RESEARCH ARTICLE OPEN ACCESS

Real-Time Speech Synthesis by Handling Amharic Gemination Words Using Statistical Technique

Walelign Tewabe Sewunetie [1], Eshete Derb Emiru [2]
Department of Information Technology
DebreMarkos University
DebreMarkos
Ethiopia

ABSTRACT
This study describes the Text-to-Speech (TTS) system for the Amharic language, using speech synthesis architecture of Festival. The system is developed based on diphone unit concatenative synthesis by applying Residual Linear Predictive Coding technique. The conversion process from input text into acoustic waveform is performed in a number of steps consisting of functional components. Procedures and functions for the steps and their components are discussed. Finally, the performance of the system is measured and the quality of synthesized speech is assessed in terms of intelligibility and naturalness. The evaluation results indicate that the majority of the words are recognizable. It is measured in MOS scale and the intelligibility and naturalness of the system is found to be 3.5 and 3.05 respectively. The overall performance of the system is found to be 79.25%. The values of performance of intelligibility and naturalness are encouraging and show that diphone speech units are good candidates to develop fully functional speech synthesizer. But, still there are areas need further investigations. Pensiveness of all type of gemination words and those ambiguities found in Amharic sentence, it may possible to handle using statistical technique based on their context. In addition the further work needed to be focus on expanding the pronunciation lexicon, the G2P or letter to sound rules to provide accurate pronunciation.
Keywords:- Amharic language, Diphone, Festival, Gemination, Speech Synthesis, TTS.

I. INTRODUCTION

Speech synthesis is the artificial production of human speech also known as a text-to-speech (TTS) system. It converts normal language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech [1]. A text-to-speech system (or "engine") is composed of two parts: a front-end (analysis) and a back-end (synthesis). A computer system used for this purpose is called a speech synthesizer, and can be implemented in software or hardware products.

As Donovan [1] discuss, speech synthesis techniques can be broadly divided into two categories: system model approach and signal model approach. System model approach tries to model directly to the human vocal system and this technique is also called articulatory synthesis. On the other side, signal model approach attempts to focus model speech signal only and with two varieties namely; rule-based formant synthesis and concatenative synthesis.

The quality of a speech synthesizer is judged by its similarity to the human voice and by its ability to be understood. An intelligible text-to-speech program allows people with visual impairments or reading disabilities to listen the written works on PC or Mobile Phone.

In the past decade, Many researches have been done on modeling speech synthesis for different language such as English, Dutch, Spanish, Finnish and Germany, and when we see attempts in case of Amharic language there are few published work done on modeling speech synthesizer. Even though, there are technological attempts on speech synthesizers for Amharic language, they are not alleviating problems related to speech naturalness and intelligibility aspects. In this study, the researcher tries to address gemination issues related to naturalness and intelligibility of synthetic speech.

Gemination is the main problem in TTS for Amharic Languages in amharictibiket/ጥብቀት/- (consonant lengthening) is not normally indicated in the Ge‟ez script. This consonant lengthening can distinguish words from one another; for example, "አለ"- alä means 'he said' and "አለ"- allä means "there is"; ይመታል-"yəmätall" means "he hits", and ይመታል-"yəmmättall" means "he is hit". Since it plays a significant role in the morphology of Amharic language, addressing this issue is important for speech applications. Hence the main objective of this research work is to study the possible methods to develop Text to Speech System for Amharic language. Therefore the researcher tried to improve naturalness of speech produced by Amharic speech synthesizer by handling germination problem, to the best of the researcher's knowledge.

In this paper, we describe the implementation and evaluation of a Amharic text-to-speech system basedon the diphone concatenation approach. The Festival framework [2] was chosen for implementing theAmharicTTS system. The Festival Speech Synthesis System is an open-source, stable and portable multilingual speech synthesis framework developed at the Center for Speech Technology Research (CSTR), of the University of Edinburgh.

The organized of this paper: The first section presents an introduction to the subject matter, statement of the problem and objectives of the research etc. The Amharic phonetics are discussed in section two. Section three provides the role of germination amharic language are discussed in detail. Section four explains about the diphone database preparation in festival toolkit. Section five discusses experimentation and testing of the Amharic speech synthesizer in detail. The last section provides conclusion, contribution and recommendations for future research.

