AMHARIC SPEECH TRAINING FOR THE DEAF

BY
DANIEL ASSEFA

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

August 2006
ADDIS ABABA

Acknowledgment

First of all, I would like to thank my Advisor Dr. Dida Midekso for his continuous support and invaluable comments. He spent his precious time in commenting my work and showing me the right directions which I found very important for my thesis.

I also would like to thank staff of Mekanissa Deaf School for their support in Sign Language Training, allowing me to have a visit even during classes and for testing our lip model. My gratitude also goes to Instructors from Institute of Language Studies, Department of Amharic Language for their support in studying Amharic language.

Finally, I would like to express my sincere thanks to my family and very close friends for being with me whenever in need.

Table of Contents

List of Tables

List of Figures

Abstract

It has been believed that all deaf persons can not make edible sound and can only communicate through Sign Language. However, Deaf people can make voices and communicate orally unless they are mute by nature. With speech training it is possible for the Deaf to learn how to speak and “listen”. Speech training can be given manually with a human trainer but it is a very tiresome task and its’ demands are more than the capability of trainers. The solution proposed for this problem is an automated speech training system which is already implemented for different languages. This thesis addresses a similar solution but for Amharic language. Due to the limitations of special equipment and software tools we can get we proposed modeling of a lip for the articulation of Amharic characters which is part of an automated speech training system. We used an Analysis-Synthesis approach to first analyze a real lip in speech making and applied the output of the analysis on our lip model to articulate different Amharic characters. The solution proposed is implemented in a prototype developed for selected Amharic characters and its efficiency is tested with some students of Mekanissa Deaf School.

Keywords: Speech Training, Sign Language, Deaf Education, Lip modeling, Talking Head

1. Introduction

Speech has two main components; auditory and visual. People suffering from a severe hearing loss depend on the visual component in verbal communication. This technique is named speech-reading. Hearing people also depend on speech reading in noisy environment where it's not possible to hear the auditory signal clearly. However, speech-reading is a difficult task, since many speech-sounds either look alike or are only partly visible. Still, these cues are essential to the understanding of speech for many people.

For a hearing-impaired person, seeing the face of the speaker considerably enhances the hearing ability. Not only real faces but also synthetic ones are a great help in increasing the understandability of natural speech. When speech-reading is a subject, the human speech organs come to the front as they are the tools of production. Of these speech organs lip is the most outer part that can easily be seen by the speech reader. In this thesis the lip movements for different Amharic characters are exploited from a human speaker and modeled to a synthesized lip to produce different Amharic words in vocabulary training for the Deaf.

1.1 Background

1.1.1 The word "Deaf"

The word deaf is used and understood in two ways. "deaf" with lower case 'd' is to mean a person who has some hearing impairment. "Deaf" which is capitalized describes a community whose members may or may not be able to hear but for which the medical condition of deafness is somehow relevant to them [26].

1.1.2 Deafness

Deafness is the inability to hear. It can affect one or both ears, either totally or partially. Deafness may be present at birth (congenital deafness), or it may occur later, suddenly or gradually. Deafness can also be classified with the degree of hearing loss in decibels (dB) ranging from mild to profound hearing loss. A loss of 35 to 70 dB is generally considered as hard of hearing, while a loss of 90 dB or greater is considered profoundly deaf [28].

Some Deaf are mute by nature; they couldn't make any voice. But others do have the ability of making voices even if they may not be heard. This is because they have a problem of controlling their pitch and volume and they don't hear what others and even they are saying, which would have been a great experience for them.

1.1.3 Sign language

Communication is an engine that drives humans in every walks' of life from birth to death. It is through language that humans can communicate. Sign language is one of the different types of languages through which the Deaf can communicate with themselves and other hearing people.

In sign language there are different signs that represent letters, numbers and frequently used words. Sign language is not spoken or heard but it's a visual language seen by the eye and expressed by the movements of hands face and the body.

Sign language is considered as mother tongue for people who are pre lingual deaf (become deaf before learning other language). These people can use oral communication as their

second language. On the other hand people who are post lingual deaf can use sign language second to oral one.

1.1.3.1 Sign language notations

Sign language notations have 4 components: hand shape, location, movement and orientation or direction of signs.

1. Hand Shape: particular orientation of the hand and fingers. In sign language, the actor is the right hand and the left hand is used only to help the right hand in some circumstances where one hand is not enough. (See Figure 1.1)

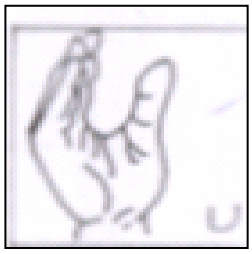
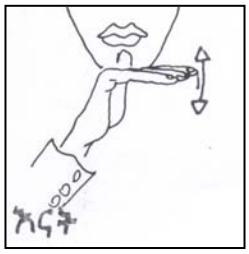


Fig 1.1 Sign for the letter "***U***"

2. Location: the position where you put your hand on. Usually you have to put your hands around your chest but sometimes it may be located in other parts. Location sometimes creates differences in the meaning of a sign (Figure 1.2 Shows the effect of location of the hand in meaning of sign).

a. Sign for the word “Father”



b. Sign for the word “Mother”

Fig 1.2 Same hand sign with difference in location of the hand representing two words

3. Movement: the movement of the hand also makes differences in the character to be represented by a given sign. In Amharic sign language all the seven orders of characters have the same sign. They are differentiated by the movement of the hand. The direction for the movement of the hand from the 2nd to the 7th order of any letter is shown in fig 1.3. Its seven orders are shown in table 1.1. Take the sign in fig 1.1 as an example; it represents the letter “***ሀ***”. To spell “**ሃ**“, which is the fourth order, we show the sign in fig 1.1 and moving our hand with the 3rd direction shown in fig 1.3. And to spell “**ህ**“ the same sign for “**ሀ**” moves with the 5th direction shown in fig 1.3.

Table 1.1 The seven orders of Amharic Letter “**ሀ**“

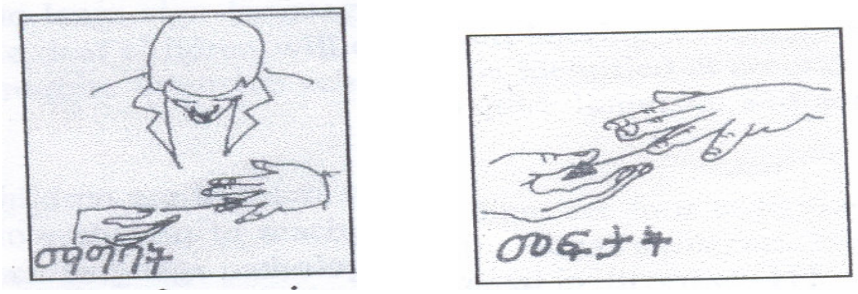
Order	1st	2nd	3rd	4th	5th	6th	7th
Symbol	**ሀ**	**ሁ**	**ሂ**	**ሃ**	**ሄ**	**ህ**	**ሆ**

Fig 1.3 From left to right the direction of movement of the arm to represent the six orders (starting from the 2nd to the 7th) of any Amharic letter with the hand showing the sign of the 1st order.

4. Orientation or direction of signs: in sign language, as a visual means of communication, the palm of the hand should face the audience with the arm being held in an easy to see position near the body.

1.1.3.2 Structure of sign language

The signs in sign language are assigned to words or actions on different basis. Some signs are descriptive; they convey ideas (for instance Marriage and Divorce, shown in fig 1.4. and other signs are of pointing to things rather than showing signs.



a. Sign for Marriage b. Sign for Divorce

Fig 1.4 Descriptive signs that convey ideas

The Deaf also use some natural signs along with sign-language. If we take “Good Bye” the sign that is used in the other society is also used in the deaf society. The other property of sign language is the use of arbitrary or local signs that are only used in small communities like a school, a village or an organization.

