Grapheme-to-Phoneme Conversion for Amharic Text-to-Speech System

Tadesse Anberbir
Ajou University, Graduate School of Information and Communication, South Korea.
tadesse@ajou.ac.kr

Michael Gasser
Indiana University, School of Informatics and Computing, USA.
gasser@indiana.edu

Tomio Takara
University of the Ryukyus, Graduate School of Engineering and Science, Japan.

Kim Dong Yoon
Ajou University, Graduate School of Information and Communication, South Korea.

Abstract

Developing correct Grapheme-to-Phoneme (GTP) conversion method is a central problem in text-to-speech synthesis. Particularly, deriving phonological features which are not shown in orthography is challenging. In the Amharic language, geminates and epenthetic vowels are very crucial for proper pronunciation but neither is shown in orthography. This paper describes an architecture, a preprocessing morphological analyzer integrated into an Amharic Text to Speech (AmhTTS) System, to convert Amharic Unicode text into phonemic specification of pronunciation. The study mainly focused on disambiguating gemination and vowel epenthesis which are the significant problems in developing Amharic TTS system. The evaluation test on 666 words shows that the analyzer assigns geminates correctly (100%). Our approach is suitable for languages like Amharic with rich morphology and can be customized to other languages.

1 Introduction

Grapheme-to-Phoneme (GTP) conversion is a process which converts a target word from its written form (grapheme) to its pronunciation form (phoneme). Language technologies such as Text-to-speech (TTS) synthesis require a good GTP conversion method.

GTP conversion comes under two main approaches: rule-based and data-driven techniques and recently some statistical techniques have been proposed (See (Damper et al., 1998) for review of the several techniques). Using these methods successful results are obtained for different languages (Taylor, 2005; Chalamandaris et al., 2005) and other. However, in many languages automatic derivation of correct pronunciation from the grapheme form of a text is still challenging. Particularly phonological features which are not shown in orthography make the GTP conversion very complex.

Amharic, the official language of Ethiopia, has a complex morphology and some phonological features are not shown in orthography. Morphology, the way morphemes in a language join to form words, influences language technology because some phonological processes cannot be modeled without proper modeling of morphological processes. For example, most geminates in Amharic language are related to grammatical processes and can be predicted from morphological processes. In general, for Semitic languages such as Amharic and Tigrinya, morphological analysis can make explicit some of the phonological features of the languages that are not reflected in the orthography and plays a very crucial role in text-to-speech synthesis. However, so far no study has been conducted in this area.

In this study, we proposed and integrated a preprocessing morphological analyzer into an Amharic Text-to-speech (AmhTTS) system mainly to automatically predict geminates and epenthetic vowels positions in a word. Our research is the first attempt to integrate a morphological analyzer called HornMorpho (Gasser, 2011) into an AmhTTS. The integrated morphological analyzer takes Amharic Unicode input and outputs Latin transcription marking geminates and the location of epenthetic vowels. Then, the output of the morphological analyzer is used by the AmhTTS system and further processed to extract all the features generated by the analyzer. AmhTTS system is a parametric and rule-based system designed based on cepstral method (Anberbir and Takara, 2006).

Conference on Human Language Technology for Development, Alexandria, Egypt, 2-5 May 2011.

The paper is organized as follows: Section 2 provides background information about Amharic language and Section 3 briefly discusses about Amharic writing system and challenges in GTP conversion. Then, Section 4 and Section 5 describe about the automatic assignment method using morphological analyzer and evaluation results, respectively. Finally, Section 6 presents concluding remarks.

2 The Amharic language

Amharic, the official language of Ethiopia, is a Semitic language that has the greatest number of speakers after Arabic. According to the 1998 census, Amharic has 17.4 million speaker as a mother tongue language and 5.1 million speakers as a second language (Ethnologue, 2004).

A set of 34 phones, seven vowels and 27 consonants, makes up the complete inventory of sounds for the Amharic language (Baye, 2008). Consonants are generally classified as stops, fricatives, nasals, liquids, and semi-vowels. Table 1 shows the phonetic representation of the consonants of Amharic as to their manner of articulation, voicing, and place of articulation.

Table 1. Categories of Amharic Consonants with corresponding IPA representation.

