions are voiced, and some types of frication noise can be difficult to distinguish from aspiration noise on these displays. For these reasons, it is preferable to rely primarily on spectrographic evidence for segmentation decisions, and to use waveforms as corroborative evidence. When using visual displays for segmentation, it is sometimes easier to see gross spectral changes when these are zoomed out, or contain longer stretches of speech. It is recommended that more zoomed out displays to determine general boundary regions, and more zoomed in displays for determining exact boundary locations [31].

The general practical suggestion for segmenting phonemes are: 1) Always observe and listen to speech environment in order to segment the target phonetic units; 2) always set the boundary in the correspondence to a significant visually discovered in the speech waveform or spectrogram and 3) label the target phone using the identified starting and ending boundaries [31].

The common adopted strategies for segmenting vowels are given below:

1. If the vowel is between stop consonants, it is preferable to use the formant energy as a segmentation criterion for determining the segmentation boundaries. When we use formant energy as a criterion, the focus should be as much as possible on the lower formants F1 and F2, since they tend to be highest in amplitude due to the tilt of the glottal spectrum. The onset or offset of F2 formant energy is often easier to identify than the onset or offset of F1, because F1 is often confusable with F0 (first harmonic energy). In simple language, in the CVC syallable structure, where C is stop context, vowel boundaries can be identified by the location in which the increase in the overall amplitude and the onset and offset of Formant energy F2 [32].

2. If the vowel is between two fricatives consonants, the offset/onset of vowel formant energy (F2) can often also be used to identify fricative constriction boundaries and the vowel boundaries, but is at times less useful than the fricative noise criterion since formant energy can often be seen in the presence of frication. Weak voiceless fricative constrictions can often be identified by the onset and offset of frication noise, meaning the vowel boundaries are determined by the constriction and release of the frication noises and offset and onset of surrounding vowels‟ F2, and corresponding dips and rises in overall amplitude [32].

3. If the vowel is adjacent to liquids, glides and other vowels, use the midpoint of transitions from a preceding glide, liquids and vowels to vowel and the midpoint of transitions from the vowel to the following glide, liquids and vowels, as criteria for segmentation. Liquids, glides and other vowels are voiced sound units which have their own formant frequency pattern. Therefore, to segment vowels from these contexts formant transition or change should be considered in order to decide the vowel boundaries and the vowel boundaries are placed at the midpoint of the formant transitions or change takes place between adjacent contexts [32].

Chapter Three

RELATED WORKS

In this section, we will be explaining the characterization of vowel speech units of different languages. The works which are important and relevant to our problem domain will be described in a brief manner. There are various works done around the acoustic characteristics of vowel speech units for different languages. Some of the works are stated as follows.

3.1 Characterization of vowel speech units for English Language

3.1.1 Control method used in a study of the vowels

One of the related works that is used in our problem domain is "control method used in a study of the vowels". This paper was written by Gorden E.Peterson and Harlod L. Barney, at Bell Telephone Laboratories, in 1952 [33]. The main aim of the paper was studying characteristics of English vowel sound units by distinguishing the factors which influence the identification process. Such factors are: language and dialect background, vocal and auditory characteristics of individuals.

The authors tried to study the vowel characteristics under control method. They collected a list of words which were /hVd/ syallable structure. These words were heed, hid, head, had, hod, hawed, hood, who‟d, hud and heard. The lists of word were recorded using magnetic tape recorder. Afterward, the authors considered two basic things to identify vowels. The first basic thing was to play back the record to the listener and the listeners tried to identify the vowels. The second measure the acoustic properties of vowels using the acoustic measurement device then classified vowels based on the acoustic properties.

The number of speakers was 76, including 33 men, 28 women and 15 children. And each record two lists of 10 words, made a total of 1520 (76 speaker * 10 words * 2 repetition) recorded words. There were also 70 listeners to identify the speakers‟ words.

The acoustic measurements from narrow-band spectral consisted of formant frequencies (F1, F2 and F3), formant amplitudes and fundamental frequency (F0). The measurements were taken at a single time slice that was judged to be "steady state." The authors described that it is difficult to estimate the formant frequency where the fundamental frequency was high because the formants are poorly defined. In addition to the acoustic measurement, the /hVd/ signals were also presented to listeners for identification purpose.

The results of the measurement study showed a strong relationship between the intended vowel and the formant frequency pattern. However, there was considerable formant frequency variability from one speaker to the next, and there was a substantial degree of overlap in the formant frequency patterns among adjacent vowels.

The listening study showed that the vowels were highly identifiable: The overall error rate was 5.6%, and nearly all of the errors involved confusions between adjacent vowels. Even if the acoustic identification seems good. The dynamic properties of vowels like duration and spectral change was not considered in this work.

3.1.2 Acoustic characteristics of American English vowels

The acoustics properties of American English vowels were studied and characterized by James Hillenbrand, LauraA. Getty, Michael J. Clark and Kimberlee Wheeler, at Western Michigan University, in 1995 [34]. The authors started this work in order to extend and replicate the classical study of Peterson and Barney.

The authors tried to study the formant contour, fundamental frequency and duration of vowels. According to this work, vowels can be easily identified and characterized by combining these parameters from the speech production systems. As stated in Peterson and Barney work, they tried to characterize the vowels based on the static properties like formant frequency, formant amplitude and fundamental frequency (F0) but they did not consider the dynamic characteristics of vowels such as duration and spectral change over time. This work could solve the problem faced to Peterson and Barney identification of vowels by suggesting both the static and dynamic properties of vowels together.

The acoustic analysis showed that the selected persons or talkers who participated in data collection process are 45 men, 48 women, and 46 ten- to 12-year-old children (27 boys, 19 girls). Then audio recordings were made of subjects reading lists containing 12 vowels. The given datasets were /hVd/ syallable structure utterances and subjects or talkers are read by randomization from a list containing the words "heed," "hid," "hayed," "head," "had," "hod," "hawed," "hoed," "hood," "who'd," "hud," "heard," "hoyed," "hide," "hewed," and "how'd.". The total number of recorded vowel signals was 1668. After recording the dataset each waveform was low-pass filtered at 7.2 kHz and digitized at 16 kHz with 12 bits of amplitude resolution. Once the recording and filtering activities were completed, the next activities were performed acoustic measurement of parameters used to characterize the vowels. These parameters are vowel duration, formant frequencies (from F1 to F3) and fundamental frequency (F0). Formant and fundamental frequencies were sampled at steady state time.

The overall acoustic measurement result of vowels is analyzed by drawing F1 vs. F2 acoustic vowel space. On this acoustic vowel space, it is impossible to identify some vowels because there is an overlap between adjacent vowels. To overcome this problem, authors considered the patterns of formant frequency change associated with these utterances with respect to duration. They take average formant frequency at 20% and 80% of vowel duration and draw F1-F0 vs. F3-F2 vowel space in Mel. Scale where F1-F0 is abscissa and F3-F2 is an ordinate part. These spectral change or formant frequency changes have a high role in vowel identification. However, there is no contribution to the quality of vowels used for various speech applications.

To interpret the vowel identification accuracy, authors selected listeners consisted of 20 undergraduate and graduate students in the Speech and Pathology and Audiology Department at Western Michigan University, none of whom had participated as talkers for listening experiment. The listening experiment shows that the total vowel identification rates were 95.4%. In addition, vowel identification rate of listener was 94.6%, 95.6% and 93.7% for vowel produced by men, women and children respectively.

Finally, according to this paper the acoustic properties of vowels which are basic for characterization are formant frequencies (F1, F2 and F3), vowel duration, fundamental frequency (F0) and spectral change. But the relationship between spectral change pattern and vowel identification (vowel classification) is less straightforward in more complex phonetic environments. The methodology, authors followed in this paper was interesting and used as a bench mark for our work.

3.2 Characterization of vowel speech units for Bangla Language

3.2.1 Acoustic classification of Bangla vowels

This paper was done by Syed Akhter Hossain, M. Lutfar Rahman, and Farruk Ahmed at World Academy of Science, Engineering and Technology, in 2007. The main goals of the paper were: to characterize vowels by segmenting from context then applied time domain and frequency domain techniques in order to represent their formant frequencies, bandwidths as well as fundamental frequencies along with the duration, to classify vowels based on vowel normalization techniques and to develop a vowel space based on the acoustic characteristic of vowels [35].

In this paper the speech corpus consisted of 120 pair of disyllabic words, drawn from three word classes: nouns (common and proper names), adjectives, and infinitives. The first syllable of each word of each pair contained a consonant having voiceless variant to avoid co-articulatory effect during analysis of vowel speech units. From the corpus data, authors tried to accomplish acoustic measurement for vowels. The measurement was applied to duration, fundamental frequency (F0) and formant frequencies (F1, F2 and F3).

Before performed the extraction of parameters, the input samples were low pass filtered at 4 kHz and digitized at a sampling rate of 10 kHz. A spectrogram of each word was made using a 256-point discrete Fourier transform (DFT) analysis with a 10-ms Hamming window with an overlap segment. The vowels in each target word were manually segmented on the waveform and the corresponding wide band spectrogram.

To measure the duration of the vowel, first distinguished the onset and offset boundaries of vowels. Vowel onset and offset positions can be determined by observing both the spectrogram and the amplitude tracing on the waveform. Formant values were both automatically computed by spectral analysis using Matlab and visually verified using the spectrographic display; these methods almost always converged. The formant frequencies measured automatically in every 10 ms with an overlap of 5 ms over the entire utterance using the developed routine in Matlab and verified using visualization tools. The fundamental frequency F0 was computed and estimated using auto-correlation method. The authors tried to analyses the relationship between vowel height and vowel frontness with the corresponding acoustic formant frequency values such as vowel height is corresponds to F1 and vowel frontness or backness is corresponds to F2.

Finally, the authors developed acoustic vowel space on two dimensional spaces in which the abscissa was the second formant frequency values and ordinate was the first formant frequency values in order to show the distribution of vowels on the space.

3.2.2 Acoustic space of Bangla vowels

This is one of the papers done for characterization of Bangla vowels in addition to the above works. The main intension of the paper was to develop a vowel space for locating vowels on acoustic vowel space. It was done by Syed Akhter Hossain, M Lutfar Rahman and Farruk Ahmed, 2005 [36].

Authors tried to develop the acoustic vowel space or graphical representation of vowel qualities. So, to develop acoustic vowel space, the authors prepared the speech corpus of words that contain vowels and filter the speech signal of words to avoid the noise from the speech. Then manually segment the vowel from the context and apply LPC based feature extraction techniques to extract the format frequencies.

In LPC analysis, the extracted speech segment was taken for Linear Predictive Coding (LPC) scheme with 14 as number of poles. The maximum formant considered was 5000 Hz and the analysis width of the window was 25 msec with an overlap of 10 msec for the accurate resolution of the formants. Among the existing methods, the Burg method was applied and along with the formant extraction, the vocal tract response characteristics were extracted for each of the vowel in the vowel space.

After obtained the format frequencies of the vowels, the authors developed the graphical representation of acoustic vowel space for individual and the whole vowels. The graph indicated that there is an overlapped between the adjacent vowels on acoustic vowel space plan. This implies that vowels are very much indistinguishable on the acoustic space. Generally, the work showed that the range of formant frequencies which are used for speech synthesis and recognition, but it did not give a solution cue for avoiding overlapped problem of the vowels on acoustic vowel space. The methodology followed by this work seems good. So, we used as a second bench mark for our work.

3.2.3 Bangla vowel characterization based on Analysis by Synthesis

This characterization of Bangla vowels was done by Syed Akhter Hossain, M. Lutfar Rahman, and Farruk Ahmed in 2007 at World Academy of Science, Engineering and Technology. Authors obtained Bangla vowels in isolated word and analyzed based on speech production model within the framework of analysis-by-synthesis approach. This had led to the extraction of spectral parameters for the production model in order to produce different Bangla vowel sound units. The extracted features produced good representation of targeted Bangla vowel [37].

The authors had tried to characterize the vowels by bringing together all pole model speech production, time dependent Fourier transform of the signal and Linear Predictive Coding (LPC) analysis techniques. The continuous speech signal should be represented using a time dependent Fourier transform from discrete time speech production model and then applied Linear predictive Coding analysis technique based on autocorrelation method with selected parameters such as prediction order 10, analysis window of 20msec and overlap segment size of 5msec on a segmented vowel from the given words.

To determine the first five formant frequencies use of the LPC co-efficient vector, the corresponding gain and the formant tracking method called peak picking is applied to each analysis frame. Then, they got the formants by applying the peak picking formant extraction to the LPC spectrum for the segmental vowel. LPC coefficients extracted from the vowel segment are used for the spectral estimates of the vowel and the same coefficients are used in the synthesis of the vowel through inverse filtering operation. The synthesis vowel obtained from the inverse filtering operation is then passed through the analysis stage to determine the parameters of the syntactic vowel. If the difference between the original segmented vowel parameters and synthetic vowel parameters are small enough considered the parameters of the synthetic vowel is the characteristics of the vowel in that context word. Otherwise, compute the synthetic vowel parameters again and again until the difference between the parameters of the original and synthetic vowels is small enough.

Finally, the authors put the first five formant frequencies and their corresponding bandwidth of the vowels for both men and women speakers. The results show that the characterization seems good but when comparing both original and synthetic vowel parameters, if they have a large difference, it takes a long time to minimize the difference. As well as the authors did not indicate the dataset used and the performance of the system. This methodology is good for generating new synthetics or artificial vowel sound units.

3.3 Characterization of vowel speech units for Amharic Language

Amharic is one of Semitic languages. It is the working language in Ethiopia. It has a set of speech sound units categorized as consonant and vowel. These speech sound units have a high contribution to speech system development for Amharic language. When researchers explore different literature reviews for the purpose of speech synthesis systems, they may focus on the properties of speech sound units unconsciously such as formant based speech synthesis for Amharic vowels [38]. Nadew in [38] tried to extract the basic acoustic characteristics of vowels like formant frequencies, amplitude and fundamental frequency to apply and use in his speech synthesis system. However, he did not focus on the dynamic characteristic of vowels such as duration. The acoustic parameters of vowels extracted for this work were not reliable because the contexts have a high coarticulation influence on the acoustic characteristics vowels.

So, to the best of the author"s knowledge, there is no work done for characterization of Amharic vowel speech sound units. This research is conducted to full fill the gap occurred in the language as well as in Amharic speech systems.

Chapter Four

VOWELS

4.1 Introduction

The term "vowel" is commonly used to mean both vowel sound units that are distinct from a consonant and the written symbols that represent them. In phonetics, a vowel (comes from the Latin word „vocalis", meaning „uttering voice" or „speaking") is a sound in spoken language that is characterized by an open configuration of the vocal tract, in contrast to consonants, which are characterized by a constriction or closure at one or more points along the vocal tract. Vowels are understood to be syllabic, meaning that they usually form the peak or nucleus of a syllable whereas consonants form the onset and/ or coda of a syllable. A syllable is a unit of organization for a sequence of speech sound units in terms of sonority [39].

In this section, we focus on vowel articulation or cardinal vowels, basic characteristics of vowels and types of vowels.

4.2 Vowel Articulation

There are three possible resonators involved in the articulation of vowel sound units: the oral cavity, the labial cavity and the nasal cavity. The numbers of resonators involved distinguish between the types of vowels created. With oral vowels, there is no nasal resonance, as the soft palate is raised and air does not enter the nasal cavity. With nasal vowels, the nasal resonator is activated by lowering the soft palate and allowing air to pass through the nose and mouth simultaneously. Amharic does not have nasal vowels but languages such as French, Polish and Portuguese have nasal vowels. In rounded vowels, a third resonator called the labial resonator which is activated from oral vowels when the lips are pushed forward and rounded. If, on the other hand, the lips are spread sideways or pressed against the teeth, as in the case of unrounded vowels, no labial resonator is formed [40].

A vowel's timbre (color or quality) depends on the number of active resonators (among the oral, labial and nasal cavities), the shape of the oral cavity and the size of the oral cavity. The articulators, primarily the tongue, jaws and lips makes the shape of the vocal tract into these different resonating cavities then give acoustic characteristic quality to each vowel sound units [40].

The shape of the oral cavity is determined by the general position of the tongue in the mouth. Based on positioning of the tongue, the vowels are divided into three classes: front, back and central vowels. In front vowels, the tongue body is held in the pre-palatal region. In back vowels, the tongue body is held in the post-palatal or velar region. In central vowels, the tongue body is in the media-palatal region.

Other factor considered in the articulatory classification of a vowel is the size of the oral cavity. It depends directly upon the degree of aperture (opening) of the mouth; that is, upon the distance between the hard palate (roof of the mouth) and the tongue's highest point. Arbitrarily, four degrees of aperture are distinguished, from the most closed (first degree) to the most open (fourth degree).

In the sound systems of many languages, vowels can be characterized by their contrastive use of vowel quality and quantity. Vowel quality refers to the relative phonological resonance or timber of the vowel sound units. In vowel quality, it is not sufficient to define all the vowels in a system such as English vowels. Some vowels can only be defined in relation to other vowels: for example, /ɛ/ in *bet* is intermediate between close /ɪ/ in *bit* and open /æ/ in *bat*. So, the most widespread method of dealing with relative vowel quality is based on the system of cardinal vowels [41].

Cardinal vowels are a set of reference vowels used by phoneticians in describing the sound units of the languages. In the early 20th century, phonetician Daniel Jones developed the cardinal vowel system to describe vowels in terms of their common features. The common features of vowels are vowel height, vowel backness and vowel roundness. Vowel height is the degree of the opening of the jaws that is the vertical dimension of the tongue which is called the size of the oral cavity. Vowel backness is the relative position of the tongue mass in the supralaryngeal vocal tract. It is also the horizontal dimension of the tongue which is called the shape of the oral cavity and vowel roundness is the configuration of the lips called lip position or the shape of the labial cavity. These three parameters are indicated in the schematic IPA vowel. There are, however, still more possible features of vowel quality, such as the velum position (which, if lowered, contributes to nasality), type of vocal fold vibration (phonation), and tongue root position [42].

4.2.1 Vowel Height

Vowel height refers to the vertical position of the tongue relative to either the roof of the mouth or the aperture of the jaw. High vowels have relatively high jaw-tongue position, whereas low vowels are more open. Raising the tongue from the floor of the mouth also tends to draw the root of the tongue forwarded and expand the pharyngeal cavity, so some authors speak in terms of tongue root position or pharyngeal width [43]. For instance in English high vowels such as [i] and [u], the tongue is positioned high in the mouth, whereas in low vowels, such as [a], the tongue is positioned low in the mouth. The IPA prefers the terms close vowel and open vowel, respectively, which describes the jaw as being relatively closed or open. However, vowel height is an acoustic rather than articulatory quality, and is defined today not in terms of tongue height, or jaw openness, but according to the relative magnitude of the first formant frequency (F1). The higher the F1 value, the lower (more open) the vowel; height is thus inversely correlated to F1 [44].

The IPA identifies seven different vowel heights: close vowel (high vowel), near-close vowel, close-mid vowel, mid vowel, open-mid vowel, near-open vowel, and open vowel (low vowel). The parameter of vowel height appears to be the primary feature of vowels cross-linguistically in that all languages use height contrastively. No other parameter, such as front-back or rounded-unrounded, is used in all languages. Some languages have vertical vowel systems in which, at least at a phonemic level, only height is used as the sole distinguishing feature to distinguish the language vowels [44].

4.2.2 Vowel Backness

Vowel backness is named for the position of the bulk tongue during the articulation of vowel sound units relative to the back of the mouth. When bulk of the tongue shifted towards the alveopalatal region the vowel is front vowel or to the velar or uvular region, the vowel is back vowel [42]. In English front vowels, such as [i], the tongue is shifted or positioned forward in the mouth, whereas in back vowels, such as [u], the tongue is positioned towards the back of the mouth. However, vowels are defined as back or front not according to actual articulation (position of the tongue), but according to the relative frequency of the second formant (F2). The higher the

F2 value, the fronter the vowel; the lower the F2 value, the more back the vowel; so backness is thus inversely correlated to F2. The IPA identifies five different degrees of vowel backness: front vowel, near-front vowel, central vowel, near-back vowel and back vowel. Although English has vowels at all five degrees of backness, there is no known language that distinguishes all five without additional differences in height or rounding.