In many cases like names of persons, countries, cities and some other common words, the signs are first letter based. The sign for the name “ኦበበ” for instance, is the sign for “ኦ” and touching part of your face. There is hand location restriction in representing the name of males and females. It is not allowed to touch above or by the side of your eyes to represent a female name and below your eyes to represent a male name. The relationship between hand location and gender is also shown in fig 1.2.

First letter based signing is not restricted to names only but common words like “እንጀራ” are also represented by using the sign of “ኦ” and adding some movement that shows method of baking “Enjera”. In some cases words may not have corresponding sign; this forces the Deaf to use finger-spelling. Finger spelling is a method of representing words using the signs of each letter the word comprises. If the word has seven characters it will be represented by the combination of seven signs, one for each letter.

1.1.4 Education and the Deaf

According to the World Federation of the Deaf (WFD) report there are over 70 million Deaf people in the world. Many of them are either Deaf from birth or become Deaf before learning the spoken language [29]. This fact has serious implication for the education of

Deaf people, because if they become Deaf before schooling it pushes them to seek knowledge from school which they should have brought from home.

A hearing person learns at least two of the language skills “Speaking” and “Listening” from home and in some cases the other two skills “Writing” and “Reading”. On the other hand, a deaf child is obliged to get all these four language skills from school as most of the families do not know sign language or they try to learn and teach the new language at the same time, by the time they know that their child is Deaf. Deafness is identified too late to have enough time to learn sign language. Until Infant hearing screening become widespread, majority of the deaf children will continue to be identified at around 18 to 24 months of age and will require intensive schooling to acquire linguistic and academic competence [21].

If the deaf children are to learn successfully there are four important things that should be fulfilled: low student to teacher ratio, instructional assistants, one-to-one instruction from the speech/language pathologist, and parental involvement [21]. This shows us that a lot of work should be done to change the success of the Deaf in education.

In [26] two teaching methods for Deaf are described:

Oralism: teaching deaf students primarily to speak and lip-read

Manualism: teaching deaf students primarily in sign language

World Federation of the Deaf (WFD) supports Manualism method. In order to involve deaf students in meaningful schooling, it is necessary that their teachers and peers all know and

use sign language. To deny sign language to the Deaf is considered as denying their basic human rights to communication and education [29].

However, sign language is not used by large part of the society. So, the Deaf miss large part of their social-learning experience. In addition to its limited domain, it has some major limitations:

1 It has a very limited number of words/ signs [21].

2 Lack of spoken response from the Deaf forces parents, teachers and the community to exclude or ignore them at home, school or work place.

3 Difficulty of the language: finger spelling is a very difficult task for the speaker as well as for the listener. In addition, the hearing person may not have the patience to learn and use each sign for letters and words. This in turn widens the gap between the Deaf and the hearing society.

4. Lack of standardization:

- A name can be differently represented by persons from different areas based on their custom; which induces misunderstanding within the Deaf.
- On a book level there is only one which was published in 1963 E.C. and nothing is added to it until now [1].

Deaf or hard of hearing people can communicate orally unless they are mute. Oral communication, if trained well, is a better option for the Deaf than sign language because it avoids all the above specified problems of sign language. However, it's not yet a means of communication. Deaf or hard of hearing people think that they are difficult to be understood

and have inappropriate pitch or volume. The cause for these is that they couldn't hear what they and other people say. Visual component of speech can be used for understanding and making of speech. It offers a lifeline to those with substantial hearing loss. The visual component includes movements of the lips, tongue, palate and jaws and it is enriched by facial expressions.

1.1.5 Computer Assisted Learning (CAL)

One of our day to day activities that can be supported by computers is education. CAL is an educational strategy that can be integrated into a particular subject. A well designed and properly integrated CAL tool presents materials in an effective way, facilitates communication between students and teachers, enables student collaboration, ease access to resources, encourages self-paced learning, and provides online assessment [3]. Therefore CAL plays a major role in enhancing the teaching-learning process.

Deaf education is no exception. It's even more useful in Deaf education as CAL satisfies the requirements listed in section 1.1.4. With the presence of CAL, one computer can be assigned for one student which avoids the high student to teacher ratio. It can be of an assistant as well, when there are pathologists. For the sake of parental involvement CAL is advantageous. It can be difficult for parents to go and assist their Deaf child at school but with CAL parents can use their time at home to involve in the child's training.

1.2 Problem Statement

In Addis Ababa there are 6 schools accredited by the city's educational council from elementary to high school and one pre-school which is giving sign-language training for

children. As we have noticed in our on-site visit all but two use co-education, mixing the Deaf with the hearing students. There are on average 25-35 students in a class with no language/speech pathologist in any of the schools. In those schools sign language is given for the students' families and other interested people, the number of trainees is incomparable with the number of deaf students. In such condition, expecting parental involvement with the education of the deaf students is unthinkable, as there will not be a common language for communication between the deaf children and their parents. There are also no instructional assistants to help the teachers. In "Mekanissa", the only school in which speech training is given, there were some letter indicator devices which are not in use now. The indicators are used to help speech training by showing some signs whether the students are correctly pronouncing a particular letter or not.

In some schools, sign language is given as one subject along with other common subjects. In the other common subjects if the teachers are able to use sign language they will use total communication: Lip-reading, sign language and facial expressions. But if the teachers don't know sign language, a student from the class is selected to translate what the teacher is saying. The translators are selected from the students, so they may not even know anything about the subject matter. The existence of loss of information in translation added to lack of enough knowledge of the translators results a very difficult condition to teach the deaf students successfully.

With such inefficient teaching methodology particularly for the deaf students, coping up with the hearing students in the same class is very difficult and it is also difficult for the

teachers to normalize the pace between the hearing and the deaf students. Due to these and other reasons the Deaf will not get enough knowledge from school.

Speech training through visual component of speech for deaf or hard of hearing people is possible and it's being used in one school in our country. But in there the school has to compromise the conflicting requirements of the deaf and hearing students. The hearing students prefer other subjects but the deaf students need speech training. If a deaf is able to speak, s/he can easily communicate with his/her friends, families and the community and express his/her feelings. Along with speech training, the deaf will also excel in lip-reading so that s/he will not be isolated in many social circumstances. These days there are some improvements in different media, like captioning in news and movies, presence of sign-language translators in some meetings and so on. But these will not be enough to combat their lack of social learning experience, which makes the importance of speech training unquestionable.

However, speech training is not an easy task:

1. It demands intensive and one-to-one training by speech and language therapists.
2. It's tiresome and time consuming task.
3. The training must be adaptive to the trainee's need.

Currently, time and financial constraints placed on families and education, the non existence of language therapists, the little attention given to developing special education and the above specified demands of speech training make it, to be done by humans, unrealistic. Consequently, there is no effective speech training in our country.

It is to address this problem that necessitated the investigation of the main research topic. In this thesis we will try to show how CAL can help the deaf to acquire vocabulary in Amharic Language.

1.3 Justification

Computers are supposed to be better than humans in speech training because they are untiring machines, can be easily adapted to the trainee's need, there will be no need of day-to-day presence of linguists to tutor, can be used any where at any time and by having a transparent face, it will be possible to display the vocal tract which will be hidden by the skin in humans.

A computer system for speech training needs a number of important components that can be developed separately and integrated to further enhance the system. Out of these components three will take major parts. One of the components is having a talking head that shows the vocal tract in speech, by having a talking head it is possible to give speech training on demand and untiringly. The second component is having a text to speech system (TTS) that makes voices along with the articulation of the talking head; this is useful for those who have some hearing capability. The third important component is a system that checks whether the trainees are making the correct speech or not on different exercises; this in turn is done on a system that is the reverse of Text to Speech System. This part of the system has great importance for the training because human trainers can make some mistakes in checking the correctness of the speech from the trainees, they may say the correct speech is wrong and/or the wrong speech is correct.