Manner of Articulation	**voicing**	**Place of Articulation**					
		Labials	**Dentals**	**Palatals**	**Velars**	**Labio-velar**	**Glottals**
Stops	Voiceless	ፐ [p]	ት [t]	ች [ʧ]	ክ [k]	ኵ [kʷ]	አ [ʔ]
	Voiced	በ [b]	ድ [d]	ጅ [ʤ]	ግ [g]	ጐ [gʷ]	
	Glottalized	ጵ [p']	ጥ [t']	ጭ [tʃ']	ቅ [q]	ቍ [qʷ]	
Fricatives	Voiceless	ፍ [f]	ስ [s]	ሽ [ʃ]			ሀ [h]
	Voiced	ቭ [v]	ዝ [z]	ዥ [ʒ]			
	Glottalized		ጽ [s']				
	Rounded						ኍ [hʷ]
Nasals	Voiced	ም [m]	ን [n]	ኝ [ɲ]			
Liquids	Voiced		ል [l] ር [r]				
Glides	Voiced	ው [w]			ይ [j]		

The seven vowels, along with their representation in Ge'ez characters, are shown in terms of their place of articulation in Table 2. In addition to the five vowels common among many languages, Amharic has two central vowels, /ə/ and /ɨ/, the latter with a mainly epenthetic function. The epenthetic vowel /ɨ/ plays a key role in syllabification. Moreover, in our study we found the epenthetic vowel to be crucial for proper pronunciation in Amharic speech synthesis.

Table 2. Categories of Amharic Vowels with IPA equivalent.

	front	central	back
High	ኢ [i]	ኧ [ɨ]	ኡ [u]
Mid	ኤ [e]	ኧ [ə]	ኦ [o]
low		ኣ [a]	

Like other languages, Amharic also has its own typical phonological and morphological features that characterize it. The following are some of the striking features of Amharic phonology that gives the language its characteristic sound when one hears it spoken: the weak indeterminate stress; the presence of glottalic, palatal, and labialized consonants; the frequent gemination of consonants and central vowels; and the use of the automatic epenthetic vowel (Bender et al., 1976). Among these, we found gemination of consonants and the use of the automatic epenthetic vowel to be very critical for naturalness in Amharic speech synthesis.

3 Amharic writing system and Problems in GTP Conversion

Amharic uses the Ge'ez (or Ethiopic) writing system which originated with the ancient Ge'ez language. In this system the symbols are consonant-based but also contain an obligatory vowel marking. Most symbols represent consonant-vowel combinations, but there may also be a special symbol for each consonant that represents the plain consonant. Each Amharic consonant is associated with seven characters (referred to as "orders") for the seven vowels of the language. It is the sixth-order character that is the special symbol, representing the plain consonant. The basic pattern for each consonant is shown in Fig. 1, where: C=Consonant and [ ] shows vowels in IPA.

1st order	2nd order	3rd order	4th order	5th order	6th order	7th order
C[ə]	C[u]	C[i]	C[a]	C[e]	C	C[o]
ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ

Figure 1. Amharic syllabic structure with example for consonant ል /l/.

Amharic writing system is partially phonetic. According to (wolf 1995), there is more or less a one-to-one correspondence between the sounds and the graphemes. However, as shown in Table 3, it has some features that make it complex from the perspective of GTP conversion.

In what follows we discuss the two main ambiguities in Amharic orthography in more detail.

Table 3. Problems in Amharic grapheme-to-phoneme (G2P) conversion.

Problem	Example
Homograph	Depending on the context, the word ገና can have the meaning 'still/yet' or 'Christmas'
Insertion of epenthetic vowel [ɨ]	in words like ትምህርት, epenthetic vowel should be inserted and pronounced as /t ɨ mhɨrt/ not /tmhrt/
Introduction of semi-vowel w, y	words like, በቅሎአችን bəqlo-aʃn 'our mule' becomes በቅሎዎችን bəqlowaʃn.
Compression of successive vowels	ለ + አኔ /lə + ɨne/ becomes ለኔ /ləne/ የአማርኛ yə-amari ɲa becomes yamari ɲa

