Syllable-Based Speech Recognition for Amharic

Solomon Teferra Abate
solomon_teferra_7@yahoo.com

Wolfgang Menzel
menzel@informatik.uni-hamburg.de

Uniformity of Hamburg, Department of Informatik Natural Language Systems Groups
Vogt-Kölln-Strasse. 30, D-22527 Hamburg, Germany

Abstract

Amharic is the Semitic language that has the second large number of speakers after Arabic (Hayward and Richard 1999). Its writing system is syllabic with Consonant-Vowel (CV) syllable structure. Amharic orthography has more or less a one to one correspondence with syllabic sounds. We have used this feature of Amharic to develop a CV syllable-based speech recognizer, using Hidden Markov Modeling (HMM), and achieved 90.43% word recognition accuracy.

1 Introduction

Most of the Semitic languages are technologically unfavored. Amharic is one of these languages that are looking for technological considerations of researchers and developers in the area of natural language processing (NLP). Automatic Speech Recognition (ASR) is one of the major areas of NLP that is understudied in Amharic. Only few attempts (Solomon, 2001; Kinfe, 2002; Zegaye, 2003; Martha, 2003; Hussien and Gambäck, 2005; Solomon et al., 2005; Solomon, 2006) have been made.

We have developed an ASR for the language using CV syllables as recognition units. In this paper we present the development and the recognition performance of the recognizer following a brief description of the Amharic language and speech recognition technology.

2 The Amharic Language

Amharic, which belongs to the Semitic language family, is the official language of Ethiopia. In this family, Amharic stands second in its number of speakers after Arabic (Hayward and Richard 1999). Amharic has five dialectical variations (Addis Ababa, Gojjam, Gonder, Wollo, and Menz) spoken in different regions of the country (Cowley, et.al. 1976). The speech of Addis Ababa has emerged as the standard dialect and has wide currency across all Amharic-speaking communities (Hayward and Richard 1999).

As with all of the other languages, Amharic has its own characterizing phonetic, phonological and morphological properties. For example, it has a set of speech sounds that is not found in other languages. For example the following sounds are not found in English: [p`], [tS`], [s`], [t`], and [q].

Amharic also has its own inventory of speech sounds. It has thirty one consonants and seven vowels. The consonants are generally classified as stops, fricatives, nasals, liquids, and semi-vowels (Leslau 2000). Tables 1 and 2 show the classification of Amharic consonants and vowels[1].

Man of Art	Voicing	Place of Articulation				
		Lab	Den	Pal	Vel	Glot
Stops	Vs	[p]	[t]	[tS]	[k]	[?]
	Vd	[b]	[d]	[d3]	[g]	
	Glott	[p`]	[t`]	[tS`]	[q]	
	Rd				[k^w] [g^w] [q^w]	
Fric	Vs	[f]	[s]	[S]		[h]
	Vd		[z]	[3]		
	Glott		[s`]			
	Rd					[h^w]
Nasals	Vd	[m]	[n]	[ɳ]		
Liq	Vd		[l] [r]			
Sv	Vd	[w]			[j]	

Table 1: Amharic Consonants

Key: Lab = Labials; Den = Dentals; Pal = Palatals; Vel = Velars; Glot = Glottal; Vs = Voiceless;

[1] International Phonetic Association's (IPA) standard has been used for representation.

Proceedings of the 5th Workshop on Important Unresolved Matters, pages 33–40,
Prague, Czech Republic, June 2007. ©2007 Association for Computational Linguistics

Vd = Voiced; Rd = Rounded; Fric = Fricatives; Liq = Liquids; Sv = Semi-Vowels.

Positions	front	center	back
high	[i]	ɨ	[u]
mid	[e]	[ə]	[o]
low		[a]	

Table 2: Amharic Vowels

Amharic is one of the languages that have their own writing system, which is used across all Amharic dialects. Getachew (1967) stated that the Amharic writing system is phonetic. It allows any one to write Amharic texts if s/he can speak Amharic and has knowledge of the Amharic alphabet. Unlike most known languages, no one needs to learn how to spell Amharic words. In support of the above point, Leslaw (1995) noted that no real problems exist in Amharic orthography, as there is more or less, a one-to-one correspondence between the sounds and the graphic symbols, except for the gemination of consonants and some redundant symbols.

