MULTIPLE PRONUNCIATION MODEL FOR AMHARIC SPEECH RECOGNITION SYSTEM

Solomon Gizaw (Msc Student In Addis Ababa University , Ethiopia)

Abstract

In this paper the research have tried to show the pattern variations of sound units in Amharic language for multiple pronunciation model. This are variation of sound units at lexical level due to dialects. After that an attempt to build a pronunciation dictionary for Automatic Speech Recognition (ASR).At last comments and recommendations are included.

1. Introduction

Amharic is an official language of Ethiopia. It is a Semitic language that has the greatest number of speakers after Arabic. Amharic has five dialectical variations spoken named as: Addis Ababa, Gojam, Gonder ,Wollo and Menz[1]. The Amharic writing system uses multitudes of ways to denote compound words and there is no agreed upon spelling standard for compounds. As a result of this and of the size of the country leading to vast dialectal dispersion, lexical variation and homophony is very common [2].

Pronunciation variation is a phenomenon observed within a speaker or within a group of speakers of the same dialect or among speakers across dialects of the same language. Pronunciation variation deals with the different ways of speaking a given word. Pronunciation variation modeling has been studied in the field of speech synthesis and recognition to improve performance of the corresponding speech systems [3].

2. Amharic Orthography

The Amharic orthography as it is represented in the Amharic character set consists of 276 distinct symbols. These symbols are classified into four groups. In the first category (33*7=231) there are thirty-three core orthographic symbols, each of which has seven different shapes, usually known as orders, to represent the seven vowels. Each consonant and the seven vowels in combination represent CV syllables[4]. Each of these consonant and vowel grapheme can appear independently or can form a combinant letter. Each consonant can form CV pattern except with the vowel /ix/ (called epenthesis vowel)[5] .

The second category (4*5=20) consists of four labio-velar symbols, which have five orders. The eighteen labelized consonant, which have only one order, are the third category. The fourth category is the representation of numbers from 1 to 10 and multiples of 10 each with different symbols [4]. The Amharic language script is called Ethiopic. Even though the vowel modification is not entirely semantic, the Ethiopic script is a syllabic structure [5].

3. Data Acquisition

Here, we describe our efforts in building up speech data, text data and lexical dictionary for Amharic speech recognition

3.1. Text Data

To create more general acoustic models we selected sentences from two different fiction books namely " Mekoya " and "Wey Addis Ababa". This fiction books are selected because they contains words which has pronunciation variations for dialects. The selection of sentences from the above books aims both phonetically rich and balanced collection of sentences with regards to variations of dialects. During the process of text selections some corrections has done like:

- Spelling and grammar errors has been corrected
- Long sentences are shortened
- Numbers and foreign words has been textually transcribed

The selected text contains 172 sentences .These selected sentences have been manually checked for grammatically, spelling, foreign words and abbreviations and have been manually corrected.

3.2. Speech Data

After selecting the text, the next step in preparation of Amharic speech corpus is recording speech. In the recording process of the selected sentences the speaker asked to read exactly what is presented to him with his own dialect. The sentences are given for the speaker in a printed format. At the process of recording the degree of control was very high, each utterance of the speaker was checked directly for errors and if an error is found, the speaker is asked to re-read the text again and again. The text has been read by 5 different native speakers in an office environment. Each speaker is selected from the different dialects of Amharic language

3.3. Building Amharic Phone Set

As defined on the Wikipedia ,phonetic transcription or notation is the visual system of symbolization of the sounds occurring in spoken Human Language. International Phonetic Alphabet (IPA) in one of the most widespread system .We have adopted Sebsibes' transliteration scheme of Amharic scripts using ASCII characters (as shown in Appendix). This transliteration scheme is designed based on the orthographic ordering of the script and the acoustic similarity of the letters.

3.4. Variations of sound units

Dialectologists have always been ready to ask language consultants (informants) questions about lexical matters ("What is your word for this?") Syntacticians readily seek out the opinion of speakers on whether some proposed strung of words is acceptable (well-formed) or not ("Can you say two furniture's?") [3].

