ISCA Archive
http://www.isca-speech.org/archive

Third Workshop on Spoken Language
Technologies for Under-resourced Languages
Cape Town, South Africa
May 7-9, 2012

SYLLABLE-BASED AND HYBRID ACOUSTIC MODELS FOR AMHARIC SPEECH RECOGNITION

Martha Yifiru Tachbelie, Solomon Teferra Abate, Laurent Besacier, Solange Rossato

Laboratoire d'Informatique de Grenoble, France
marthayifiru,solomon_teferra_7@yahoo.com
laurent.besacier,solange.rossato@imag.fr

ABSTRACT

This paper presents the results of our experiments on the use of hybrid acoustic units in speech recognition and the use of syllable and hybrid acoustic models (AM) in morpheme-based speech recognition. Although hybrid AMs did not bring improvement in speech recognition performance when words are used as dictionary entries and units in a language model (LM), we observed a significant word error rate (WER) reduction (compared to triphone-based systems) in morpheme-based speech recognition. Syllable AMs also led to a significant WER reduction over the triphone-based systems. It was possible to obtain a 3% absolute WER reduction as a result of using syllable acoustic units. Generally, our result shows that syllable and hybrid AMs are best fitted in morpheme-based speech recognition.

***Index Terms*—** syllable-based acoustic models, hybrid acoustic models, morpheme-based speech recognition, Amharic

1. INTRODUCTION

A speech recognition task requires segmentation of speech into fundamental acoustic units. The ideal and natural acoustic unit is a word. However, the use of words as acoustic unit in large vocabulary continuous speech recognition (LVCSR) systems is impractical because of the need for a very large data to train models adequately. Thus, sub-word acoustic units such as phones or context dependent phones are generally used in LVCSR. Nevertheless, the phone-based units are not efficient in modeling long-term temporal dependencies. The solution is the use of multi-phone sub-word units, particularly syllables in LVCSR systems. Many researchers ([1], [2], [3], [4], [5], [6]) investigated the use of syllables in acoustic modeling.

For Amharic, the first experiment on syllable-based speech recognition is due to [4]. They experimented on the use of consonant vowel (CV) syllables (corresponding to the Amharic orthography) in acoustic modeling. Models with different HMM topologies have been developed. A model with 5 states per HMM and without skips found to be the best one in terms of accuracy. Compared to a triphone-based recognizer, the context independent syllable-based one performed slightly worse in terms of accuracy. However, recognition speed and storage requirement are improved. Thus, they concluded that the use of CV syllables is a promising alternative in the development of automatic speech recognition for Amharic.

Recently, we [7] have further investigated the use of syllables for Amharic speech recognition. Like [4], we used CV syllables for acoustic modeling since CV syllables cover the majority of syllables in the language and correspond to the Amharic orthographic symbols. In addition to context independent syllable acoustic modeling, we have experimented with context-dependent syllable acoustic modeling which was not addressed by [4]. Our result showed that syllable acoustic units lead to a better word recognition accuracy than the triphone acoustic units for Amharic speech recognition.

From our previous work [7], we noticed the existence of rare CV syllables in the training data. Since it is not feasible to record audio data in order to enrich the rare syllables with the time and resource available for our experiment, we decided to decompose the rare CV syllables into phones and train hybrid (CV syllable and phones) AMs. Moreover, manual analysis of recognition results of the syllable-based and triphone-based systems showed that most short (consisting of only two CVs) words are correctly recognized by the syllable-based recognizers while deleted by the triphone-based ones. This inspired us to investigate the use of syllable AMs in morpheme-based speech recognition where morphemes are used as entries in the pronunciation dictionary as well as units in the LM. Since morphemes are usually shorter than words, we expected that syllable AMs lead to a better recognition performance in morpheme-based speech recognition system.

