

Amharic Speech Recognition: Past, Present and Future

¹Solomon Teferra Abate, Martha Yifiru Tachbelie,
Wolfgang Menzel

1. Introduction

Human beings have been dreaming of and struggling for developing intelligent machines. An important function of these machines is that they master speech input and output to communicate with us naturally. They understand human speech and generate human like speech (Chollet, 1994). A field that aims at the development of technologies that enable human beings to talk naturally with computers and or other devices is called spoken language processing. It refers to technologies related to speech recognition (converts speech into words), text-to-speech (converts text to speech), and speech understanding (maps words into action and plans system-initiated actions) (Huang et al., 2001). Abiding to our title, in this paper, we concentrate only on speech recognition.

1.1. Automatic Speech Recognition

Automatic Speech Recognition (ASR), as described by Junqua and Haton (1996), is the “decoding of the information conveyed by a speech signal and its transcription into a set of characters.” The resulting characters can be used to perform various tasks such as controlling a machine, accessing a database, or producing a written form of the input speech. ASR is a difficult task and as a result, different constraints may need to be imposed during the development of speech recognition systems. Based on the constraints, ASR systems may be categorized as follows:

- Speaker dependent or independent: Speaker-independent systems are capable of recognizing speech from any new speaker. Speaker dependent systems, on the other hand, recognize speech from those people whose speech is used during the development of the recognizer.
- Isolated or continuous speech: In isolated word recognizers, a user must pause between words that means artificial pause should be inserted before and after each word whereas in continuous speech recognizers, a user speaks naturally.

¹ All the Authors are affiliated to the University of Hamburg, Department of Informatics, natural language systems division. Vogt-Kölln-str. 30, D-22527 Hamburg, Germany. email: solomon_teferra_7@yahoo.com, {tachbeli, menzel}@informatik.uni-hamburg.de.

- Small, Medium or Large Vocabulary system: small vocabulary recognition systems are those which have a vocabulary size of 1 to 99 whereas medium and large vocabulary systems have a vocabulary size of 100 - 999 and 1000 or more words, respectively.
- Read or spontaneous speech: a speech recognizer may be designed to recognize only read speech or to recognize spontaneous speech which might have disfluencies.
- Noisy or Noise free speech: recognizers may recognize speech in noisy or only in noise free environment.

Although we have different kinds of speech recognition systems, they have similar components. Figure 1. depicts the components of a statistical speech recognition system.

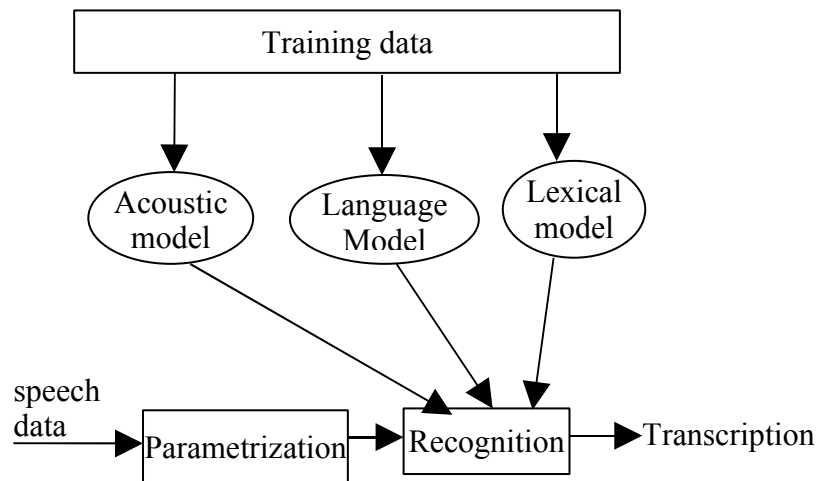


Figure 1. Components of ASR system (Adapted from Zue et al. (1996))

As it can be seen from the above figure, training data (speech and text corpus) are important resources in the process of developing ASR systems. But, the speech data can not be used directly. There is a need for parameterization which involves converting the speech wave form to some type of parametric representation (Markowitz, 1996). Developing ASR systems, specifically large vocabulary, speaker independent ones, involves the development of the acoustic model (from the speech data), the lexical model and the language model (from text data). During recognition the three models – acoustic, lexical and language model – work together. The acoustic model provides the probability that when the speaker utters a word sequence the acoustic processor produces the representation of the word sequence, the lexical model specifies the words that may be output by the speech recognizer and the language model provides an estimate of the probability of a word sequence W for a give recognition task.

