UNIT SELECTION VOICE FOR AMHARIC USING FESTVOX

Sebsibe H/Mariam †, S P Kishore †‡§, Alan W Black ‡, Rohit Kumar †, and Rajeev Sangal †

† Language Technologies Research Center
International Institute of Information Technology, Hyderabad

‡ Language Technologies Institute, Carnegie Mellon University
§Institute for Software Research International, Carnegie Mellon University

Abstract

In this paper, we try to describe the issues to be considered in developing a concatenative speech synthesizer for Amharic language. The complexity of the syllable structure of the language, the phonetic nature of the language and the result of the perceptual test of the synthesizer will be discussed. Comments and recommendations for further research are included.

1. INTRODUCTION

Amharic is the official language of Ethiopia. Among 73 languages which are registered in the country, 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. In this paper we discuss the development of unit selection voice for Amharic using Festvox [1].

Festvox is a voice building framework which offers general tools for building unit selection voices in new languages. The unit selection paradigm is a cluster based technique where units of the same type are clustered based on the acoustic differences. The clusters are then indexed based on higher level phonetic and prosodic context. During synthesis an appropriate unit is chosen from multiple instances of that unit based on minimization of joining cost and concatenation cost. Voices generated by this system may be run in the Festival speech synthesis system [2].

This paper is organized as follows: Section 2 focuses on the nature of the Amharic script. Section 3 explains issues like representation of the phone set, the letter to sound rules, syllable structure of the language and syllabification rules. In section 4 we described the voice building process. Section 5 presents the results of perceptual testing conducted on the voice. Conclusion and recommendation are given in Section 6.

2. NATURE OF AMHARIC LANGUAGE SCRIPT

The script of Amharic language is phonetic in nature. It has 32 consonants and 7 vowels. The orthographic representation of the language is organized into orders. Each of the 32 consonants has seven orders (derivatives). Six of them are CV combinations while the seventh is the consonant itself. 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.

The phonetic features of these groups of symbols are not clearly studied. The vowels also find one line in the ordering list except /e/[1]. For each consonant C, the orthographic ordering is as follows:

C/e/ C/u/ C/i/ C/a/ C/ie/ C C/o/

Unlike the orthographic representation, Amharic language has one special property in its spoken form (CV sequence of the acoustic form of the orthographic representation). The sixth order orthographic symbols, which do not have any vowel unit associated to it in the written form (CV transcription of the orthographic form), may associate the vowel /ix/ in its spoken form which has important role during syllabification of the word in the language which allows splitting impermissible consonant clusters.

3. AMHARIC PHONE SET AND SYLLABIFICATION

3.1. Building Amharic Phone Set

To work with Amharic scripts, we defined a transliteration scheme using ASCII characters (as shown in Appendix A). This transliteration scheme is designed based on the orthographic ordering of the script and the acoustic similarity of the letters. It also covers all phonemes under

[1] The transliteration scheme (I-X notation) is mentioned in appendix A.

Table1. Consonant with their features (mainly adopted from [3])

		Labials		*Alveolar*		*Palatals*		*Velars*		*Labio-Velar*		*Glottals*	
Stops	Voiceless	p	ፕ	t	ት			k	ክ	kx	ኳ	ax	ዕ
	Voiced	b	ብ	d	ድ			g	ግ	gx	ጓ		
	Glottalized	px	ጵ	tx	ጥ			q	ቅ	qx	ቋ		
Fricatives	Voiceless	f	ፍ	s	ስ	sx	ሽ					h	ህ
	Voiced	v	ቭ	z	ዝ	zx	ዥ						
	Glottalized			xx	ጽ							hx	ኋ
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						

consideration in this work and avoids all possible ambiguities for sentence parsing.

In Festvox, the phone set of the language is described with the corresponding features like voicing, tongue position, tongue height, place of articulation, and manner of articulation. From the studies reported in [3], we derived a set of phonetic features for the 39 phones. The lists of the phone sets are mentioned in table 1 (consonant) and figure 1 (vowels).

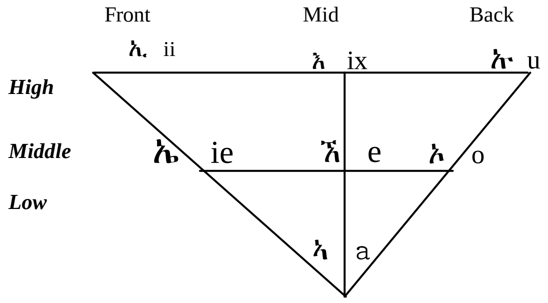

Figure 1. Vowels with their features (mainly adopted from [3])

3.2. Letter to Sound Rules

The way Amharic orthographic characters are written is very similarly to the way they are spoken. It means Amharic is a phonetic language. The mapping of the written form and the spoken form is one to one except the epenthetic vowel which is mentioned above in transliteration scheme. The syllabification rule for Amharic mentioned below will decide the presence or absence of such vowel in the spoken form of the language.