One of the main challenges in modeling a pronunciation is to know which variation we are trying to model. There is no research or study made regarding multiple pronunciations variations for Amharic Language. There is only two article which has found in Addia Ababa University, Language and literature department .The first article was written by Tesfaye Mechi in titled "Amharic Pronunciation words" which is written in 1965 and the second article was written by Yonas Admasu in titled "Characteristics of Amharic Language "which is written in 1966.I have used this two research papers to get some of the words which has variations in pronunciations in order to model the dictionary. These two articles generally identified the pronunciation variations among dialects of the language and show many words and sentences which has variations.

The other system which I have used to get variations of pronunciation is giving for each dialect speakers a list of words in the standard dialect and asking them to find the variations in their specific dialect. We will use the Addis Ababa Dialects as a standard. The reason we choose is that Addis Ababa dialects is most of the time peoples from different dialects are living in Addis Ababa and it is used by all other peoples of the country and also it is used by the national media and news papers.
The following table shows the pattern variations of sound unit in different dialects.

Standard	Gondar	Gojam	Wollo	Menz
? sx ?	? tx ?	? tx ?	? tx ?	? tx ?
? q ?	? a / cx ?	? cx ?	? a ?	? a ?
? zx ?	? j ?	? j ?	? j ?	
? n ?	? n e ?			
? y a ?	? a ?	? a ?	? a ?	? a ?
? s i ?	? t i ?			
? consonant a m #		? consonant t o m #		
? consonant u m #		? consonant c ie m #		
? consonant ie #	? consonant e #			
# consonant e ?	# consonant ie ?			
# consonant e ?	# consonant h a ?	# consonant h a ?	# consonant h a ?	# consonant h a ?

Note that to show variation of sounds in the table we used the following representation:
? :- any phonemes
:- beginning or ending of the words
consonant :- consonant

3.5. Building pronunciation Dictionary

We have used the HTK lexical dictionary format to create lexical for Amharic language. In order to build a dictionary the first thing to do is create a sorted list of the required words. To build robust acoustic models, it is necessary to train them on a large set of sentences containing many words .The word lists are extracted from the data which we have found from variation patterns. Before using HTK toolkit, one would need to edit the text into a suitable format. For example, it would be necessary to change all white space to new lines and then to sort the words into a unique alphabetically ordered set, with one word per line.
The general format of each dictionary entry in HTK is WORD [outsym] p1 p2 p3 ..... and these pronunciations have transcribed manually.
A pronunciation dictionary can be classified into two , canonical or multiple dictionary. For each word a canonical pronunciation dictionary includes only the standard phone sequence assumed to be pronounced in read speech. It does not consider pronunciation variations. On the other hand, multiple pronunciation dictionary is a pronunciation dictionary that uses the actual phone sequences pronounced in speech variations can be included.

To add alternate pronunciation in the standard lexicon , it only need modification of the pronunciation dictionary .We have added the alternate pronunciation in the dictionary shown as follows and a number after the word indicates the alternate pronunciation due to dialect variations:

Words	**Pronunciation**
ABESHA	h a b e sx a
ABESHA(2)	a b e sx a
ABWARA	a b a r a
ABWARA(2)	a b w a r a
ACHAMNA	a c a m n a
ACHAMNA(2)	t a c a m n a
ADENAGARI	a d e n a g a r ii
ADENAGARI(2)	a q e b a j a r ii
AGER	h a g e r
AGER(2)	a g e r
ALBELAHUM	a l b e l a h u m
ALBELAHUM(2)	a l b e l c ie m
ALBELAM	a l b e l a m
ALBELAM(2)	a l b e l t o m

4. Implementation and Results

4.1. Implementation

The study needed speech data, transcription of the speech and acoustic model (grammar) to train the HMMs. These were prepared before and a dictionary was needed to define the valid words and their pronunciation for the recognition. The dictionary for our task is prepared based on the set of lexicon in the training and test data. This dictionary prepared to support multiple pronunciations.

After feature extraction was done in advance for the training data set. The phone-based acoustic model was trained for all of the 35 Amharic monophones using the feature extracted speech. The parameters of the HMMs were trained using the EM-algorithm. The EM-algorithm is an iterative algorithm; each iteration was invoked in HTK with a call of the function HERest. To find initial parameters the HTK tool HCompV can be used.