4.2.3 Vowel Roundedness

Roundedness refers to the position of the lips whether the lips are rounded or unrounded /spread. In most languages, roundedness is a reinforcing feature of mid to high back vowels, and is not distinctive. Usually the higher a back vowel is, the more intense the rounding. However, some languages treat roundedness and backness separately. Nonetheless, there is usually some phonetic correlation between rounding and backness: front rounded vowels tend to be less fronter than front unrounded vowels, and back unrounded vowels tend to be less back than back rounded vowels. That is, the placement of unrounded vowels to the left of rounded vowels on the cardinal vowel chart is reflective of their typical position [44].

Rounding is generally realized by a complex relationship between F2 and F3 that tends to reinforce vowel backness. One effect of this is that back vowels are most commonly rounded while front vowels are most commonly unrounded. Another effect is that rounded vowels tend to plot to the right of unrounded vowels in vowel charts.

In mid to high rounded back vowels the lips are generally protruded („pursed‟) outward, a phenomenon known as exolabial rounding because the insides of the lips are visible. In mid to high round front vowels, however, the lips are generally „compressed‟, with the margins of the lips pulled in and drawn towards each other. This latter phenomenon is known as endolabial rounding. (In English, [u] is exolabial.) In many phonetic treatments, both are considered types of rounding, but some phoneticians do not believe that these are subsets of a single phenomenon of rounding, and prefer instead the three independent terms rounded (exolabial), compressed (endolabial), and spread (unrounded) [40].

4.3 Characteristics of Vowels

The acoustics of vowels are fairly well understood for analyzing the basic vowel characteristics. The different vowel qualities and vowel quantities are realized in acoustic analyses of vowels by the relative values of the formants, acoustic resonances of the vocal tract which show up as dark bands on a spectrogram and the duration of the vowels extracted from the spectrogram by labeling the given contexts. The acoustics of vowels can be visualized using spectrograms, spectrum and waveform. A spectrogram is a three-dimensional plot of the intensity of the frequency content of a signal as it changes over time. Usually the vertical axis indicates frequency, the horizontal axis indicates time, and the color or gray scale indicates intensity. Spectrogram used to measure formants, especially vowel formants and analyzes voiced and unvoiced sound units, periodic and aperiodic sound units, glottal pulses, and fundamental frequency overtones as well as the duration of the phones. Formants appear as dark bands in a spectrogram [45].

In this section, we focus on the acoustic characteristics of vowels which are very relevant and critical for characterization and perception or identification of vowel sound units.

4.3.1 Vowel Formants

Each phoneme is distinguished by its own unique pattern in the spectrogram. Spectrograms display the formants, which appear as dark bands at each frequency, and how this changes with time. For voiced phonemes, the signature involves large concentrations of energy (formants). During voicing, which is a relatively long phase, the spectral or frequency characteristics of a formant evolves as phonemes unfold and succeed one another. Formants that are relatively unchanging over time are found in the monophthongs or pure vowels and the nasals; formants which are more variable over time are found in the diphthong vowels and the semi vowels (approximants) but in all cases the rate of change is relatively slow [44].

The vocal tract acts as a resonant cavity, and the positions of the jaw, lips, and tongue used in the articulation of specific vowel sound units which affect the parameters of the resonant cavity, resulting in different formant values. The information that humans require to distinguish between vowels can be represented purely quantitatively by the frequency content of the vowel sound units; that is, the different vowel qualities are realized in acoustic analyses of vowels by the relative values of the formants which is the acoustic resonances of the vocal tract. Each vowel has its own distribution of acoustic energy that distinguishes it from all other vowels. Vowels will almost always have four or more distinguishable formants. However, the first two formants are the most important in determining vowel quality and in differentiating it from other vowels.

Each vowel, therefore, has its own „fingerprint", which is defined or characterized by its unique frequencies at the first and second formants. These formants are usually referred to as the vowel formants. They are not adjustable for each given vowel or variant of each vowel. For example, the formant frequencies for vowel [i] in any given voice are more or less constant and remain within very specific limits in the frequency range. For this reason, these vowel formants may also be called fixed formants. If these vowel formants are not produced by the vocal tract, the particular vowel cannot exist. Conversely, whenever the formants for a particular vowel are present, that vowel is heard.

4.3.2 Vowel length

Vowel length is the length of time it takes to pronounce the vowels. It is a term sometimes used in phonetics to refer the subjective impression of speaker that is distinct from physically measurable duration [39]. It is the phonological correlate of durational differences between sound units, tied to the phonological concept called quantity. Vowel quantity refers to the phonologically distinctive length of a vowel relative to one or more vowels of similar quality in the language [41]. A distinction in vowel quantity is generally realized acoustically by the duration of a vowel, with a phonologically long vowel having a duration which extends over more time than a phonologically short vowel. In addition, the greater amount of time associated with a phonologically long vowel may also be associated with an articulation using greater extremes of the vocal space than phonologically short vowels, and consequently may also affect the vowel spectrum [41]. The concept of vowel length usually considers duration segmentally, attributing length to particular segments (the vowel or consonant is the locus of the property "long" or "short", and not the syllable, foot or word). Length being a phonological attribute, distinctions is discrete mental categories, not physical measurement, and like most phonological attribute in generative theories, it is traditionally treated binarily [46].

As with all phonological distinctions, so also with the feature of vowel length the question arises as to the likelihood that the distinction rests on a single phonetic feature or phonemic feature. On the face of it, distinctive length seems to be little more than a matter of relative timing. Thus, in a given context, members of a "long" category are longer than their counterparts in the "short" category [47].

Phonetic duration does not automatically translate into phonological length, and durational differences, especially ternary oppositions, are not necessarily of a continuous type (1, 2, 3, 4 units), but may instead reflect multiple, intersecting phonological phenomena [46]. Vowel duration is used to differentiate the vowel length. The vowel duration of long vowel is twice of short vowel. Many languages make a phonemic distinction between long and short vowels in terms of vowel length such as: Sanskrit, Japanese, Finnish, Hungarian, etc. Long vowels may or may not be separate phonemes. In Latin and Hungarian, long vowels are separate phonemes from short vowels, thus doubling the number of vowel phonemes. For example: some dialects of English language have phonemic vowel length. Other example that shows the phonemic vowel length distinction is Japanese vowel length. In Japanese long vowels are analyzed as two same vowels or a vowel + the pseudo-phoneme /H/ [48] and the number of vowels are five as shown in Table 4.1:

Table 4.1: Japanese vowels (mainly adopted from [48])

	Front		Central		Back	
	short	long	short	long	short	long
High	/i/	/ii/ or /iH/			/u/	/uu/ or /uH/
Mid	/e/	/ee/ or /eH/			/o/	/oo/ or /oH/
Low			/a/	/aa/ or /aH/		

4.3.3 Vowel duration

The scientific measure of the amount of time that an event takes is called duration [39]. Measuring duration of events in a speech signal can be useful in any of the following types of linguistic analysis [45]:

- Duration of pauses between clauses in discourse studies.
- Duration of vowels for differeciate contrastive vowel length.
- Time between the release of a plosive or stop and the beginning of voicing (voice onset time) for the purpose of distinguishing unaspirated, aspirated and voiced stop consonants.
- Duration of the closure phase in order to distinguishing voiced and voiceless stops.

Vowel duration is the amount of time that measured for phoneme sound units within different contexts such as words and sentences. The amount of time that a sound lasts for is a very important feature of that sound unit. In the study of vowel speech sound units, it is usual to use the term length for the listener's impression of how long a sound units lasts for, and duration for the physical, objectively measurable time. For example, we might listen to a recording of the following syllables and judge that the first two contained short vowels while the vowels in the second two were long: [bIt, bet, bi:t, bɔːt] ; that is a judgment of length. However, if we use a laboratory instrument to measure those recordings and find that the vowels last for 100, 110, 170 and 180 milliseconds respectively, we have made a measurement of duration [39].

4.4 Types of Vowels

Vowels are classified into three based on the change in vowel quality over the duration of the vowels. These are monophthongs, diphthongs and triphthongs. A vowel sound whose quality does not change over the duration of the vowel is called monophthongs. Monophthongs are sometimes called pure or stable vowels. Apart from listening to them, acoustically, most monophthongs can be identified by looking at their acoustic characteristic formant patterns for the first and second formants. In order to identify the frequency of the individual formants, it is best to try and identify steady state periods, which are times in which the formant patterns remain relatively stable. Usually those patterns occur somewhere near the middle of the vowel because the beginning and ending may be influenced by the effect of neighboring consonants due to coarticulation or assimilation process [49].

Diphthongs are unitary vowels (contour vowels) that change quality during their pronunciation, or glide, with a smooth movement of the tongue from one articulation to another. They often form when separate vowels are run together in rapid speech. Such as in the English words „eye", „boy", and „cow", there is a glides of vowels , so these vowels are diphthongs. A vowel sound that glides successively through three qualities is triphthongs. Diphthongs and triphthongs are essentially just combinations of monophthongs, with the main difference that the formant patterns associated with them usually show distinct movements from one vowel pattern to another [49].

Chapter Five

THE AMHARIC LANGUAGE PHONOLOGY

5.1 Introduction

Amharic is the working language of Ethiopia. It is the Semitic language family that has the largest number of speakers next to Arabic [50]; as per the 1998 census, it is spoken by 17.4 million people as a mother tongue and 5.1 million people as a second language [51]. The other Semitic languages spoken in Ethiopia include: Geez (the language of liturgy); Tigre and Tigrinya in the north; Gurage, Harari, Argobba and Gafat in the south.

Owing to political, economical and social conditions, Amharic is gaining ground throughout the whole country. Amharic is used in business, government and instruction. It has four dialectical variations spoken in different Amhara region: Gojjam, Gonder, Wollo and Menz [52]. The dialect of Addis Ababa has emerged as the fifth standard dialect and has wide acceptance across all Amharic-speaking communities [50].

Amharic has its own alphabets and about 33 consonants and 7 vowel sound units. The orthographic representation of the language is organized into orders. Each of the 33 consonants has seven orders (derivatives). Six of them are CV combinations while the seventh is the consonant itself. Table 5.1 shows the Amharic alphabets. The first column characters are called Ge‛‛ez (ግዕዝ) that is to say first in Ge‛‛ez. Similarly, the remaining characters are named as Ka‛‛ib (ካብዕ), Sali‛‛s (ሳልስ), Rab‛ï (ራብዕ), Hami‛‛s (ሃምስ), Sadi‛‛s (ሳድስ) and Sab‛ï (ሳብዕ) respectively [53]. Moreover, there are extra orthographic symbols in the language that are not organized as above. As a whole, the total number of orthographic symbols of the language exceeds 230 [54].

Although it is the Semitic language like Hebrew and Arabic, its writing is a syllabic left-to-right script. In addition to the alphabets, Amharic also has its own numerals and punctuation symbols [55].

Even if Amharic may have different phonetic representation of phonemes in some other literatures [9], in this work we preferably follow the transliteration presented in [54] because this representation is clear and it makes our research implementation easy as well as most of the researches that are conducted on Amharic language follow this phonetic representation.

In this chapter, we will have a brief look on phonology of the Amharic words and phonological rules of Amharic phonemes. In the phonology of Amharic words section, we will explain about vowels, consonants and the Amharic word transcription. In the phonological rules of Amharic phonemes section, the discussion focuses on the different phonological process such as vowel sequences, gemination, and consonant clustering, stress and phoneme length.

Table 5.1: Amharic alphabets with their seven orders

First	Second	Third	Fourth	Fifth	Sixth	Seventh
ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ
ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ
ሐ	ሑ	ሒ	ሓ	ሔ	ሕ	ሖ
መ	ሙ	ሚ	ማ	ሜ	ም	ሞ
ሠ	ሡ	ሢ	ሣ	ሤ	ሥ	ሦ
ረ	ሩ	ሪ	ራ	ሬ	ር	ሮ
ሰ	ሱ	ሲ	ሳ	ሴ	ስ	ሶ
ሸ	ሹ	ሺ	ሻ	ሼ	ሽ	ሾ
ቀ	ቁ	ቂ	ቃ	ቄ	ቅ	ቆ
በ	ቡ	ቢ	ባ	ቤ	ብ	ቦ
ተ	ቱ	ቲ	ታ	ቴ	ት	ቶ
ቸ	ቹ	ቺ	ቻ	ቼ	ች	ቾ
ኀ	ኁ	ኂ	ኃ	ኄ	ኅ	ኆ
ነ	ኑ	ኒ	ና	ኔ	ን	ኖ
ኘ	ኙ	ኚ	ኛ	ኜ	ኝ	ኞ
አ	ኡ	ኢ	ኣ	ኤ	እ	ኦ
ከ	ኩ	ኪ	ካ	ኬ	ክ	ኮ
ኸ	ኹ	ኺ	ኻ	ኼ	ኽ	ኾ
ወ	ዉ	ዊ	ዋ	ዌ	ው	ዎ
ዐ	ዑ	ዒ	ዓ	ዔ	ዕ	ዖ
ዘ	ዙ	ዚ	ዛ	ዜ	ዝ	ዞ
ዠ	ዡ	ዢ	ዣ	ዤ	ዥ	ዦ
የ	ዩ	ዪ	ያ	ዬ	ይ	ዮ
ደ	ዱ	ዲ	ዳ	ዴ	ድ	ዶ
ጀ	ጁ	ጂ	ጃ	ጄ	ጅ	ጆ
ገ	ጉ	ጊ	ጋ	ጌ	ግ	ጎ
ጠ	ጡ	ጢ	ጣ	ጤ	ጥ	ጦ
ጨ	ጩ	ጪ	ጫ	ጬ	ጭ	ጮ
ጰ	ጱ	ጲ	ጳ	ጴ	ጵ	ጶ
ጸ	ጹ	ጺ	ጻ	ጼ	ጽ	ጾ
ፀ	ፁ	ፂ	ፃ	ፄ	ፅ	ፆ
ፈ	ፉ	ፊ	ፋ	ፌ	ፍ	ፎ
ፐ	ፑ	ፒ	ፓ	ፔ	ፕ	ፖ

5.2 Phonology of the Amharic words

In phonology, phoneme is the fundamental class of sound units that describes how speech conveys linguistic meaning. The meaning of a word is dependent on the phoneme that it contains [56]. But, all the letters of the Amharic script are not necessary for the pronunciation patterns of the spoken language. Generally, phonemes are categorized into consonants and vowels. A set of 39 phonemes, seven vowels and thirty two consonants, make up the complete inventory of sound units for the Amharic language.

Consonant and vowel sound units are produced in fundamentally different ways. While consonants are articulated with a substantial degree of obstruction in the oral cavity, vowels are produced with a relatively free airflow. We give a brief overview of each of these major categories of Amharic phonemes in the subsequent sections.

5.2.1 Vowels

The tongue is the main articulator in the mouth to determine the nature of the vowels; other articulators also come into play as well. The most important secondary vowel articulation mechanism for Amharic language is lip rounding. For example, if you say the Amharic word /በሉ/ [Belu] which means *they ate*, your lips begin to round out, ending in a more puckered position. The word contains the Amharic vowel /ኡ/ [u] at the end of the utterance associated with /ሉ/. Consider another word like /ጋቢ/ [Gabii] that is to say a *cloth,* when you say /ኢ/ [ii] of /ቢ/ your lips will become flat, opened, and somewhat spread. So, depending on the position of the lip the Amharic vowels (ኧ, ኡ, ኢ, ኣ, ኤ, እ and ኦ) are broadly categorized into rounded (ኡ and ኦ) and unrounded (ኧ, ኢ, ኣ, ኤ and እ). On the other hand, the Amharic vowels are classified into lax and tense based on the muscular tension occurred when the vowels are produced. So, vowels (ኧ, ኣ, ኤ, እ and ኦ) are lax and vowels (ኢ and ኡ) are tense vowels [38].

Out of the 33 basic forms, two of them represent vowels in isolation (አ and ዐ) [8]. In this paper, we consider the Alfa-A/አ/ and its variation to deal with the Amharic vowels as the others have the same shape. These vowels are found with a combination of each consonant in consonant vowel (CV) manner. Six of them have this CV combination while the seventh is the consonant itself [54]. For each consonant C, the orthographic ordering is as follows:

C[e] C[u] C[ii] C[a] C[ie] C C[o]

From the above representation, we can see that the sixth order in the orthographic symbols, which does not have any vowel unit associated to it in the written form (CV transcription of the orthographic form), may associate the vowel **/እ/** [ix**]** in its spoken form, which has important role during syllabification of the word in the language which allows splitting impermissible consonant clusters [54].

A chart depicting the Amharic vowels in the IPA representation are shown in Figure 5.1 [54]. The IPA maps the vowels according to the position of the tongue. The vertical axis of the chart is mapped by vowel height. Vowels pronounced with the tongue lowered are at the bottom, and vowels pronounced with the tongue raised are at the top.

In a similar fashion, the horizontal axis of the chart shows vowel backness. Vowels with the tongue moved towards the front of the mouth (such as the /**ኤ**/ [ie] vowel in /**ቤት**/ [B ie t]) are to the left in the chart, while those in which the tongue moved to the back (such as the vowel /**ኦ**/ [o] in /**ሶስት**/ [Sost]) are placed to the right in the chart.

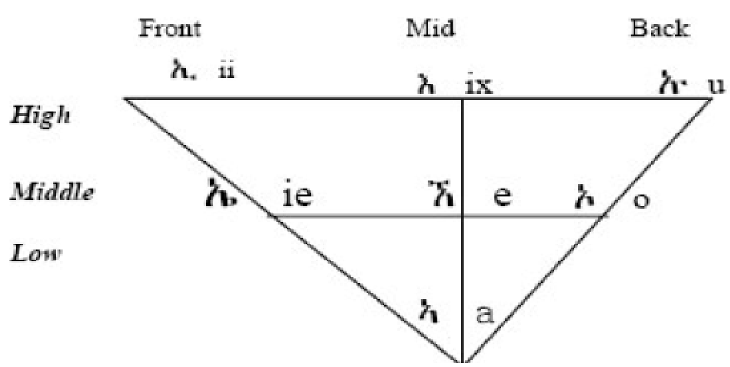


Figure 5.1: The seven Amharic vowels with their features (mainly adopted from [54])

5.2.2 Consonants

Consonants are characterized by significant constriction or obstruction in the pharyngeal and/or oral cavities [57]. They involve constrictions, or gestures that narrow the vocal tract at a particular point [38]. Most consonants use only one constriction, but some have more than one. In order to fully describe a consonant sound unit; we need to describe each constriction as well as some other properties of the vocal tract. For this, there are parameters of a constriction gesture that shows the degree of constriction and how one consonant is different from the other. These parameters are described in terms of active articulator which is the articulator that moves to make the constriction, passive articulator which is articulator that the active articulator touches or approaches, degree of constriction is the extent in which the active and passive articulators close to each other, and laterality are sound units that have airflow along the sides of the tongue.

Generally, consonant sound units are described in terms of the following parameters. These are:

1. Place of articulation: Where is the sound constricted?
2. Manner of articulation: How is the airstream constricted?
3. State of the glottis: are the vocal folds open?
4. Nasality: Is air escaping through the nose?

Answering each of these questions about the state of the vocal tract is enough to uniquely identify any consonant. If two consonants are different, they must differ in their answers for at least one of the above questions. The active and passive articulators are collectively referred to as the **place of articulation (POA)**. The other parameters are often lumped together and referred to as **manner of articulation.** The active articulator usually moves in order to make the constriction. The passive articulator usually just sits there and gets approached. Amharic consonants are generally classified as stops, fricatives, affricatives, nasals, liquids, and semi-vowels in their manner of articulation and as labial, alveolar, palatal, velar, and labio-velar and glottal based on their place of articulation shown in Table 5.2 [9].

As mentioned above, place of articulation refers to the place where an obstruction occurs in the vocal tract - which active articulator gets close to which passive articulator. This is expressed in constriction degree, which refers to how close they get. In **stops**, the active articulator touches the passive articulator and completely cuts off the airflow through the mouth. Amharic stops include / **ፕ**/ [p], /**ክ**/ [k] and /**ት**/ [t].

The other manner of articulation is **fricative**, here the active articulator does not touch the passive articulator, but gets close enough that the airflow through the opening becomes turbulent. Amharic fricatives include /ፍ/ [f], /ስ/ [s], and/ሽ/ [sx]. Affricative consonants /ጅ/ [j], /ች/ [c] and /ጭ/ [cx] are produced when the active articulator touches the passive articulator the air cuts off completely, then such articulators are immediately separate to each other airflow through the opening becomes turbulent.