Many (Bender 1976; Cowley 1976; Baye 1986) have claimed the Amharic orthography as a syllabary for a relatively long period of time. Recently, however, Taddesse (1994) and Baye (1997), who apparently modified his view, have argued it is not. Both of these arguments are based on the special feature of the orthography; the possibility of representing speech using either isolated phoneme symbols or concatenated symbols.

In the concatenated feature, commonly known to most of the population, each orthographic symbol represents a consonant and a vowel, except for the sixth order[2], which is sometimes realized as a consonant without a vowel and at other times a consonant with a vowel. This representation of concatenated speech sounds by a single symbol has been the basis for the claim made of the writing system, as syllabary.

Amharic orthography does not indicate gemination, but since there are relatively few minimal pairs of geminations, Amharic readers do not find this to be a problem. This property of the writing system is analogous to the vowels of Arabic and Hebrew, which are not normally indicated in writing.

The Amharic orthography, as represented in the Amharic Character set - also called [fidəlI] consists of 276 distinct symbols. In addition, there are twenty numerals and eight punctuation marks. A sample of the orthographic symbols is given in Table 3.

	ə	*u*	*i*	*a*	*e*	ɨ	*o*
h	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ
l	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ
m	መ	ሙ	ሚ	ማ	ሜ	ም	ሞ
r	ረ	ሩ	ሪ	ራ	ሬ	ር	ሮ

Table 3: Some Orthographic Symbols of Amharic

However, research in speech recognition should only consider distinct sounds instead of all the orthographic symbols, unless there is a need to develop a dictation machine that includes all of the orthographic symbols. Therefore, redundant orthographic symbols that represent the same syllabic sounds can be eliminated. Thus, by eliminating redundant graphemes, we are left with a total of 233 distinct CV syllable characters. In our work, an HMM model has been developed for each of these CV syllables.

3 HMM-Based Speech Recognition

The most well known and well performing approach for speech recognition are Hidden Markov Models (HMM). An HMM can be classified on the basis of the type of its observation distributions, the structure in its transition matrix and the number of states.

The observation distributions of HMMs can be either discrete, or continuous. In discrete HMMs, distributions are defined on finite spaces while in continuous HMMs, distributions are defined as probability densities on continuous observation spaces, usually as a mixture of several Gaussian distributions.

The model topology that is generally adopted for speech recognition is a left-to-right or Bakis model

[2] An order in Amharic writing system is a combination of a consonant with a vowel represented by a symbol. A consonant has therefore, 7 orders or different symbols that represent its combination with 7 Amharic vowels.

because the speech signal varies in time from left to right (Deller, Proakis and Hansen 1993).

An HMM is flexible in its size, type, or architecture to model words as well as any sub-word unit.

3.1 Sub-word Units of Speech Recognition

Large Vocabulary Automatic Speech Recognition Systems (LVASRSs) require modeling of speech in smaller units than words because the acoustic samples of most words will never be seen during training, and therefore, can not be trained. Moreover, in LVASRSs there are thousands of words and most of them occur very rarely, consequently training of models for whole words is generally impractical. That is why LVASRSs require a segmentation of each word in the vocabulary into sub-word units that occur more frequently and can be trained more robustly than words. Using sub-word based models enables us to deal with words which have not been seen during training since they can just be decomposed into the sub-word units. As a word can be decomposed in sub-word units of different granularities, there is a need to choose the most suitable sub-word unit that fits the purpose of the system.

Lee et al. (1992) pointed out that there are two alternatives for choosing the fundamental sub-word units, namely acoustically-based and linguistically-based units . The acoustic units are the labels assigned to acoustic segment models, which are defined on the basis of procuring a set of segment models that spans the acoustic space determined by the given, unlabeled training data. The linguistically-based units include the linguistic units, e.g. phones, demi-syllables, syllables and morphemes.

It should be clear that there is no ideal (perfect) set of sub-word units. Although phones are very small in number and relatively easy to train, they are much more sensitive to contextual influences than larger units. The use of triphones, which model both the right and left context of a phone, has become the dominant solution to the problem of the context sensitivity of phones.