3.3. Syllable Structure

Amharic words are characterized by weak, indeterminate stress; presence of glottal, palatal and labialised consonants; frequent geminate consonants; high frequency of the central vowels, and use of an automatic helping vowel /ix/ [4]. Though strict definition of syllable is difficult [5], a word in Amharic could be monosyllabic like “na” (meaning come) or polysyllabic like “al.me.ta.ciim” (meaning she didn’t come), which consists of four syllables. All syllables have a vowel nucleus.

Several researchers studied the syllable structure of Amharic language and came up with different syllable template. For example, [3] states the six possible syllabic structures in Amharic as V, VC, VCC, CV, CVC, and CVCC and [4] states the syllable structure of Amharic as CV and CVC only.

In this paper we use the following templates in the Amharic speech synthesis: V, VC, VCC, CV, CVC, and CVCC. Moreover rarely initial cluster could exist when the second consonant in the cluster is liquid (and form CCV and CCVC). Depending on the context the nucleus may be simple or complex.

The syllabification of a given Amharic word into its syllable set needs:

- Compression of two successive vowels into one nucleus
- Insertion of epenthetic vowel /ix/

[4] has also pointed out the possibility of having consecutive vowels. If the back rounded vowel (/o/, /u/) appears at the same morpheme boundary, before the middle lower vowel /a/, then the preceding consonant gets labialised. For example “samuat”, “ruac”, “huala”, “fuafuatie”, “quanqua”, “txuat”, etc. In this case, both

vowel phonemes (/ua/) act as a nucleus of the corresponding syllable, which is compressed into one vowel. In all other cases when two successive vowels come together, the first vowel will be the nucleus of the left syllable and the second will be the nucleus of the next syllable for example se.at , me.at, be.hua.la, te.sxua.mi.

3.4. Syllabification Rules

A recursive algorithm is used to identify the set of syllables in a word. This algorithm assumes inter independence of the left most syllable to the rest syllables.

The algorithm to identify the left most syllable make use of the following basic rules of the language after compressing successive vowels into one based on the above compression rule.

1. Consonant between two vowels is always an onset for the second vowel
2. Word with VC, VCC, and CVC phone sequence are monosyllable. Other words that start with vowel take the left vowel as a left most syllable.
3. A word, which consists of only CC phonemes, will insert the epenthetic vowel and form a monosyllable word CVC.
4. If the left most part of the word match the template CVCCV then CVC is taken as a left syllable.
5. If the left most part of the word satisfies the template CVCC then the left syllables may be CVC or CV depending upon the sonority of the last consonant cluster.
6. A word with CCVC sequence, where the second consonant is a liquid is monosyllable.
7. All words with consonant clusters and liquid at second position have left most syllable of type CCV.
8. In all other cases, the left most syllable is CV syllable and covers a larger portion of the syllable distribution of the language. This may apply insertion of the epenthetic vowel if consonant cluster exist.

We used simple stress pattern of 1 (primary stress) for initial syllable and 0 (secondary stress) for all of the remaining syllables in the word.

4. BUILDING THE VOICE

4.1. Creation of Speech Database

To build Amharic speech database, the prompt-list is selected from different sources such as newspapers (“Addis Zemen, Reporter, Sixmixax xxdixq, etc.”), fictions (“Fikir Eske Mekabir, keadmas Bashager, etc”), and publications (different publication of Addis Ababa University (AAU), and other institute) all exist in hard copy form. Selection is done manually to have complete phone coverage of the language.

The total number of phones instances in the training is 27,153 excluding silences. The most dominantly used phones in the spoken form of the corpus are /e/, /a/ and the epenthetic vowel /ix/.

The training corpus consists of a total of 29,480 diphone instances made up of 801 unique diphones. The corpus covers 52.3% of the theoretically possible diphones in Amharic. Out of these unique diphones 14% occurs only once in the corpus.

Moreover, the corpus consists of a total of 12,724 syllables instances and 1317 unique syllables. Out of these, the first hundred high frequency syllables cover 70% of the total distribution. Moreover among the 12,724 syllables instances: 316 are monosyllables, 3752 are front syllable (word initial), 4904 are middle syllable (word middle) and 3752 are back syllable (word final), which shows us the language has small number of monosyllabic words and most of the words consist of a minimum of 3 syllables.