II. PHONOLOGY OF AMHARIC LANGUAGE

Amharic (አማርኛ) is a Semitic language and it is one of the most widely spoken languages in Ethiopia. Being a Semitic Language of the Afro-Asiatic Language Group, this language is related to Hebrew, Arabic, and Syrian. Unlike Arabic, Hebrew, or Syrian, Amharic language is written from left to right [3]. It is one of the most widely spoken languages in Ethiopia. It has its own non Latin based syllabic script called "Fidel" or "Abugida". Among 73 languages which are registered in the country, Amharic is the official language of Ethiopia. Amharic is the widely spoken language and is one of the Semitic languages having its own script. The scripts are more or less orthographic representation of the phonemes in the language [4].

Amharic uses a unique script, which has originated from ancient language, the Ge'ez alphabet, which is the liturgical language of the Ethiopian Orthodox Church [5]. The script of Amharic language is phonetic in nature. On the other hand the orthographic representation of the language is organized into orders (derivatives) as shown in Table 2-1. Six of them are CV (C is a consonant, V is a vowel) combinations while the sixth order is the consonant itself. In total there are 32 consonants and 7 vowels with 7x32= 224 Syllables. Moreover there are extra orthographic symbols in the language that are not organized as above. The total number if orthographic symbols of the language exceed 230 [6].

Table 2-1: Amharic consonants /መ/ and /በ/ with their associated vowels. [31]

Order	1st	2nd	3r d	4th	5th	6th	7th
V / C	e	u	ii	a	ie	ix	o
/m/	መ	ሙ	ሚ	ማ	ሜ	ም	ሞ
/b/	በ	ቡ	ቢ	ባ	ቤ	ብ	ቦ

Consonants, as opposed to vowels, are characterized by significant constriction or obstruction in the pharyngeal and/or oral cavities, that some consonants are voiced; others are not. Amharic consonants according to manner of articulation are generally classified as stops/ዝግ/, fricatives/ተሽሎክላኪ/- formed when the structure is very narrow (but without total closure) so that when air flows out, a hissing noise is made, lateral- is made when air flows out of the sides of the mouth, nasals/የአፍንጫ/- air to flow out of the nose, liquids/የምላስ/, and semivowels/ግማሽአናባቢ/. Table3-2 shows the phonetic representation of the consonants of Amharic as to their manner of articulation, voicing, and place of articulation are voiceless /የማይነዝር/, voiced/ነዛሪ/, and glottal/የጉሮሮ/ [3].

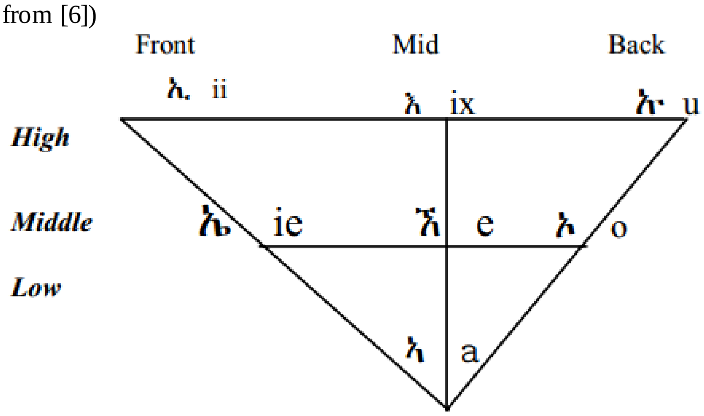
Table 3 2: Phonetic representation of Amharic consonants [3].

		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	ኅ
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						

Vowels are always voiced sounds and they are produced with the vocal cords in vibration [7]. The tongue shape and positioning in the oral cavity do not form a major constriction of air flow during vowel articulation [8]. Vowels have considerably higher amplitude than consonants and they are also more stable and easier to analyze and describe acoustically [7].

Vowels can be classified by the position of the tongue and the lips. The tongue and the lips produce different vowels by altering the shape of the vocal trace and enable the vibrating air produce sound in which different frequencies are emphasized [8]. The IPA (International Phonetic Association)- responsible for standardizing representation of the sounds of spoken language defines a vowel as a sound, which occurs at a syllable center [9]. A chart depicting the Amharic vowels in the IPA representation is shown Figure 2-1.

Figure 2-1: Vowels with their features (mainly adopted from [6])



According to the above Figure 3-3, depending on the position of the lip the Amharic vowels (አ፣ኡ፣ኢ፣ ኣ፣ኤ፣እ፣ and ኦ) are broadly categorized into rounded (ኡ and ኦ) and unrounded (ኣ፣ኢ፣ኣ፣ኤ and እ) [3].