Different types of speech recognition systems are required in different application domains. For instance, there is no need to have a large vocabulary speech recognition system for an automatic phone dialing task, small vocabulary systems might be sufficient. The following section presents the application of speech recognition systems.

1.1.1. Application of automatic speech recognition

ASR systems are useful in various aspects of life. In fact, any task that involves interfacing with a computer can potentially use ASRs. The following are the most common application areas of speech recognition systems.

- Dictation systems: This includes medical transcriptions, legal and business dictation, and general word processing;
- Command and Control systems: These are systems that use speech input to perform functions and actions. Using utterances like "Open file", the systems will do just that;
- Telephony systems: ASR systems allow callers to speak commands instead of pressing buttons to dial a number;
- Meeting transcription systems;
- As an assistive technology for disabled people;
- Computer Aided Instruction systems;
- Information retrieval systems;
- Broadcast news transcription systems.

The desire to automate the above mentioned and other tasks that inherently require human-machine interactions as well as technological curiosity about the mechanisms underlying the human speech capability attracted researchers' attention to speech recognition research (Juang and Rabiner, 2005). Research in automatic speech recognition by machine has been undertaken for more than fifty years. The following is a brief description of the development in the area of speech recognition.

1.1. 2. Historical Overview and state of the art²

Research in ASR started in the late 1940s and early 1950s simultaneously in Europe with Dreyfus-Graf who developed the first "Phonetographe" and in U.S.A with K-H Davis (and his colleagues at Bell Laboratory) who built the first speaker dependent isolated digit recognizer. These systems were electronic devices. Research in computer-based ASR were started in the late 1950s and early 1960s. The speaker independent vowel recognition system at MIT Lincoln and a system to recognize 10 syllables of a single speaker at RCA laboratories are among the early recognition systems. It was at this time that several

² The description is based on Juang and Rabiner (2005) and Junqua and Haton (1996) unless otherwise specified.

Japanese laboratories built special purpose hardware to perform a speech recognition task. In general, the 1960s were the time for small vocabulary, isolated word recognition systems developed based on simple acoustic-phonetic properties of speech sound.

The 1970s were the time for medium vocabulary recognition system built using simple template-based, pattern recognition methods. The first speech recognition commercial company, which developed VIP-100 system, was also founded during this time. Although VIP-100 has been applied only in a few simple applications (for quality control, package sorting), it greatly influenced Advanced Research Projects Agency of the U.S. department of Defense (DARPA) to fund the Speech Understanding Research (SUR) program. Harpy (which was shown to be able to recognize speech using a vocabulary of 1,011 words with reasonable accuracy), Hearsay-II and HWIM (Hear What I Mean) are among the systems developed under the DARPA's SUR program. It is also at this period that IBM researchers started studying large vocabulary speech recognition for different tasks, notably for voice activated type writer, and AT&T Bell laboratory researchers started to study a speaker independent speech recognition to provide automated telecommunication service.

In 1980s researchers started to tackle large vocabulary speaker-independent, continuous speech recognition problems based on statistical methods using technologies such as hidden Markov models (HMM). Speaker-independence has been possible because of HMM's ability of capturing and modeling speech variability. Artificial neural networks (ANN) were also reintroduced in the late 1980's. Recording of large speech databases such as TIMIT, which contributed to the advances made in ASR, was done in this period.

In 1990s it was possible to build large vocabulary systems with unconstrained language models, and constrained task syntax for continuous speech recognition and understanding. The success of statistical methods revived the interest from DARPA leading to several speech recognition systems such as the Sphinx system, the BYBLOS system and the DECIPHER system from CMU, BBN and SRI respectively. Real speech-enabled applications such as AT&T's VRCP (automated handling of operator-assisted calls), Universal Card Service (customer service line), "How May I Help You? (HMIHY)" system for call routing of consumer help line calls and United Airlines' automatic flight information system have been realized in the late 1990s. As a result, researchers developed an increasing interest for speech processing under noisy or adverse conditions, spontaneous speech recognition and dialog management.