3.1 Gemination

Gemination in Amharic is one of the most distinctive characteristics of the cadence of the speech, and also carries a very heavy semantic and syntactic functional weight. Unlike English language in which the rhythm of the speech is mainly characterized by stress (loudness), rhythm in Amharic is mainly marked by longer and shorter syllables depending on gemination of consonants, and by certain features of phrasing (Bender and Fulass, 1978). In Amharic, all consonants except /h/ and /ʔ/ may occur in either a geminated or a non-geminated form. Amharic, and other languages written with the Ge'ez script, differs from most other languages that feature gemination, such as Japanese, Arabic, Italian, and Tamil, in that gemination is not shown in the orthography.

Table 4. Minimal pairs with Singleton vs. Geminate Consonants

Singleton			geminate	
Orth.	**Pronunc.**	**Gloss**	**Pronunc.**	**Gloss**
ገና	gəna	still/yet	gənna	Christmas
ለጋ	ləga	fresh	ləgga	he hit
ስፊ	səfi	tailor	səffi	wide
ሸፍታ	ʃifta	outlaw	ʃiffɨta	rash
ይሰማል	yɨsəmal	he hears	yɨssəmmal	he/it is heard

Amharic gemination is either lexical or morphological. As a lexical feature it usually cannot be predicted. For instance, ገና may be read as /gəna/, meaning 'still/yet', or /gənna/, meaning 'Christmas'. (See Table 4 for some similar example of minimal pairs). Although this is not a problem for Amharic speakers because minimal pairs are relatively infrequent, it is a challenging problem in Amharic speech synthesis. In fact, the failure of the orthography of Amharic to show geminates is the main challenge in GTP conversion that we found in our research. Without a context, it is impossible to disambiguate such forms.

On the other hand, when gemination is morphological, rather than lexical, it is often possible to predict it from the orthography of the word alone. This is especially true for verbs (Bender and Fulass, 1978). For example, consider two words derived from the verb root consisting of the consonant sequence sbr 'break', ሰበረው and ይሰበራሉ. The first is unambiguously /sɨbərəw/ 'break (masc.sing.) it!', the second unambiguously /yɨssəbbərallu/ 'they are broken'. The fact that the /s/ and /b/ are not geminated in the first word and are both geminated in the second and that the /r/ is geminated in neither word is inferable from the prefix, the suffix, and the pattern of stem vowels. That is, within the verb there is normally some redundancy. Therefore, with knowledge of the lexical and morphological properties of the language, it is possible to predict gemination.

3.2 Epenthesis

Epenthesis is the process of inserting a vowel to break up consonant clusters. Epenthesis, unlike gemination is not contrastive and it is not surprising that it is not indicated in the orthography of Amharic and other languages. But, although it carries no meaning, the Amharic epenthetic vowel /ɨ/ (in Amharic 'ሰርጎ ገባ' Baye, 2008) plays a key role for proper pronunciation of speech and in syllabification. However, the nature and function of the epenthetic vowel has been a problem in Amharic studies and so far no study has been conducted on the phonetic nature of this vowel.

As noted above, Amharic script does not distinguish between consonants that are not followed by a vowel and consonants that are followed by the high central vowel /ɨ/, and as shown in Fig.1, both are represented by the sixth order (ሳድስ) character in a series. The sixth order characters are ambiguous; depending on their position in a word; they can be either voweled (with epenthic vowel /ɨ/) or unvoweled. For example, in ልብ /lɨbb/ 'heart', the first character, ል, represents the CV sequence /lɨ/(voweled), whereas in ስልክ /sɨlk/ 'telephone', the same character represents the bare consonant /l/(unvoweled). Because such differences are crucial for speech synthesis, a TTS system needs access to the epenthesis rules.

4 Proposed Amharic GTP Conversion Method

In this section, first we discuss about Amharic morphology and the 'HornMorpho' morphological analyzer which we integrated into the AmhTTS system. Then, we briefly discuss our proposed GTP conversion method.

4.1 Amharic Verb Morphology

Amharic is a morphologically complex language. As in other Semitic languages such as Arabic, Amharic verbs consist of a stem – analyzable as a combination of a lexical root and a pattern representing tense, aspect, mood, and various derivational categories – and various affixes representing inflectional categories (Bender and Fulass, 1978). Verb roots consist of consonant sequences, most of the three consonants. The patterns that combine with roots to form verb stems consist of vowels that are inserted between the consonants and gemination of particular root consonants.