1.4 Objectives

1.4.1 General objective

The purpose of this thesis research is to study how computers can help deaf or hard of hearing people to enhance their Amharic language vocabulary acquisition.

1.4.2 Specific objectives

The specific objectives are to:

- Study Amharic language structure, especially its characters and pronunciation.
- Study current teaching methodology of different schools of deaf or hearing-impaired people.
- Study how speech is made for Amharic characters that can easily be understood by lip reading.
- Develop a prototype that enables deaf or hard of hearing people to acquire some vocabularies through lip reading.

1.5 Scope and Limitation

1.5.1 Scope

Effective computer assisted speech training can be given by developing a full fledged system that includes all the three components specified in section 1.3. But this project is limited only to modeling a lip that can articulate Amharic characters. It covers those characters whose pronunciation can be understood by the articulation of the mouth. In our

language if a person can pronounce a character s/he can pronounce the character in any word it occurs. So, training the characters means training the vocabulary.

1.5.2 Limitations

Visual information elicitation is the core for the development of the talking head, and it can effectively be done by using machines that show the internal vocal organs of human during speech. However, the machines that can be used for this purpose like an Ultrasound with a special kind of probe designed for such task and Magnetic Resonance Indicator (MRI) are difficult to get. There were also some software tools developed for tracking but it was not possible to find any. Due to this we are obliged to use marking method on the lip and video camera to track the lip movement in speech.

1.6 Methodology

The current instructional method used in the different deaf schools in Addis Ababa was studied on-site. The study was only conducted within the 6 primary and secondary schools in Addis Ababa. Addis Ababa is selected because there is no significant difference in schools with other cities as they are governed by the Ministry of Education.

The structure of Amharic vocabulary was studied so that logical categories were created based on the similarity and difference of the characters in their articulation. The study is made through group discussions and/or interview with linguists and teachers from deaf

schools that use total communication[!]. In addition, we also studied similar systems that are developed for speech training and other purposes

The visual information or the movement of lip is then tracked from two persons: male and female who both are native Amharic speakers. Then the information from tracking the lip articulation is fed into computer animated graphical object (lip model). The graphical object is developed using "JAVA" programming language. Java is selected because it has good graphics capabilities and it is object oriented. The object oriented feature is needed because the lip model should be integrated with models of other articulators to form a full fledged talking head. In the entire system, each articulator's model can be considered as an object to ease the combination.

1.7 Thesis Overview

This thesis is organized into five chapters including the current one. The first chapter discusses about the Deaf and their problems. The second chapter is a literature review part that discusses about the state of the art in the technology of intelligent animated language tutors and other related works. The third chapter is detailed discussion of elicitation of visual information from speech and the design of the lip model. The fourth chapter discusses the prototype of the lip model and testing. The last chapter presents conclusions and future works.

[!] Total communication is a combination of sign-language, facial expression and lip-reading.

2. Literature Review

This chapter discusses related works in the area of computer assisted language learning systems for the Deaf. As the system is related to many other areas of knowledge we will also cover all those related matters without going too far from our main subject.

2.1 Phonemes and Visemes

2.1.1 Phonemes

Phonemes are the smallest units of speech. They are the set of sounds that we use to make the words we speak. The science that studies about phonemes is phonetics. Phonetics has two branches: Articulatory and Acoustic Phonetics. Articulatory phonetics studies about the physical process of human speech production. On the other hand acoustic phonetics deals with the physical property of the sound waves of speech [9]. Articulatory phonetics is also described as a science that studies about the articulators, their positions when creating sounds and place of articulation. And acoustic phonetics is a science that studies about the aspects of speech signal representation, analysis and processing.

2.1.2 Visemes

The term viseme is coined as an abbreviation of the terms “Visual” & “Phoneme”. Viseme is what we see on the vocal tract when speech is read. Phonemes with similar visual appearance form a viseme group. The utterance of a phoneme does not generate exactly the same lip position and movement all the time. Various factors such as person, situation and mood cause these variances in phoneme production [23]. Age and sex of the person also

result in variation in articulation. Due to these factors visemes are not constant units. However, groups can be created based on the phonemes' frequency in a particular viseme group. For instance, in French language the vowels and consonants are classified into six groups based on their articulatory realizations [18]:

1. rounded vowels [y, o, …]
2. non rounded vowels [a, i, …]
3. bilabial plosives [p, b, m]
4. labio-dental fricatives [f, v]
5. post-alveolar fricatives with labial protrusion [ʃ, ʒ]
6. alveolar fricatives [s, z]

One phoneme can have one key mouth shape but a mouth shape may be used to represent different phonemes. Because of this, phonemes are frequently confused within a viseme group but the groups are mutually exclusive [18]. If we take one viseme group from above /p, b, m/, the three letters have similar mouth shape but they have distinctive mouse representation with other viseme groups.

2.2 Amharic Language

Amharic language consists of 33 basic and other additional symbols. Each basic symbol has seven different orders. The orders are created by combination of each consonant with the seven vowels (see table 2.1). In most languages there are five vowels /a, ë, i, o, u/. In addition to these 5 vowels Amharic has two additional vowels /e, ï/ [3].

Table 2.1 The seven orders of the Amharic letter "ሀ"

	Orders						
	1st	***2nd***	***3rd***	***4th***	**5th**	**6th**	***7th***
Symbol	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ
	/le/	/lu/	/li/	/la/	/l ë/	/l ï/	/lo/

According to (Henock, 2003) the seven vowels can be classified based on the position of the tongue. When pronouncing vowels the tongue may be vertically positioned high, mid, or low and as front, central or back horizontally. The lips may be rounded or un-rounded on each orientation of the tongue (see table 2.2). Bender, 1974 classifies the vowels into two: central and peripheral but we prefer Henock's classification [9] because it considers the lip gesture along with the orientation of the tongue.

Table 2.2 Orientations of tongue and lip when pronouncing Amharic vowels

	Front		**Central**		**Back**	
	Rounded	*Unrounded*	*Rounded*	*Unrounded*	*Rounded*	*Unrounded*
High		/i/		/ï/	/u/	
Mid		/ ë/		/e/	/o/	
Low				/a/		

In Amharic one symbol can represent consonant and vowel (CV) combination. This CV-syllable is pronounced similar wherever it's placed in words. A vowel cannot be pronounced differently with different words. For instance, in Amharic /ሐ/ is pronounced the same in any word appearance. However in English "a" is pronounced as /a/ in "cat" and as / ë/ in "Care".

Such variation induces a complexity in identification of manner of articulation for each existence of vowels in different words.

2.2.1 Amharic Visemes

As shown in table 2.1 each consonant has 7 orders in which it's pronounced differently with different vowels. Out of these 7 orders the sixth order is considered as basic and any order of a consonant is pronounced by integrating this order's pronunciation with the respective vowel of the order. For instance the third order of ሰ (ሲ) is the combination of the consonant "L" (ል) and the vowel "i" ሂ. Due to this, the classification of Amharic consonants is done to the 6th order. The common classifications based on place of articulation, manner of articulation and voicing are shown in table 2.3 and table 2.4. In Amharic there are no specifically defined viseme groups with the intention of the letters visual identity. We can only use the classification shown in the following two tables as our basis. However, some differences can easily be detected from the classifications. For example Henock, 2003 considers /ም/ has labio-dental [9] and Bender, 1974as bilabial [3]. I agree with Bender as /m/ is articulated by the labials. We can also consider the English viseme groups as cue, because the Amharic /ም/ and the English /m/ are similar. With such differences however, researches that depend on it may not be standardized. So, viseme groups formation is a topic that needs further research.

Table: 2.3 Classification of Amharic Consonants with respect to place and manner of articulation [3].