In this paper we present the results of our experiments conducted on hybrid acoustic units for Amharic speech recognition and the use of syllable as well as hybrid AMs in morpheme-based speech recognition. To our knowledge, this work is the first in investigating the use of longer (syllables) acoustic units and shorter (morphemes) lexical units in a speech recognition system. The next section gives a brief overview of the Amharic language. In section 3, we present the speech and text corpora used in our experiments. The different (triphone, CV syllable and hybrid) speech recognition systems are described in section 4. Section 5 deals with morpheme-based speech recognition experiments. Finally, in section 6, conclusions and future directions are given.

2. THE AMHARIC LANGUAGE

Amharic is a member of the Ethio-Semitic languages, which belong to the Semitic branch of the Afroasiatic super family [8]. It is related to Hebrew, Arabic, and Syrian. Amharic, which is spoken mainly in Ethiopia, is the second most populous Semitic language, after Arabic. According to the 1998 census, it is spoken by over 17 million people as a first language and by over 5 million as a second language throughout different regions of Ethiopia. Currently, Amharic is the official working language of the federal democratic republic of Ethiopia and of several states within the federal system.

2.1. Amharic Phonetics

Amharic has a total of 38 phonemes (31 consonants and 7 vowels) [9]. On the basis of their manner of articulation, the consonants are classified into stops, fricatives, nasals, liquids and semivowels. The phonetic transcription of most of the Amharic consonants corresponds to that of the English ones. However, the ejective sounds (t' k' p' s' ʧ') are peculiar to Amharic and they do not have correspondents in English.

Amharic has seven vowels: (ə u i a e ɨ o). [9] said that there is no precise correspondence between Amharic and English vowels. The vowel /ə/ is pronounced like 'er' in 'bigger' and the vowel /u/ is pronounced like 'o' in 'who'. The vowels /i/ and /a/ are pronounced like 'ee' in 'feet' and 'a' in father, respectively. The other vowels /e/, /ɨ/ and /o/ are pronounced as 'a' in 'state', as 'e' in 'roses' and as 'o' in 'shore'.

2.2. Amharic Morphology

Like other Semitic languages such as Arabic, Amharic exhibits a root-pattern morphological phenomenon. A root is a set of consonants (called radicals) which has a basic 'lexical' meaning. A pattern consists of a set of vowels which are inserted (intercalated) among the consonants of the root to form a stem. In addition to this non-concatenative morphological feature, Amharic uses different affixes to form inflectional and derivational word forms.

Some adverbs can be derived from adjectives but, they are not inflected. Nouns are derived from other basic nouns, adjectives, stems, roots, and the infinitive form of a verb by affixation and intercalation. Case, number, definiteness, and gender marking affixes inflect nouns. Adjectives are derived from nouns, stems or verbal roots by adding a prefix or a suffix. Adjectives can also be formed through compounding. Like nouns, adjectives are inflected for gender, number, and case [10].

Unlike the other word categories, the derivation of verbs from other parts of speech is not common. The conversion of a root to a basic verb stem requires both intercalation and affixation. For instance, from the root *gdl* 'kill' we obtain the perfective verb stem *gəddəl-* by intercalating pattern ə-ə. From this perfective stem, it is possible to derive the passive stem *təgəddəl-* and the causative stem *asgəddəl-* using prefixes tə- and as-, respectively. Other verb forms are also derived from roots in a similar fashion. Verbs are inflected for person, gender, number, aspect, tense and mood [10]. In this work, only the concatenative morphemes are considered.

2.3. Amharic Writing System

Amharic is written in its own script known as fidəl. The Amharic script is syllabary since each symbol represents a consonant combined with a vowel and the vowel has no independent existence [9]. In other words, each symbol in Amharic orthography represents a consonant and vowel (CV) syllable. Exceptions are the glottal stop and the 6th order fidəl. The glottal stop consonant may or may not be pronounced and, therefore, the fidəl may represent only a vowel (when the glottal stop is not pronounced). Each fidəl has seven different shapes or orders according to the vowels combined to it. However, 6th order fidəls may represent only a consonant since the vowel /ɨ/ (combined to the consonants to form 6th order fidəls) serves as an epenthetic vowel.