Consider the role of gemination in distinguishing the three main categories of three-consonant (triradical) roots, traditionally referred to as types A, B, and C (Bender and Fulass, 1978). Table 5 shows the pattern of gemination and the positions of vowels for triradical roots in the five basic tense-aspect-mood categories of the language. In type A, except in the perfective form, there is no gemination at all. Type B is characterized by the gemination of the second consonant throughout the conjugation. Type C differs from the other two types in that the second consonant is geminated in the perfective and imperfective aspects only (Amsalu and Demeke, 2006). The system is complicated by the fact that each root can also occur in combination with up to 10 derivational categories, such as passive and causative.

Table 5. Vowel-gemination patterns for triradical roots (adopted from Amsalu and Demeke, 2006)

Verb Types	Type A	Type B	Type C
Perfective	C1V**C2C2**C3-	C1V**C2C2**VC3	C1V**C2C2**VC3-
Imperfective	-C1VC2C3(-)	-C1V**C2C2**C3(-)	-C1V**C2C2**C3(-)
Imperative	C1C2VC3(-)	C1V**C2C2**C3(-)	C1VC2C3(-)
Gerund	C1VC2C3-	C1V**C2C2**C3-	C1VC2C3-
Infinitive	-C1C2VC3	-C1V**C2C2**C3	-C1VC2C3

The patterns shown in Table 5 are for the simple derivational category, which can be seen as the default for most roots, but each cell in the table could be augmented with up to 9 other patterns. These different derivational categories also affect gemination. For example, for the passive category, the imperfective pattern for Type A becomes -C1C1VC2C2VC3(-), with both first and second consonants geminated.

Based on the gemination patterns, a morphological analyzer can be used to locate geminates in an input word form. For example, the Amharic word ይሰበራል 'it is broken' is correctly pronounced with gemination (lengthening) of the first and second consonants: yissəbbəral. The gemination in this case is grammatical, and a morphological analyzer can infer it based on its knowledge of Amharic verb roots and the particular patterns that they occur with.

4.2 'HornMorpho' Morphological Analyzer

Amharic morphology, especially verb morphology, is extremely complex (Baye, 2008), but it is relatively well understood. Thus it is possible, with considerable effort, to create a morphological analyzer for the language using finite state techniques. HornMorpho (Gasser, 2011) is such a system. Given a word in conventional Amharic orthography, it infers the gemination of consonants in the word wherever this is possible (as well as extracting grammatical and lexical information from the word). The rules for epenthesis in Amharic are also quite complex and not completely understood, but a first approximation has been implemented in HornMorpho. That is, the output of HornMorpho includes the epenthetic vowel /ɨ/ wherever it is expected, according to the rules in the program, effectively disambiguating the sixth-order orthographic characters and simplifying the overall GTP conversion.

When analyzing, the word is first Romanized using a variant of the SERA romanization system. Next the program checks to determine whether the word is stored in a list of unanalyzable or pre-analyzed words. If not, it attempts to perform a morphological analysis of the word. It first does this using a "lexical" approach, based on a built-in lexicon of roots or stems and its knowledge of Amharic morphology. If this fails, it tries to guess what the structure of the word is, using only its knowledge of morphology and the general structure of roots or stems. If the "guesser" analyzer fails, the program gives up. Both the lexical and guesser analyzers operate within the general framework known as finite state morphology[1].

[1] For a more in-depth introduction to finite state morphology, see Beesley & Karttunen(2003), for another application of

4.3 Finite State Morphology

Finite state morphology makes use of transducers that relate surface forms with lexical forms. In the case of HornMorpho, a surface form is the orthographic representation of an Amharic word, for example, ይሰበራል *yɨsəbəral*, and the corresponding lexical form is the root of the word and a set of grammatical features. For ይሰበራል, the lexical representation is *sbr* + [sbj=3sm, asp=imprf, vc=ps], that is, the root *sbr* 'break', third person singular masculine subject, imperfective aspect, and passive voice, in English, 'he/it is broken'.