Triphones are also relatively inefficient sub-word units due to their large number. Moreover, since a triphone unit spans a short time-interval, it is not suitable for the integration of spectral and temporal dependencies.

An other alternative is the syllable. Syllables are longer and less context sensitive than phones and capable of exploiting both the spectral and temporal characteristics of continuous speech (Ganapathiraju et al. 1997). Moreover, the syllable has a close connection to articulation, integrates some co-articulation phenomena, and has the potential for a relatively compact representation of conversational speech.

Therefore, different attempts have been made to use syllables as a unit of recognition for the development of ASR. To mention a few: Ganapathiraju et al. (1997) have explored techniques to accentuate the strengths of syllable-based modeling with a primary interest of integrating finite-duration modeling and monosyllabic word modeling. Wu et al. (1998) tried to extract the features of speech over the syllabic duration (250ms), considering syllable-length interval to be 100-250ms. Hu et al. (1996) used a pronunciation dictionary of syllable-like units that are created from sequences of phones for which the boundary is difficult to detect. Kanokphara (2003) used syllable-structure-based triphones as speech recognition units for Thai.

However, syllables are too many in a number of languages, such as English, to be trained properly. Thus ASR researchers in languages like English are led to choose phones where as for Amharic it seems promising to consider syllables as an alternative, because Amharic has only 233 distinct CV syllables.

4 Syllable-Based Speech Recognition for Amharic

In the development of syllable-based LVASRSs for Amharic we need to deal with a language model, pronunciation dictionary, initialization and training of the HMM models, and identification of the proper HMM topologies that can be properly trained with the available data. This section presents the development and the performance of syllable based speech recognizers.

4.1 The Language Model

One of the required elements in the development of LVASRSs is the language model. As there is no usable language model for Amharic, we have trained bigram language models using the HTK statistical language model development modules. Due to the inflectional and derivativational morphological feature of Amharic our language models have relatively high perplexities.

4.2 The Pronunciation Dictionary

The development of a large vocabulary speaker independent recognition system requires the availability of an appropriate pronunciation dictionary. It specifies the finite set of words that may be output by the speech recognizer and gives, at least, one pronunciation for each. A pronunciation dictionary can be classified as a canonical or alternative on the basis of the pronunciations it includes.

A canonical pronunciation dictionary includes only the standard phone (or other sub-word) sequence assumed to be pronounced in read speech. It does not consider pronunciation variations such as speaker variability, dialect, or co-articulation in conversational speech. On the other hand, an alternative pronunciation dictionary uses the actual phone (or other sub-word) sequences pronounced in speech. In an alternative pronunciation dictionary, various pronunciation variations can be included (Fukada et al. 1999).

We have used the pronunciation dictionary that has been developed by Solomon et al. (2005). They have developed a canonical and an alternative pronunciation dictionaries. Their canonical dictionary transcribes 50,000 words and the alternative one transcribes 25,000 words in terms of CV syllables.

Both these pronunciation dictionaries do not handle the difference between geminated and non-geminated consonants; the variation of the pronunciation of the sixth order grapheme, with or without vowel; and the absence or presence of the glottal stop consonant. Gemination of Amharic consonants range from a slight lengthening to much more than doubling. In the dictionary, however, they are represented with the same transcription symbols.

The sixth order grapheme may be realized with or without vowel but the pronunciation dictionaries do not indicate this difference. For example, the dictionaries used the same symbol for the syllable [rI] in the word [dƷəmərInI] 'we started', whose vowel part may not be realized, and in the word [bərIzo] 'he diluted with water' that is always realized with its vowel sound. That forces a syllable model to capture two different sounds: a sound of a consonant followed by a vowel, and a sound of the consonant only. A similar problem occurs with the glottal stop consonant [?] which may be uttered or not.

A sample of pronunciations in the canonical and alternative pronunciation dictionaries is given in Table 4[3]. The alternative pronunciation dictionary contains up to 25 pronunciation variants per word form. Table 5 illustrates some cases of the variation.