The writing system consists of 276 distinct symbols, 20 numerals and 8 punctuations. There are 33 core consonants each of which have seven shapes or orders. This makes 231 (33x7) distinct symbols (CV syllables) out of the 276. The remaining symbols include labiovelars (20), labialized consonants (18) and the consonant /v/ (which appears only in modern loan words like /viza/ meaning visa) in its seven orders.

2.4. Amharic Syllable Structure

Most Amharic linguists ([10], [11]) agree that the syllable structure of Amharic is (C)V(C)(C) where C represents a consonant and V a vowel. That means the syllable types of Amharic are V, CV, CVC, VC, CVCC, VCC. Since CV syllables cover the large majority of syllable distribution in the language [12] and the Amharic writing system is syllabary (representing CV syllables), only CV syllables have been considered in our syllable-based AMs.

As indicated in section 2.3, Amharic has 276 distinct CV syllabic symbols. However, some of the symbols are duplicate in a sense that they represent the same syllabic sounds. Like [4], redundant symbols that represent the same syllabic sounds have been eliminated and a total of 233 CV syllables have been considered as acoustic units in our study.

3. SPEECH AND TEXT DATA

The speech corpus used to develop the different speech recognition systems (described in Section 4) is an Amharic read speech corpus [13] that contains 20 hours of training speech collected from 100 speakers who read a total of 10,850 sentences (28,666 tokens). Compared to other speech corpora that contain hundreds of hours of speech data for training, our models obviously suffer from a lack of training data. Although the corpus includes four different test sets (5k and 20k both for development and evaluation), for the purpose of the current investigation we have used the 5k development test set, which includes 360 sentences (4,106 tokens or 2,836 distinct words) read by 20 speakers.

The ATC_120k text corpus [14] consisting of 120,262 sentences (2,348,150 tokens or 211,120 types) has been used to derive the vocabulary for the pronunciation dictionary and to train LMs.

4. AMHARIC SPEECH RECOGNIZERS

4.1. Experimental Setup

The acoustic features used (in all the AMs described in Sections 4.2, 4.3 and 4.4) consist of 13 dimensional Mel Frequency Cepstral Coefficient (MFCC). A window size of 25ms with an overlap of 10ms has been used in the estimation of the MFCCs. The AMs have been trained using Sphinx[1], one of the most widely used open source speech recognition toolkits.

A word trigram LM has been developed using the ATC_120k corpus and the SRILM toolkit [15]. The LM is smoothed with modified Kneser-Ney smoothing technique and made open by including a special unknown word token. Moreover, since the amount of the training text is small, all trigrams (regardless of their number of occurrence) have been included in the model. This LM has been used in triphone, context-dependent CV syllable and hybrid recognition systems described in Sections 4.2, 4.3 and 4.4, respectively.

4.2. Triphone-Based recognizers

The triphone-based systems have been developed for comparison purpose. The number of phones used in these systems is 41 (inclusing a silence). The number is more than the number of Amharic phones indicated in Section 2.1 since we have separately modeled the rounding feature of labiovelars and labialized consonants as vowels. The systems' vocabulary (pronunciation dictionary) consists of 65k most frequent words taken from the ATC_120k text corpus. We have automatically generated the pronunciation of each word based on the transcription of the fidəls.

An HMM (Hidden Markov Model) of 3-state Bakis topology (with an additional non-emitting last state) has been used. We have trained several cross-word triphone models with varying number of tied state triphones (senones) and Gaussian mixtures. The best performing model, which has the smallest word error rate (WER), namely 18.0%, consists of 2,500 senones with 16 Gaussian mixtures. A WER of 17.9% has been obtained as a result of using user defined tree questions (see Triphone_3states_UDQ in Table 1).

4.3. Syllable-Based recognizers

The entries in the syllable-based recognizers' pronunciation dictionary are the 65k words used in the triphone-based recognizers. However, each word has been automatically segmented into CV syllables taking the advantage of the syllabary nature of the Amharic writing system. Although, there are irregularities (with the 6th order and the glottal stop fidəl) in the Amharic writing system, we have uniformly used CV syllable representation for all the Amharic fidəls.