A male speaker using a normal microphone in a quiet room environment recorded the set of prompts. 183 sentences were recorded at 22050 Hz. A speech corpus of 40 minutes duration was generated. The 183 utterances were hand-labeled at phone level using EMU Labeler tool.

4.2. Feature Extraction and Clustering

The labeled speech database was processed by applying simple power normalization on each utterance. The maximum and minimum pitch value of the speaker was determined using the KTH Wavesurfer Free Pitch Marker Tool. The Festvox pitch extraction parameters were adjusted accordingly to obtain pitch features for the utterances. Mel Frequency Cepstral Coefficients were also extracted.

The Unit Selection Amharic voice was built by applying unit clustering algorithm on the units of the database. Further details of this algorithm can be found in [6].

5. PERCEPTUAL EVALUATION OF AMHARIC VOICE

To evaluate the quality of Amharic synthesizer, we conducted perceptua1 tests on 11 college students who are native speakers of the language: 2 females and 9 males. All subjects are 20 to 30 years old in age. Each subject listens to 5 sentences and gives a ranking value for the naturalness of the speech and its intelligibility. They evaluate the system based on the quality of the speech output by giving a measure of quality as follows:

Table 2: Perceptual Evaluation Categories

Category	Measure
Excellent	5
Very Good	4
Good	3
Fair	2
Poor	1
Very poor	0

The results show that the average score of the Amharic synthesizer is 2.9 (which is categorized as good). The summary of the result is shown in table 3 and table 4.

6. CONCLUSION AND RECOMMENDATION

In continuation with our efforts to build synthesizers and recognizers for new languages, in this paper, we discussed the development of unit selection voice for Amharic language. We defined a transliteration scheme to work with Amharic scripts and incorporated Amharic phone set, syllabification rules, letter to sound rules into Festvox. We selected the prompt-list from various sources and built a unit selection voice for Amharic. Perceptual evaluation of the synthesizer showed that the quality of the voice is good (as categorized in the above section).

The following are the recommendations to further improve the quality of the Amharic synthesizer. The epenthetic vowel is used mainly to split impermissible consonant clusters. There is scope for further improving the algorithm we have used for handling the epenthetic vowel. The epenthetic vowel duration is usually much smaller than the same vowel that exists in the written form representation of the text. Modeling identification of the epenthetic vowel improves speech synthesis process as well as automatic syllabification of speech waveforms.

Though the quality of the speech synthesizer is not high, it can be improved by:

- Proper selection of unit. Since the language is phonetic, syllable as a basic unit may outperform the phone as a basic unit.
- Optimal selection of corpus, which proportionally covers all basic units and variations, will give better quality.

11. REFERENCES

[1]. Alan W. Black and Kevin A. Lenzo, “Building Synthetic Voices - for FestVox 2.0 Edition,” 2003 http://www.festvox.org/bsv/

[2]. Alan W. Black , Paul Taylor and Richard Caley, “The Festival Speech Synthesis System -for The Festival Speech Synthesis System,” Edition 1.4, 1999 http://www.speech.cs.cmu.edu/festival/

[3]. Getahun Amare. “**ዘመናዊ የአማርኛ ሰዋሰው በቀላል አቀራረብ፡፡** (“Modern Amharic Grammar in a simple approach”) 96

[4]. Mulugeta Seyoum, “The syllable Structure and Syllablification in Amharic,” Masters of philosophy in general linguistic thesis, Department of Linguistics, Trondheim, Norway, 2001

[5]. Andrew Radfors et.al. “Linguistics: An Introduction,” Cambridge University Press. 1999

[6]. Alan W. Black and Paul Taylor, “Automatically clustering similar units for unit selection in speech synthesis,” in proceedings of EUROSPEECH’97, page-601-604, 1997

Table 3. Result by sentence

Rank	Excellent	Very Good	Good	Fair	Poor	Very poor
Sentence 1	0	1	6	2	2	0
Sentence 2	0	3	2	5	1	0
Sentence 3	0	4	4	3	0	0
Sentence 4	2	3	4	2	0	0
Sentence 5	0	3	4	3	1	0

Table 4. Result by total average

Rank	Excellent	Very good	Good	Fair	Poor	Very poor
Number of Sentence	2 3.6%	14 25.5%	20 36.4%	15 27.3%	4 25.4%	0 0%

APPENDIX A: (I – X NOTATION)

Amharic Phonetic List, IPA Equivalence and its ASCII Transliteration Table

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