In the first half of 2000s, the attention of researchers has been attracted towards spontaneous, robust, multimodal speech recognition. Although accuracy of more than 95% can be achieved using read speech, the accuracy drastically decreases for spontaneous speech. On the other hand, speech recognition systems applied for real applications should have high accuracy for spontaneous speech and should also be

robust. “Spontaneous Speech: Corpus and Processing Technology” is a national project in Japan working on spontaneous speech recognition. The world largest spontaneous speech corpus, Corpus of Spontaneous Japanese (CSJ), is one product of the project.

Human beings use multimodal communication when they speak with each other especially in noisy environments. Speech recognition systems developed using multimodal (audio and visual) information showed better recognition performance than systems developed using only speech information. During this period, DARPA conducted The Effective Affordable Reusable Speech-to-Text (EARS) program so as to develop automatic transcription with the aim of achieving more accurate output (Furui, 2005).

As it can be observed from the above description, speech recognition technology has gone through significant progress. The use of HMM, the development of large speech corpora for system development training and testing, establishment of standards for performance evaluation, and advances in computer technology are the factors that fueled the progress.

Recognition errors decrease from time to time. As indicated by (Zue et al., 1996), a small vocabulary continuous speaker independent digit recognition system can achieve an error rate of 0.3%. A medium vocabulary speaker independent recognition system has an error rate of less than 4% using word-pair language model on the resource management (RM) task, in which inquiries can be made concerning various naval vessels in the Pacific Ocean. An error rate of less than 3% has been reported for large vocabulary speech recognition system which uses a bigram language model in the Air Travel Information Service (ATIS) domain. In 1994 the best very large (20,000 words or more) speaker independent continuous speech recognition system had an error rate of 7.2%. Recently, Fellbaum and Freitas (2007) indicated the fact that very large vocabulary (in order of 50,000 words or more) speaker independent (with the use of speaker adaptation) systems with error rate of less than 2% are available on the market.

At present, several very large vocabulary dictation systems are available for document generation. Dragon Naturally Speaking is, for example, one commercial dictation system from Nuance that is claimed to have almost perfect (as high as 99%) accuracy.

Although lots of work has been conducted in the area of ASR for technologically favored languages for more than 50 years, research in Amharic speech recognition started only recently, 5 years ago. The following section pinpoints past attempts in Amharic ASR.

2. Attempts in Amharic ASR

Research in automatic speech recognition for Amharic started in 2001 when Solomon (2001) developed isolated Consonant-Vowel syllable recognition system. As we can see from section 1.1.2, for other languages this kind of system has been developed in 1960s, and 2001 is

a time where so many speech recognition systems and applications have been developed with an accuracy of 98% and above. In this section we give a review of works done so far in Amharic ASR.

Solomon (2001) developed a speech recognizer that Amharic that recognizes a sub-set of isolated consonant-vowel (CV) syllable using the HTK (Hidden-Markov Modeling Toolkit). He selected 41 CV syllables of Amharic language out of 234. Speech data of the selected CV syllables has been recorded from eight people (4 male and 4 female) with the age range of 20 to 33 years. The average recognition accuracies he could achieve were 87.68% and 72.75% for speaker dependent and independent systems, respectively. As indicated by the researcher, the result seems to be low compared to other systems developed for other languages. This might be due to problems of the recording environment and insufficient training speech data.

Kinfe (2002) developed sub-word based isolated word recognition systems for Amharic using HTK. The sub-word units used in his experiment are phones, triphones, and CV-syllables. He considered 20 phones (out of 39) and 104 CV syllables, which are formed using the selected phones. Speech data of 170 words, which are composed of the selected sub-word units, have been recorded from 20 speakers (speech of 15 speakers for training and the remaining for testing). Speaker dependent phone-based and triphone-based systems have an average recognition accuracy of 83.07% and 78% respectively. Phone-based and triphone-based speaker independent systems have an average recognition accuracy of 72% and 68.4% respectively. In addition, a comparison of the different sub-word units revealed that the use of CV syllables has led to relatively poor performance. This is a contradiction with recent findings of Solomon (2006) who attributed the poor performance of Kinfe's CV syllable-based recognizers to the use of insufficient speech data.