A 5-state Bakis topology with no skip and with an additional non-emitting last state has been used. As we did with the triphone-based systems, different models have been developed using different number of senones and Gaussian mixtures. The model with 1,500 senones and 24 Gaussian mixtures is the best one with a WER of 17.1%. No WER reduction has been obtained as a result of using user defined questions. Rather, as it is shown in Table 1, a slight increase in WER (17.1% to 17.3%) has been observed. This can be explained with the simplicity[2] of the clusters that we have defined for the syllable-based recognizers.

4.4. Hybrid recognizers

From our experiments on syllable-based acoustic modeling, we have noticed that some of the syllables are relatively rare in the training data and, therefore, not trained very well. As a solution for this problem, we have decomposed the rare syllables (based on their frequency in the training transcription) into constituent phones and trained hybrid (phone/syllable) AMs. We have prepared two versions of training pronunciation dictionaries: HyridDict_FL100 and HyridDict_FL500. In HyridDict_FL100, all the CV syllables with a frequency of less than 100 have been decomposed into their constituent phones. The number of distinct pronunciation units considered in this dictionary is 204 (31 phones, 172 CV syllables and a silence). CV syllables that appeared less than 500 times in the training transcription have also been decomposed (into phones) to form HyridDict_FL500 dictionary. The total number of distinct pronunciation units in this dictionary is 170 (41 phones, 128 CV syllables and a silence). Hybrid AMs have, then, been developed using these dictionaries (HyridDict_FL100 and HyridDict_FL500) with different HMM topologies. Since it is difficult (in Sphinx toolkit) to

[1]http://cmusphinx.sourceforge.net

[2]The clusters (Nasals, fricatives, etc.) are defined based only on the phonetic category of the consonant irrespective of the vowel of the syllable

use different number of states for various units (phones and syllables) in one system, we used an HMM topology of 5 states with skips. We assume that this HMM topology handles the irregularities (in length) of the hybrid acoustic units. However, for comparison purpose, we have also developed hybrid AMs with 4 and 5 states without skips.

The hybrid AMs have been evaluated using the 65k pronunciation dictionary used in the syllable-based recognizers. However, the CV syllables that are decomposed into phones in the training dictionaries have also been decomposed in the 65k dictionary. Table 1 presents the performance of the hybrid system evaluated on the 5k development test set.

Models	WER in %
Triphone	18
Triphone_3states_UDQ	17.9
CD_Syllable	17.1
CD_Syllable_UDQ	17.3
Hybrid_204Units_5statesWS[a]	17.8
Hybrid_204Units_5statesWOS	16.9
Hybrid_204Units_4statesWOS	17.1
Hybrid_170Units_5statesWS[b]	18.9
Hybrid_170Units_5statesWOS	17.0
Hybrid_170Units_4statesWOS	17.5

Table 1. WER of Several AMs using Word-based LM

[a]Models trained with a HyridDict_FL100 dictionary.
[b]Models trained with a HyridDict_FL500 dictionary.

As it can be seen from the table, decomposing rare syllables into phones did not bring significant performance improvement. Rather, in some cases, the result is even worse compared to the pure context dependent CV syllable-based speech recognition systems. The use of 5 states with skip model topology led to the worst performance (17.8% and 18.9% for the AMs developed with HyridDict_FL100 and HyridDict_FL500 dictionaries, respectively). Although this topology enables us to capture irregularities in acoustic units' length, it requires much training data as the number of parameters (transition matrices) to be estimated are larger than AM without skips. This is why hybrid models with skips did not perform well compared to that of 5 states without skips (see Hybrid_204Units_5statesWOS and Hybrid_170Units_5statesWOS in the table).

5. MORPHEME-BASED RECOGNITION

From the manual analysis of recognition outputs of syllable- and triphone-based recognizers, we have observed that most short words are recognized by the syllable-based systems but not by the triphone-based ones. Based on our observation, we investigated the use of syllable and hybrid AMs in morpheme-based recognition system with the assumption that these AMs will lead to a better recognition performance in such system since morphemes are usually shorter than words. This section presents the results of morpheme-based recognition experiments following a brief description of the morphological segmentation methods used in our experiments. The results reported are word error rates computed after reconstructing words from the recognized morpheme sequences. They are, therefore, directly comparable with the results reported in Section 4.