Nasal consonants /ም/ [m], /ን/ [n] and /ኝ/ [nx] are a mixed bag: the oral cavity has significant constriction (by the tongue or lips), yet the voicing is continuous, like that of the sonorant[1], because, with the velar flap open, air passes freely through the nasal cavity, maintaining a pressure differential across the glottis. **Liquids** (/ል/ [l] and /ር/ [r]) are also sonorant and are quite vowel-like. In fact, it may become syllabic or act entirely as vowels in certain positions, such as the /ል/ [l] at the end of /ከፍል/. Phones that typically have voicing without complete obstruction or narrowing of the vocal tract are called ***semivowels***. In Amharic, they are usually grouped with consonants [8]. They named so because they function in a similar way to consonants but are phonetically similar to vowels: In Amharic, for example, the sound units /ው/ [w] and /ይ/ [y] (as found in /ውሃ/ and /ይምጣ/) are of this type. The term semivowel has been in use for a long time for such sound units, though it is not a very helpful or meaningful name. The term ***approximant*** is more often used today, meaning that the tongue approaches the top of the oral cavity, but does not completely contact so as to obstruct the air flow [57].

On the other hand, Table 5.2 indicated that Amharic consonants are also classified in to six categories based on their place of articulation and stated as follows. The **Labial** consonants have their major constriction at the lips and it produced by creating a closure with both lips. This includes /ብ/ [b], /ፕ/ [p] (these two differ only by manner of articulation), /ም/ [m] and /ፍ/ [f]. **Alveolar** consonants are produced when tongue tip against or close to the superior alveolar ridge, behind the top teeth. This includes /ት/ [t], /ጥ/ [tx], /ስ/ [s], and /ል/ [l].

---

[1] Sonorant sound units are those that have continuous voicing. Liquids, gilds (semivowels) and vowels are sonorant.

Palatal consonants have approximation or constriction on or near the roof of the mouth, called the palate. The members include /ጅ/ [j] and /ች/ [c] [54]. **Velar** consonants bring the articulator (generally the back of the tongue) up to the rearmost top area of the oral cavity, near the velar flap [57]. In Amharic, the velar sound units are /ክ/ [k], /ግ/ [g], and /ቅ/ [q].

Labio-velar consonants are produced at two place of articulation, one at the lips and other at the soft palate. The members include: /ኵ/ [kx], /ጕ/ [gx], and /ቍ/ [qx]. The last group of Amharic place of articulation is **Glottal**. Actually, it is not strictly a place of articulation, but they had to put it in the chart somewhere. Glottal sound units are made in the larynx. For the glottal stop, the vocal folds close momentarily and cut off all airflow through the vocal tract. Amharic uses the glottal stop in the utterance, for which we need some sort of repetition of what is said by someone or it also serves during surprising movement. The sound units are represented by /ዕ/ [ax] and /ህ/ [h]. In /ህ/ [h], the vocal folds are open, but close enough together that air passing between them creates friction noise. Table 5.2 shows the phonetic representation of the consonants with their corresponding manner and place of articulations.

Table 5.2: Categories of Amharic consonants (adopted from [54])

		Labials		*Alveolar*		*Palatals*		*Velars*		*Labio-Velar*		*Glottals*	
Stops	Voiceless	p	ፕ	t	ት			k	ክ	kx	ኵ	ax	ዕ
	Voiced	b	ብ	d	ድ			g	ግ	gx	ጕ		
	Glottalized	px	ጵ	tx	ጥ			q	ቅ	qx	ቍ		
Fricatives	Voiceless	f	ፍ	s	ስ	sx	ሽ					h	ህ
	Voiced	v	ቭ	z	ዝ	zx	ዥ						
	Glottalized			xx	ጽ							hx	ኃ
Africatives	Voiceless					c	ች						
	Voiced					j	ጅ						
	Glottalized					cx	ጭ						
Nasals	Voiced	m	ም	n	ን	nx	ኝ						
Liquids	Voiced			l	ር								
				r	ል								
Glides		w	ው			y	ይ						

5.2.3 Amharic word Transcription

The other important issue in relation to vowels is to study how to transcribe Amharic words or contexts. Transcription is used to produce the phonetic representation of a given word. In Amharic writing system, the vowels are found masked with the consonants. They are important to make variation in meaning for each consonant in the seven orders. One of the means to find out the vowels is to label each character of a word based on its acoustic information. For example, in Amharic word /ሰሩ/ "Seru" there is no vowel at all in the writing system, but still there are two vowels(/ኧ/ [e] and (/ኡ/ [u]), one in each character, embedded in the two consonants. But, if the vowels appeared to be in the word, they will be written in the transcription with the same vowel without any addition.

In Amharic, there are some characters which have different orthographic representation but still have the same sound. For example, the Amharic characters /ሀ/, /ሐ/ and /ኀ/ have different orthographic form, but still having the same meaning and representing the same phoneme. In writing the transcription, we will follow the same sound pattern for all redundant characters. For instance, the Amharic characters /ሀ/ [h] in /ሀሳብ/ [hesab] and /ሐ/ [h] in /ሐቅ/ [heq] have different representation but still have the same sound pattern and transcription. This holds true for all the redundant characters in the Amharic character set [55].

For Amharic language, before transcription is made the transliteration (representation of an alphabet with letters from a different alphabet) of each character is made by using the ASCII value of each of the character. The complete list of this transliteration is available in (Appendix A). When we write phonetic transcription, it is conventional to use brackets at the beginning and end of the item or passage to indicate the nature of the symbols. Usually, square brackets are used to represent the phonetic transcription.

5.3 Phonological rules of Amharic Phonemes

The phonological rules of Amharic phonemes include vowel sequence, gemination, and consonant clustering, stress and phoneme length which are described in the following sections.

5.3.1 Vowel sequences

As a rule, Amharic avoids the coming together of two vowels in pronunciation as well as in writing. For morphological and phonological reasons, if two vowels should come together, it may occurs either elision of one of the vowels or introduction of a glide (semi vowel) /**ው**/ [w] or /**ይ**/ [y] between two vowels [58].

5.3.1.1 Elision

Elision is a process where one or more phonemes are dropped, usually in order to simplify the pronunciation. It may occur for both vowels and consonants, although it is much more common for consonants. Where it occurs for vowels, we have extreme cases of vowel reduction or weakening to the point that the vowel is no longer pronounced at all [59].

In Amharic language, vowel elision occurs: 1) if the two vowels are the same and 2) if the first vowel is central vowel and the second vowel is a vowel other than a back vowel. In the first case, one of the two vowels is removed [58]. For example**:** For vowel **/አ/-/አ/**, /**ሰማቸው**/, "semma-accew", "he heard them", the correct pronunciation is "semmaccew. For vowel: **/እ/-/እ/**, **ልሰብር**, "lix-ixsebixr", "so that I break", pronounced as" lixsebixr". In the second case, if the two vowels are different and the first is central (/**ኧ**/, /**አ**/ and /**እ**/) while the second is other than back vowels (/**ኡ**/ and /**ኦ**/), one of the vowels is eliminated and if the second vowel is /**ኡ**/ or /**ኦ**/, there is no elision [60]. For example:

1. If the vowels /**ኧ**/ and /**እ**/ come together, the second vowel is removed.

   /**ኧ**/ +/**እ**/⟶ /**ኧ**/: Such as, /**የጅ**/ is not pronounced as /**ይኧእጅጅ**/, but the correct pronunciation is /**ይኧጅጅ**/

2. If the vowels /ኧ/ and /አ/ come together, the first vowel is discarded.

   /ኧ/ +/አ/⟶/አ/: Such as, the word /ለአንተ/ is not pronounced as /ልኧአንትኧ/ rather it is

   pronounced as /ልአንትኧ/.

3. If the vowels /አ/ and /እ/ come together, the second vowel is removed.

   /አ/ + /እ/ ⟶ /አ/

4. If the vowels /አ/ and /ኧ/ come consecutively, the first vowel is removed.

/አ/ + /ኧ/ → /ኧ/

5. If the vowels /እ/ and /አ/ come respectively, the first vowel will be elided.
6. If the vowels /ኡ/ and /አ/ come consecutively, the vowel spoken as /ኡኣ/ or /ዉኣ/.
7. If the vowels /ኦ/ and /አ/ come successively, the vowel spoken as /ኦኣ/ or /ዉኣ/
8. If the vowels /ኢ/ and /አ/ come respectively, the vowel spoken as /ኢኣ/ or /ይኣ/

5.3.1.2 Insertion of semivowel /ዉ/ [w] or /ይ/ [y]

When the vowels come together in the following way, the semivowel is inserted [58].

1. If the first vowel is a back vowel /ኡ/ [u] or /ኦ/ [o], the semivowel /ዉ/ [w] is inserted.
2. If the first vowel is a front vowel /ኢ/ [ii] or /እ/ [ix] and the second vowel is the central vowel /አ/ [a], the semivowel /ዉ/ [w] is inserted.
3. If the first vowel is a front vowel /ኢ/ [ii] or /እ/ [ix] and the second vowel is a back vowel /ኦ/ [o], a semivowel /ይ/ [y] is inserted.
4. If the first vowel is a central vowel and the second vowel is a back vowel /ኦ/ [o], a semivowel /ዉ/ [w] is inserted.

Generally, when a semivowel is inserted, the vowel preceding it may remain or may elide it.

5.3.2 Gemination

Gemination is most conventionally described as the lengthening of the consonants. All the consonants, except /ህ/ [h], can occur either in geminated or non-geminated form. In Amharic language gemination never occurs at initial position. It occurs in medial and final positions of the words [58]. Example: In medial position, /ዋና/, [wanna] "Main, Principal" is containing the geminated consonant /ን/ [n]. At the final position, note the contrast between the non- geminated and geminated consonants in /ወጥ/, [wet] "stew" which contain a single consonant /ጥ/ [t] as against /ውጥ/ [wett] "someone or something that is likely to go out" which contain geminated consonant /ጥ/ [t].

The nature of gemination is phonemic and often results from the assimilation of one consonant to another. Thus /የትሰበር/ [yetsebber] "it is broken" is derived from /ተሰበረ/ [tesebbere] by assimilating [yet] in place of [te] [58]. Gemination has not a specific symbol to represent in the word, but to show simply, doubling the consonant.

Gemination is the main super segmental in Amharic language. It differentiates lexical and grammatical (morphological features) or meanings in contrast to vowel length and stress. Gemination as a lexical item cannot predict. But it can make a difference between the lexical meanings occurring in minimal pairs such as in /ገና/ [genna] against [gena]. In addition, gemination also varies grammatical meaning occurring in minimal pairs such as /መሳም/ [mesam] "to kiss" and /መሳም/ [messam] "to be kissed" have different meaning due to gemination [60].

The analysis of gemination through speech analysis tools is conducted using praat in order to support the theoretical perception with practical acoustic analysis to see the effect of gemination on consonant phonemes. The spectrogram and waveform description of geminated context shows that, gemination gives the length variation in Amharic consonant phonemes. The duration of geminated consonant is longer than non-geminated consonant. For example: The duration of geminated consonant (/ር/ [r]) within /በራ/ [b e r r a] context shown in Figure 5.2 is longer than non geminated consonant (/ር/ [r]) within /በራ/ [b e r a] context shown in Figure 5.3.

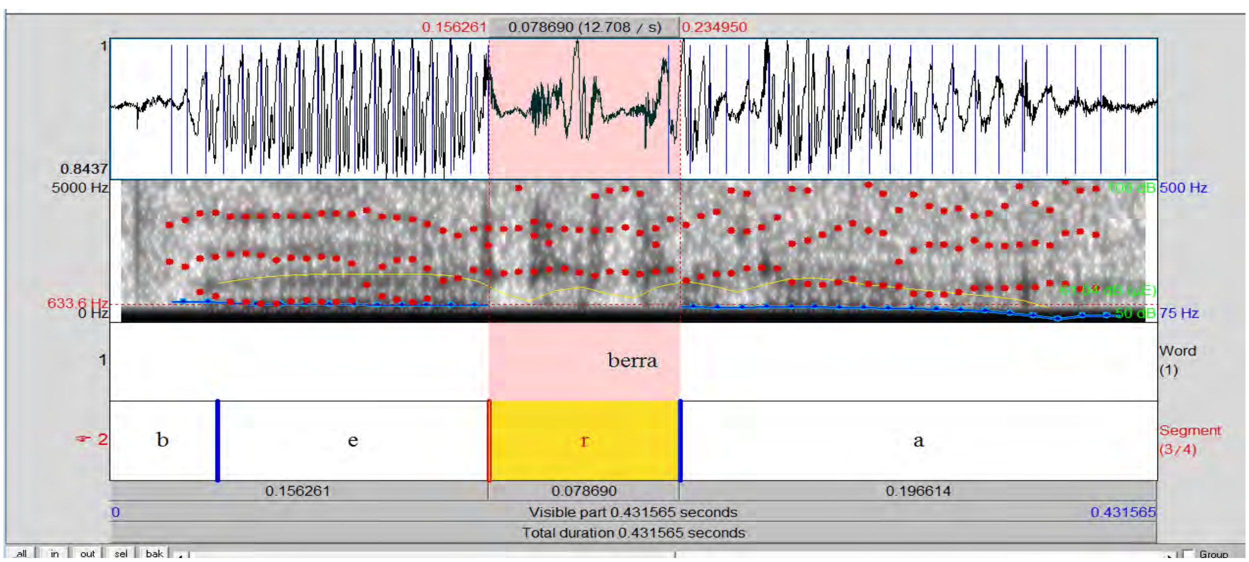


Figure 5.2: The word /በራ/- [b e r r a] contain geminated consonant (/ር/ [r])

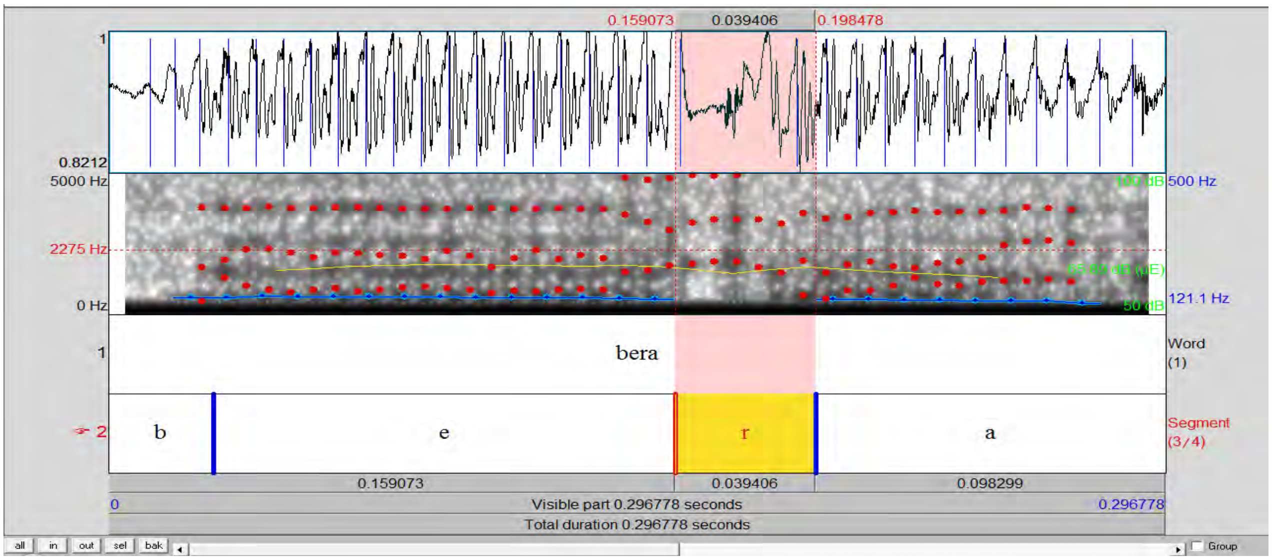


Figure 5.3: The word /በራ/- [b e r a] contain nongeminated consonant (/ር/ [r])

5.3.3 Consonant clustering

Consonant clustering occurs when a succession of two consonants that are not separated by a vowel. There are conditions in which consonant clustering occur in Amharic language. At the fist position, Amharic has not consonant clustering but in medial and final positions it has clusters of two consonants but there is no word that has a cluster of three or more consonants [58].

5.3.3.1 Medial consonant clusters

Amharic has clusters of two consonants in medial position, when two consonants meet in the middle of a word, the first consonant closes the syallable and the second consonant opens the next syallable. For example, /መንግስት/ pronounced as [men-gixst]. In writing, the first consonant is written in the sixth order, the second one either in the sixth order or in any other order [58, 61].

Whatever the word pattern is CVCixCV/C, the vowel /እ/ [ix] in this pattern is eliminated. As a result of it, a noun such as /መቅደስ/ can be read as only [meqdes] and not [meqedes]. The syallable pattern mentioned above is also helpful in reading words that have consonants in the sixth order. Thus, for instance, /ድንግል/ can be read only as [dixngixl] as sown in the Figure 5.4.

The consonant clustering in medial position is valid only for nouns. For verbs the clustering depends on the form of the verb or of the verbal form. For instance, /ይስማል/ „he kisses" is to be read [yixsixmal] and not as [yixsmal]. A consonant in the sixth order preceding a geminated consonant is to be followed by /እ/ [ix] such as the word /ብብት/ is read as [bixbbixt].

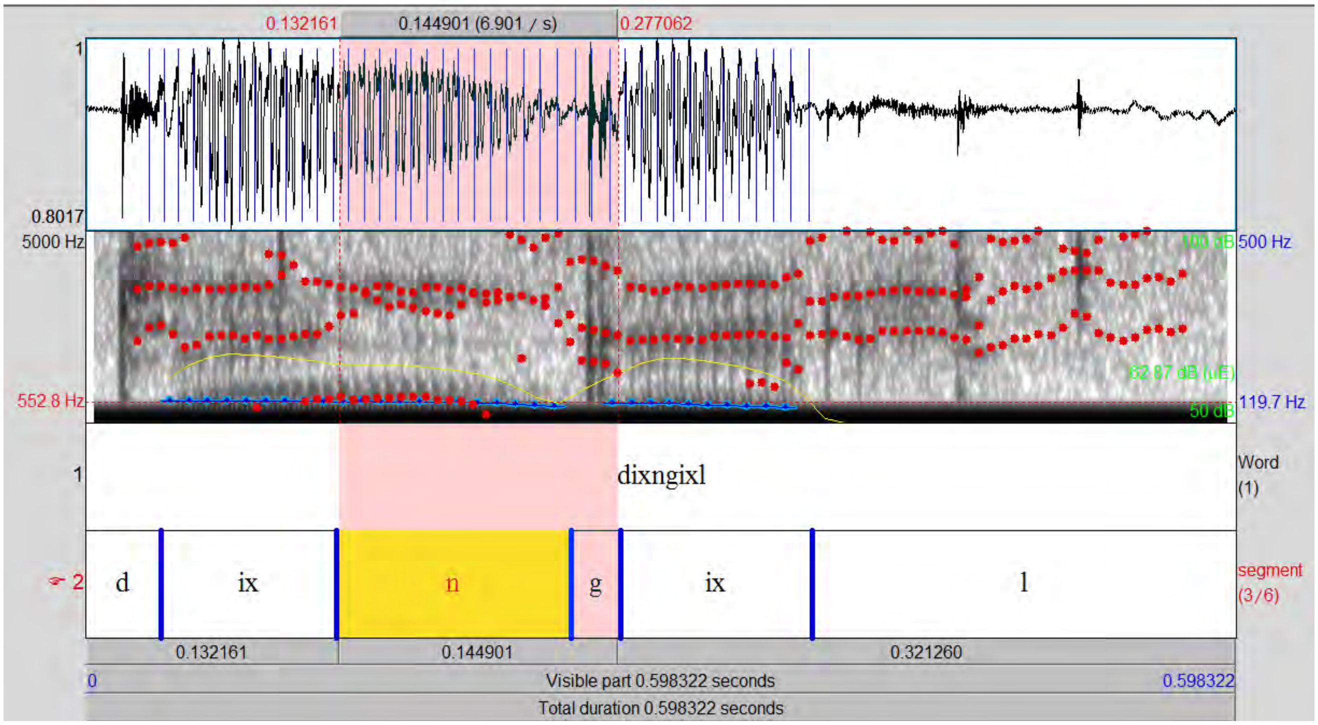


Figure 5.4: Medial consonant clusters of a word /ድንግል/ [d ix n g ix l]

5.3.3.2 Final consonant clusters

Amharic has final clusters of two consonants in the verbal forms regardless of the nature of the consonants. For example: /ሊለብስ/ read as [liilebs] as shown in Figure 5.5. For the nouns, the occurrence or non- occurrence of final consonant clusters depend on the nature of the consonants and the types of nouns. However, if a noun is written with two final consonants in the sixth order, the last one being /ት/ [t], there is a final cluster. So, the nouns /ጥቅምት/, /ንግስት/ and /ከብት/ are read as [txixqixmt], [nixgixst] and [kebt] respectively. In addition, the final consonant cluster occurs if and only if when the sonority of the first consonant must be higher than the last consonant.