Zegaye (2003) investigated the possibility of developing large vocabulary, continuous speech and speaker independent Amharic speech recognizer. He built both phone-based and tri-phone based recognizers using HTK. For the training and testing of his recognizers, he used the speech data, which consists of 8000 sentences read by 80 people, recorded by Solomon et al. (2005). The best recognizer was a tri-phone based recognizer which has 76.2% word recognition accuracy.

Martha (2003) explored the possibility of developing an Amharic speech input interface to command and control Microsoft Word. It requires a speech recognizer and a communication interface between the recognizer and the application. 50 command words were selected from different menus (File, View, Insert, Tools, Table, Window, and Help), translated to Amharic and used to develop the prototype system. Speech data of these 50 command words were recorded from 26 people (10 female and 16 male) in the age range of 20 to 35. 76.9% of the recorded data were used to train the recognizers and the remaining for testing the performance of recognizers. The communication interface has been written using Visual Basic 6. To test the performance of the system as a

whole, 18 randomly selected command words were given to 6 people (3 command words for each) and these people were asked to command Microsoft Word orally. The system performed 16 commands accurately and only two command words were wrongly recognized and thus Microsoft Word performed wrong actions.

Molalgne (2004) tried to compare HMM-based small vocabulary speaker dependent continuous speech recognizers built using two different toolkits: HTK, and the MSSTATE Toolkit from Mississippi State. He collected a speech data of 50 sentences with ten words (the digits) from a single speaker. The speaker dependent recognition system developed using the HTK has 82.5% word recognition accuracy whereas the system developed using MSSTATE has 79.0% word recognition accuracy. He also found out that HTK was faster than MSSTATE Toolkit.

Hussein and Gambäck (2005) developed an Amharic speaker independent continuous speech recognizers based on an HMM/ANN hybrid approach. The recognizer was constructed at a context dependent phone level with the help of the CSLU Toolkit. They used part of the data (5000 sentences) recorded by Solomon et al. (2005). The recognizers were tested with a total of 20 sentences read by 10 speakers (2 sentences each) who also read the training data. The best result obtained with this test data was 74.28% word recognition accuracy. When tested with test data read by speakers who were not in the training, performance degradation was observed. There was 4.28% absolute word recognition accuracy reduction.

Since there was no usable speech corpus for the development of a large vocabulary speech recognizer, Solomon (2006) developed an Amharic speech corpus that can be used for various kinds of investigations in the development of ASR for Amharic. He explored various possibilities for developing a Large Vocabulary Speaker Independent Continuous Speech Recognition System for Amharic. Solomon assumed that, due to their highly regular Consonant Vowel (CV) structure, Amharic syllables lend themselves to be used as a basic recognition unit. Indeed, he has been able to show that syllable models can be used as a competitive alternative to the standard architecture that is based on triphone models. The acoustic model of the syllable-based recognizer requires 15MB memory. Together with the language model and use of speaker adaptation it has a word recognition accuracy of 90.43% on the 5,000 words evaluation test set at a speed of 2.4 minutes per sentence. While the acoustic model of the triphone-based recognizer requires 38MB memory and has a word recognition accuracy of 91.31% on the same test set at a speed of 3.8 minutes per sentence. Although this is the state-of-the-art recognition performance in Amharic, we still see the room for improvement because the word recognition accuracy of ASR in the technologically favored languages is approaching to 100% .

Solomon also mentioned limitations of his work in the areas of speech corpora, language model, acoustic models, and the application of the recognizers. The developed speech corpus is a read speech corpus

and cannot be used to develop, e.g. recognizers for spontaneous speech or telephone-based applications. It also requires modifications for the development of speech recognizers for specific purposes like legal, medical, meeting and other similar transcriptions. The acoustic models of the speech recognition system have also suffered from a shortage of training speech data. The irregular realization of the sixth order vowel and the glottal stop consonant, as well as the gemination of the other consonants are not handled by the pronunciation dictionaries. Due to the rich inflection of Amharic, the language model has a relatively high perplexity. So far, no effort has been made towards the application of the recognizers.