5.1. Morphological Segmentation

Two segmentation methods have been applied to obtain morphologically segmented text: unsupervised and finite state-based supervised segmentation. Morfessor [16], which is a freely available language independent unsupervised morphology learning tool that tries to identify all the morphemes found in a given word, has been used for unsupervised segmentation. In order to obtain a grammatical morpheme segmentation, we used a segmented text described in [17] to train a finite state machine (FSM) based morphological segmenter (a composition of morph transducer and 12gram CV syllable-based LM) using the AT&T FSM Library and GRM Library (Grammar Library) [18]. This segmenter segments only prepositions (mostly prefix) and conjunctions (mostly suffix) from words.

Morfessor and the FSM-based segmenter have been used to segment the ATC_120k text corpus. That means we have two versions of morphologically segmented corpora: Morfessor segmented and FSM segmented. The Morfessor segmented corpus consists of 4,035,592 morpheme tokens or 15,933 types. The FSM segmentation, on the other hand, resulted in a morpheme-based text corpus consisting of 3,104,474 and 142,855 morpheme tokens and types, respectively. These corpora have been used to prepare pronunciation dictionaries and LMs for morpheme-based speech recognition experiment.

In order to facilitate the conversion of morpheme sequences to words after recognition, a special morpheme boundary marker has been attached to the left and right of morphemes obtained in unsupervised manner. For the morphemes obtained using the FSM-based segmentation, two boundary markers "#" and "+" have been attached to prefixes and suffixes, respectively. This made the morphemes (in Morfessor and FSM segmented texts) context sensitive and consequently increased the number of distinct morphemes to 45,338 for Morfessor and 144,024 for FSM segmented text.

5.2. Acoustic, Language and Lexical Models

The AMs used in morpheme-based recognition are the triphone, context dependent CV syllables and hybrid models described in Section 4. Morpheme-based trigram LMs have been developed (in similar fashion as the word trigram LM described in Section 4.2) with the unsupervised and supervised morphologically segmented corpora. For the latter corpus, a 65k morpheme vocabulary has been prepared by taking the most frequent morphemes. This vocabulary has been used to prepare three types of pronunciation dictionaries according to the units (phone, CV syllable, hybrid) used in the AMs.

Recognition experiment has, then, been performed using the 5k development test set. Table 2 gives the WER of different AMs in FSM segmented morpheme-based recognition. As the table shows, the CV syllable-based AMs outperformed the triphone-based ones. A 2.4% absolute WER reduction (cf. Triphone_3states and CD_Syllable) has been obtained as a result of using syllable acoustic units in morpheme-based recognition. This improvement is statistically significant with p-value of less than 0.001. The use of user defined question for decision tree-based clustering has positive influence (resulted in a 0.8% absolute WER reduction) on the triphone-based AM. However, using user defined question did not bring WER reduction for the syllable-based AMs as it is also true in word-based recognition experiments described in Section 4.3. Nevertheless, the syllable-based AM with user defined questions resulted in a significant (at p value of 0.005) WER reduction compared to the equivalent triphone-based system. Generally, all hybrid AMs performed significantly (at p-value of 0.001) better than the triphone-based systems, the best performing (with a WER of 13.9%) being the Hybrid_170units_5statesWOS[3]. This system has also a slightly lower WER compared to the pure CV syllable-based systems. Although the model topology is crude in representing the acoustic units (not state of the art for phones), this model achieved the lowest WER of all the others. This indicates the potential of the hybrid AM for even higher performance provided that proper topologies are used for each of the units.