Amharic Consonants

	Labials	Dentals	Palatals	Velars	Labio Velars	Glottals
Stops & Affricatives						
Voiceless	P	T	c	k	k^w	?
Voiced	B	D	j	g	g^w	
Glottalized	P	T	C	K	K^w	
Fricatives						
Voiceless	F	S	Š			h
Voiced	V	Z	ž			
Glottalized		S				
Nasals	M	N	ñ			
Liquids		L, r				
Glides	W		y			

Table2.4. The Amharic consonants and their corresponding place and manner of articulation [9]

ARTICULATION PLACE									
			Bilabial	Labio-dental	Alveolar	Palatal	Velar	Labio-velar	Glottal
Manner of Articulation		Voiced	ብ /b/		ድ /d/		ግ /g/	/g ʷ/	
		Voiceless	ኘ /p/		ት /t/		ክ /k/	/kʷ/	ዕ (?)
		Ejective	ጵ /p'/		ጥ /t'/		ቅ /k'/	/k' ʷ/	
		Voiced		ቭ /V/	ዝ /z/	ዠ / ž /			
		Voiceless		ፋ /f/	ስ /s/	ሽ /Š /			H/h/ፙ /hw/
		Plosive			ጽ /s'/				
		Voiced				ጅ /j /			
		Voiceless				ች / ĉ/			
		Plosive				ጭ /c'/			
	Nasals			ም/m/	ን /n/	ኝ / ň /			
	Liquids					ል /l/	P /r/		
	Glides			ው /w/			/y/		

2.3 Text to Visual-Speech System (TtVS)

TtVS is a system that changes a text input into its viseme equivalents or visible speech. It has a close connection with Text to Speech System (TTS) [15, 19]. In systems like ArtSS [7] they are already integrated and considered as two modules of a single system.

TtVS has two phases:

1. Analysis phase: analyzing the text input which is used to select the components of the visual speech
2. Synthetic phase: synthesizing visual speech from the selected components in the first phase.

2.3.1 Analysis Phase

In this phase, each letter of the text is identified and converted into the corresponding linguistic representation or visemes. Conversion may be simple or difficult depending on the specific language. Amharic, as described in section 2.2.1, is a type of language in which a letter is pronounced similar in any word. This makes the conversion easier when compared to languages like English.

A text input alone has some drawback. It doesn't give detailed information on how to synthesize the visual speech other than text, prosodic information is very important. According to Henock, 2003 intonation, stress and duration are prosodic features. Prosody gives additional information that the uttered words cannot give alone and it can even change

the meaning of a sentence. Including prosodic information to written text makes the system more efficient than depending only on the written text.

2.3.2 Synthesis Phase

This phase can also be named viseme mapping. It can be performed by video rewriting and image morphing on real images or graphic approaches [2]. The first two are appropriate for systems like “Face translation” [13, 19] in which the spoken utterances of one person will be translated into another language which, the person is not capable of speaking. During translation the target language utterances will be integrated with an audio-visual speech of a synthesized face to be destined at the speaker’s face. (See figure 2.1). Face translation is applicable for different web-based customer service systems and telephone system with the help of audio-visual speech recognition [15].

In telephone systems when a person speaks on the phone his/her speech is identified and displayed for the other party as a visual speech. The same with web based customer service systems. In both cases rather than transferring data about face of the speaker for each articulation it is possible to transfer only the text and show the speech by morphing the corresponding articulation of the text on to a permanently stored face in the destination.

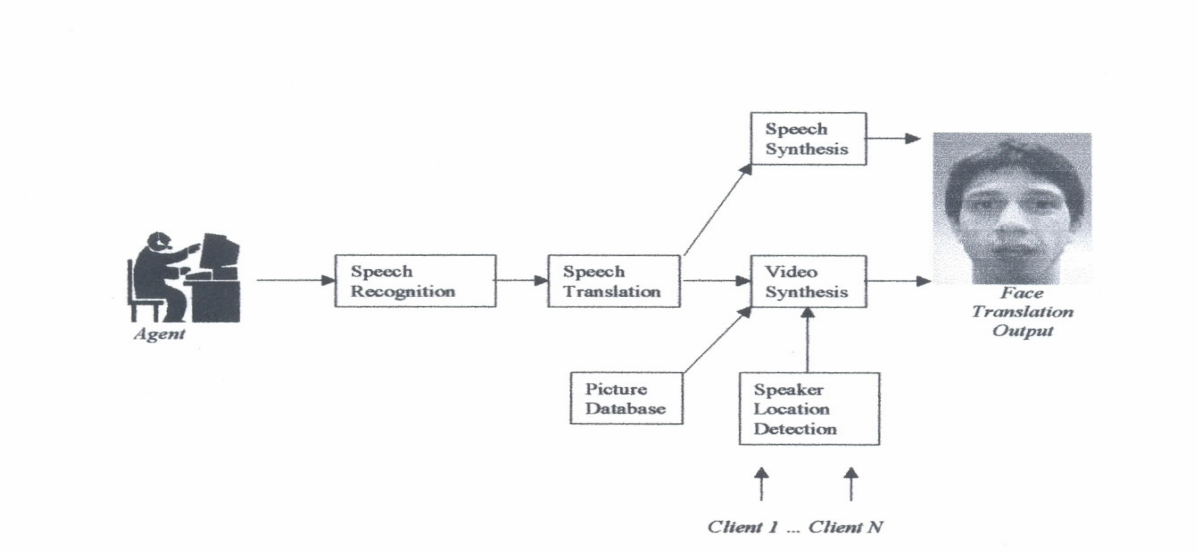


Figure 2.1 A picture from [19] showing the structure of Face Translation system

Even if video rewriting and image morphing technologies have such important and special capabilities, they cannot be used for vocabulary training purpose as they show only the external part of the face, a view of the inside of the mouth is highly informative in language training [6] (see figure 2.2). As skin hides much of the vocal tract, there must be a transparent face [12] to show the internal motions of articulators to facilitate the speech reading.

This hypothesis is supported by different researchers, especially in CSLU toolkit the talking head “Baldi” has shown a great efficiency in vocabulary tutor [10]. It is not difficult to conclude that the ICATIANI toolkit can also have same result as it follows a similar approach to CSLU but for Mexican-Spanish [10]. There are also other graphics-based talking heads for English language and one for Arabic named Baldr. For Amharic language there is no system developed to produce visual speech.

2.4 Talking Head

Language perceivers are viewed as having available multiple sources of information supporting the identification and production of language [13]. Here lays the basis for the importance of talking heads.

Talking heads are models of the human face to show visual-speech and developed with detailed design consideration on the vocal tract and the eye. The vocal tract is the part that shows place and manner of articulation and the eye is used to show some prosodic information to show the feelings of the speaker.

The development history of 3D talking heads is dated back to1975 since Parke developed a talking head using his software [16]. Since then different types of talking heads have been and being developed. Of these, the talking head Baldi has shown a great promise in language training for deaf students [5, 6, 12, 13, 14, 21, 30].

Developing talking heads comprises visual speech analysis and synthesis techniques. For making visual speech it is possible to use methods described in 2.3.2. However, as described in that section, for language training purpose graphics based visual speech synthesis is a better way. That's one of the reasons that make Baldi effective in vocabulary training. Such interactive language technologies offer very promising and cost effective means of increasing the quality of education for deaf children, particularly as it pertains to developing and refining spoken communication skills [21].

Some of the reasons for the efficiency of talking heads in language training are: they are well trained with different data and perform the tasks of a never tiring instructional

assistance. In addition, as the skin hides much of the vocal tract a transparent face is a means to uncover the hidden part in talking heads.

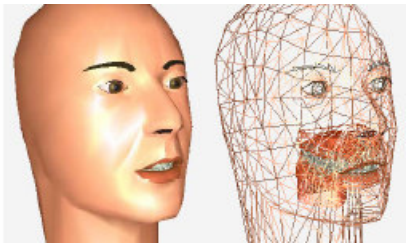


Fig. 2.2 Baldi, a computer animated talking head, in normal and wire-frame presentations [12]

2.4.1 The Human Vocal Tract

The human vocal tract plays a core role in producing speech. In the source/filter theory, the larynx is considered as the source and the vocal tract as filter to produce an output. The output of the filtered source is a speech.