Currently, Solomon is working on speaker adaptation and tying (for the syllable-based model) to improve his speech recognizers. Martha is working on language modeling to find the best way of modeling the Amharic language. But, still the speech recognition area is almost untouched and there is much to be done. The following section indicates the areas that need the attention of researchers and developers.

3. Areas of R&D Calling for our Attention

In this section we will try to indicate the gaps that should be filled in the area of ASR for Amharic. We generally categorized the gaps into 5 categories as follows.

- ◆ Speech Corpus
- ◆ Pronunciation dictionary
- ◆ Language Model
- ◆ Acoustic Model
- ◆ Application of ASR

3.1. Speech Corpus

A speech corpus is defined as a collection of speech recordings which is accessible in computer readable form, and which has an annotation and documentation sufficient to allow re-use of the data in-house, or by scientists in other organizations (Gibbon et al., 1997). A speech corpus is one of the fundamental requirements for any speech recognition research. Recording of large speech corpora contributed a lot to the advances made in ASR for other languages. Therefore, if advances in Amharic ASR are to be made, the need for adequate speech corpora is unquestionable. It is important to note that a speech corpus is a language specific resource and should be prepared for each and every language.

There are various kinds of speech corpora. Depending on the style of speech, Schiel et al. (2003) categorized speech corpora into the following groups:

- Read Speech
- Answering Speech

- Command / Control Speech
- Descriptive Speech
- Non-prompted Speech
- Spontaneous Speech
- Neutral vs. Emotional Speech

Each kind of corpus may be required for a different purpose. For example, a command and control speech corpus can only be used to develop speech recognition systems that can be effectively used for command and control applications. Thus to have speech recognition systems for different applications, the development of various kinds of corpora is a must.

To our knowledge, so far there is only one medium size read speech corpus for Amharic (Solomon et al., 2005). All other kinds of corpora still need to be developed. In addition, the existing read speech corpus is not complete. Solomon (2006) said that “we make no claim that complete work has been done in the area of developing a read speech corpus for Amharic.” Therefore, there is also a need to extend the existing speech corpus.

3.2. Pronunciation dictionary

A pronunciation dictionary is a machine-readable transcription of words in terms of sub-word units. It specifies the finite set of words that may be output by the speech recognizer and gives, at least, one pronunciation for each. The pronunciation dictionary, i.e. the lexical model, is one of the most important blocks in the development of large vocabulary speaker independent recognition systems.

A pronunciation dictionary can be classified as a canonical or alternative one on the basis of the pronunciations it includes. For each word 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 is a pronunciation dictionary that uses the actual phone (or other sub-word) sequences pronounced in speech (Fukada et al., 1999).

Like speech corpora, a pronunciation dictionary is also a language specific resource that is required in the course of ASR system development. For Amharic language, Solomon (2006) prepared canonical and alternative pronunciation dictionaries based on phones and syllables. The canonical pronunciation dictionaries have been developed using the phonetic feature of the Amharic writing system and transcribe about 50,000 words each. The syllable-based alternative pronunciation dictionary, which includes all pronunciation variations for each word, transcribes about 25,000 words. None of these pronunciation dictionaries handles the problems of gemination of consonants and irregular realization of the sixth order vowel and the glottal stop

consonant. These problems have a direct effect on the quality of sub-word transcriptions.

Therefore, proper editing of the pronunciation dictionaries is one work that needs attention of developers and researchers. In addition, extending the pronunciation dictionaries by adding more words is also a task on the agenda.

3.3. Language modeling

In language modeling, the problem is to predict the next word given the previous words (Manning and Schütze, 1999). It is fundamental not only to ASR but also to statistical machine translation (SMT), question answering, text summarization, spelling checkers, information retrieval, etc., (Gao and Lin, 2004). Statistical language models are the most commonly used types of language models (Juqua and Haton, 1996). Since different languages have different structure, language models are also basic language specific resources.