As a rule, if a noun ends in three consonants written in the sixth order, the last consonant being /ት/ [t], C6C6t6 has to be read as [CixCt] [58]. Thus, /መንግስት/ has to be read as [mengixst] shown in Figure 5.6.

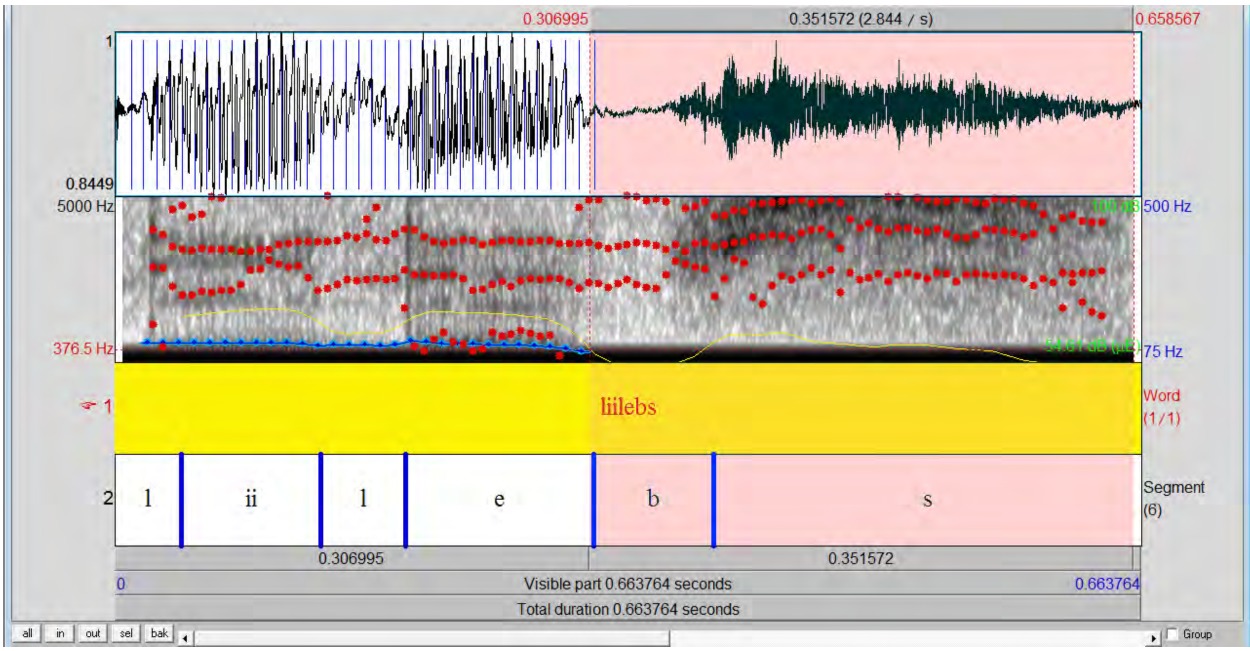


Figure 5.5: Final consonant clusters of a word /ሊለብስ/ [l ii l e b s]

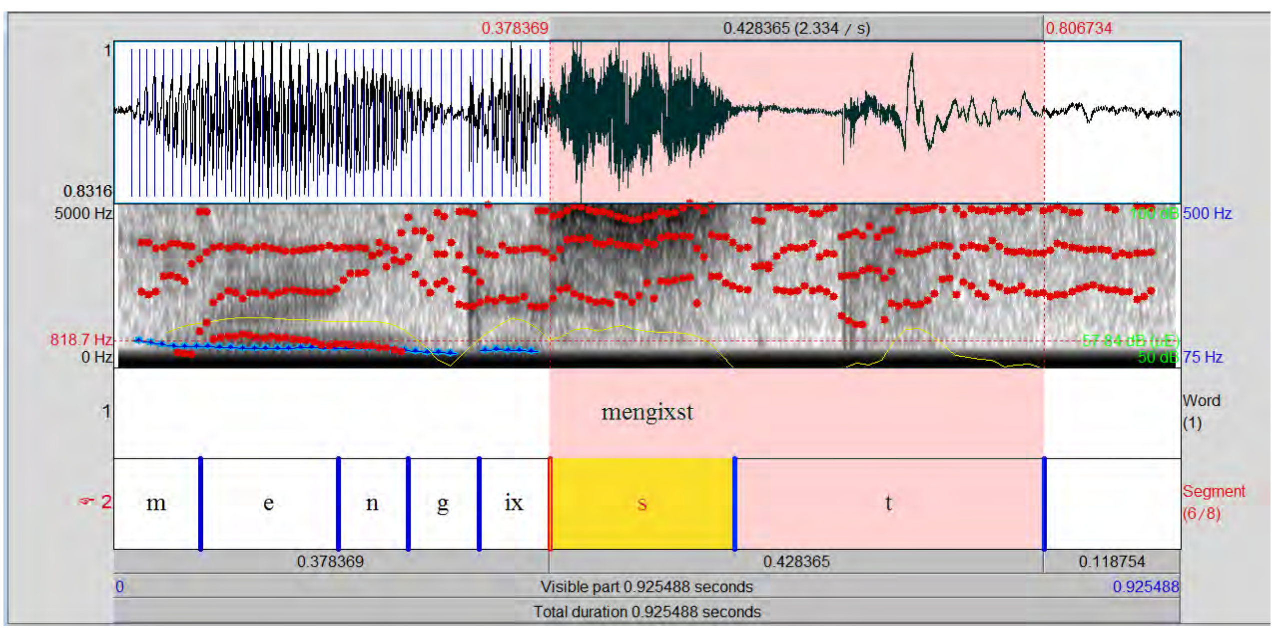


Figure 5.6: Final consonant clusters of a word /መንግስት/ [m e n g ix s t]

5.3.4 Stress

Amharic stress has many representations, and many things can be said about it, but the first thing to say is that "it" does not really exist [62].

Amharic stress does not exist in phonemically spoken. It does not create lexical contrasts, in spite of such interesting differences as between *be'KKele* "Bekele (name)" with stress on the second syllable - and *'beKKele* "it sprouted" with stress on the first [62].

Amharic stress is hardly visible on the phonetic level. When a list of words is read, every item will take word final and demarcative pitch/accent - as noted by Alemayehu Haile and Tsegaye Woldeyesus. In other studies, Amharic stress is characterized by statements about its being instable [63], its being difficult to detect [64], playing a lesser role [65], and being relatively weak [66].

Because of this, and because of further complications due to the syllabic script and its ambiguities of syllabicity, the term "stress" is sometimes employed as meaning nothing more than "this fidel is syllabic" (i.e., it is pronounced with both consonant and vowel) [62].

5.3.5 Phoneme length

The length of the phonemes can be described in terms of the time duration when the phones are produced. The length may be short or long. All Amharic consonants occur short (indicated by a single symbol; /ር/ [r], /ም/ [m], /ል/ [l] and etc) and long (indicated by doubling symbols; /ርር/ [rr], /ምም/ [mm], /ልል/ [ll] and etc) [67].

Amharic does not have the feature of vowel length for expressing lexical and grammatical meanings. Vowel length does not occur phonetically in Amharic language. However, a vowel can be lengthened in a normal speech for expressing emotional feelings [60]. This implies that, vowel length of Amharic language is not phonemic [68]. It is simply lengthened in the context when there is the influence of coarticulation and assimilation from the adjacent phonemes.

Chapter Six

CONTEXT SPECIFICATION AND ARCHITECTURE OF AMHARIC VOWEL CHARACTERIZATION

6.1 Introduction

This chapter demonstrates the context specification to prepare the text corpus used in the characterization of Amharic vowel sound units in the first section and the general architecture of this characterization in the second section.

6.2 Context specification for vowel characterization

Context specification is a process of choosing or specifying the contexts (consonants) that are relevant to accomplish the vowel sound unit characterization. In Amharic language, there are no well defined contexts to perform this kind of research. So, before going to the development and implementation of the architecture of Amharic vowel sound unit characterization, the first activity should be selecting the best contexts which make the characterization appropriate.

Under this section, the context selection rule and regulation as well as the preparation of text corpus based on the selected contexts, and their respective speech corpus and the listening test conducted on the speech files are described.

6.2.1 Corpus preparation

The main purpose of a corpus preparation is to verify a hypothesis about the language. Both text and speech corpora were prepared for this work as explained in the following sections.

6.2.1.1 Text corpus preparation

Before proceeding to the preparation of text corpus, the contexts (consonants) which are relevant to make a text corpus should be selected. In order to select these contexts, we considered the following factors:

- Consonant coarticulation effect on the vowels
- Segmentation challenges

- Manner and place of articulation of consonants

Consonants may have a higher or lower coarticulation effect on the adjacent vowels. The higher coarticulation effect comes due to voiced consonants [38]. If the vowels are connected to the voiced consonants, there will be a change in formant frequencies due to coarticulation influence [23]. When coarticulation effect is high, there is a difficult situation to decide the steady state time and segmentation boundaries of the vowels [49]. In contrast, voiceless consonants have less coarticulation effect on the adjacent vowels because the formant frequency values for voiceless consonants are not well defined [18, 23]. So, it is easy to distinguish the vowels and their segmentation boundaries and the steady state time within voiceless consonants than voiced consonants.

The second factor used to select the consonants is segmentation challenges of vowels from the contexts. Segmentation challenge depends on the properties of consonants connected to the vowels [49]. If the consonants are voiced, there is a high coarticulation effect. This implies that it is difficult to state the segmentation boundaries when segmenting the required vowels from such voiced contexts. On the other hand, if the consonants are voiceless, there is a lower coarticulation effect on adjacent vowels. This makes the identification of vowel segmentation boundaries easy in these contexts.

The third factor considered in context specification is the distribution of contexts or consonants with their place and manner of articulation. This factor is used in order to choose the best contexts from the selected contexts for the purpose of making fair context distribution with respect to their manner and place of articulations.

Based on context selection rule and regulation stated above, we carry out a preliminary context specification experimentation in order to choose and select appropriate contexts used to conduct the characterization of Amharic vowel sound units. This experiment is conducted on vowel /ኣ/ [a] within different consonants. We collect 100 words that contain vowel /ኣ/ [a] in both written and spoken form. Then we perform the experiment on the acoustic data of words to see the coarticulation effect of contexts and the level of challenge when segmenting the target vowel /ኣ/ [a] from its contexts. The experiment is conducted through segmenting (labeling) these acoustic data to analyze both the segmentation challenge and influence of coarticulation. If the vowel /ኣ/ [a] in a given context is segmented easily, the contexts have less coarticulation influence and segmentation challenging. Otherwise, difficulty of identification of segmentation boundary indicates complexity of coarticulation effect and inappropriateness of the contexts.

The combination of selected contexts and vowels is a CVC syallable structure to make the text corpus, where C stands for consonants and V stands for the target vowel. According to the assessment made, we found that the first consonants (onset) should be voiceless and glottalized contexts and the last consonants (coda) can be voiced, voiceless and glottalized contexts which are selected from the assessment. The contexts used in this work as per the preliminary observation are given in Table 6.1.

Table 6.1: The possible combination of contexts to make CVC syallable structure data

No.	By Manner of articulation		By Place of articulation		By Consonants	
	Left context	Right context	Left context	Right context	Left	Right
1	Voiceless stop	Voiced fricative	Alveolar	Alveolar	[t]	[z]
2	Voiceless stop	Voiced liquid	Alveolar	Alveolar	[t]	[r]
3	Voiceless fricative	Voiced stop	Labial	Velar	[f]	[g]
4	Voiceless fricative	Voiced liquid	Labial	Alveolar	[f]	[r]
5	Voiceless stop	Glottalized fricative	Labial	Alveolar	[p]	[xx]
6	Glottalized stop	Voiceless fricative	Alveolar	Labial	[tx]	[f]
7	Glottalized stop	Voiced fricative	Alveolar	Alveolar	[tx]	[z]
8	Voiceless fricative	Voiceless stop	Alveolar	Labial	[s]	[p]
9	Glottalized fricative	Voiceless stop	Alveolar	Velar	[xx]	[g]
10	Glottalized fricative	Voiceless stop	Alveolar	Labial	[xx]	[p]

As shown in Table 6.1, 10 different contexts were considered and in each context, only one onset and one coda are used (see the last two columns). Each context is applied to the seven Amharic vowels. Hence we have 70 CVC formatted corpus. In addition, the seven vowels are also recorded separately without any context that makes the total distinct working text corpus size is 77.

The collected text corpus based on the selected contexts and the CVC syllable structure combination stated above is given in Table 6.2. In the table v stands for the target vowel.

Table 6.2: The collected text corpus for the research

		Vowels						
No.	**Contexts**	**ኧ [e]**	**ኡ [u]**	**ኢ [ii]**	**ኣ [a]**	**ኤ [ie]**	**እ [ix]**	**ኦ [o]**
1	No context	ኧ	ኡ	ኢ	ኣ	ኤ	እ	ኦ
2	[t] v [z]	ተዝ	ቱዝ	ቲዝ	ታዝ	ቴዝ	ትዝ	ቶዝ
3	[t] v [r]	ተር	ቱር	ቲር	ታር	ቴር	ትር	ቶር
4	[f] v [g]	ፈግ	ፉግ	ፊግ	ፋግ	ፌግ	ፍግ	ፎግ
5	[f] v [r]	ፈር	ፉር	ፊር	ፋር	ፌር	ፍር	ፎር
6	[p] v [xx]	ፐጽ	ፑጽ	ፒጽ	ፓጽ	ፔጽ	ፕጽ	ፖጽ
7	[tx] v [f]	ጠፍ	ጡፍ	ጢፍ	ጣፍ	ጤፍ	ጥፍ	ጦፍ
8	[tx] v [z]	ጠዝ	ጡዝ	ጢዝ	ጣዝ	ጤዝ	ጥዝ	ጦዝ
9	[s] v [p]	ሰፕ	ሱፕ	ሲፕ	ሳፕ	ሴፕ	ስፕ	ሶፕ
10	[xx] v [g]	ጸግ	ጹግ	ጺግ	ጻግ	ጼግ	ጽግ	ጾግ
11	[xx] v [p]	ጸፕ	ጹፕ	ጺፕ	ጻፕ	ጼፕ	ጽፕ	ጾፕ

6.2.1.2 Speech corpus preparation

The 77 text corpus prepared with and without context is recorded to produce speech corpus for the characterization of the distinctive vowels. Recording is made using a creative as well as Skype certified microphone at sampling rate of 22,050 Hz and sample size of 16 bits and with two different speaking rates: fast and normal.

Fast speaking rate is defined as the fastest tempo possible without making an excessive number of errors and the normal rate is defined as a tempo that is relaxed and comfortable for the speaker [69]. The speaker records the CVC contexts and unique vowels in a fast rate in order to analyze the effect of duration on formant frequency values through comparing and contrasting with the similar contexts recording within normal speaking rate. On the other hand, the CVC contexts and unique vowels are recorded in normal speaking rate mainly used to characterize the Amharic vowel sound units.

The speech corpus is prepared from a single speaker. The speaker recorded each of the sample text corpus 20 times where 10 of them are at fast speaking rate and the rest with normal speaking rate.

6.2.2 Listening test

Listening test is the way of perceiving the CVC and vowel signals recorded by a speaker. The main aim of listening test was to obtain aural classification or perception of each signal to supplement the speaker‟s classification or perception as the CVC is not taken arbitrary to fit the required context. Before going to the preparation of labeled speech corpus, we perform the listening test through listeners. Five fluent speakers were participated in the listening test. Two of them are females and the rest are male listeners. The number of utterances presented to the listeners is 10*77 which are recorded at normal speaking rate. The way of presenting the utterances to the listener is in random manner to minimize successful guessing. The listing test shows the speaker identified all the speech utterances without failure (100% accuracy).

6.3 Architecture of Amharic vowel sound units characterization

Architecture is the graphical representation of procedures and processes used to conduct the expermentation of the thesis. The Architecture of Amharic vowel sound units characterization looks like the one shown in Figure 6.1.

The architecture development for Amharic vowel sound unit characterization is based on the speaking rates consideration used in this work. As we disscused in section 6.2.1.2, the normal and fast speaking rates are used in this research. The contexts spoken at normal rate are used for characterization purpose while the contexts spoken at fast rate are used to investigate the effect of duration on formant frequency values. Therefore, we should consider the speaking rate when we develop the architecture because the investigation of contexts spoken at the two speaking rates is conducted in a separate manner as well as the purpose of the contexts spoken at the two rates are different.

The architecture contains basic components such as segmentation, feature extraction, aggregate feature extractor (context dependent), duration effect on formant frequency values analyzer, aggregate feature extractor (context independent), duration and fundemental frequency extractor, acoustic vowel spaces analyzer and formant frequency variability analyzer.

The aggregate feature extractor (context dependent) component investigates the sample vowel features at normal rate and sample vowel fearures at fast rate independently. However, aggregate feature extractor (context independent), duration and fundemental frequency extractor, acoustic vowel spaces analyzer and formant frequency values varability analyzer are components applied on contexts acoustic data parameters spoken at normal rate whereas duration effect on formant frequency values analyzer focuses on contexts acoustic data parameters spoken at both normal and fast rates. The description of the individual components of the archiecture is given in the subsequent section.

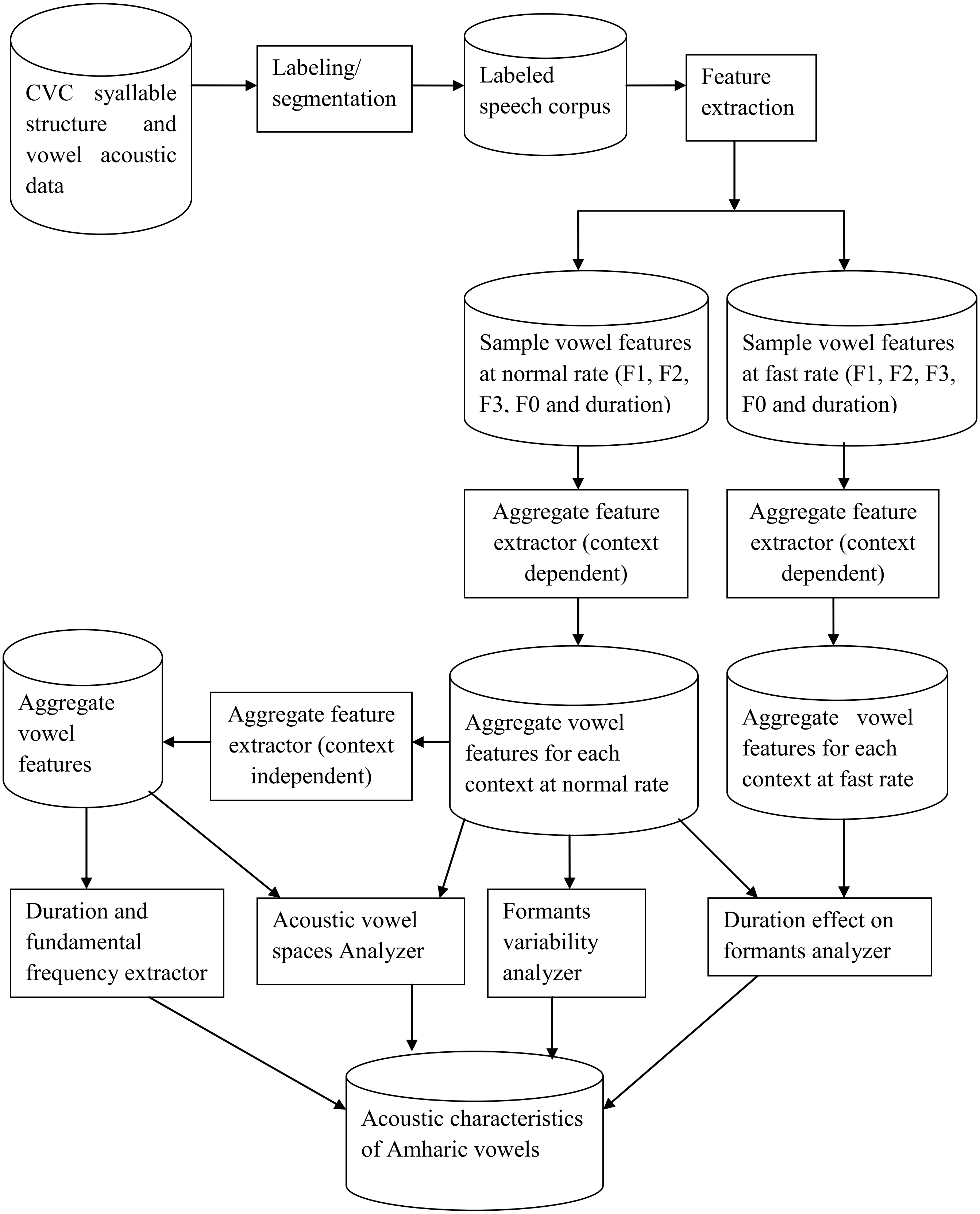


Figure 6.1: General Architecture of Amharic vowel sound units characterization

6.3.1 Labeling / Segmentation

Once the speech files (acoustic data) were prepared and the listening test is completed, the next step is segmenting or labeling the speech files to get a labeled speech corpus. To make easy the segmentation of speech files, we used the text corpus transcription to identify phones found in the speech files (Appendix B), a high resolution gray scale spectrogram is used as a main segmentation environment and waveform is used as a segmentation corrective environment to increase the reliability of the labeling. In addition to these environments, the criteria which are stated in the literature review of this work were used to perform the segmentation.