III. ROLE OF GEMINATION, EPENTHESIS AND SYLLABIFICATION OF SPEECH SYNTHESIS

Gemination is most conventionally described as thelengthening of the consonants. Gemination in Amharic is one of the most distinctive characteristics of the cadence of the speech, and also carries a very heavy semantic and syntactic functional weight. All the consonants, except /ሀ/ [h], can occur either in geminated or non-geminated form [10]. In most Ethiopic based scripts, no special representation is used for a geminate consonant different from the non-geminated one [11]. The famous Ethiopian novelist HaddisAlemayehu, who was an advocate of Amharic orthography reform, indicated gemination in his novel "FəqərƎskäMäqabər"-"ፍቅርእስከመቃብር", by placing a dot above the characters whose consonants were geminated, but this practice has not caught on [3].

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 [10].

Table 3-1: Some example of Gemination words with Minimal pairs of Singleton vs. Geminate Consonants

Singleton			Geminate	
Orth.	Pronunc	Gloss	Pronunc.	Gloss
ገና	gəna	still/yet	gənna	Christmas
ሽፍታ	ʃɨfta	outlaw	ʃɨffita	rash
አለ	Ale	He says	alle	He present
ይሰማል	yɨsəmal	he hears	yɨssəmmal	he/it is heard
በላ	Bela	Continuous to speek	bella	He eat
ለጋ	ləga	fresh	ləgga	he hit
ሰፊ	səfi	tailor	səffi	Wide
ወጥ	Wet	stew	wett	Constant or (someone/ something) that is likely to go out
በራ	bera	bald	berra	Blow/burn out
ዋና	wana	swimming	wanna	Main, Principal
መሳም	mesam	to kiss	messam	To be kissed
ይበላል	yibelal	(he) will eat	yebbelal	(ht) can be eaten
ይመታል	yimetal	(he) can hit	yimmetal	(he) can be beaten
ይጠጣል	yetetal	(he) will drink	yettetal	(it) can be drunk

To analyze the frequency of gemination words in amharic corpus, we have prepared Amharic corpus which have a total 475,860 words. Most of the words were collected from a blog called www.danielkibret.com, which is social, political, economical, philosophical, cultural, spiritual issues raised.

Table 3 6: The total number of some gemination words in our corpus.

Word	ገና	ይመታል	አለ	ሽፍታ	ይመታል	ለጋ	ሰፊ
Frequency	183	2	757	4	2	1	31
Word	ወጥ	ይጠጣል	ዋና	መሳም	ይበላል	በላ	በራ
Frequency	52	2	133	0	12	8	4

Total	1191

As shown in the above table 3-2 the total number gemination words are insignificant, out of 14 gemination words 3 of them (አሉ, ገና and ዋና) are more frequently occurred than others. So due to time constraint the researcher tried to create a mechanism to handle gemination problem only for top three more frequently occurred gemination words.

IV. DIPHONE DATABASE CONSTRUCTION

The basic idea behind building diphone databases is to clearly list all possible phone-phone transitions in a language. This makes the incorrect but practical and simplifying assumption that co-articulatory effects never go over more than two phones. We do not always pronounce a phone the same way, its pronunciation depends on its neighbouring phones. This is known as the co-articulatory effect. Festival relies on the simplifying assumption that the co-articulatory effect does not extend across more than a pair of phones.

The first step in building a diphone database for Amharic is to determine all possible diphone pairs. In general, the typical diphone size is the square of the phone number for any language [12] [8]. As discussed above, Amharic has 33 basic consonant phonemes, five of them consonants have the same sound and seven vowels. This results in 34 possible phonemes. Since we are interested in possible phoneme combinations, i.e. diphones, we get 34 times 34 = 1156 diphone pairs prepared. However not all phone-phone pairs occur physically in a language.

Designing, recording, and labeling a complete diphone database is a laborious and a time consuming task. The overall quality of the synthesized speech is entirely dependent on the quality of the diphone database. This section describes the methodology adopted in the construction of amharicdiphone database. TTS development was primarily carried out on linuxredhat family Fedora 8 operating system. The latest versions of the tools (Edinburgh Speech Tools, festival, festvox, festvox_CMU, festvox_OALD, and festvox_POSLEX) were downloaded, compiled, configured and tested with the pre-packaged voices.

At a high level, developing a new open domain voice requires the specification of a model for the language concerned along with a recording of all diphones that can possibly occur within that language. In festival, the language modelling is specified within a set of Scheme files. The development of these language specific files and the diphone database is assisted by the festvox scripts.The main steps in the process are discussed below.