Solomon (2006) developed statistical language models (bigram) using the HTK toolkit and pointed out that they have a high perplexity³, which is mainly due to the morphological richness of the language.

Currently, Martha is working on morphology-based language modeling to identify the best language model for Amharic which takes the morphological richness of the language into consideration.

Although statistical language models work fine in reducing the search space and reducing acoustic ambiguity, language models with much constraints (syntactic and semantic) might be required as the vocabulary size grows and other constraints are relaxed to have more habitable systems. There is also a lot to be done in this area, which actually requires researchers from different areas.

Thus, development of appropriate language models for different NLP applications in general and ASR applications in particular, is also an open area for research.

3.4. Acoustic Model

The acoustic model (AM) models the sound units of a language based on speech features extracted from the audio signal. The acoustic models that have been developed as part of the research efforts reviewed above need further performance improvement. To this end, taking the state-of-the-art performance that is achieved by Solomon (2006) as a starting point, new methods for initializing the models, a refinement of the models by means of parameter tying, and speaker adaptation techniques, for example, need to be investigated. The use of more speech data is also an obvious option for performance improvement.

Using segmented speech data for the initialization is a method for ASR, but any segmentation error also has a significant negative impact on the performance of the models. Unfortunately we do not have

³ Perplexity is one of the metrics used to evaluate language models. Low perplexity means better, i.e. more restrictive, language model.

speech data that is segmented at sub-word level except the corpus developed by Solomon et al. (2005) that has been segmented using a poorly trained speech recognizer and which was edited only by non-linguists. Therefore, the preparation of properly segmented speech data for model initialization is also an area that requires our attention.

Since tying is one way of minimizing the problem of shortage of training speech data, tying the syllable models would possibly result in a gain of some degree of performance improvement. The CV syllable model is a model of a consonant and a vowel. Syllable models have less context information than the tri-phone models. This can be addressed by exploring methods of modeling contexts between syllables. The development of a bi-syllable and/or tri-syllable model together with the proper tying is recommended, to gain a considerable performance improvement for the syllable-based recognizers.

Performance improvement can be gained by the use of different algorithms of speaker adaptation. The attempts that have been reviewed in this work used only maximum likelihood linear regression (MLLR) technique, and have applied only the procedure that is described in the HTK tutorial. But, the MLLR technique itself is flexible and can be used in different ways, like varying the number of transformations to fit the available adaptation speech data (Young et.al., 2002). Furthermore, the maximum a posteriori (MAP)(Young et.al., 2002) and other adaptation approaches (Woodland, 2001; Chen, 1997), for example, have not been investigated so far. We, therefore, recommend working on the application of a speaker adaptation technique.

3.5. Application of ASR

The development of an ASR system is not an end by its own. There remains much work in the area of incorporating the Amharic ASR system in different areas that benefit from the use of ASR. Section 1.1.1 indicates the various application areas of speech recognition systems. As far as we know Amharic speech recognition technology has not been applied in any of the areas except Martha's study of technical feasibility (Martha, 2003). Therefore, this untouched area awaits our attention.

4. Conclusion

In this paper we tried to show the current status of research in ASR for Amharic and the gaps that need to be filled by future research. The attempts made in the area of Amharic ASR are encouraging. But in view of the wider realm of ASR that has been experienced for the technologically favored languages, a lot of work remains to be done. Some of the open questions are mentioned in section 3. To this end, researchers and financial supporters, including the Ethiopian government and commercial enterprises, need to work together.