The Viterbi decoder performs recognition. It matches speech signals against a network of HMMs and returns a transcription for each speech signal. Finally, evaluates the performance of the speech recognizer using the HResults tool. It reads in a set of label files (typically output from the recognition tool such as Hvite) and compares them with the corresponding reference transcription.

4.2. Result

The recognition results were displayed in terms of Word Error Rate (WER), including the number of insertions, deletions and substitutions that would align the reference and recognized transcriptions.

The following Tables show the evaluations of speech recognizer categorized by single pronunciation dictionary and multiple pronunciation dictionary. It also shows the results of the test and training data evaluation separately.

Test Sets	% Correct	% Accuracy	H	D	S	I	N
Training	83.28	20.91	478	3	93	358	574
Test	66.00	10.00	66	0	34	76	100

Table 5-1: WORD EVALUATION for Single Pronunciation

Test Sets	% Correct	% Accuracy	H	D	S	I	N
Training	84.67	21.25	486	2	86	364	574
Test	68.00	13.00	68	0	32	81	100

Table 5-2: WORD EVALUATION for Multiple Pronunciations

4.3.Conclusion

This read speech has tried to cover pronunciation variations of all dialects of the language. The study also used the corpus to develop phone-based multiple pronunciation ASRS for Amharic which have word and sentence recognition accuracy of above 84.67 %.

Therefore, we believe the potential of developing Amharic ASRS using HTK that applies the HMM frame work was demonstrated by the study. It was also proved that it is possible to use multiple pronunciation dictionary as a basic unit of phone-based speech recognition for Amharic. For the study's experiment it was found that the optimal HMM topology for Amharic phone-based model with five emitting states and twelve Gaussian Mixtures.

Based on the findings of the study's experiments, it can be concluded that multiple pronunciation model can improve the performance of the phone-based recognizer. Furthermore it is expected that more improvement in the performance of recognizer if other pronunciation variations properly included.

Reference

[1] Solomon Tefera , "Automatic Speech Recognition for Amharic" , Hamburg university , Dec, 2005

[2] H. Seid and B. Gamback , "A speaker independent continuous speech recognizer for Amharic " , INTERSPEECH 2005 Lisbon.

[3] B.Byrne and M.Finke , " Pronunciation Modeling for conversational speech recognition" , A status report from ws97 (1997) Johns Hopkins University.

[4] Solomon Teferra , Wolfgang Menzel and Bairu Tafila, "An Amharic Speech Corpus for Large Vocabulary Continuous Speech Recognition " , Hamburg University , 2004

[5] Sebsibe Haile Mariam , " Extraction of linguistic information from acoustic data for building speech systems " , International Institute of information technology ,India , August 2007

[6] J.C. Wells , " Pronunciation research by written questionnaire" , J.C. Wells University College London , 1985

[7] Sebsibe H/Mariam , S P Kishore , Alan W Black , Rohit Kumar and Rajeev Sangal , " UNIT SELECTION VOICE FOR AMHARIC USING FESTVOX " , 5th ICSA speech synthesis workshop –Pittsburgh , 2004

APPENDIX A: (I – X NOTATION)

Amharic Phonetic List, IPA Equivalence and it's ASCII Transliteration Table

IPA	Transcription	Amharic equivalence
Consonants		
[p]	[p]	ý
[t]	[t]	*f*
[k]	[k]	¡
[ʔ]	[ax]	°
[b]	[b]	w
[d]	[d]	É
[g]	[g]	Ó
[p']	[px]	å
[t']	[tx]	Ø
[c']	[cx]	ß
[q]	[q]	p
[f]	[f]	õ
[s]	[s]	e
[ʃ]	[sx]	i
[h]	[h]	I
[s']	[xx]	ê
[tʃ]	[c]	‹
[g']	[j]	Ï
[m]	[m]	U
[n]	[n]	”
[n']	[nx]	⁋
[l]	[l]	M
[r]	[r]	`
[j]	[y]	Ã
[w]	[w]	⍹

[v]	[v]	{
[z]	[z]	ዝ´
[z’]	[zx]	»
Vowels		
[E]	[e]	ኧ
[U]	[u]	ኡ
[I]	[ii]	ኢ
[A]	[a]	ኣ
[e]	[ie]	ኤ
[^]	[ix]	ኧ
[o]	[o]	ኦ