Words	Canonical Pronunciation	Alternative Pronunciation
CAmA	CA mA sp	CA mA sp
		Ca mA sp
HiteyoPeyA	Hi te yo Pe yA sp	Hi te yo Pe yA sp
		Hi te yo Pi yA sp
		Hi to Pe yA sp
		te yo Pe yA sp
		to Pe yA sp

Table 4: Canonical and Alternative Pronunciation

Words	Number of pronunciation variants
HiteyoPeyAweyAne	25
HiheHadEge	16
yaHiteyoPeyAne	7
miniseteru	7
yaganezabe	6
HegeziHabehEre	6
yehenene	5

Table 5: Number of Pronunciation variants

Although it does not handle gemination and pronunciation variabilities, the canonical pronunciation dictionary contains all 233 distinct CV syllables of Amharic, which is 100% syllable coverage.

Pronunciation dictionaries of development and evaluation test sets have been extracted from the canonical pronunciation dictionary. These test dictionaries have 5,000 and 20,000 words each.

4.3 The Acoustic Model

For training and evaluation of our recognizers, we have used the Amharic read speech corpus that has been developed by Solomon et al. (2005).

The speech corpus consists of a training set, a speaker adaptation set, development test sets (for 5,000 and 20,000 vocabularies), and evaluation test sets (for 5,000 and 20,000 vocabularies). It is a medium size speech corpus of 20 hours of training speech that has been read by 100 training speakers who read a total of 10850 different sentences. Eighty of the training speakers are from the Addis

[3] In tables 4 and 5, we used our own transcription

Ababa dialect while the other twenty are from the other four dialects.

Test and speaker adaptation sets were read by twenty other speakers of the Addis Ababa dialect and four speakers of the other four dialects. Each speaker read 18 different sentences for the 5,000 vocabulary (development and evaluation sets each) and 20 different sentences for the 20,000 vocabulary (development and evaluation sets each) test sets. For the adaptation set all of these readers read 53 adaptation sentences that consist of all Amharic CV syllables.

Initialization: Training HMM models starts with initialization. Initialization of the model for a set of sub-word HMMs prior to re-estimation can be achieved in two different ways: bootstrapping and flat start. The latter implies that during the first cycle of embedded re-estimation, each training utterance will be uniformly segmented. The hope of using such a procedure is that in the second and subsequent iterations, the models align as intended.

We have initialized HMMs with both methods and trained them in the same way. The HMMs that have been initialized with the flat start method performed better (40% word recognition accuracy) on development test set of 5,000 words.

The problem with the bootstrapping approach is that any error of the labeler strongly affects the performance of the resulting model because consecutive training steps are influenced by the initial value of the model. As a result, we did not benefit from the use of the segmented speech, which has been transcribed with a speech recognizer that has low word recognition accuracy, and edited by non-linguist listeners. We have, therefore, continued our subsequent experiments with the flat start initialization method.

Training: We have used the Baum-Welch re-estimation procedure for the training. In training sub-word HMMs that are initialized using the flat-start procedure, this re-estimation procedure uses the parameters of continuously spoken utterances as an input source. A transcription, in terms of sub-word units, is also needed for each input utterance. Using the speech parameters and their transcription, the complete set of sub-word HMMs are re-estimated simultaneously. Then all of the sub-word HMMs corresponding to the sub-word list are joined together to make a single composite HMM. It is important to emphasize that in this process the transcriptions are only needed to identify the sequence of sub-word units in each utterance. No boundary information is required (Young et al. 2002).

The major problem with HMM training is that it requires a great amount of speech data. To overcome the problem of training with insufficient speech data, a variety of sharing mechanisms can be implemented. For example, HMM parameters are tied together so that the training data is pooled and more robust estimates result. It is also possible to restrict the model to a variance vector for the description of output probabilities, instead of a full covariance matrix. Rabiner and Juang(1993) pointed out that for the continuous HMM models, it is preferable to use diagonal covariance matrices with several mixtures, rather than fewer mixtures with full covariance matrices to perform reliable re-estimation of the components of the model from limited training data. The diagonal covariance matrices have been used in our work.

HMM Topologies: To our knowledge, there is no topology of HMM model that can be taken as a rule of thumb for modeling syllable HMMs, especially, for Amharic CV syllables. To have a good HMM model for Amharic CV syllables, one needs to conduct experiments to select the optimal model topology. Designing an HMM topology has to be done with proper consideration of the size of the unit of recognition and the amount of the training speech data. This is because as the size of the recognition unit increases and the size of the model (in terms of the number of parameters to be re-estimated) grows, the model requires more training data.