Larynx	+	Vocal Tract	=	Output
(Source)		(Filter)		

The vocal tract consists of a number of articulators used in speech synthesis (i.e tongue, lips, teeth and palate). The main visually informative articulators are lips, tongue and teeth. The source of speech is thought to be the larynx (voice box) whose folds oscillate when air exhaled from the lung. In the vocal tract the acoustic energy is filtered to be an audible sound.

The vocal tract is not an interest only for acoustic output but for the articulatory view too. The different articulators of the vocal tract give information about the output speech even without listening to the sound output.

2.4.2 Modeling a Talking Head

Currently a number of researches are being done to model the human face in computers. The main component of the talking head is the vocal tract which comprises different articulators. Each of these articulators is very complex and often developed independently of other structures [24].

As mentioned in section 2.3.2 facial modeling and animation research falls into two major categories, those based on geometric manipulations and those based on image manipulations. Even if the existing models are not rated they differ in naturalness, usability and implementation. One can select modeling technique depending on the purpose of the model. For web based application for instance, a model that minimizes the size of data transfer is preferable. In such cases it's possible to store small regions of visemes in the database so that images between visemes can be synthesized using interpolations. On the other hand for speech learning tools, a model that is capable of displaying the internal of the vocal tract is preferable, which is a geometric model.

Most of the vocal tract modeling techniques is driven by data or images retrieved by medical equipment like ultrasound, electropalatography and MRI (Magnetic Resonance Indicator) and synthetic articulatory parameters [7, 22, 24]. Colored markers painted on the face or lips are also used to aid in tracking facial expressions or recognizing speech from video

sequences. The medical equipment has the capability of penetrating the skin and show the articulation of the tongue as well as the teeth even when covered by the lip. The quality of the model depends on the base image data. The vocal tract as described in the previous section consists of different substructures and each has a dozen of subcomponents that should directly be mapped into controlling parameters in the model. So the more quality and extensive data you get the more closer the model gets to perfection.

3. Designing a 3D-Lip model for Amharic Visual-Speech Synthesis

Lips are the most outer parts of the human vocal tract and they play a major role in visual speech synthesis. Developing a realistic 3DLip model means increasing understandability of speech, visually [18]. Lips can be modeled by one to one mapping of each and every component like muscles; but the controlling parameters will be very large in number and it will make the model too complicated. But the main concern in lip modeling is to make it articulate the visemes correctly and in an easily understandable way. Taking this in to consideration we tried to use a simple, with respect to controlling parameters, modeling technique and increase its intelligibility by deeply analyzing a real human lip in speech. The method we used is analysis-synthesis approach first to analyze the human lip in speech and to synthesis a lip model that can articulate as it is fetched from the real lip. We tried to show the analysis of a real lip from human speaker in section 3.1 and a detailed description of the model we developed in section 3.2.

3.1 Analyzing Lip poses in speech

There are two main approaches to analyze the lip in speech explicit and implicit feature extraction methods. Implicit feature extraction methods use the raw image data input to a recognition engine which learns the pixel patterns associated with certain lip movements. The recognition engine may use Hidden Markov Model, Quadtree Splines or Finite Elements model. All these models are suitable for visual speech recognition. In a noisy environment the auditory input of a speaker may not be enough for the listener; in such

cases visual speech analysis models can be used to understand what the speaker is saying in addition to the auditory signal. Speech-to-text systems also use these models to minimize errors in speech identification. On the other hand explicit feature extraction methods use image processing techniques, to find the location of certain feature points (e.g. like lip corners) in the image [8]. Most visual speech synthesis techniques use this method to get input data to develop models of articulators.

Reveret et al. 1998 classify explicit feature extraction methods in to two: texture-based and model based approaches. The texture based approach tries to separate the lip from the rest of the face without applying make-up or special lightning. This has some limitations as the color of the lip may resemble that of the tongue and the rest of the face. Model based approach considers the shape of the lip a priori. This approach is usually implemented by putting some marks on selected points on the lip. The location and the number of points differ with different models [8, 18].

We use marking method to analyze the lip gestures by putting some control points on the lip and track their coordinates when the lip moves. From the control parameters fetched on the lip, we build a synthetic lip that can articulate Amharic characters in a similar manner. [18] uses three contours (inner, median and outer) and 10 points on each contour to analyze the lip. But most of the points on the inner contour get covered during lip movement. To counter this, rather than taking three contours we used only two (inner and outer). The inner contour in our model is between the inner and median contours used by Reveret et al. 1998.

Each contour has a total of 12 control points (see fig 3.1); the two side points (left and right), the three points on the upper V-shaped part (cupidon arc), two points between the two side

points and the cupidon arc, the lower lip's three points in the middle and two intermediate points between the middle and the side ones. Along with points that lie on the lip corners we use eight intermediate points four in each contour and on each contour two on the upper lip and two in the lower one. These points are required to minimize the gap between the control points in the corners so that we can make the curves smoother than a curve without their appearance.

Based on this point distribution, there will be a total of 24 points to be tracked of; but the two side points on the inner contour are hidden in normal (the lip gesture when the person is not speaking) as well as most other lip gestures so we used the coordinates of the outer curve's side points for the side points in the inner contour. This makes the number of points to be tracked of to be 22.

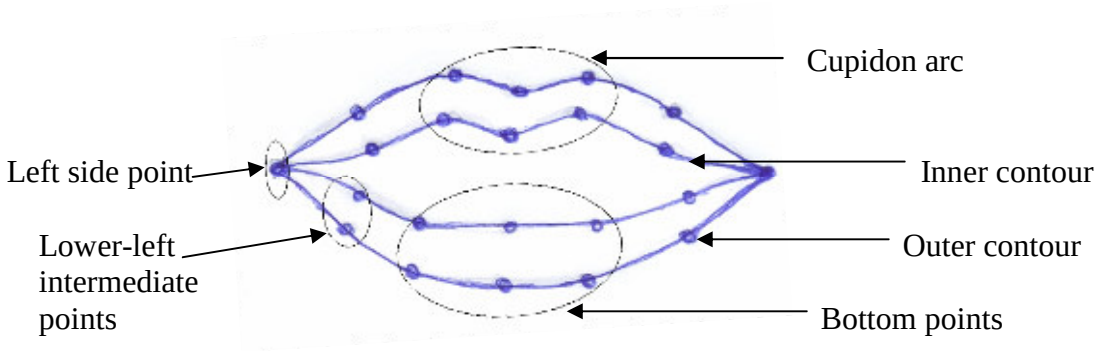


Fig 3.1 Outer and inner contours in our lip model along with selected control points

Along with this shape information we need the exact coordinates of the 22 control points in both inner and outer contours in normal lip pose and during speech. First we put marks on the selected lip corners of a real human speaker and analyzed the coordinates of the control points in normal lip pose to initialize the lip model and then we analyzed the respective

location of the control points on different visemes of any given corpus by a native Amharic speaker.

The coordinates of the points on the lip are measured taking the head position as our reference (see fig. 3.2). When we speak, the lip and its surrounding parts are not fixed. To have correct measurement the coordinates of the points on the lip are adjusted according to the coordinates of the point on the head which is relatively static.