A. Defining the Phone Set

Before advancing to other steps in the database creation process, Amharic language phoneset definition is essential as it is used in later stages. Accordingly, all the phonemes, including phones which are representing silences are prepared; it includes the type of vowel, consonant phoneme descriptions along with their length, height, place of articulation and other essential elements for each of the phonemes.

B. Recording the diphones

Diphone recording is one of the important works in speech synthesis because of the speech quality will be highwhen the recording is good. The recording environment such as background noise, recording computer noise must be kept to a minimum. Different Literatures advice sounds preferable to be recorded in appropriate sound studio in order to reduce noise level. As a result, the recording was done in a video conferencing room of Arba Minch University which is a noise free environment, to avoid the noise of recording device we used a small sound recording device called Sony ICD-UX533F Digital Voice Recorder instead of computer. Which is perfect for recording for speech synthesis and for classes or lectures with intelligent noise cut and high quality sound playback.

Figure 4-1: Sony ICD-UX533F Digital Voice Recorder

Everything should be as well-defined as possible, as far as gain settings. A head mounted mike has been used to keep this constant; the SOLIC stereo headphones SLR-M821 headset is used for recording. Because the distance between the speaker and the microphone is crucial, it works well with the mic positioned at 8mm from the lips. The better the recording the better the voice will have, the native Amharic speaker, writer's voice has recorded. Many researchers recommend to record all diphones in one day, in order to minimize the variation in voice quality [12], to record all diphones it takes approximately 1.5 hours in one days. More than one thousand one hundred non-sense words were recorded.

Fabricated words were chosen to be recorded where only one occurrence of each diphone is recorded. For best result, the words should be pronounced with little prosodic variation, as monotone as possible. The advantage with recording nonsense words according to Black and Lenzo [12], is that one does not need to search for natural examples that have the desired diphone that means easy to get all diphones, likely to be pronounced consistently and no lexical interference. The diphone list can then be easily checked and the presentation is less prone to pronunciation errors than if real words were presented. Disadvantages of recording nonsense words is that, possibly bigger database created, speaker becomes bored. On the other hand advantages of designing a diphone inventory using natural words will be pronounced naturally, Easier for speaker to pronounce, but the main disadvantages are may not be pronounced consistently.

Then the speech signal recorded with stereo microphone at default sampling rate 44 kHz of Sony ICD-UX533F Digital Voice Record. Sampling rate of recorded speech signals were dawn to 8 kHz using the speech analysis software tool wave surfer, to minimize the space and processing time. Normalize the recorded signal and split into individual wave form files and finally we got a high quality and noise free signal.

C. Labelling the Carrier Phrases

Once the speaker specific carrier phrases had been recorded, the next phase required the labelling of these utterances in order to identify phones and diphone boundaries. Labelling is the hardest and one of the most important part of making a voice. Once the blocks of speech have been recorded, the words that contain diphones are cut out and stored in separate files, and the diphone boundaries are marked. Labelling nonsense words is *much* easier than labelling continuous speech, whether it is by hand or automatic. With nonsense words it is completely defined which phones are there (and if not it is an error) and they are clearly articulated. Each wave file needs a corresponding label file, basically a label file is a set of times showing where in the wave file various phones are located.

This eventually allows the synthesizer to lookup and extract the required diphones from the recordings. Complete manual labelling would be a very laborious task. For this reason, within the Festvox framework this process is partially automated. This process was initiated by launching the following command:

wale@localhost>bin/make_labs prompt-wav/*.wav

The script make_labs works by comparing the synthesized and the recorded speech signals in the spectral domain. It first extracts the Mel Frequency Cepstral Coefficients (MFCC) of both signals and stores them in individual files within the directories prompt-cep (for the synthesized prompts) and cep (for the recorded utterances). A Dynamic Time Warping (DTW) algorithm is then used to compare the two sets of coefficients and, using the labelling information for the synthesized prompts, a best fitting label alignment is assigned to each of the recorded carrier phrases[13]. The new label files are stored within a folder named lab.

This process was carried out for the recorded utterances in order to create the label files. However, correct labelling is very important for good synthesis results. Once the labelling process was completed, an index of where each diphone can be located was created in a new file. The command used to achieve this was:
wale@localhost>bin/make_diph_indexetc/amdiph.listdic/waldiph.est

The resulting text file waldiph.est consists of a header section followed by a number of lines, one per diphone, specifying a diphone name, the name of the WAV source file, as well as the start-time, mid-point and end-time of the diphone, the latter three values specified in seconds.