5. References

- Chollet, Gerard. 1994. "Automatic Speech and Speaker Recognition: Overview, Current Issues and Perspectives." In *Fundamentals of Speech Synthesis and Speech Recognition: Basic Concepts, State of the Art and Future Challenges*. pp 129-147. Edited by Eric Keller. Chichester: John Willy & Sons.
- Fellbaum, Klaus and Diamantino Freitas. 2007. *Speech Processing*. In *Towards an Inclusive Future: Impact and Wider Potential of Information and Communication Technologies*. Edited by Patrick R.W. Roe. Available at: http://www.tiresias.org/cost219ter/inclusive_future
- Fukada, T., T. Yoshimura and Y.Sagisaka. 1999. Automatic generation of multiple pronunciation based on neural networks. *Speech Communication* 27:63-73. Available at: <http://citeseer.ist.psu.edu/fukada99automatic.html>
- Furui, Sadaoki. 2005. 50 Years of Progress in Speech and Speaker Recognition Research. *ECTI Transactions on Computer and Information Technology* Vol. 1. No. 2. Available at: http://www.ecti.or.th/~explorers/EEC/0511_cit/Content_Furui1.pdf.
- Gao, Jianfeng and Lin Chin-Yew. 2004. "Introduction to the Special Issue on Statistical Language Modeling" *ACM transactions On Asian Language Information Processing*. 3(2): 87-93.
- Gibbon, Dafydd, Roger Moore and Richard Winski. ed. 1997. *Handbook of Standards and Resources for Spoken Language systems*. Berlin and New York: Walter de Gruyter Publishers.
- Huang, Xuedong, Alex Acero and Hsiao-Wuen Hon. 2001. *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*. New Jersey: Prentice Hall PTR.
- Hussien Seid and Björn Gambäck. 2005. A Speaker Independent Continuous Speech Recognizer for Amharic. Available at: <http://eprints.sics.se/170/01/eurospeech05.pdf>.
- Juang, B. H. and Lawrence R. Rabiner. 2005. *Automatic Speech Recognition--A Brief History of the Technology Development*. Elsevier Encyclopedia of Language and Linguistics, Second Edition.
- Junqua, Jean-Claude and Jean- Paul Haton. 1996. *Robustness In Automatic Speech Recognition: Fundamentals and Applications*. Boston: Kluwer Academic Publishers.
- Kinfe Tadesse. 2002. Sub-word based Amharic speech recognizer: An experiment using Hidden Markov Model (HMM). MSc Thesis, School of Information Studies for Africa, Addis Ababa University, Ethiopia, June.
- Manning, Christopher D. and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. London: The MIT Press.
- Markowitz, Judith A. 1996. *Using Speech Recognition*. Upper Saddle River, New Jersey: Prentice Hall, Inc.
- Martha Yifiru. 2003. Automatic Amharic Speech Recognition System to Command and Control Computers. MSc Thesis, School of Information Studies for Africa, Addis Ababa University, Ethiopia.

- Molalgne Girmaw. 2004. An Automatic Speech Recognition System for Amharic. MSc Thesis, Dept. of Signals, Sensors and Systems, Royal Institute of Technology, Stockholm, Sweden.
- Schiel, Florian and Christoph Draxler. 2004. The Production of Speech Corpora. Available at: <http://www.phonetik.uni-muenchen.de/Forschung/BITS/TP1/Cookbook/TP1.html>
- Solomon Berhanu. 2001. Isolated Amharic Consonant-Vowel syllable recognition: An Experiment using the Hidden Markov Model. Msc Thesis, School of Information Studies for Africa, Addis Ababa University, Ethiopia.
- Solomon Tefera Abate, Wolfgang Menzel and Bahiru Tefla. 2005. An Amharic speech corpus for large vocabulary continuous speech recognition. In Proc. 9th Europ. Conf. on Speech Communication and Technology, Interspeech-2005, Lisbon, Portugal.
- Solomon Teferra Abate. 2006. Automatic Speech Recognition for Amharic. Ph.D. Thesis. Available at: <http://www.sub.uni-hamburg.de/opus/volltexte/2006/2981/pdf/thesis.pdf>.
- Zegaye Seifu. 2003. HMM based large vocabulary, speaker independent, continuous Amharic speech recognizer. Msc Thesis, School of Information Studies for Africa, Addis Ababa University, Ethiopia.
- Zue, Victor, Ron Cole and Wayne Wand. 1996. "Speech Recognition" In Survey of the State of the art in Human Language Technology. Edited by Ronald A. Cole, Joseph Mariani, Hans Uszkoreit, Annie Zaenen and Victor Zue. Available at: http://www.lt-world.org/HLT_Survey/master.pdf