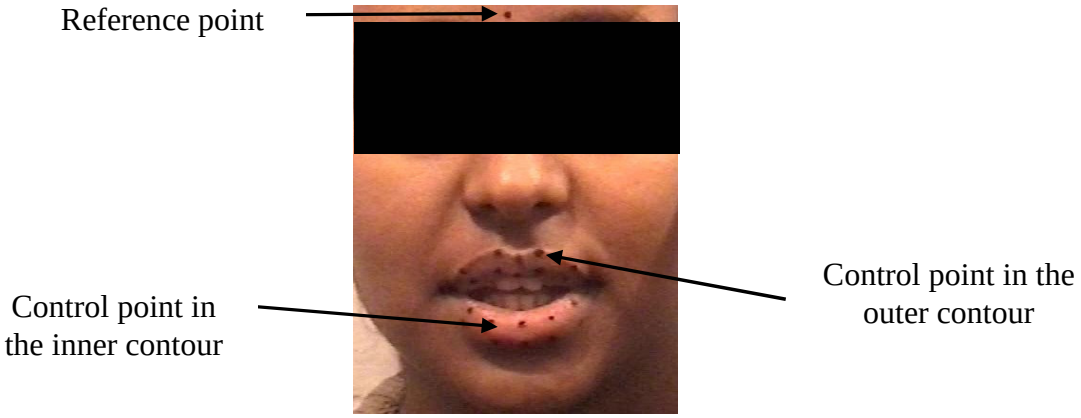


Fig 3.2 The points on the two contours of the lip and the reference point on the head of a speaker.

As described in section 2.3.1, Amharic consonants are pronounced by combining the sixth order of any consonant with one of the seven vowels. So we have to analyze the articulation of the seven Amharic vowels and the sixth order of the consonants that are combined in every possible consonant-vowel (CV) syllables.

For tracking the coordinates of the control points the speaker will be filmed by a video camera. To have 3D coordinates of the points we need two views of the speaker: front view and profile view. One method of getting these two views of the speaker is using two video

cameras one in each side and track the coordinates of the points from the films. This may incur some inconsistency on our measurement. To avoid this problem, rather than using two video cameras in front and by the side of the speaker, a mirror is placed at an angle of 45 degree to the speaker's face (see fig. 3.3) and filmed the speaker with only one video camera. With this arrangement it is possible to get the coordinates of the control points on both views using a single video camera which will avoid the possible mismatch in control points coordinate measurement if it had been from two video cameras.

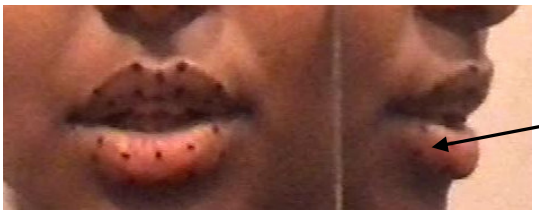


Fig 3.3 Mirror image of a speaker to show profile view of the lip

After we get a video film of the speaker in speech making for the consonants as well as the seven vowels, a video de-compiler tool named "OSS Video De-compiler" is used to extract frames in JPEG format from the video file. From the images we extract using the de-compiler the coordinates of the control points can be tracked using MS-Paint program. During coordinate extraction we encounter some hidden points during speech, especially points in the inner contour when the lip protrudes to the inside of the mouth. This induces the problem of missing data (see fig. 3.4).

Fig. 3.4 The right & left intermediate points between the side and cupidon arc ones in the upper inner contour are hidden as the lip protrude to inside of the mouth.

The coordinates of the hidden points can be retrieved by calculus of finite difference. In calculus of finite difference if there are m functional values and we are given n of them where m=n+1, it is possible to get the missing one value. One letter can be represented by using ‘k’ number of frames. Each frame contains a total of 22 control points in the inner and outer contours. Out of these ‘k’ frames a point may not be visible in one of them; so it is not possible to directly measure its coordinates. At this time we can use calculus of finite difference and the coordinates of similar points on the other frames to generate its coordinate on the frame in which the point is hidden. The coordinate of the hidden points is calculated from the coordinate of a similar point but on a different frame not from the other points on a similar frame. Then the coordinates of all control points on lip poses are stored in a database specifying to which letter and to which frame of the letter the points belong to. This ends the analysis phase of a real lip whose output will be used to synthesize the lip model and make it articulate different characters in a similar manner to the human speaker for all vowels and some selected consonants.

3.2 Synthesizing 3D-Lip model

This phase deals with developing a lip model driven by the fetched data from the analysis phase, described in the previous section. As described in section 3.1, the lip can be modeled by the inner and outer contours. To generate these contours if we draw lines in between the control points it will not look like a curve, but a polygon as the points are placed wide apart. This will degrade the realistic nature of our model.

If the points were very near to each other, it would have been possible to draw lines in between and generate the required contours. But it is not feasible to increase the number of control points to be marked, as their selection takes the shape of the lip into consideration. To counter this problem we used Bernstein polynomial to generate intermediate points with a very small interval, to the extent required, in between the extracted control points. The sequence of very short lines between these mathematically generated points however can create a Bezier curve, which gives the lip model a reality.

Bezier curve is a curve which is exactly determined by a set of control and generated points. Each point of the curve is calculated from a parametric mathematical function which uses the coordinates of the control points as parameters. Bezier curves do have a number of advantages: firstly they are very smooth and if we want to zoom in it is only a matter of increasing the space between the control points and we can redraw a perfectly smooth curve again. Secondly they are the adopted solutions to describe an image in terms of its mathematical representation; this has a positive impact in computer graphics especially with data volume. If we can describe a graphical object mathematically it’s possible to generate it

using the mathematical function on small number of parameters compared to what is needed without using such functions. This decreases the volume of data to be stored.

3.2.1 Graphical Rendering

In this synthesis phase we have to do two graphical renderings: Interpolation and Projection. When we de-compile the video of the speaker it will result in to many frames. Storing the coordinates of the control points on all these frames is not practical so we select some frames from the entire set. The frames are selected with an interval of five. So if there are 45 frames for the articulation of one character we choose 9 of them. However, when we synthesize the lip model, we need all the frames. So to regenerate the frames we use interpolation on the selected or sample frames.

3.2.2 Articulation variation

It was mentioned in section 2.1.2 that different factors affect the articulation of characters. Gender is one of the major factors that can affect it. Taking one sample speaker was not found to be enough so we analyzed the articulation of two persons male and female. In order to have the effect of both samples on our model, during synthesis we first base our model on the normal lip gesture of any of the speakers. Then during articulation of selected character we first retrieve the coordinates of the control points when both speakers articulate the same character. This gives us the deviation of each point from the normal one. The average of the deviation values then can be used on the lip model to show the articulation of the character.

3.3 System Architecture

The system accepts user input character for which a lip articulation is to be shown. In Amharic language one letter represents CV combination. So, whenever a person chooses a letter the system identifies the consonant as well as the vowel to be articulated. Once the consonant and vowel are identified, the next step is finding which lip gesture can make up the required articulation.

Then the database class identifies the coordinates of control points on the lip poses that correspond to the chosen letter and fetch it to the drawing class in which the curves will be painted using the method described in the previous section. The system then shows different modeled lip gestures of the chosen letter in sequence with some time interval. (Fig 3.5 shows the architecture of the system)

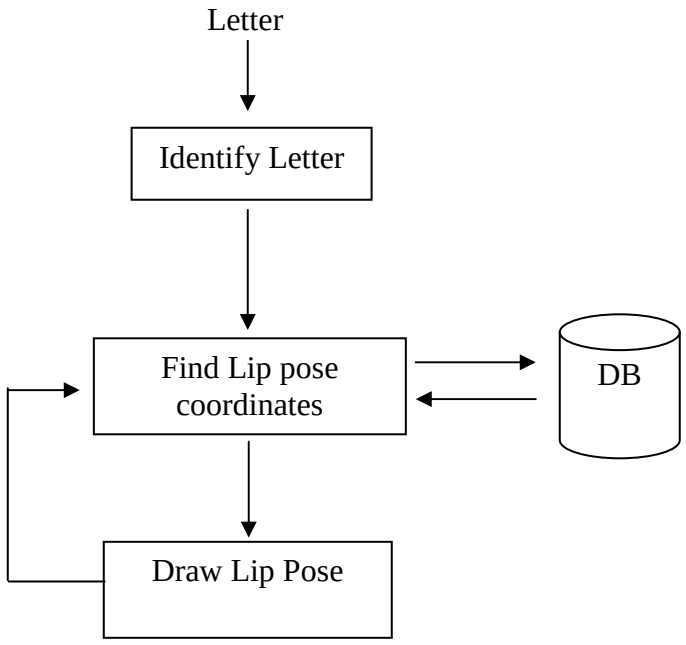


Fig 3.5 Architecture of the system

The relational DBMS, Microsoft Access is selected for persistent data storage. The database comprises tables to store the inner and outer contours' control point coordinates. For each lip pose of any given letter we store the letter's x, y and z coordinates of each control point, and lip pose number as a given letter has more than one lip pose.