D. Extracting Pitchmarks and LPC coefficients

As discussed above, Festival supports both residual excited Linear Predictive Coding (LPC) and Pitch Synchronous Overlap Add (PSOLA) for diphoneresynthesis, although only the former is freely distributed with the package, and was hence the one utilized. Both of these techniques are pitch synchronous, implying a prior requirement to identify where the pitch periods occur within a speech signal in order to operate.

The Festvox scripts allow for the pitchmarks to be extracted from the speech wave file. The script make_pm_wave is used for this purpose, while another utility, make_pm_fix, is used to move the predicted pitchmarks to the nearest peak in the waveform:

wale@localhost>bin/make_pm_wave wav/*.wav
wale@localhost>bin/make_pm_fix pm/*.pm

Internally, the script make_pm_wave calls an Edinburgh Speech Tools application pitchmark in order to band-bass filter the speech signal and identify the pitch mark peaks.

Having identified the location of the pitch marks, it is possible, then, to extract the LPC coefficients. However, a recommended prior step in this case is to normalize the power of the recorded diphones. This is useful since a speaker, even with the best of efforts, is unlikely to maintain a constant vocal amplitude throughout the whole of the recording process. The scripts used to perform these last two steps are as follows:

wale@localhost>bin/find_powerfactors lab/*.lab
wale@localhost>bin/make_lpc wav/*.wav

E. Adding Lexicon and LTS Support

The next step in the development process was to specify how Amharic free text should be translated into a sequence of phones. Festival supports both a static lexicon format and an LTS rule system. Support is also provided to automatically learn LTS rules from a tagged corpus.

As discussed in section two, the phonetic correlate for Amharic text can be derived, with certain limitations, by rule. The LTS rule format supported by Festival is quite similar to that utilized in formulating the Amharic LTS system, consisting of a prioritized list of rules of the following form:

(Left Context [orthography] Right Context => phones)

Implementing the rule in Festival was thus a relatively straightforward exercise,

```
(lex.add.entry
    '("televixzxn"
        nil
        (((teh) 1) ((leh) 0)
            ((vih) 1) ((zhi h n)
        0))))
```

F. Prosodic analysis

As for the case of tokenization, no major implementation work can be claimed for this work, reliance having been made upon the default modelling provided by Festvox. Having the appropriate information about the diphone being concatenated, the DSP component finally produces a speech signal uttered through the computer's speaker.

G. Packaging for Distribution

As a final implementation detail, it is worth noting that Festvox provides a mechanism for packaging the voice developed in format suitable for easy distribution, containing only the relevant files for use. A group file is first created, containing only the parts of the spoken utterances which contain the diphones:

festival> (am_make_group_file "group/walelpc.group" nil)

The choice of diphone database is then changed from separate to grouped by editing the file amu_am_wale_diphone.scm and changing the line:

```
(set! amu_am_wale _db_name
        (am_diphone_initamu_am_wale_lpc_sep
))
to
(set! amu_am_wale _db_name
        (am_diphone_initamu_am_wale
_lpc_group))
```

The voice can then be distributed by bundling the Scheme files under the festvox directory along with the grouped diphone database. The top level folder can also be linked to Festival's voice directory in order to provide direct access to the voice within the system.

V. TESTING AND EVALUATION OF THE NEW SYSTEM

The two most important criteria that are measured when evaluating a synthesized speech are the intelligibility and the naturalness of the speech. The evaluation has been at several levels, such as phoneme, word or sentence level, depending what kind of information is needed. The performance of the Amharic synthesizer is evaluated for both normal and geminated words of a language.

Due to time consideration of respondents only 10 randomly selected phone, word and sentences used for evaluation for both two parts of intelligibility and naturalness. In order to evaluate the quality of the synthesized speech developed Mean Opinion Score (MOS) methods has been utilized. We prepared on line evaluation system and sent a link for more than 40 respondents via email and shared for more than 300 facebook friends.

Based on the suggestion of evaluators, the result of speech synthesizer is 3.5 and 3.05 scored for intelligibility and naturalness, respectively. Accordingly, as per MOS score test the intelligibility of synthesized speech by the synthesizer is "very good" and naturalness is also "good".

VI. CONCLUSION

This thesis work describes the development and evaluation of diphone based TTS system for Amharic language, with a consideration of amharicgemination. The system is developed based on the framework of Festival, using Residual Excited Linear Predictive (RELP) coding synthesizer. In this study the design of a diphone database and the natural language processing modules developed has been described.