We, therefore, carried out a series of experiments using a left-to-right HMM with and without jumps and skips, with a different number of emitting states (3, 5, 6, 7, 8, 9, 10 and 11) and different number of Gaussian mixtures (from 2 to 98). By jump we mean skips from the first non-emitting state to the middle state and/or from the middle state to the last non-emitting state. Figure 1 shows a left-to-right HMM of 5 emitting states with jumps and skips.

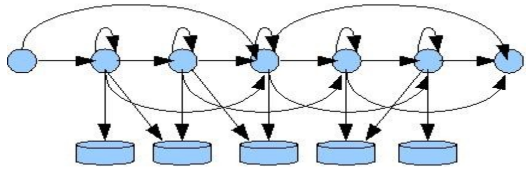

Figure 1: An example of HMM topologies

We have assumed that the problem of gemination may be compensated by the looping state transitions of the HMM. Accordingly, CV syllables containing geminated consonants should have a higher loop probability than those with the non-geminated consonants.

To develop a solution for the problem of the irregularities in the realization of the sixth order vowel [I] and the glottal stop consonant [?], HMM topologies with jumps have been used.

We conducted an experiment using HMMs with a jump from the middle state to the last (non-emitting) state for all of the CV syllables with the sixth order vowel, and a jump from the first emitting state to the middle state for all of the CV syllables with the glottal stop consonant. The CV syllable with the glottal stop consonant and the 6th order vowel have both jumps. These topologies have been chosen so that the models recognize the absence of the vowel and the glottal stop consonant of CV syllables. This assumption was confirmed by the observation that the trained models favor such a jump. A model, which has 5 emitting states, of the glottal stop consonant with the sixth order vowel tends to start emitting with the 3rd emitting state with a probability of 0.72. The model also has accumulated a considerable probability (0.38) to jump from the 3rd emitting state to the last (non-emitting) state.

A similar model of this consonant with the other vowels (our example is the 5th order vowel) tend to start emitting with the 3rd emitting state with a probability of 0.68. This is two times the probability (0.32) of its transition from the starting (non-emitting state) to the 1st emitting state.

The models of the other consonants with the sixth order vowel, which are exemplified by the model of the syllable [jI], tend to jump from the 3rd emitting state to the last (non-emitting) state with a probability of 0.39, which is considerably greater than that of continuing with the next state (0.09).

Since the amount of available training speech is not enough to train transition probabilities for skipping two or more states, the number of states to be skipped have been limited to one.

To determine the optimal number of Gaussian mixtures for the syllable models, we have conducted a series of experiments by adding two Gaussian mixtures for all the models until the performance of the model starts to degrade. Considering the difference in the frequency of the CV syllables, a hybrid number of Gaussian mixtures has been tried. By hybrid, we mean that Gaussian mixtures are assigned to different syllables based on their frequency. For example: the frequent syllables, like [nI], are assigned up to fifty-eight while rare syllables, like [p`i], are assigned not more than two Gaussian mixtures.

4.4 Performance of the Recognizers

We present recognition results of only those recognizers which have competitive performance to the best performing models. For example: the performance of the model with 11 emitting states with skips and hybrid Gaussian mixtures is more competitive than those with 7, 8, 9, and 10 emitting states. We have also systematically left out test results which are worse than those presented in Table 6. Table 6[4] shows evaluation results made on the 5k development test set.

States	Transition Topologies	Mix.	Models		
			AM	AM + LM	AM + LM + SA
3	No skip and jump	18	62.85	88.82	
		Hy	60.87	87.63	88.50
	skip	12		69.20	
	jump	12	43.74	79.94	
5	No skip and jump	12	69.29	88.99	**89.80**
		Hy	60.04		
	skip	12		85.77	
	jump	12	54.53	84.60	
11	skip	12	55.04		
		Hy	71.83	89.21	**89.04**

Table 6: Recognition Performance on 5k Development test set

From Table 6, we can see that the models with five emitting states, with twelve Gaussian mixtures, without skips and jumps has the best (89.80%) word recognition accuracy. It has 87.69% word recognition accuracy on the 20k development test set.