4. Implementation and Testing

The design of the system is described in the previous chapter; in this chapter we will look at the prototype in detail.

We used Java Programming Language and MS-Access for the implementation. We used “jdbc:odbc” connection to access the database. The database stores the coordinates of control points on the inner and outer lip contours which we get from the analysis of human lip. There are four tables that store such coordinates for the male and female speakers used as samples. For one speaker we use two tables, the first one to store the inner lip contour coordinates and the second for outer lip contour coordinates.

The coding part is to first identify the selected character and retrieve the corresponding coordinates from the database. Then it applies some graphical rendering and generates a lip and displays it on the screen. The lip always starts from normal pose and ends up with the same. In the program there are 3 classes: DrawLip, Database and Projection.

Class DrawLip: This class has the four applet methods “init”, “paint”, “update”, and “run“ which we override some of them to suite our model and additional action listener method that listens to user action. “init” method initializes’ all the required variables and elements of the screen. “paint” displays the lip given the coordinates of the lip poses. By lip we mean the two contours and the space between them. In order to draw the contours it uses drawLine method with a very small length. “update” is called whenever there is a change in the lip. “run” method is used to retrieve the appropriate

coordinates of lip control points when the user selects a character and invokes the paint method. In addition it also interpolates in-between frames, because the database stores only coordinates for some frames of a given character.

Class Database: this class is used to connect with the data base through “jdbc:odbc” connection. It provides a method named “getCord”. This method is invoked by the methods that require coordinates of lip control points. This method is given a letter and lip pose number and returns coordinates.

Class Projection: the projection class is used for 3D to 2D mapping. It’s usually invoked when other methods need to paint the lip on the screen. It accepts 3D coordinates of points and converts them through its constructor method.

4.1 Demonstration

To show the functionality of the system we selected the letter “ሙ“/M/. It is selected because of two reasons. First it is a bilabial consonant, the major articulators are lips, and next it is the most frequently used letter in Amharic.

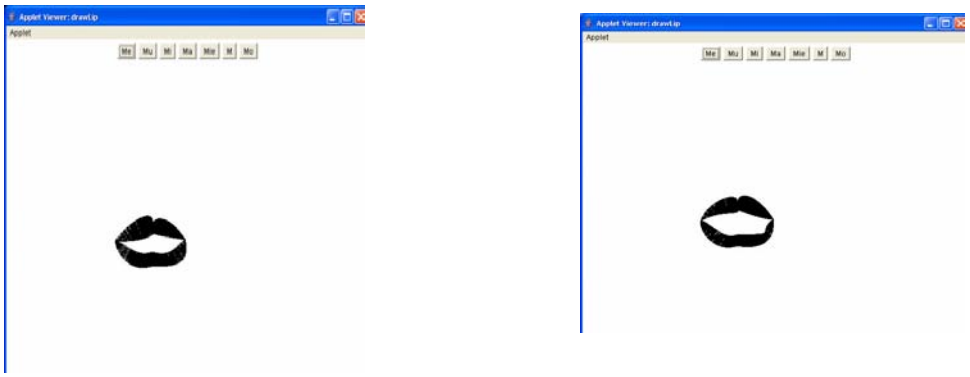


Fig. 4.1 Different poses of the lip model in articulation of the letter “ሙ”/M/

The system here shows fully painted lip. However, in full fledged talking heads, we may need to show the internal vocal tract organs. In order to allow such functionalities, the model has wire frame representation (See fig 4.2). With the wire frame representation the articulation doesn't differ but the lip is not fully painted. It has some openings to show the inside of the lip. The size of the opening can be made small or large.

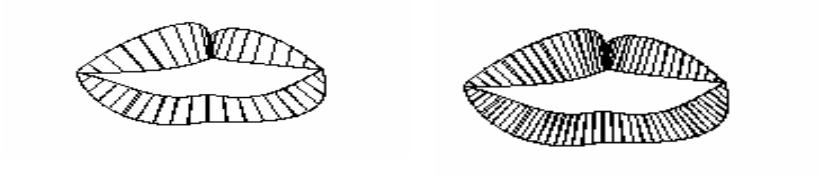


Fig 4.2 Wire-frame view of the lip model with different spacing

4.2 Testing

The Lip model is tested with 15 male and 15 female hearing impaired students of "Mekanissa Deaf School". The group comprises students from KG level up to grade 8. We showed the lip model to the students articulating each of the seven orders of the Amharic letter "ሙ". Then we asked the students which letter is articulated. They were allowed to try three times for each articulation by the lip model. We gave them three chances out of the seven choices because the orders have similar articulation, and can create ambiguity unless other vocal tract organs are shown. For example, the articulation of "ሙ" and "ማ", "ሚ", "ም", and "ሚ " looks similar and the same with "ሙ" and "ሞ". The result of the test is shown in tables 4.1 and 4.2.

Table 4.1 Result of Students

	ሙ	ሙ	ሚ	ማ	ሚ	ም	ሞ
No. of Students with correct articulation	17	23	15	20	24	21	27
Percentage of Students with correct articulation	56.7%	76.7%	50%	66.7%	80%	70%	90%

This result shows that the letter "ሞ" is the most correctly identified letter from the seven orders. And the letter "ሚ" is the least correctly identified one. As the articulation of "ሞ" resembles only "ሙ" the students can easily identify it within two or three trials. But for "ሚ" it resembles other four letters and the probability of identification is less. In addition, when we articulate this letter the tongue will be vertical in the middle of our mouse and it is clearly seen to the listener. As our model doesn't include the tongue the students will have less chance to identify this letter.

Table 4.2 Result of Students by gender

		ሙ	ሙ	ሚ	ማ	ሚ	ም	ሞ
No. of Students with correct articulation	**Male**	10	10	6	11	12	8	13
	Female	7	13	9	9	12	13	14

There is some difference in the result of male and female students but as the sample is very small we couldn't conclude about a difference in the ability of males and females in speech reading and it needs further research.

5. Conclusion and future works

In this research we tried to explore the problems with the Deaf and hard of hearing people in language perception, education and social life. We tried to look at the existing solutions, for the problems of the Deaf, along with their limitations. The problems with Deaf are mainly due to lack of enough communication with the hearing society. It is better for the Deaf to get adapted to the customs and norms of the hearing society. Having their own culture in hand it will be better for them to get mixed.

The main means of communication for the Deaf is sign language, but its use by the other society is very limited or not significant. If a deaf person is not mute s/he can speak and listen. Communicating with oral language enables them to share knowledge and experience. But they couldn't listen to what others say and couldn't speak as they lack their hearing capability. Speech, however, has visual component in addition to auditory signal. If the Deaf get speech training these problems can be solved.

However, speech training is not an easy task. It demands intensive and one-to-one training by speech and language therapists, it is tiresome and time consuming task and the training must be adaptive to the trainee's need.

Taking all these into consideration we proposed vocabulary training system to enable the Deaf to listen (visually) and speak. The entire training system is very complex and can not be done with the limited equipment and tools we have access to. The system will be effective if we have different medical equipment that can analyze the articulation of the

vocal tract, the eye gaze and part of the face around mouth. In addition, it needs special tools to interpret the data that we get from the equipment.