To design the TTS, there is a serious of steps considered: Text Analysis which is capable of converting raw text to pronounceable words, Phonetic Analysis which converts text in orthographic form to phonemes, Prosodic Analysis where certain properties of the speech signal are processed, Diphone database Creation which provides diphone speech units to be concatenated and uttered and Diphone Concatenation where the speech is generated.

Based on the evaluation, the system performance has been 79.25% on the average; 3.5 MOS score in intelligibility and 3.05 MOS score for naturalness. The result looks encouraging and further improvement of intelligibility and naturalness depend on proper works in different context. In this research we prepared diphone inventory in consultation with the domain experts that is linguistic people. As reviewed different literatures and the indication of our result shows, if recording quality is good, the bitterness of the diphone units will be very high. In addition, improvement of intelligibility and naturalness also depends on in particular proper lexical stress assignment and a more sophisticated generation of prosodic features.

VII. CONTRIBUTION

From this research the main contribution reported in this thesis are the following:

- ✓ The use of RELP coding synthesizer based on festival framework is shown to be effective.
- ✓ A Prototype Amharic TTS synthesizer having a gemination issue handling mechanism has been developed.
- ✓ The use of multilevel speech synthesis evaluation method is demonstrated.
- ✓ The mechanism to define the pronunciation of germination words is described.
- ✓ All steps of the Amharic TTS System Development in festival have been discussed.

VIII. RECOMMENDATION

Based on the findings, the following recommendations are forwarded for further work to improve total quality of the system.

- ✓ This study may be extended by including all gemination words and non-standard words.
- ✓ Further work needed to be focus on expanding the pronunciation lexicon, the G2P or letter to sound rules to provide accurate pronunciation.
- ✓ By considering accent, all possible contexts (allophones) may get good speech synthesizer to be applicable for reading different literatures such as poetry, poem, and fiction e.t.c.
- ✓ Finally, we investigate from different reviewed literatures the quality of most system is affected by unavailability of sound laboratory. This is approved by our result, since we has been recording in noise free video conferencing with high quality digital sound recording device. That means recording in appropriate sound studio for having better quality output is contributing significant role.

ACKNOWLEDGMENT

Our special thanks goes to Ato Alula Tefera who is lecturer in Adama Institute of Technology, Dr. Sebsbie Hailmariam who is lecturer in Addis Abeba University and AtoYonas Demeke who is lecturer in Debre Birhan University for their invaluable technical help and they never hesitated to help me in any circumstances.

REFERENCES

[1] E. Donovan, "“Trainable Speech Synthesis”," UK, 1996.

[2] A. J. e. al., "From Text to Speech: The MITalk system," 1987.

[3] M. Shiferaw, "Syllable-Based Text-To-Speech Synthesis," Addis Ababa University, Addis Ababa, Ethiopia, June, 2012.

[4] K. JANNA, "PHONETIC GRAMMARS FOR GF," in Chalmers University of Technology.

[5] Y. Tafere, "Formant-Based Speech Synthesis: A Case of Amharic Words," Addis Abeba University, Addis Abeba, June, 2008.

[6] S. H. e. al., "Unit Selection Voice for Amharic Using Festvox," 5th ISCA Speech Synthesis Workshop-Pittsburgh, pp. 103-107, 2005.

[7] S. Lemmetty, "Review of Speech Synthesis T echnology," Helsinki University of Technology, Helsinki, March 30, 1999.

[8] A. Tefere, "A Generalized Approach to Amharic Text To Speech (TTS) Synthesis System," Addis Abeba University, Addis Abeba, July, 2010.

[9] N. Tademe, "Formant Based Speech Synthesis for Amharic Vowels," Addis Abeba University, Addis Abeba, Ethiopia, January, 2008.

[10] T. Geteye, "Characterization of Amharic Vowel Sound Units," Addis Abeba University, Addis Abeba, Ethiopia, November, 2011.

[11] S. A. a. G. A, "Non-concatinative Finite-State Morphotactics of Amharic Simple Verbs," ELRC Working Papers , vol. 2, no. 3, September, 2006.

[12] A. e. a. Lenzo, "Building Synthetic Voices," 2003. [Online]. Available: http://Festvox.org/. [Accessed Accessed on 15 March 2014].

[13] D. S. Blarowski Mike, Time Segment Processing, University of Victoria, BC, Canada: ELEC 484 Project.