6.3.2 Feature extraction/ Acoustic Measurement

Acoustic measurement is the process of extracting the features or parameters of a target vowel sound units from the given labeled speech corpus. The acoustic measurement is done for both static and dynamic characteristics of vowels. The static characteristics of vowels are formant frequencies (F1, F2 and F3) and fundamental frequency (F0) which are taken at steady state time where as vowel duration is the dynamic characteristic of vowels.

Vowel duration is calculated as a time taken by vowel from initial boundary up to the final boundary in the labeled speech files. It was measured by hand from a high resolution gray scale digital spectrograms and waveform description of vowel sound units using measurement criteria presented in [31]. After conducting the duration measurement for the target vowel, the next step was estimating the steady state time or period. Steady state time is a stable condition that the characteristics of vowels do not change over time. It was measured as the central point of the vowel duration. Meaning, it is half of the total duration of the vowel. The main reason in which this measurement is conducted was that the characteristics of vowels do not vary at this time and to minimize (avoid) the influence of the preceding and following consonants on the adjacent vowel. The measurement of duration and steady state time were made while viewing waveform and gray scale spectrogram description of the vowel sound units like shown in Figure 6.2 and Figure 6.3 respectively. Figure 6.2 shows the duration of vowel /ኤ/ [ie] within /ጼፕ/ context, it is measured from the initial boundary near to /ጽ/ [xx] up to the final boundary near to /ፕ/ [p] and the value is 0.35595 seconds (356 ms). Figure 6.3 indicates the steady state time of vowel /ኤ/ [ie] within /ጼፕ/ [xx ie p] context. The value is taken from the mid of the total vowel duration which is 0.35595/2= 0.177975 seconds.

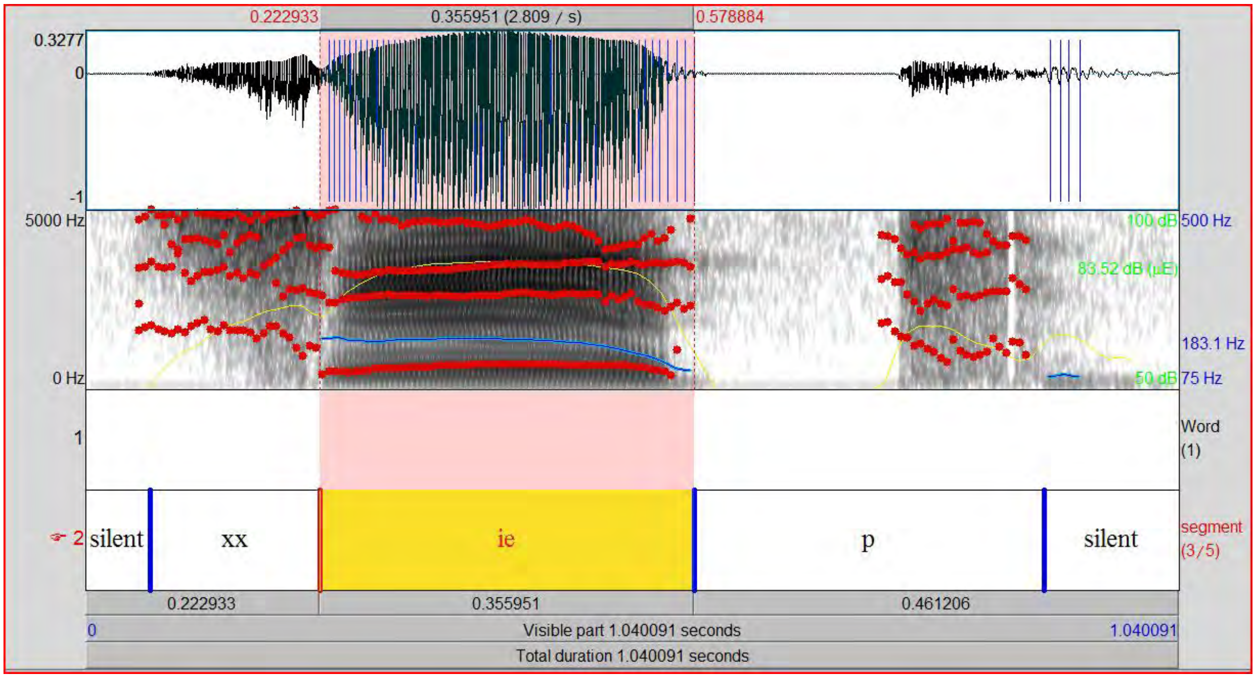


Figure 6.2: Duration measurement of vowel /ኤ/ [ie] in /ዤፕ/ [xx ie p] context

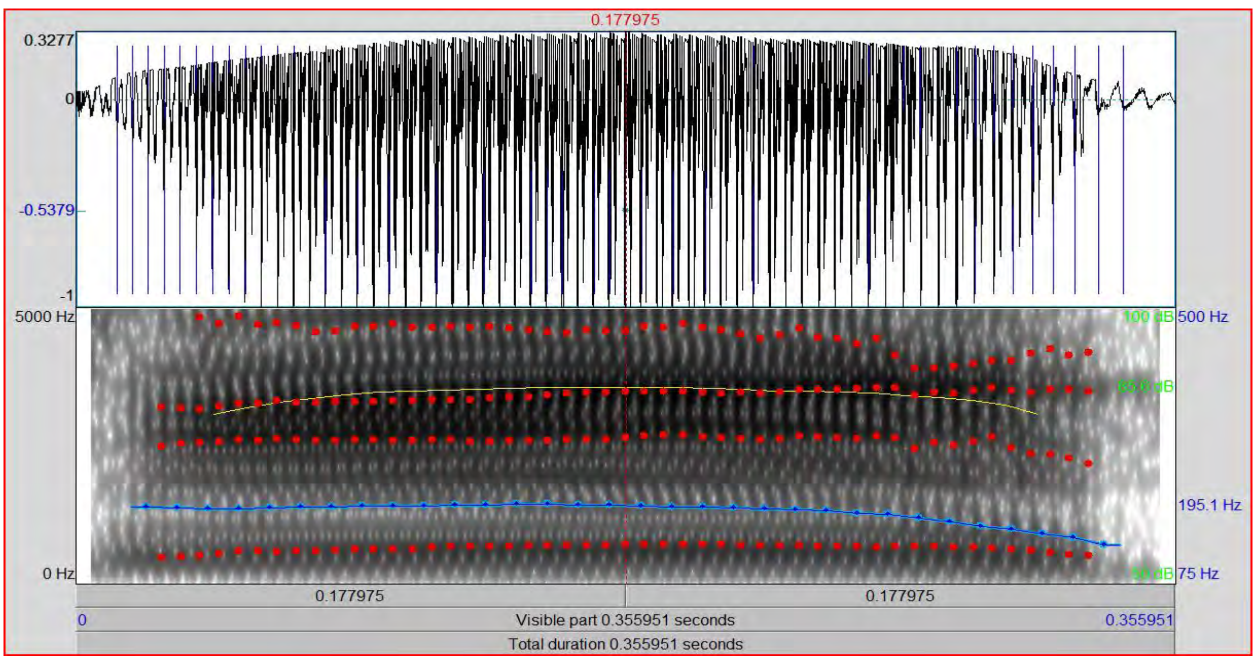


Figure 6.3: Steady state time measurement of vowel /ኤ/ [ie] in /ዤፕ/ [xx ie p] context

The most important parameters which are used to characterize the vowel sound units are formant frequencies. These frequencies are measured by using a modified LPC model called a bank of first order LPC tracking filters with 20-ms Hamming window based on [70]. This method is based on pre-filtering speech using a time-varying adaptive filter for each formant before spectral peak estimation. This spectral peak peaking algorithm is embedded on the speech analyzer tool. Therefore, formants are extracted using a speech analyzer tool. The first three formant frequencies are extracted for vowel characterization purpose. Because, F1 corresponds to the tongue height, F2 corresponds to the tongue length and the graphical representation of F1 vs. F2 and F2 vs. F1 corresponds to the position of the tongue shape. Formant extraction is conducted on the steady state time of the vowel. For instance, the steady state time of vowel /ኢ/ [ii] is shown in Figure 6.4 as a green line over the waveform and spectrogram description of the vowel and formant frequencies are extracted at this time. As shown in Figure 6.4, the first three formant frequency values of vowel /ኢ/ [ii] in /ጢዝ/ [tx ii z] context were found at the right button of the figure. The values are F1= 436 Hz, F2= 1958 Hz and F3= 3189 Hz.

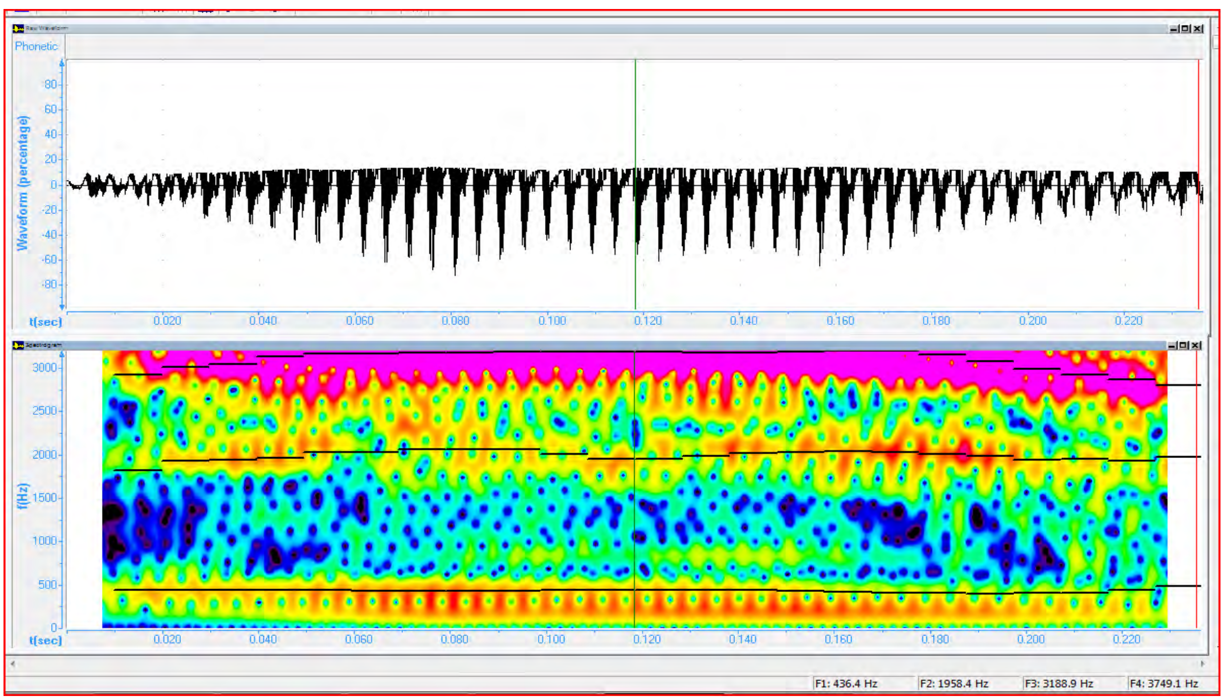


Figure 6.4: Formant frequency extraction of vowel /ኢ/ [ii] in /ጢዝ/ [tx ii z] context using speech analyzer tool

In addition to duration, steady state period and formant frequencies, we also estimate and extract the fundamental frequency (pitch) of the vowel sound units. It was extracted from a steady state period of the vowel signals using speech analyzer tool. It can be computed using the average magnitude difference function (AMF), which is a measure of how well a signal matches a time-shifted version of itself, as a function of the amount of time shift. The match is calculated using subtraction (instead of multiplication, as in the autocorrelation), which reduces computational time.

Based on the feature extraction methods stated above, this component can extract both the static and dynamic features of vowels spoken at both speaking rates and produce the sample vowel features at normal rate which contains vowel features extracted from CVC contexts and vowels that are spoken at normal rate and sample vowel features at fast rate which contains vowel features extracted from CVC contexts and vowels that are spoken at fast speaking rate, independently.

6.3.3 Aggregate feature extractors

As pointed out on corpus preparation, there are 77 text corpus contexts for the seven Amharic vowels. This implies that each vowel contains 11 different contexts. Similarly, the contexts acoustic data spoken at normal rate for each vowel is 11 contexts * 10 times = 110 utterances. As well as contexts acoustic data spoken at fast rate for individual vowel is also 11 contexts * 10 times =110 utterances. Here each context records 10 times. Therefore, the aggregate feature extractor (context dependent) use the descriptive statistics such as average and standard deviation to apply on each sample vowel features at normal and at fast rates extracted from similar context utterances separately. In simple language, the aggregate feature extractor (context dependent) components accepts as an input the 10 repeated utterances of a single context and then calculate the average and standard deviation of the utterances for the vowel connected to that context. This procedure works for both contexts spoken at normal and fast speaking rates as well as it also works for all vowels in order to produce the aggregate vowel features for each context at normal and at fast rate independently.

In aggregate feature extractor (context independent), the extractor takes the aggregate vowel features for each context at normal rate as an input and conducts other extraction on them. This extractor uses the descriptive statistics to obtain the average formant frequency values, duration and fundamental frequency of vowels extracted from the ten CVC contexts and one unique vowel in normal rate to come up with the final and singe features value for each vowel.

6.3.4 Duration and fundamental frequency extractor

This component can retrieve the duration and fundamental frequency characteristics of the seven Amharic vowel sound units. This extraction is applied after performing the aggregate feature extractor (context independent) on the aggregate features of vowels for each context at normal rate. Then duration and fundamental frequency features or characteristics of vowels are extracted from the aggregate vowel features.

6.3.5 Acoustic vowel spaces Analysis

Acoustic vowel space is the graphical representation of the formant frequency values of vowels on different planes. However, the acoustic space used to characterize the vowels is represented using either F1 vs. F2 or F2 vs. F1 planes. The basic inputs of the planes are the first and the second formant frequencies values of vowels. Acoustic vowel space can represent the shape of the tongue acoustically. So, to characterize Amharic vowel sound units acoustically, we develop acoustic vowel spaces using the sample vowel features obtained at normal rate, aggregate vowel features for each context at normal rate and aggregate vowel features to reach with the acoustic characteristics of vowel sound units.

The acoustic vowel space component consists of three different spaces. The first acoustic space takes the sample vowel features at normal rate and display all vowels extracted from 770 utterances on F1 vs. F2 plane. On this acoustic space, all of the individual utterances of vowels are distributed on the space. It is used to observe the vowel distribution on the space and the space provided for the vowels. The second space is developed for the aggregate vowel features for each context at normal rate. It contains 77 vowels extracted from 770 utterances through the aggregate feature extractor (context dependent) component and displayed on F1 vs. F2 space.

The third acoustic vowel space is developed only for the seven Amharic vowels on F2 vs. F1 plane. The inputs of this space are the aggregate features of vowels which are the average formant frequency values of vowel extract from eleven contexts by applying the aggregate feature extractor (context independent) on aggregate vowel features for each context at normal rate. This space is called acoustic vowel space of Amharic language. It shows the acoustic characteristics of Amharic vowel sound units implicitly.

6.3.6 Duration effect on formant frequency values analyzer

As stated in the literature review under formant based characterization, factors which affect formant frequency values are linguistics and non- linguistics factors. Duration is one of the linguistic factors that affect the formant frequency values of vowel sound units. The formant frequency values of vowels either increase or decrease due to duration when the speaker records the contexts with various speaking rates.

This component investigate the effect of duration on formant frequency values of vowels by considering the two duration variations such as long and short duration which come due to the normal and fast speaking rates of the speaker, respectively.

The analyzer accepts the aggregate vowel features for each context at normal rate and aggregate vowel features for each context at fast rate as an input for both duration levels such as long and short durations respectively, then compare and contrast the vowel formant frequencies obtained from the two duration levels. Finally, the analyzer gives the relationship between duration and formant frequency values as an output.

6.3.7 Formant frequency variability analyzer

Formant frequency variability analyzer performs two separate analyses on sample vowel features at normal rate and aggregate vowel features for each context at normal rate. These are formant frequency variability within similar CVC context utterances and within different contexts (CVC contexts and unique vowel).

To conduct the formant frequency variability within similar CVC context utterances, consider one vowel and one CVC context utterances for that vowel from the sample vowel features at normal rate component, since one CVC context has ten repetitions of similar utterances, the main intention was to see the formant frequency variability of selected vowel within selected CVC context utterances. Then compute the mean and standard deviation values for the first three formant frequency values individually. Similarly, the formant frequency variability within different contexts can be conducted by using the aggregate vowel features for each context at normal rate. Aggregate vowel features for each context at normal rate consists of the mean and standard deviations of the first three formant frequency values individually for all CVC contexts and unique vowel separately.

The analysis of formant frequency variability within similar CVC context utterances as well as within different contexts is done following the same procedures. To both analyses, based on the average and standard deviation values of the formant frequencies, we compute the formant frequency distribution classes known as normal, above normal and below normal. Normal distribution is the class when formant frequency values lie between [(mean value - standard deviation) and (mean value + standard deviation)] while above normal distribution is the formant values greater than [mean value + standard deviation] and below normal distribution is the formant values less than [mean value – standard deviation].

To the formant variation within similar CVC context utterances, after the formant frequency distribution classes are identified, the next procedure is analyzing the number of vowel utterances found in the three formant distribution classes and the respective utterances percentage value computed from the total number of 10 utterances of a single context considered for each vowel.

To the formant variation within different contexts, after the formant frequency distribution classes are identified, the next procedure is analyzing the number of vowel instances found in the three formant distribution classes and the respective instance percentage value computed from the total number of 10 CVC and 1 vowel contexts considered for each vowel.

Finally, the CVC utterances as well as CVC contexts and unique vowel found in the above and below normal distribution classes contribute to the variation of formant frequency values of vowels for similar CVC context utterances and different contexts analyses respectively. This component analyses and identifies the factors which affect the formant frequency variability within similar CVC context utterances and within different contexts as an output.

Chapter Seven

EXPERIMENTATION AND DISCUSSION

7.1 Introduction

Experimentation and discussion focuses on the result of the experiment and analysis of the result. The basic procedures used to conduct the experimentation are context specification, corpus preparation, labeling/ segmentation, feature extraction and aggregate feature extractors. A number of aggregate feature extractors were applied on sample vowel features to achieve the acoustic characteristics of vowels.

This chapter discusses the acoustic characteristics of Amharic vowel sound units such as duration, fundamental frequency and formant frequencies acquired from the experiment. It also explains about the first three formant frequency ranges of vowels, effect of duration on formant frequency values, acoustic vowel space of Amharic language and formant frequency variability of vowel sound units occurred due to the loudness/ intensity variation and adjacent contexts.

7.2 Acoustic measurement result and analysis

The averages of acoustic measurements shown in the tables, and the data displayed in the subsequent figures, are based on the measurements from individual samples. The aggregate features of the seven Amharic vowels from CVC signals and unique vowels are shown in Table 7.1.