Since the most commonly used number of HMM states for phone-based speech recognizers is three emitting states, one may expect a model of six emitting states to be the best for an HMM of

[4] In tables 6 and 7, States refers to the number of emitting states; Mix refers to the number of Gaussian mixtures per state; Hy refers to hybrid; AM refers to acoustic model; LM refers to language model; and SA refers to speaker adaptation.

concatenated consonant and vowel. But the result of our experiment shows that a CV syllable-based recognizer with only five emitting states performed better than all the other recognizers.

As we can see from Table 6, models with three emitting states do have a competitive performance with 18 and hybrid Gaussian mixtures. They have the least number of states of all our models. Nevertheless, they require more storage space (33MB with 18 Gaussian mixtures and 34MB with hybrid Gaussian mixtures) than the best performing models (32MB). Models with three emitting states also have larger number of total Gaussian mixtures[5] (30,401 with 18 Gaussian mixtures and 31,384 with hybrid Gaussian mixtures) than the best performing models (13,626 Gaussian mixtures).

The other model topology that is competitive in word recognition performance is the model with eleven emitting states, with skip and hybrid Gaussian mixtures, which has a word recognition accuracy of 89.21%. It requires the biggest memory space (40MB) and uses the largest number of total Gaussian mixtures (36,619) of all the models we have developed.

We have evaluated the top two models with regard to their word recognition accuracy on the evaluation test sets. Their performance is presented in Table 7. As it can be seen from the table, the models with the better performance on the development test sets also showed better results with the evaluation test sets. We can, therefore, say that the model with five emitting states without skips and twelve Gaussian mixtures is preferable not only with regard to its word recognition accuracy, but also with regard to its memory requirements.

States	Mix.	Models			
		AM + LM		AM + LM + SA	
		5k	20k	5k	20k
5	12			**90.43**	87.26
11	Hy	89.36	87.13		

Table 7: Recognition Performance on 5k and 20k Evaluation test sets

For a comparison purpose, we have developed a baseline word-internal triphone-based recognizer using the same corpus. The models of 3 emitting states, 12 Gaussian mixtures, with skips have the best word recognition accuracy (91.31%) of all the other triphone-based recognizers that we have developed. This recognizer also has better word recognition accuracy than that of our syllable-based recognizer (90.43%). But tying is applied only for the triphone-based recognizers.

However the triphone-based recognizer requires much more storage space (38MB) than the syllable-based recognizer that requires only 15MB space. With regard to their speed of processing, the syllable-based model was 37% faster than triphone-based one.

These are encouraging results as compared to the performance reported by Afify et al. (2005) for Arabic speech recognition (14.2% word error rate). They have used a trigram language model with a lexicon of 60k vocabulary.

4.5 Conclusions and Research Areas in the Future

We conclude that the use of CV syllables is a promising alternative in the development of ASRSs for Amharic. Although there are still possibilities of performance improvement, we have got an encouraging word recognition accuracy (90.43%). Some of the possibilities of performance improvement are:

- The pronunciation dictionary that we have used does not handle the problem of gemination of consonants and the irregular realization of the sixth order vowel and the glottal stop consonant, which has a direct effect on the quality of the sub-word transcriptions. Proper editing (use of phonetic transcription) of the pronunciation dictionaries which, however, requires a considerable amount of work, certainly will result in a higher quality of sub-word transcription and consequently in the improvement of the recognizers' performance. By switching from the grapheme-based recognizer to phonetic-based recognizer in Arabic, Afif et al. (2005) gained relative word error rate reduction of 10% to 14%.
- Since tying is one way of minimizing the problem of shortage of training speech, tying the syllable-based models would possibly result in a gain of some degree of performance improvement.

[5] We counted the Gaussian mixtures that are physically saved, instead of what should actually be.

5 References

Afif, Mohamed, Long Nguyen, Bing Xiang, Sherif Abdou, and John Makhoul. 2005. Recent progress in Arabic broadcast news transcription at BBN. In *INTERSPEECH-2005*, 1637-1640

Baye Yimam and TEAM 503 students. 1997. "ፊደል እንደገና" Ethiopian Journal of Languages and Literature 7(1997): 1-32.