Unit	Models	WER (%)
Phone	Triphone_3states	16.7
	Triphone_3states_UDQ	15.9
CV syllable	CD_Syllable	14.3
	CD_Syllable_UDQ	14.6
Phone + CV syllable	Hybrid_170units_5statesWS	16
	Hybrid_170units_5statesWOS	13.9
	Hybrid_170units_4statesWOS	14.3
	Hybrid_204units_5statesWS	14.6
	Hybrid_204units_5statesWOS	14.2
	Hybrid_204units_4statesWOS	14.3

Table 2. WER of Several AMs with FSM Segmented Morpheme-based LMs

For the unsupervised morpheme-based recognition, all the distinct morphemes in the segmented text (45k) have been considered as entries for pronunciation dictionaries. As we did for the FSM segmented text, we prepared three versions of pronunciation dictionaries according to the type of units used in the AMs. Performance of the morpheme-based recognition using the different AMs on the 5k development test set is presented in Table 3. As it is true in the FSM segmented morpheme-based recognition experiment, the use of syllable AMs led to greater WER reductions in morpheme-based recognition than the others. 3% and 2% absolute WER reductions have been obtained compared to triphone-based ones that use automatically and user defined tree questions, respectively. These error rate reductions are statistically significant with the p-value of less than 0.001. The result clearly shows that syllable-based AMs are best fitted for morpheme-based recognition. The hybrid AMs (except Hybrid_170units_5statesWS) outperformed the triphone and CV syllable-based models although the improvement over the syllable-based model with user defined question (CD_Syllable_UDQ) is not statistically significant. In all the other cases the WER reduction is statistically significant at p-value of less than 0.001 compared to the triphone-based systems and with minimum p-value of 0.002 compared to the syllable-based one. This shows that the hybrid models are the best for Amharic morpheme-based recognition systems.

Unit	Models	WER (%)
Phone	Triphone_3states	17.8
	Triphone_3states_UDQ	15.9
CV Syllable	CD_Syllable	14.8
	CD_Syllable_UDQ	13.9
Phone + CV syllable	Hybrid_170units_5statesWS	15.5
	Hybrid_170units_5statesWOS	13.7
	Hybrid_170units_4statesWOS	13.5
	Hybrid_204units_5statesWS	13.7
	Hybrid_204units_5statesWOS	13.7
	Hybrid_204units_4statesWOS	13.3

Table 3. WER of Several AMs with Morfessor Seg. Morpheme-based LMs

Although the difference is not big, Morfessor-based segmentation led to a lower WER than FSM-based segmentation. The results presented in this section also showed that using morphemes (instead of words) as entries in pronunciation dictionary and units in LM brings improvement in Amharic speech recognition system.

6. CONCLUSION AND FUTURE DIRECTIONS

In this paper we have presented the results of our experiments on the use of hybrid units in acoustic modeling and the use of syllable as well as hybrid AMs in morpheme-based speech recognition. The hybrid AMs did not bring significant improvement when word units are used in the pronunciation dictionary and LM. However, they brought significant WER reductions in morpheme-based speech recognition.

[3]The model with HMM topology of 5 states without skips and in which CV syllables with a frequency less than 500 have been decomposed into phones

The syllable-based AMs also outperformed the triphone-based models in morpheme-based speech recognition. This enables us to conclude that the use of syllables and hybrid units in acoustic modeling and morphemes in lexical and language modeling is the best for Amharic speech recognition. We recommend investigation of the use of syllable AMs in morpheme-based speech recognition for other morphologically rich languages.

In the future, other Amharic syllables (in addition to CV syllables) will be used in syllable based acoustic modeling. Moreover, the epenthetic vowel and the glottal stop will be realized in their proper places in the pronunciation. In hybrid acoustic modeling, we have used an HMM topology of 5 states with skips assuming that this topology can handle the irregularities (in length) of the acoustic units. However, since the parameters estimated in this model topology are very large, big training data is required to adequately train such models. As we have used only 20 hours of training speech data, we could not see the benefit of using such a topology in hybrid AM. Instead a high WER has been observed. Thus, in order to get the real benefit of using models with skips, large training data has to be used. However, since acquiring data is not easily achievable, an alternative approach that is using different model topologies for syllable and phone units in hybrid acoustic modeling will be investigated.