Table 7.1: The aggregate features of the seven Amharic vowels

No.	Vowel	F1 (Hz)	F2 (Hz)	F3 (Hz)	F0 (Hz)	Duration (ms)
1	[e]	717	1441	2681	195	393
2	[u]	523	1297	2548	209	329
3	[ii]	443	2005	3102	207	225
4	[a]	804	1426	2705	173	375
5	[ie]	598	1972	2821	196	361
6	[ix]	536	1554	2648	197	223
7	[o]	624	1378	2740	203	366

7.2.1 Vowel duration and fundamental frequency

Table 7.2 shows the average duration of each of the seven vowels in the two different types speaking rate. The table shows that fast speaking rate duration of all vowels are shorter than the corresponding duration in the normal speaking rate.

Table 7.2: The average duration of the seven Amharic vowels at normal and fast speaking rates

Number	Vowel	Average duration (ms)		
		Fast	Normal	Average
1	[e]	217	393	305
2	[u]	201	329	265
3	[ii]	215	225	220
4	[a]	256	375	316
5	[ie]	252	361	307
6	[ix]	214	223	219
7	[o]	245	366	306

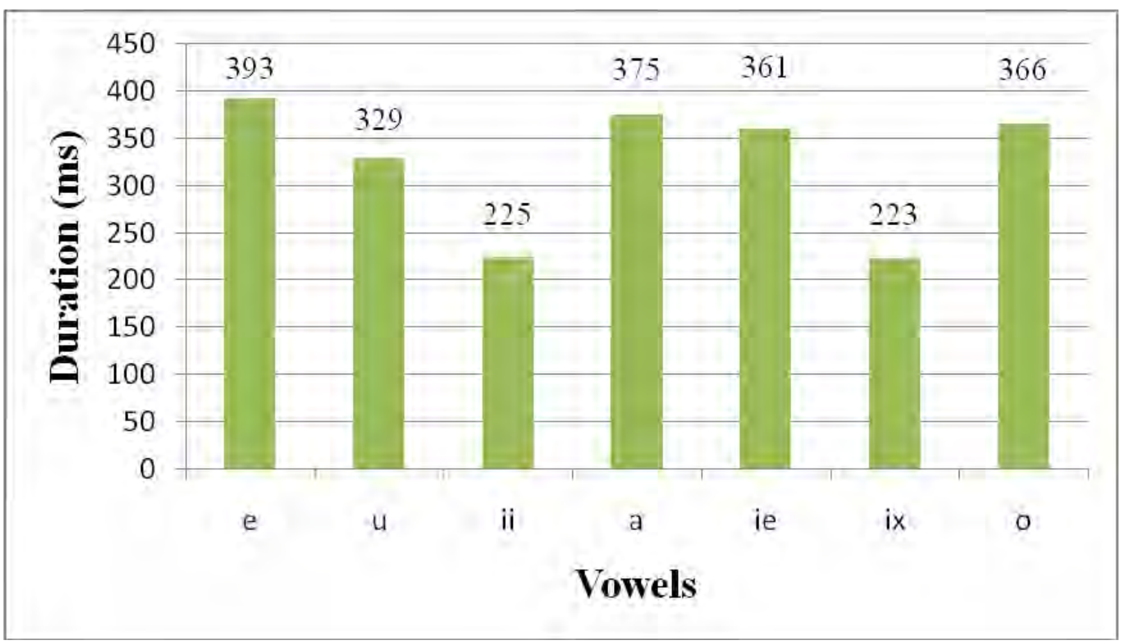


Figure 7.1: The average duration of the seven Amharic vowels

Figure 7.1 shows that the duration variation of the seven Amharic vowels within normal speaking rate. The measurement and aggregate feature extractor results show that vowel /ኧ/ [e] is the longest of all, vowel /አ/ [a] is longer than the rest five vowels and /ኦ/ [o] and /ኤ/ [ie] are long vowels. In contradictory, vowel /እ/ [ix] is the shortest of all, vowel /ኢ/ [ii] is shorter than the others and vowel /ኡ/ [u] is the short individual.

The fundamental frequency values of the seven Amharic vowels that are extracted from the aggregate vowel features were shown in Table 7.3. As shown in Table 7.3, the fundamental frequency values of all vowels were near to each other. This is true because F0 is a factor which depends on only the rate of vibration of the vocal folds of the speaker.

Table 7.3: The average fundamental frequency values of the seven Amharic vowels

Number	Vowel	Fundamental frequency (Hz)
1	[e]	195
2	[u]	209
3	[ii]	207
4	[a]	173
5	[ie]	196
6	[ix]	197
7	[o]	203

7.2.2 Formant Frequency

From acoustic measurement result (sample vowel features at normal rate), we filtered out the formant frequency values of the vowels and draw on F1 vs. F2 plane to see the distribution of vowels on the space as shown in Figure 7.2. Figure 7.2 shows the individual data points for 770 utterances on the vowel space. It shows that vowel /ኢ/ [ii] and /ኤ/ [ie] were not overlapped to each other as well as to other vowels. The overlaps occurred between other vowels such as between vowel /ኡ/ [u] and /እ/ [ix], /ኡ/ [u] and /ኦ/ [o], /ኧ/ [e] and /ኦ/ [o] and /ኧ/ [e] and /አ/ [a]. These all overlaps occurred between adjacent vowels which have similar behavior. Even though the vowels are overlaps to each other, they have their own central tendency formant frequency value and specific space on the plane which is a dense area on Figure 7.2.

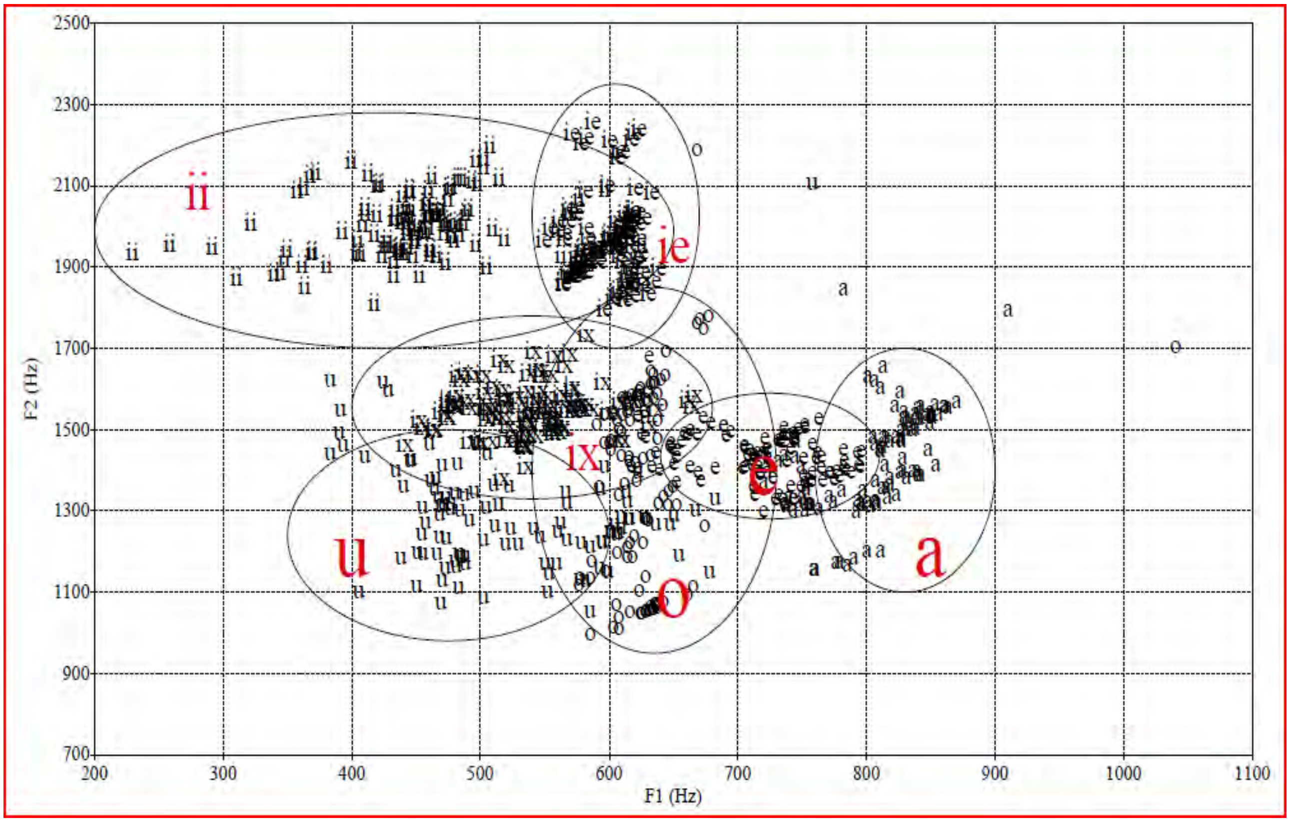


Figure 7.2: F1 vs. F2 space for 770 individual data points or utterances

After getting the measurement values of formant frequencies, fundamental frequency and duration of the total 770 utterances at normal rate, the aggregate feature extractor (context dependent) was applied to get the aggregate vowel features for each context at normal rate in CVC syallable structure contexts and unique vowels. The aggregate vowel features for each context at normal rate were shown in Table 7.4. It shows that the average formant frequency, fundamental frequency and duration values of each vowel extracted from 11 contexts independently.

Table 7.4: Aggregate features of the seven Amharic vowels for each context at normal rate extracted from 77 contexts

No.	Contexts	Target vowel	F1 (Hz)	F2(Hz)	F3(Hz)	F0(Hz)	Duration(ms)
1	[e]	[e]	690	1490	2710	213	450
2	[xx e p]	[e]	716	1411	2716	202	401
3	[xx e g]	[e]	722	1465	2718	183	394
4	[tx e f]	[e]	772	1411	2705	197	386
5	[tx e z]	[e]	736	1483	2689	185	392
6	[s e p]	[e]	726	1352	2711	184	358
7	[p e xx]	[e]	733	1461	2645	183	334
8	[t e r]	[e]	700	1452	2666	193	380
9	[f e r]	[e]	744	1432	2649	204	428
10	[f e g]	[e]	712	1405	2641	198	430
11	[t e z]	[e]	632	1493	2642	199	368
12	[u]	[u]	473	1290	2602	226	449
13	[xx u p]	[u]	528	1258	2672	208	401
14	[xx u g]	[u]	528	1231	2662	193	379
15	[tx u f]	[u]	498	1393	2618	207	357
16	[tx u z]	[u]	575	1226	2517	202	399
17	[s u p]	[u]	612	1253	2306	222	199
18	[p u xx]	[u]	482	1291	2488	207	304
19	[t u r]	[u]	]482	1198	2385	228	327
20	[f u r]	[u]	604	1275	2573	202	307
21	[f u g]	[u]	492	1541	2821	191	178
22	[t u z]	[u]	481	1306	2374	217	321
23	[ii]	[ii]	419	2104	3235	185	338
24	[xx ii p]	[ii]	469	2038	3053	206	239
25	[xx ii g]	[ii]	458	2028	3032	205	204
26	[tx ii f]	[ii]	493	2113	3199	236	214
27	[tx ii z]	[ii]	442	1984	3205	211	247
28	[s ii p]	[ii]	406	1932	3068	210	209
29	[p ii xx]	[ii]	446	1996	3089	227	228
30	[t ii r]	[ii]	459	1998	2995	224	181
31	[f ii r]	[ii]	380	1904	3107	162	165
32	[f ii g]	[ii]	479	1966	3082	213	213
33	[t ii z]	[ii]	425	1990	3062	203	235

34	[a]	[a]	795	1309	2688	190	431
35	[xx a p]	[a]	753	1344	2822	186	335
36	[xx a g]	[a]	735	1401	2805	180	349
37	[tx a f]	[a]	772	1327	2820	179	378
38	[tx a z]	[a]	806	1410	2788	163	357
39	[s a p]	[a]	817	1408	2771	166	332
40	[p a xx]	[a]	818	1487	2491	165	369
41	[t a r]	[a]	826	1493	2796	166	393
42	[f a r]	[a]	845	1517	2453	170	391
43	[f a g]	[a]	837	1469	2511	169	381
44	[t a z]	[a]	836	1522	2803	169	407
45	[ie]	[ie]	600	2215	2804	194	416
46	[xx ie p]	[ie]	580	1946	2782	189	357
47	[xx ie g]	[ie]	565	1972	2818	184	364
48	[tx ie f]	[ie]	616	2025	2835	204	322
49	[tx ie z]	[ie]	583	1974	2848	192	377
50	[s ie p]	[ie]	574	1894	2787	189	327
51	[p ie xx]	[ie]	617	1979	2871	205	323
52	[t ie r]	[ie]	622	1915	2828	202	352
53	[f ie r]	[ie]	610	1917	2790	201	377
54	[f ie g]	[ie]	605	1957	2855	196	363
55	[t ie z]	[ie]	609	1895	2807	202	392
56	[ix]	[ix]	491	1554	2655	192	347
57	[xx ix p]	[ix]	557	1525	2686	192	235
58	[xx ix g]	[ix]	519	1580	2624	181	238
59	[tx ix f]	[ix]	545	1554	2714	192	204
60	[tx ix z]	[ix]	538	1537	2793	187	244
61	[s ix p]	[ix]	520	1531	2657	185	212
62	[p ix xx]	[ix]	541	1571	2699	189	183
63	[t ix r]	[ix]	571	1601	2632	220	187
64	[f ix r]	[ix]	560	1498	2563	199	187
65	[f ix g]	[ix]	550	1562	2495	216	213
66	[t ix z]	[ix]	510	1580	2608	211	202
67	[o]	[o]	653	1613	2611	215	436
68	[xx o p]	[o]	632	1158	2805	207	376
69	[xx o g]	[o]	624	1301	2815	204	392
70	[tx o f]	[o]	630	1379	2863	207	315

71	[tx o z]	[o]	627	1082	2876	204	367
72	[s o p]	[o]	623	1596	2907	200	350
73	[p o xx]	[o]	602	1440	2678	190	263
74	[t o r]	[o]	627	1383	2786	204	374
75	[f o r]	[o]	625	1502	2496	199	401
76	[f o g]	[o]	611	1451	2578	196	363
77	[t o z]	[o]	612	1254	2724	209	390

Formant frequency values were the main parameters to characterize vowels on the acoustic vowel spaces (on F1 vs. F2 plane or on F2 vs. F1 plane). The aggregate vowel features for each context at normal rate shown in Table 7.4 were plotted on the acoustic vowel space given in Figure 7.3. It shows that all of the seven vowels within eleven different contexts have a specific space on the plane. Even if there are different contexts for a single vowel, the acoustic space of the vowel lays in one ellipse because the formants are extracted from steady state time. Even though there is some effect on the vowels due to the contexts surrounding it, it does not affect the position of vowel on the space. In Figure 7.3, the blue colored vowel inside all ellipses indicates the position of the vowels without any context whereas the remaining black color vowels were extracted from different contexts. When we see the vowels position without context, vowels /ኢ/ [ii], /ኤ/ [ie] and /ኦ/ [o] are far from the central tendency position on the space. However, the other vowels like /ኡ/ [u], /እ/ [ix], /ኧ/ [e] and /አ/ [a] are near to the central tendency. Therefore, the acoustic characterization of vowels was not conducted without the context consideration. The reason is the characteristics of vowels obtained in the absence of context did not represent the characteristics of vowels in different contexts. This means, the properties of vowels vary with respect to the contexts surrounding it. So, during the characterization we should consider the contexts. For this reason, we use 10 different contexts in our research. As shown in Figure 7.3, the vowels within different contexts were somewhat perfectly distributed over the acoustic space. Vowel /ኢ/ [ii], /ኤ/ [ie] and /እ/ [ix] were not overlapped to other vowels as well as to each other. The rest vowels overlap to each other such as vowel /ኡ/ [u] with /እ/ [ix], vowel /ኦ/ [o] with /ኡ/ [u], vowel /ኧ/ [e] with /ኦ/ [o] and vowel /ኧ/ [e] with /አ/ [a]. These overlaps are expectable because overlapping is occurred between adjacent vowels because adjacent vowels have similar characteristics at least on one of the tongue position either tongue height or tongue length.

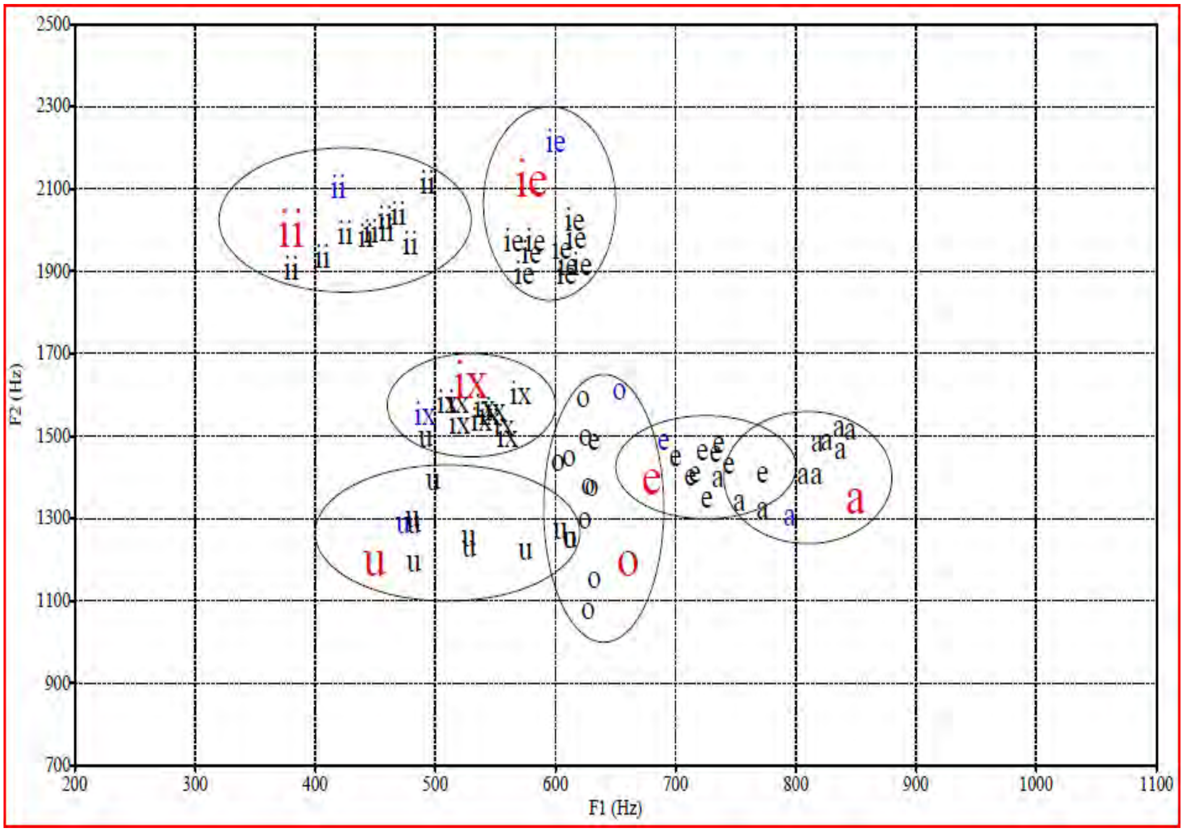


Figure 7.3: F1 vs. F2 space of the seven vowels extracted from 77 utterances

7.2.2.1 Formant Frequency Ranges

To give a clear explanation about the acoustic characteristics of Amharic vowels, we describe the formant frequency ranges of vowels shown in Figure 7.4. Figure 7.4 shows the formant frequency measured in terms of hertz in the y-axis and the vowels in the x-axis to indicate the vowels with their corresponding formant frequency ranges.