Baye Yimam. 1986. "የአማርኛ ሰዋሰው". Addis Ababa. ት.መ.ማ.ማ.ድ.

Bender, L.M. and Ferguson C. 1976. The Ethiopian Writing System. In Language in Ethiopia. Edited by M.L. Bender, J.D. Bowen, R.L. Cooper, and C.A. Ferguson. London: Oxford University press.

Cowley, Roger, Marvin L. Bender and Charles A. Fergusone. 1976. The Amharic Language-Description. In Language in Ethiopia. Edited by M.L. Bender, J.D. Bowen, R.L. Cooper, and C.A. Ferguson. London: Oxford University press.

Deller, J.R. Jr., Hansen, J.H.L. and Proakis, J.G., Discrete-time Processing of Speech Signals. Macmillan Publishing Company, New York, 2000.

Fukada, Toshiaki, Takayoshi Yoshimura and Yoshinori Sagisa. 1999. Automatic generation of multiple pronunciations based on neural networks. Speech Communication 27:63—73 http://citeseer.ist.psu.edu/fukada99automatic.html.

Ganapathiraju, Aravind; Jonathan Hamaker; Mark Ordowski; and George R. Doddington. 1997. Joseph Picone. Syllable-based Large Vocabulary Continuous Speech Recognition.

Getachew Haile. 1967. The Problems of the Amharic Writing System. A paper presented in advance for the interdisciplinary seminar of the Faculty of Arts and Education. HSIU.

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

Hu, Zhihong; Johan Schalkwyk; Etienne Barnard; and Ronald Cole. 1996. Speech recognition using syllable like units. Proc. Int'l Conf. on Spoken Language Processing (ICSLP), 2:426-429.

Kanokphara, Supphanat; Virongrong Tesprasit and Rachod Thongprasirt. 2003. Pronunciation Variation Speech Recognition Without Dictionary Modification on Sparse Database, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003, Hong Kong).

Kinfe Tadesse. 2002. Sub-word Based Amharic Word Recognition: An Experiment Using Hidden Markov Model (HMM), M.Sc Thesis. Addis Ababa University Faculty of Informatics. Addis Ababa.

Lee, C-H., Gauvain, J-L., Pieraccini, R. and Rabiner, L. R.. 1992. Large vocabulary speech recognition using subword units. Proc. ICSST-92, Brisbane, Australia, pp. 342-353.

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

Martha Yifiru. 2003. Application of Amharic speech recognition system to command and control computer: An experiment with Microsoft Word, M.Sc Thesis. Addis Ababa University Faculty of Informatics. Addis Ababa.

Rabiner, L. and Juang, B. 1993. Fundamentals of speech recognition. Englewood Cliffs, NJ.

Hussien Seid and Björn. Gambäck 2005. A Speaker Independent Continuous Speech Recognizer for Amharic. In: INTERSPEECH 2005, 9th European Conference on Speech Communication and Technology. Lisbon, September 4-9.

Solomon Birihanu. 2001. Isolated Amharic Consonant-Vowel (CV) Syllable Recognition, M.Sc Thesis. Addis Ababa University Faculty of Informatics. Addis Ababa.

Solomon Teferra Abate. 2006. Automatic Speech Recognition for Amharic. Ph.D. Thesis. University of Hamburg. Hamburg.

Solomon Teferra Abate, Wolfgang Menzel and Bairu Tafla. 2005. An Amharic Speech Corpus for Large Vocabulary Continuous Speech Recognition. In: INTERSPEECH 2005, 9th European Conference on Speech Communication and Technology. Lisbon, September 4-9.

Tadesse Beyene. 1994. The Ethiopian Writing System. Paper presented at the 12th International Conference of Ethiopian Studies, Michigan State University.

Wu, Su-Lin. 1998. Incorporating Information from Syllable-length Time Scales into Automatic Speech Recognition. PhD thesis, University of California, Berkeley, CA.

Young, Steve; Dan Kershaw; Julian Odell and Dave Ollason. 2002. The HTK Book.

Zegaye Seyifu. 2003. Large vocabulary, speaker independent, continuous Amharic speech recognition, M.Sc Thesis. Addis Ababa University Faculty of Informatics. Addis Ababa.