Due to these limitations, we have developed only model of a lip. We selected lip because it is the outer and main vocal tract part. The complexity of the muscular structure of the lip however makes it difficult to completely model physiologically and it also needs medical equipment that can penetrate the skin to show the inner tissues and muscles. However, we get at least some external parameters from the outer part of the lip that can be used as a basis for developing the model.

We have seen different talking head models and specifically lip modeling techniques. Our modeling approach is analysis-synthesis which is used to first analyze lip articulation from human and synthesize the lip model. This approach is implemented by developing a system that shows the model articulating a selected Amharic letter. Our model is tested with 30 students (15 male and 15 female) from “Mekanissa Deaf School” and it shows 70% accuracy. From the result we have also seen that letters with less similarity than the others and can be seen only by lips are easily readable. And the letter that includes visible tongue movement along with the lips has some difficulty to be identified. This shows us that our lip model can be more efficient if it is integrated with other articulators.

This thesis work is the start for automated Amharic vocabulary training system for the Deaf. It has great impact in the education of the Deaf and their life as well. As said above, to make the system complete there should be additional vocal tract models to be integrated with this one. The result of the test also shows us that the developed lip model is more efficient in letters that can be articulated by only the lip.

Our model is developed taking the future work into consideration. We also identified most of the important elements needed for the implementation of the entire system and it’s an open research. In the future, our work can be further extended and strengthened to be used for the training of the Deaf.

References

[1] "ሀ" የአማርኛ የምልክት ቋንቋ መስማትና መናገር ለተሳናቸው፣ በትምህርት ሚኒስቴር ፣ በመካኒሳ የደናቁርት ትምህርት ቤት ፣ ቦሌ በዕዝነ ጽሙማን ትምህርት ቤትና በኢትዮጵያ ደናቁርት ማህበር የተዘጋጀ (1963ዓ.ም)

[2] Barker L. J., (2003) ***Computer Assisted Vocabulary Acquisition: The CSLU vocabulary tutor in oral-deaf education.*** Journal of Deaf Studies and Deaf Education, 8, 187-198

[3] Bender, M.L. (1974). ***Amharic Verbal Types from Text & Lexicon***. P. 132-139

[4] Chen X.W. and Yang J., ***Visual Speech Synthesis Using Quadtree Splines***. *Proceedings of ICASSP 2001*, Salt Lake City, May 2001.

[5] Cohen M. M., Beskow J., and Massaro D.W. ***Recent Developments in Facial Animation: An Inside View.*** Proceedings of Auditory Visual Speech Perception '98 (pp 201-206). Terrigal-Sydney Australia, December, 1998.

[6] Cole R., Carmell T., Connors P., Macon M., Wouters H., Jacques de Villiers, Tarachow A., Massaro D. W., Cohen M., Beskowt J., Yang J., Meier U., Waibel A., Stone P., Fortier G., Davis A., Soland C. (1998). ***Intelligent Animated Agents for Interactive Language Training***. Proceedings of Speech Technology in Language Learning, Stockholm, Sweden

[7] Fels S. S., Vogt F., Gick B., Jaeger C. and Wilson I. ***User-centered Design for an Open-source 3-D Articulatory Synthesizer.*** In ICPhS, 2003.

[8] Göcke R, Tran Q. N., Millar J. B., Zelinsky A., and Robert-Ribes J. ***Validation of an Automatic Lip-Tracking Algorithm and Design of a Database for Audio-Video Speech Processing.*** Proceeding at the 8th Australian International Conference on Speech Science and Technology SST-2000, December 2000

[9] Henock L. (2003) ***Concatnetive Text-to-Speech Synthesis for Amharic Language***. MSc Thesis AddisAbabaUniversity 2003

[10] Kirschning I., Phd, ***CSLU Toolkit-based Vocabulary Tutors for the Jean Piaget Special Education School***. Proceeding of InSTIL/ICALL2004-NLP and Speech Technologies in Advanced Language learning Systems-Venice 17-19 June, 2004

[11] LOTTIE. L, R.(1963) ***TALK TO THE DEAF*** 20th edn. Gospel Publishing House/ Springfield, Missouri

[12] Massaro D. W. (2003) ***A computer-animated tutor for spoken and written language learning.*** Proceedings of the 5th international conference on Multimodal interfaces. Vancouver, British Columbia, Canada. Pages: 172 - 175

[13] Massaro D.W. ***From theory to practice: Rewards and challenges.*** In Proceedings of the International Conference of Phonetic Sciences, San Francisco, CA, August 1999.

[14] Massaro D. W. ***Symbiotic Value of an Embodied Agent in Language Learning.*** Proceedings of the 37th Hawaii International Conference on System Sciences-2004

[15] Massaro D.W., Beskow J., Cohen M. M., Fry C. L., Rodriguez T. (1999) ***"Picture My Voice: Audio to Visual Speech Synthesis using Artificial Neural Networks"***, in *Proc. AVSP'99*, Santa Cruz, USA.

[16] Massaro D.W., Ouni S., Cohen M.M., Clark R.. (2005) ***A Multilingual Embodied Conversational Agent***. IEEE proceedings of 38th Annual Hawaii International Conference on System Sciences (HICCS '05)

[17] MICHAEL K. (1990) ***COMPUTER ASSISTED LEARNING, Selected Proceedings from the CAL'89 Symposium.*** 11-14 April 1989, Pergamon Press.

[18] Reveret, L*, Benoit, C. . ***A New 3D Lip Model for Analysis and Synthesis of Lip Motion in Speech Production.*** Proceedings of the Intrnaitonal

Conference on Auditory-Visual Speech Processing, December 1998, Terrigal, Australia.

[19] Ritter M., Meier U., Yang J., Waibel A., ***Face Translation: A Multimodal Translation Agent.*** In proceedings of Auditory-Visual Speech Processing (AVSP 99), 1999.

[20] ROBERT L. BLOMEYER JR, DIANNE MARTIN (1991) ***CASE STUDIES IN COMPUTER AIDED LEARNING***. The Falmer Press, London

[21] Stone P. (1999) ***Revolutionizing Language Instruction In Oral Dead Education.*** Proc. 14th International Congress of Phonetic Sciences; San Francisco, U.S.A.

[22] Tiddeman B., Perrett D. ***Prototyping and Transforming Visemes for Animated Speech.*** *Proc. Computer Animation 2002, Geneva, Switzerland*. 2002

[23] Visser M. [2], Poel M. [1], Nijholt A.[1] (1999) ***Classifying Visemes for Automatic Lipreading.*** Springer Berlin / Heidelberg p. 349

[24] Vogt F., Fels S.S., Gick B., Jaeger C., Wilson I. ***Extensible infrastructure for a 3D face and vocal-tract model***. In ICPhS, 2003.

[25] Vogt F., McCaig G., Ali M. A., Fels S. ***Tongue 'n' Groove: an Ultrasound based Music Controller.*** Proceedings of 2002 Conference on New Instruments for Musical Expression (NIME-02), Dublin, Ireland, May 24-26, 2002.

[26] The free dictionary, encyclopedia article about Deafness URL: http://encyclopedia.thefreedictionary.com/deafness page accessed on 12/2/2004

[27] Issues on hearing loss URL: www.hearinglossweb.com page accessed on 12/2/2004

[28] Types of Deafness. URL: www.idcs.info/childhood_deafness/types_of.html page accessed on 13/01/2004

[29] World Federation of the Deaf. URL: www.wfdeaf.org/pdf/policy_adult_ed.pdf
page accessed on 10/4/2005

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.

Declared by:
Name: _____
Signature: _____
Date: _____

Confirmed by advisor:
Name: _____
Signature: _____
Date: _____

Place and date of submission: Addis Ababa, August 2006.