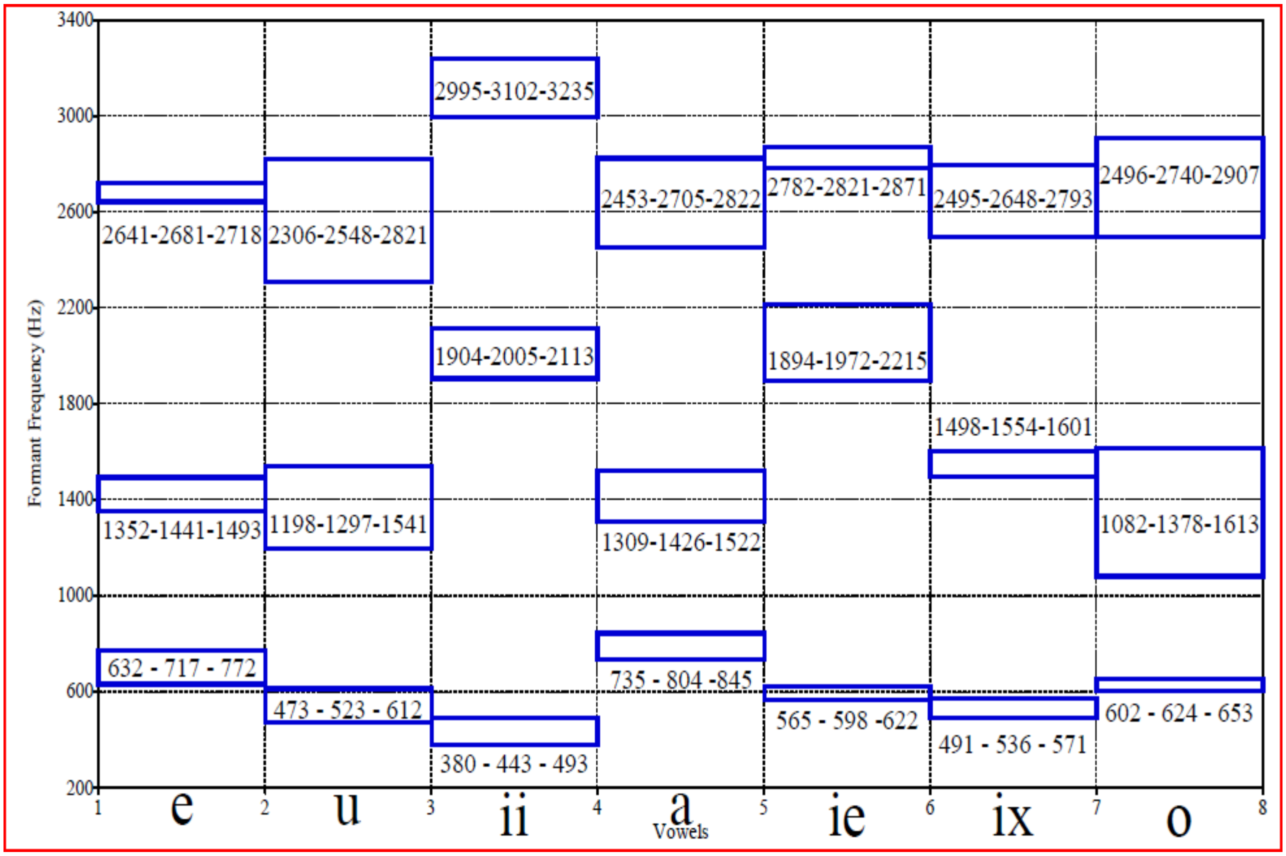


Figure 7.4: Formant frequency ranges of the seven Amharic vowels

Figure 7.4 indicates the first three formant frequency ranges for the seven Amharic vowels. The formant frequency values below 1000 Hz was the first formant frequency range whereas the formant frequency between 1000 Hz and 2220 Hz were the second formant frequency range and the formant frequency value above 2300 Hz was the third formant frequency range. In Figure 7.4, the blue colored rectangle shows the frequency ranges from minimum up to the maximum frequency. The label of the rectangle tells us the minimum, average and maximum formant frequency value of the vowels.

F1 range analysis showed that vowel /ኢ/ [ii] has minimum F1 value whereas vowel /አ/ [a] has a maximum F1 value. F1 values of the vowels were increased in the following manner: vowel /ኢ/ [ii], /ኡ/ [u], /እ/ [ix], /ኤ/ [ie], ኦ/o/, /ኧ/ [e] and /አ/ [a] respectively. The F1 range indicates that vowel /ኢ/ [ii] is a high vowel because it has the smallest F1 value. However, vowel /አ/ [a] is a low vowel because it has a largest F1 value. Similarly, vowel /ኡ/ [u] and /እ/ [ix] has smaller F1 value than /ኤ/ [ie], /ኧ/ [e] and /ኦ/ [o] vowels. This is due to the height of the tongue, /ኡ/ [u] and /እ/ [ix] are high vowels corresponds to minimum F1 values and /ኤ/ [ie], /ኧ/ [e] and /ኦ/ [o] are middle vowels corresponds to the moderate F1 values.

F2 range analysis indicates that front vowels have a maximum F2 value. However, back vowels have a minimum F2 values and middle vowels have medium F2 values [49]. So, Figure 7.4 tells us vowel /ኢ/ [ii] and /ኤ/ [ie] have high F2 value, vowel /ኦ/ [o] and /ኡ/ [u] have low F2 values and vowel /እ/ [ix], /ኧ/ [e] and /ኣ/ [a] have medium F2 values.

7.2.2.2 Effect of Duration on Formant frequency values

Based on the facts described on section 6.3.6, to conclude about the relationship between duration and formant frequency values, we conducted an experiment on all Amharic vowels. For instance, the result of evaluation for vowel /ኤ/ [ie] is given in Tables 7.5, 7.6 and 7.7. Table 7.5 shows the relationship between F1 and duration. Likewise Table 7.6 and 7.7 show the relationship between F2 and duration, F3 and duration respectively. The comparison of parameters from these tables showed that the duration and F1 are inversely proportional to each other which means, the shorter the vowel duration, the higher the F1 value and vice versa. In contrast, duration and F2 and duration and F3 are directly proportional to each other. This implies that the longer the vowel duration, the higher the F2 and F3 values and vice versa. This relationship between duration and formant frequency values is also the same for the seven Amharic vowels. Even if the duration affects the formant frequency values of vowels extracted from different contexts, this effect does not change the acoustic vowel space rather than showing a little shift in the vowel position on the space. However, the relationship between duration and formant frequency values should be considered during characterization of vowel sound units.

Table 7.5: F1 values of vowel /ኤ/ [ie] in both short and long duration

No.	Context	Average Duration(ms)		Average F1 value (Hz)	
		Short	Long	At short duration	At long duration
1	[tx ie f]	225	322	803	616
2	[p ie xx]	261	323	677	617
3	[s ie p]	231	327	610	574
4	[t ie r]	255	352	682	622
5	[xx ie p]	230	358	602	560
6	[f ie g]	230	363	517	605
7	[xx ie g]	252	364	623	565
8	[tx ie z]	287	377	746	583
9	[f ie r]	238	377	634	610
10	[t ie z]	265	392	636	609
11	[ie]	302	416	822	600

Table 7.6: F2 values of vowel /ኤ/ [ie] in both short and long duration

No.	Context	Average Duration(ms)		Average F2 value (Hz)	
		Short	Long	At short duration	At long duration
1	[tx ie f]	225	322	2193	2215
2	[xx ie p]	230	358	1970	1974
3	[f ie g]	230	363	1845	1894
4	[s ie p]	231	327	1952	1972
5	[f ie r]	238	377	1911	1917
6	[xx ie g]	252	364	1839	1979
7	[t ie r]	255	352	1910	2025
8	[p ie xx]	261	323	1897	1946
9	[t ie z]	265	392	1891	1957
10	[tx ie z]	287	377	1863	1915
11	[ie]	302	416	1827	1895

Table 7.7: F3 values of vowel /ኤ/ [ie] in both short and long duration

No.	Context	Average Duration(ms)		Average F3 value (Hz)	
		Short	Long	At short duration	At long duration
1	[tx ie f]	225	322	2800	2804
2	[xx ie p]	230	358	2723	2782
3	[f ie g]	230	363	2786	2818
4	[s ie p]	231	327	2697	2835
5	[f ie r]	238	377	2751	2848
6	[xx ie g]	252	364	2711	2787
7	[t ie r]	255	352	2782	2871
8	[p ie xx]	261	323	2720	2828
9	[t ie z]	265	392	2739	2790
10	[tx ie z]	287	377	2746	2855
11	[ie]	302	416	2745	2807

7.2.2.3 Acoustic Vowel Space

Table 7.1 shows the aggregate features of the seven Amharic vowels such as the first three formant frequency values of each vowel. Based on this result, the acoustic vowel space for Amharic language was developed as shown in Figure 7.5. The acoustic vowel space of Amharic language shows that vowel /ኢ/ [ii], /እ/ [ix] and /ኡ/ [u] are high, vowel /ኤ/ [ie], /ኧ/ [e] and /ኦ/ [o] are middle and /አ/ [a] is low vowel based on the F2 vs. F1 acoustic space which corresponds with the tongue height. Similarly, vowel /ኢ/ [ii] and /ኤ/ [ie] are front, vowel /እ/ [ix], /ኧ/ [e] and /አ/ [a] are middle and /ኡ/ [u] and /ኦ/ [o] are back vowels according to F2 Vs. F1 acoustic space which corresponds with the tongue length. This result definitely and perfectly matches with the articulatory vowel space developed by International Phonetic Alphabet (IPA) through the position of the tongue.

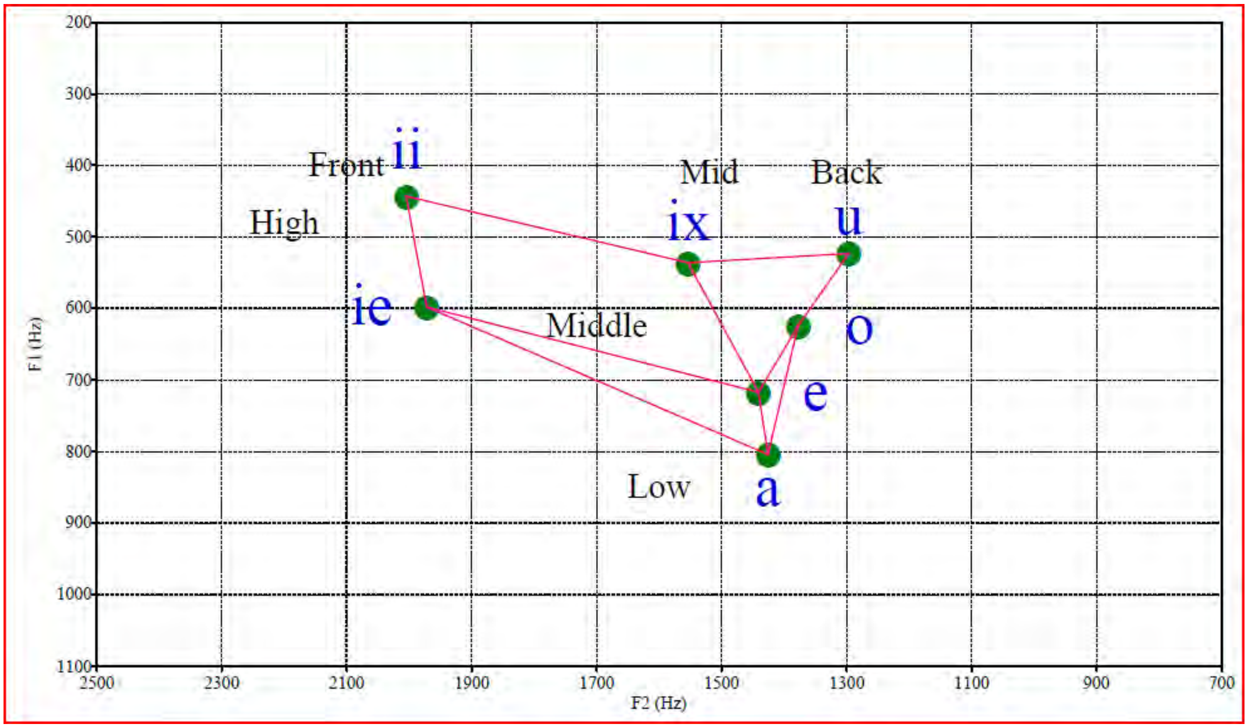


Figure 7.5: Acoustic vowel space of the seven Amharic vowels

7.2.2.4 Formant frequency variability

Based on the procedures pointed out in section 6.3.7, the formant frequencies distribution of vowels among similar CVC context utterances lays in three formant frequency distribution classes. This is true for all Amharic vowels. For example, Table 7.8 indicated that the variation of formant frequency values of vowel /ኢ/ [ii] among similar CVC context utterances. For instance, the variability of F1 in [xx ii p] context utterances indicated that the numbers of utterances in the normal class are eight out of ten whereas the number of utterances in the above normal class is one and below normal class is one. The analysis of the remaining formant frequency values and CVC contexts of vowel /ኢ/ [ii] were conducted in the similar manner. Similarly, the variability of F1, F2 and F3 of other vowels are also done based on the same procedure. As a result of these analyses, the distribution above and below the normal class were contributing to the formant frequency variation of vowels. This variation comes due to the loudness or intensity difference during context recording process. Even though the formant frequency is affected by intensity level but there was less influence on the acoustic vowel space.

Table 7.8: The number of utterances and their percentage value which makes formant frequency variation among similar CVC context utterances of vowel /ኢ/ [ii].

			F1 Number of utterances & Percentage value			F2 Number of utterances & percentage value			F3 Number of utterances & percentage value		
No.	**Context**	**Vowel**	**Normal**	**Above**	**Below**	**Normal**	**Above**	**Below**	**Normal**	**Above**	**Below**
1	[xx ii p]	[ii]	8, 80%	1, 10%	1, 10%	8, 80%	1, 10%	1, 10%	8, 80%	1,10%	1, 10%
2	[xx ii g]	[ii]	7, 70%	2, 20%	1, 10%	7, 70%	2,20%	1, 10%	6, 60%	2, 20%	2, 20%
3	[tx ii f]	[ii]	8, 80%	1, 10%	1, 10%	7, 70%	2,20%	1, 10%	7, 70%	2, 20%	1, 10%
4	[tx ii z]	[ii]	9, 90%	1, 10%	0%	8, 80%	2, 20%	0%	8, 80%	1, 10%	1, 10%
5	[s ii p]	[ii]	5, 50%	3, 30%	2, 20%	8, 80%	0%	2, 20%	8, 80%	1, 10%	1, 10%
6	[p ii xx]	[ii]	9,90%	0%	1, 10%	6, 60%	2, 20%	2, 20%	7, 70%	1, 10%	2, 20%
7	[t ii r]	[ii]	6, 60%	2, 20%	2, 20%	8, 80%	1, 10%	1, 10%	6, 60%	2, 20%	2, 20%
8	[f ii r]	[ii]	8, 80%	1,10%	1, 10%	8, 80%	0%	2, 20%	8, 80%	1, 10%	1, 10%
9	[f ii g]	[ii]	9, 90%	1, 10%	0%	7, 70%	2, 20%	1, 10%	6, 60%	2, 20%	2, 20%
10	[t ii z]	[ii]	9, 90%	1, 10%	0%	6, 60%	2, 20%	1, 10%	8, 80%	2, 20%	0%

Similarly, the formant frequency variability of vowels within different contexts is explained under section 6.3.7, this idea also access through our experimentation. The complete experiment result is given in Tables 7.9 and 7.10.

Table 7.9: The number of instances and their percentage values for the seven Amharic vowels extracted from 11 contexts.

		F1 Number of instance			F2 Number of instance			F3 Number of instance		
No.	**Vowel**	**Normal**	**Above**	**Below**	**Normal**	**Above**	**Below**	**Normal**	**Above**	**Below**
1	[e]	9, 82%	1, 9%	1, 9%	8, 73%	2, 18%	1, 9%	6, 55%	2, 18%	3, 27%
2	[u]	8, 73%	3, 27%	0, 0%	9, 82%	1, 9%	1, 9%	6, 55%	1, 9%	4, 36%
3	[ii]	7, 64%	2, 18%	2, 18%	7, 64%	2, 18%	2, 18%	7, 64%	3, 27%	1, 9%
4	[a]	8, 73%	1, 9%	2, 18%	6, 55%	2, 18%	3, 27%	8, 73%	0, 0%	3, 27%
5	[ie]	8, 73%	1, 9%	2, 18%	10,91%	1, 9%	0, 0%	6, 55%	2, 18%	3, 27%
6	[ix]	8, 73%	1, 9%	2, 18%	9, 82%	1, 9%	1, 9%	8, 73%	1, 95	2, 18%
7	[o]	9, 82%	1, 9%	1, 9%	7, 64%	2, 18%	2, 18%	7, 64%	2, 18%	2, 18%

Table 7.10: Contexts which varies the formant frequency values of the seven Amharic vowels

		Context makes F1		Context makes F2		Context makes F3	
No.	**Vowel**	**Above normal**	**Below Normal**	**Above normal**	**Below normal**	**Above normal**	**Below normal**
1	[e]	[tx e f]	[t e z]	[e], [t e z]	[s e p]	[xx e p], [xx e g]	[p e xx], [f e g], [t e z]
2	[u]	[tx u z], [s u p], [f u r]	No context	[f u g]	[t u r]	[f u g]	[s u p], [p u xx], [t u r], [t u z]
3	[ii]	[tx ii f], [f ii g]	[f ii r], [s ii p]	[tx ii f], [ii]	[s ii p], [f ii r]	[tx ii f], [tx ii z], [ii]	[t ii r]
4	[a]	[f a r]	[xx a p], [xx a g]	[f a r], [t a z]	[tx a f], [xx a p], [a]	No	[p a xx], [f a r], [f a g]
5	[ie]	[t ie r]	[s ie p], [xx ie g]	[ie]	No context	[p ie xx], [f ie g]	[f ie r], [s ie p], [xx ie p]
6	[ix]	[t ix r]	[t ix z], [ix]	[t ix r]	[f ix r]	[tx ix z]	[f ix g], [f ix r]
7	[o]	[o]	[p o xx]	[o], [s o p]	[xx o p],[tx o z]	[tx o z], [s o p]	[f o g], [f o r]

Table 7.9 shows the number of instances and percentage values of vowels based on the three distribution classes. For instance, F1 distribution showed that vowel /ኢ/ [ii] has 7 instances or contexts with 65% out of the eleven contexts in the normal class whereas in above and below normal classes, there are only two contexts with 18% percentage value and two instances with 18% percentage value respectively. According to Table 7.9, the first formant frequency distribution of the seven Amharic vowels was consistent. The minimum number of instances within normal distribution class of the first formant frequency occurred to vowel /ኢ/ [ii]. There are only seven contexts out of the total eleven contexts. The other vowels contain eight or nine instances in their normal distribution classes for F1. The description of the formant frequency values variability of other vowels also analyses in the similar manner from Table 7.9. As a conclusion, the formant frequency values variability which comes due to the contexts is medium.

Table 7.10 shows that the contexts which make formant frequency variability for the seven Amharic vowels. These contexts are classified according to the distribution classes as either above or below normal distribution classes and with respect to the formant frequencies which varies due to the contexts. For example, the context that increases the first formant frequency value of vowel /ኧ/ [e] is /ጠፍ/ [tx e f] whereas F1 value decreases within /ተዝ/ [t e z] context. The second formant frequency value is increase in /ተዝ/ [t e z] context and without context /ኧ/ [e] while the context which decreases F2 value is /ሰፕ/ [s e p]. The third formant frequency increases within /ጸፕ/ [xx e p] and /ጸግ/ [xx e g] contexts. There are also contexts which decrease F3 value of vowel /ኧ/ [e] such as /ፐጽ/ [p e xx], /ፈግ/ [f e g] and /ተዝ/ [t e z].

Generally, there are contexts for each seven vowels that make a variation of formant frequency values either increase or decrease from the normal distribution classes. But this variability has not a large influence on the acoustic vowel space rather simply change the frequency values of vowels. However, when the speech systems such as speech synthesis and speech recognition are developing, the characteristics of vowels depend on the context in which the vowels are located. Therefore, in speech system development the contexts in which the vowel is placed should be considered. If the vowel is found in CVC syllable structure this characterization is used as an input for the speech system development.

Even if there is a change in formant frequency values of vowels occurred due to the specific context influence and the loudness/intensity difference when recording such contexts, there is not huge change on the acoustic vowel space rather than appearing a little shift or movement in the position of vowels on the space.

Chapter Eight

CONCLUSION AND FUTURE WORKS

8.1 Conclusion

Acoustic speech sound units characterization is a means of studying the detail acoustic characteristics or behaviors of the speech units. It has many advantages such as in formant based speech synthesis used to determine appropriate parameter selection that applied to synthesis, in concatenative speech synthesis used to describe acoustic and phonetic features of the respective speech units and in speech recognition systems used to determine the phonological rules. Therefore, we conducted this research in order to get the aforementioned advantages and to avoid the problems occurred in speech systems such as performance problem due to lack of best parameters.