7. REFERENCES

[1] A. Ganapathiraju, J. Hamaker, J. Picone, M. Ordowski, and G. R. Doddington, "Syllable-based large vocabulary continuous speech recognition," *IEEE Transactions on speech and audio processing*, vol. 9, no. 4, pp. 358–366, 2001.

[2] A. Sethy, S. Narayanan, and S. Parthasarthy, "A syllable based approach for improved recognition of spoken names," in *Proceeding of the 5th ISCA Pronunciation Modeling Workshop*, 2002, pp. 30–35.

[3] A. Hämäläinen, L. Boves, and J. de Veth, "Syllable-length acoustic units in large-vocabulary continuous speech recognition," in *SPECOM 2005*, 2005, pp. 499–502.

[4] Solomon Teferra Abate and Wolfgang Menzel, "Syllable-based speech recognition for amharic," in *Proceedings of the 2007 Workshop on Computational Approaches to Semitic Languages: Common Issues and Resources*, 2007, pp. 33–40.

[5] R. Thangarajan and A.M. Natarajan, "Syllable based continuous speech recognition for tamil," *South Asian Language Review*, vol. 17, no. 1, pp. 71–85, 2008.

[6] M.M. Azim, H. Tolba, S. Mahdy, and M. Fahsal, "Syllable-based automatic arabic speech recognition in noisy-telephone channel," *WSEAS Transactions on Signal Processing*, vol. 4, no. 4, pp. 211–220, 2008.

[7] Martha Yifiru Tachbelie, Laurent Besacier, and Solange Rossato, "Comparison of syllable and triphone based speech recognition for Amharic," in *Proceedings of the LTC 2011*, 2011, pp. 207–211.

[8] R. M. Voigt, "The classification of central semitic," *Journal of Semitic Studies*, , no. 32, pp. 1–21, 1987.

[9] Wolf Leslau, *Introductory Grammar of Amharic*, Harrassowits Verlag, Wiesbaden, 2000.

[10] B. Yimam, *yäamarIña säwasäw*, EMPDE, Addis Ababa, 2nd. ed. edition, 2007.

[11] Alemayehu Haile, "Is syllable weight distinction relevant for amharic stress assignment?," *Journal of Ethiopian Studies*, vol. 28, no. 2, pp. 15–25, 1995.

[12] Sebsibe H/Mariam, K.S. Prahallad, A.W. Black, R. Kumar, and Rajeev Sangal, "Unit selection voice for amharic using festvox," in *Proceeding of the 5th ISCA Speech Synthesis Workshop*, 2004, pp. 103–107.

[13] Solomon Teferra Abate, Wolfgang Menzel, and Bairu Tafila, "An amharic speech corpus for large vocabulary continuous speech recognition," in *Proceeding of INTERSPEECH 2005*, 2005, pp. 1601–1604.

[14] Martha Yifiru Tachbelie, *Morphology-Based Language Modeling for Amharic*, Ph.D. thesis, University of Hamburg, Germany, 2010.

[15] A. Stolcke, "SRILM — an extensible language modeling toolkit," in *Proceedings of International Conference on Spoken Language Processing*, 2002, vol. II, pp. 901–904.

[16] M. Creutz and K. Lagus, "Unsupervised morpheme segmentation and morphology induction from text corpora using morfessor 1.1," Tech. Rep. A81, Neural Networks Research Center, Helsinki University of Technology, 2005.

[17] Martha Yifiru Tachbelie, Solomon Teferra Abate, and Laurent Besacier, "Part-of-speech tagging for under-resourced and morphologically rich languages - the case of amharic," in *Proceedings of the HLTD 2011*, 2011, pp. 50–55.

[18] Mehryar Mohri, Fernando Pereira, and Michael Riley, "A rational design for a weighted finite-state transducer library," in *Automata Implementation*, Derick Woodand Sheng Yu, Ed., vol. 1436 of *Lecture Notes in Computer Science*, pp. 144–158. Springer Berlin / Heidelberg, 1998.