Amharic language has articulatory vowel space that is adopted from IPA. However, it doesn"t have acoustic vowel space. Therefore, this work performed the acoustic characterization of vowel sound units in order to develop the acoustic vowel space of the language using the formant based characterization approach and the duration based characterization approach is also used to see the influence of vowel duration on formant frequency values.

Before characterizing the Amharic vowel sound units, the contexts which are relevant and important for characterization purpose were specified or selected based on the context selection rules and regulations. Afterward, the overall procedures to achieve acoustic characterization of Amharic vowels were: preparation of text and speech corpora, segmentation of speech corpus, extraction of vowel features and analysis of the extracted vowel features.

Based on the procedures pointed out in the above, the experiment is conducted to obtain the acoustic characteristics of Amharic vowel sound units. The experiment mainly focused on the analysis of the relationship between speaking rate and duration, the development of acoustic vowel space of Amharic vowels, computing the formant frequency ranges of vowels; analysis of the relationship between duration and formant frequency values and the investigation of formant frequency variability.

The experimentation result showed that fast speaking rate duration of all vowels are shorter than the corresponding duration in the normal speaking rate, duration and first formant frequency value are inversely proportional to each other while duration and second formant frequency value as well as duration and third formant frequency value are directly proportional to each other, the formant frequency values of vowels varies due to the context and intensity variations.

According to the experiment result, the acoustic characteristics of Amharic vowel sound units are influenced by different factors. Especially, the formant frequency values and acoustic vowel space of vowels are affected by the adjacent contexts, duration and the loudness/intensity. However, these factors had not a huge influence on the acoustic vowel space rather than appearing a little shift in the position of vowels on the space.

Even if the position of the vowels on the acoustic vowel space is shifted due to different factors, there are seven Amharic vowel sound units on the acoustic vowel space. Similarly, there are also seven Amharic vowels on the articulatory vowel space which is adopted from IPA. When we compare and contrast the two characterizations of Amharic vowels, they are equivalent to each other. This implies that, the presence of seven vowel sound units in Amharic language phonology is acceptable and logical.

8.2 Future Works

In this study, the characterization of vowel sound units for Amharic language in CVC contexts was done. Here, the vowels are extracted from the contexts to get the appropriate characterization. Still, there are also some works to be done in the future to develop this work into a full fledged characterization and acquire better results. The works that will be conduct in the future are given below.

- In this work, characterization of vowel sound units were done but the characterization of consonant sound units was not considered. In order to have a full information about Amharic sound units, characterization of Amharic consonant sound units will be done as a future work.

- In this research, the characterization of vowel sound units was done in CVC syallable structure contexts. But, in Amharic there are also other basic syallable structures such as Consonant-Vowel(CV), Vowel-Consonant(VC), Vowel-geminated Consonant(VC^{-}) and Vowel-Consonant-Consonant(VCC) where **C^{-}** stands for a geminated consonants, the characteristic of vowels in these syllable structure should be studied to see the difference and similarity between them and with CVC contexts. So, characterization of vowel sound units in the remaining syllable structures will be considered as a future work.

- There is no standard Amharic corpus to serve such kinds of researches. One can engage in the preparation of standard corpus for Amharic language that is important for speech characterization processes.

- The extraction of formant frequency values of vowels was conducted using LPC with peak picking algorithm, as we have seen in different researches, there is a variation of formant frequency values in different formant extraction algorithm. So, doing the characterization using other formant frequency extraction algorithm such as burg, LPC cepstrum and direct spectrum analysis will be considered as a future work in order to compare and contrast to this characterization.

9. REFERENCES

[1] Liddy, E.D., “Natural Language Processing,” In Encyclopedia of Library and Information Science, 2nd Ed. NY. Marcel Decker, Inc., 2001.

[2] Gobinda G. Chowdhury, “Natural Language Processing,” University of Strathclyde, Department of Computer and Information Sciences, Glasgow G1 1XH, UK, 2005.

[3] Liddy, E., “Enhanced text retrieval using natural language processing,” Bulletin of the American Society for Information Science, 24, 14-16, 1998.

[4] Lehmam, A., “Text structuration leading to an automatic summary system”: RAFI. Information Processing & Management, 35, 181-191, 1999.

[5] J. Deller, et. al., Discrete-Time Processing of Speech Signals, MacMillan Publishing Co., ISBN: 0-7803-5386-2, 2000.

[6] Harrington, J. & Cassidy, S., Techniques in Speech Acoustics, Kluwer Academic Publishers: Foris, Dordrecht. ISBN: 0-7923-5731-0, 1999.

[7] Masur EF., Infants' early verbal imitation and their later Lexical development, Merrill Palmer Quarterly, 41, 286-306, 1995.

[8] Hussien Seid and Bj¨orn Gamb¨ack “A Speaker Independent Continuous Speech Recognizer for Amharic”, INTERSPEECH, Lisbon Portugal, 2005.

[9] Solomon Teferra Abate, Wolfgang Menzel, Bairu Tafila, “An Amharic Speech Corpus For Large Vocabulary Continuous Speech Recognition”, Fachbereich Informatik, Universit¨ at Hamburg, 2005.

[10] F. Martinez, A. Guillamon and J.J. Martinez, “Vowel and Consonant Characterization using fractal dimension in natural speech,” ISCA Non-Linear speech production and characterization Workshop, Spain, page 1-4, 2003.

[11] Masaaki Honda, NTT CS Laboratories, “Human speech production mechanisms”, Journal of the Acoustical Society of Japan, Vol. 1, No. 2, pp. 24-29, 2003 (in Japanese).

[12] Giegerich, Heinz J., English Phonology: An introduction, Cambridge: Cambridge University Press, 1992.

[13] Coleman, John, The vocal tract and larynx, Available from http://www.phon.ox.ac.uk/~jcoleman/phonation.htm., 2001.

[14] Steven K Smith, Digital Signal Processing, A Practical Guide for Engineers and Scientists, ISBN 0-75067444-X, USA, 2003

[15] Styger, T., & Keller, E. "Formant Synthesis. In E. Keller (ed.), Fundamentals of Speech Synthesis and Speech Recognition: Basic Concepts, State of the Art, and Future Challenges (pp. 109-128)" Chichester: John Wiley, 1994.

[16] Diemo Schwarz, Spectral Envelopes in Sound Analysis and Synthesis**,** Version 98.1 p1 release, March 2nd, 1998.

[17] Fant, G., Acoustic Theory of Speech Production, Mouton & Co, The Hague, Netherlands, 1960.

[18] Syed Akhter Hossain, M. Lutfar Rahman, and Farruk, "Bangla Vowel Characterization based on Analysis by Synthesis," Ahmed World Academy of Science, Engineering and Technology, 2007.

[19] Benade, Arthur H., Fundamentals of Musical Acoustics, Oxford University Press, 1976

[20] Sundberg, Johan, The acoustics of singing voice, scientific American, page 82, March 1977.

[21] Syed Akhter Hossain, Mlutfar Rahman and Farruk Ahmed, "Acoustic Space of Bangla Vowels", Proceedings of the 5th WSEAS Int. Conf. on signal, speech and image processing, Corfu, Greece (pp138-142), 2005.

[22] Chanwoo Kim, Kwang-deok Seo, and Wonyong Sung, "A Robust Formant Extraction Algorithm Combining Spectral Peak Picking and Root Polishing", Hindawi Publishing Corporation, EURASIP Journal on Applied Signal Processing Volume, Pages 1–16, 2006.

[23] Minkyu Lee, Jan van Santen, Bernd Möbius, and Joseph Olive, "Formant Tracking Using Context-Dependent Phonemic Information", IEEE Transactions on Speech and Audio Processing, vol. 13, no. 5, pp 741-750, September 2005.

[24] TOM McARTHUR, "VOWEL QUALITY", Concise Oxford Companion to the English Language, 1998, Encyclopedia.com, 23 April, 2011.

[25] Myers, S. and B.Hansen, "The origin of vowel length neutralization in vocoid Sequences": Evidence from Finnish speakers. Phonology 22. 317-344, 2006.

[26] Klatt, D., “Linguistic uses of segment duration in English: Acoustic and perceptual evidence”, JASA 59, 1208-1221, 1976.

[27] Lindblom, B., B. Lyberg, and K. Holmgren, “Durational Patterns of Swedish Phonology: Do They Reflect Short-Term Memory Processes?” Indiana University Linguistics Club, Bloomington, 1981.

[28] Mandelbrot BB, The fractal geometry of nature, free man, New York, 1983.

[29] Rosana Esteller, Student Member, IEEE, George Vachtsevanos, Senior Member, IEEE, Javier Echauz, Member, IEEE, and Brian Litt, Member, IEEE,” A Comparison of Waveform Fractal Dimension Algorithms,” IEEE transactions on circuits and systems-i: fundamental theory and applications, vol. 48, no. 2, February 2001.

[30] Chouard CH, Meyer B, “Fractal dimension of speech”, Article in French, Paris, 2001.

[31] Karen Croot and Belinda Taylor, “Criteria for Acoustic-Phonetic Segmentation And Word Labeling in the Australian National Database of Spoken Language”, Speech, Hearing and Language Research Centre, Macquarie University, 1995.

[32] Stevens, K. N., “Toward a model for lexical access based on acoustic landmarks and distinctive features”, Journal of the Acoustical Society of America, 111(4), 2002.

[33] Gordon E. Peterson and Harold L. Barney, “Control methods used in a study of the vowels”, the journal of the acoustical society of Amharica, Bell laboratories, Inc, Murray Hill New Jersey, 1952.

[34] James Hillenbrand, Lura A. Getty, Michael J. Clark, and Kimberlee Wheeler, “Acoustic characteristics of American English vowels”, Department of speech pathology and audiology, Western Michigan University, 1995.

[35] Syed Skhter Hussain, M. Lutfar Rahman, and Farruk Ahmed, “Acoustic classification of Bangla vowels”, World Academy of Science, Engineering and Technology, 2007.

[36] Syed Akhter Hossain, Mlutfar Rahman and Farruk Ahmed, “Acoustic Space of Bangla Vowels”, Proceedings of the 5th WSEAS Int. Conf. on signal, speech and image processing, Corfu, Greece (pp138-142), 2005.

[37] Syed Akhter Hossain, Mlutfar Rahman and Farruk Ahmed, “Bangla vowel characterization based on Analysis by Synthesis”, World Academy of Science, Engineering and Technology, 26, 2007.

[38] Nadew Tademe, “Formant based speech synthesis for Amharic vowels”, MSc Thesis, Faculty of Informatics, Department of Computer Science, Addis Ababa University, Ethiopia, 2008.

[39] Peter Roach, “A little encyclopedia of phonetics”, professor of phonetics, University of reading, UK, 2002

[40] http:// www.singwise.com, karyn O‟Connor, 2011, last date of access, July, 2011

[41] Behne, D. and Moxness, B. and Nyland, A., “Acoustic-phonetic evidence of vowel quantity and quality in Norwegian,” journal: TMH-QPSR, volume: 37, number: 2, pages: 013-016, KTH computer science and communication, 1996

[42] Laura L. Koeing, “Towards a physical definition of the vowel systems of languages”, hard- science linguistics, continuum, pp 49-66, 2003

[43] Lindau M., “The feature expanded”, journal of phonetics, 7, 163-176, 1979.

[44] Ladefoged, Peter, A Course in Phonetics (Fifth Edition), Boston: Thomson Wadsworth, p. 189, 2006

[45] Baart, Joan, Acoustic Phonetics, CD-ROM. In LinguaLinks 5.0. Dallas: SIL, 2001

[46] David Odden, “The Representation of Vowel Length,” Department of linguistics, OSU, Columbus, 2010.

[47] Arthur S. Abramsont and Nianqi Rent, “Distinctive Vowel Length: Duration vs. Spectrum in Thai,” Haskins Laboratories Status Report on Speech Research,” SR-101 /102,256-268, 1990.

[48] Akamatsu, Tsutomu, Japanese phonology: A functional approach, München: Lincom Europa, 2000.

[49] Ladefoged, P., Vowels and Consonants, Oxford: Blackwell, 2005.

[50] Hayward, Katrina and Richard J. Hayward, “Amharic, In Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet”, Cambridge: the University Press, 1999.

[51] Ethnologue, Languages of the World, 14th Edition. http: ==www:ethnologue:com=14=showlanguage:asp?code = AMH, 2004.

[52] Cowley, Roger, et al., The Amharic Language, in Bender, M. et al (eds.),

Languages in Ethiopia. London: Oxford University Press, 1976.

[53] Bekrie Ayele, “Ethiopic: An African Writing System; its History and Principles”, Canada: RSP,1997.

[54] Sebsibe H/Mariam, S P Kishore, Alan W Black, Rohit Kumar, and Rajeev Sangal, “Unit Selection Voice for Amharic using Festivox”, 5th ISCA Speech Synthesis Workshop – 99 Pittsburgh, page 103-107, 2005

[55] Samuel Eyassu and Bj¨orn Gamb¨ack, “Classifying Amharic News Text Using Self-Organizing Maps”, Proceedings of the ACL Workshop on Computational Approaches to Semitic Languages, pages 71–78, Ann Arbor, June 2005.

[56] Laine Berhane, “Text-to-Speech Synthesis of the Amharic Language,” MSc Thesis Faculty of Technology, Addis Ababa University, Ethiopia, 1998

[57] Huang, X., Acero, A., Hon, H. “Spoken Language Processing. Prentice Hall”, Upper SaddleRiver, New Jersey, 2001.

[58] Leslau, W., Introductory Grammar of Amharic, Wiesbaden: Harrassowitz, 2000.

[59] Roach, P., English Phonetics and Phonology: a Practical Course, Cambridge: CUP, 2000.

[60] Lulseged Erkihun, “A contrastive Analysis of the phonology of Gedio and Amharic”, Addis Ababa University, Master of Arts in linguistics, April 1981.

[61] Leslau, W., Reference Grammar of Amharic, Wiesbaden: Harrassowitz, 1995.

[62] Anne and Klaus Wedekind, “Amharic stress (beat) rules of linguists, poets and singers: which beat rules beat which?”, paper presented at the 11^{th} international conference of Ethiopian studies, Addis Ababa, Aril 1-6, 1991.

[63] Ullendorff, E., “The Semitic language of Ethiopia,” A comparative phonology”, London, 1995.

[64] Armbruster, C. H., Initia Amharica, An introduction to spoken Amharic, Cambridge university press, 1908.

[65] Richter, R., A practical course in Amharic Language, Washington, DC, 1987.

[66] Alemayehu Haile, 1987, “An auto segmental approach to Amharic intonation,” PhD thesis, SOAS, University of London, 1987.

[67] Foreign Service Institute Language Program, “FSI Amharic Basic Course,” 1.36MB, 12 Nov 2007, retrieved from http://www.multilingualbooks.com/fsi-Amharic.html/Amharic1-TOCUnit1.pdf on April 21, 2011.

[68] Keith Brown, Sarah Ogilvie, “The encyclopedia of world languages”, University of London, UK, 2006.

[69] Hirata, Yukari and Kimiko Tsukada, “The Effects of Speaking Rates and Vowel Length on Formant Movements in Japanese”, In Proceedings of the 2003 Texas Linguistics Society Conference, ed. Augustine Agwuele et al., 73-85. Somerville, MA: Cascadilla Proceedings Project, 2004

[70] Mustafa, Kamran, and Ian C. Bruce, "Robust Formant Tracking for Continuous Speech With Speaker Variability," IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, No. 2, March 2005.

APPENDIXES

Appendix A: (I- X NOTAION) Amharic Phonetic List, IPA Equivalence and its ASCII Transliteration Table

IPA	Transcription	Amharic equivalence
Consonants		
[p]	[p]	ፕ
[t]	[t]	ት
[k]	[k]	ክ
[?]	[ax]	ዕ
[b]	[b]	ብ
[d]	[d]	ድ
[g]	[g]	ግ
[p’]	[px]	ጵ
[t’]	[tx]	ጥ
[c’]	[cx]	ጭ
[q]	[q]	ቅ
[f]	[f]	ፍ
[s]	[s]	ስ
[ʃ]	[sx]	ሽ
[h]	[h]	ህ
[s’]	[xx]	ጽ
[tʃ]	[c]	ች
[g’]	[j]	ጅ
[m]	[m]	ም
[n]	[n]	ን
[n’]	[nx]	ኝ
[l]	[l]	ል
[r]	[r]	ር
[j]	[y]	ይ
[w]	[w]	ው
[v]	[v]	ቭ
[z]	[z]	ዝ
[z’]	[zx]	ዥ
Vowels		
[ε]	[e]	ኧ
[ʊ]	[u]	ኡ
[ɪ]	[ii]	ኢ
[ɑ]	[a]	ኣ
[e]	[ie]	ኤ
[ɨ]	[ix]	እ
[o]	[o]	ኦ

Appendix B: Contexts used in the thesis with their target vowel and transcription

For vowel /ኧ/ [e]:

No.	Contexts	Transcription
1	ኧ	[e]
2	ተዝ	[t e z]
3	ተር	[t e r]
4	ፈግ	[f e g]
5	ፈር	[f e r]
6	ፐጽ	[p e xx]
7	ጠፍ	[tx e f]
8	ጠዝ	[tx e z]
9	ሰፕ	[s e p]
10	ጸግ	[xx e g]
11	ጸፕ	[xx e p]

For vowel /ኡ/ [u]:

No.	Contexts	Transcription
1	ኡ	[u]
2	ቱዝ	[t u z]
3	ቱር	[t u r]
4	ፉግ	[f u g]
5	ፉር	[f u r]
6	ፑጽ	[p u xx]
7	ጡፍ	[tx u f]
8	ጡዝ	[tx u z]
9	ሱፕ	[s u p]
10	ጹግ	[xx u g]
11	ጹፕ	[xx u p]

For vowel /ኢ/ [ii]:

No.	Contexts	Transcription
1	ኢ	[ii]
2	ቲዝ	[t ii z]
3	ቲር	[t ii r]
4	ፊግ	[f ii g]
5	ፊር	[f ii r]
6	ፒጽ	[p ii xx]
7	ጢፍ	[tx ii f]
8	ጢዝ	[tx ii z]
9	ሲፕ	[s ii p]
10	ጺግ	[xx ii g]
11	ጺፕ	[xx ii p]

For vowel /አ/ [a]:

No.	Contexts	Transcription
1	አ	[a]
2	ታዝ	[t a z]
3	ታር	[t a r]
4	ፋግ	[f a g]
5	ፋር	[f a r]
6	ፓጽ	[p a xx]
7	ጣፍ	[tx a f]
8	ጣዝ	[tx a z]
9	ሳፕ	[s a p]
10	ጻግ	[xx a g]
11	ጻፕ	[xx a p]

For vowel /ኤ/ [ie]:

No.	Contexts	Transcription
1	ኤ	[ie]
2	ቴዝ	[t ie z]
3	ቴር	[t ie r]
4	ፌግ	[f ie g]
5	ፌር	[f ie r]
6	ፔጽ	[p ie xx]
7	ጤፍ	[tx ie f]
8	ጤዝ	[tx ie z]
9	ሴፕ	[s ie p]
10	ጼግ	[xx ie g]
11	ጼፕ	[xx ie p]

For vowel /እ/ [ix]:

No.	Contexts	Transcription
1	እ	[ix]
2	ትዝ	[t ix z]
3	ትር	[t ix r]
4	ፍግ	[f ix g]
5	ፍር	[f ix r]
6	ፕጽ	[tx ix xx]
7	ጥፍ	[tx ix f]
8	ጥዝ	[tx ix z]
9	ስፕ	[s ix p]
10	ጽግ	[xx ix g]
11	ጽፕ	[xx ix p]

For vowel /ኦ/ [o]:

No.	Contexts	Transcription
1	ኦ	[o]
2	ቶዝ	[t o z]
3	ቶር	[t o r]
4	ፎግ	[f o g]
5	ፎር	[f o r]
6	ፖጽ	[p o xx]
7	ጦፍ	[tx o f]
8	ጦዝ	[tx o z]
9	ሶፕ	[s o p]
10	ጾግ	[xx o g]
11	ጾፕ	[xx o p]

Declaration

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

____

TESSFU GETEYE FANTAYE

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

____

SEBSBIE HAILEMARIAM (PhD)

Addis Ababa, Ethiopia

November, 2011