Each of the arcs joining the states in a finite state transducer (FST) represents an association of an input character and an output character. Successful traversal of an FST results in the transduction of a string of input characters into a string of output characters. In HornMorpho, using a technique developed by Amtrup (2003), we add grammatical features to the arcs as well, and the result of traversing the FST is the unification of the features on the arcs as well as the output string. This allows us to input and output grammatical features as well as character strings.

A simple FST can represent a phonological or orthographic rule, a morphotactic rule (representing possible sequences of morphemes), or the form of a particular root. Because germination is both lexical and grammatical in Amharic, it plays a role in all three types. By combining a set of such FSTs using concatenation and composition, we can produce a single FST that embodies all of the rules necessary for handling words in the language. Since FSTs are reversible, this FST can be used in either direction, analysis (surface to lexical forms) or generation (lexical to surface forms). HornMorpho uses a single large FST to handle the analysis and generation of all verbs. This FST includes a lexicon of verb roots as well as all of the hundreds of phonological, orthographic, and morphological rules that characterize Amharic verbs.

4.4 Proposed GTP conversion Method

For Amharic text-to-speech, we use the existing HornMorpho analysis FST and a modified version of the HornMorpho generation FST. We first run the analysis FST on the input orthographic form. For example, for the word ይሰበራል, this yields sbr + [sbj=3sm, asp=imprf, vc=ps]. Next we run a phonological generation FST on this output, yielding the phonological output yɨssəbbəral for this example. The second and third consonants are geminated by an FST that is responsible for the passive imperfective stem form of three-consonant roots.

Figure 2 shows the architecture our proposed method. First the morphological analyzer accepts Unicode text from file and generates the corresponding phonemic transcription with appropriate prosodic markers such as geminates and indicates the location of epenthetic vowel. Then, the output of the analyzer will be an input for AmhTTS system and further annotated by the text analysis module to extract syllables and prosodic marks. Finally the speech synthesis part generates speech sounds.

AmhTTS synthesis system is a parametric and rule based system designed based on the general speech synthesis system (Takara and Kochi, 2000). The text analysis in AmhTTS system extracts the linguistic and prosodic information from the output of a morphological analyzer and extracts the gemination and other marks and then converts into a sequence of syllables using the syllabification rule. The syllabification rule uses the following syllables structure (V, VC, V'C, VCC, CV, CVC and CV'C and CVCC).

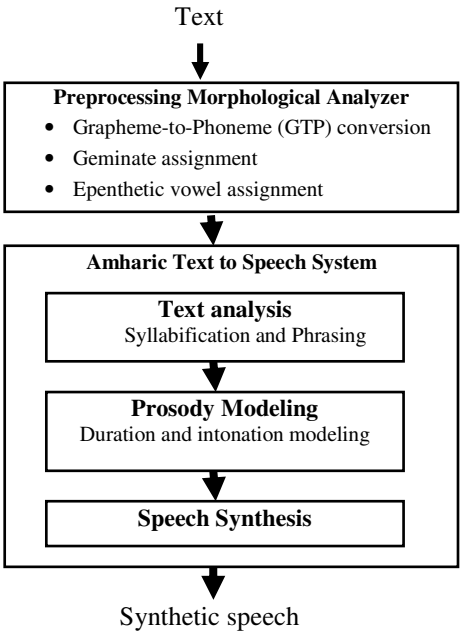

Figure 2. Proposed architecture for GTP conversion and automatic geminate and epenthetic vowel assignment

finite state technology to Amharic morphology, see Saba & Girma (2006).

5 Evaluation

A preliminary evaluation of the proposed automatic geminate assignment method was made by analyzing 666 words and we found 100% correct assignment/restoration of gemination. The words were selected from Armbruster (1908) verb collections, where gemination is marked manually, representing all of the verb conjugation classes (Type A, B and C). In Type A penultimate consonant geminates in Perfect only, in Type B penultimate consonant geminates throughout the conjugation and in Type C penultimate consonant geminates in Perfect and Contingent.

However, the analyzer does not analyze words that contain non-Ge'ez characters or Ge'ez numerals, and also there is an incomplete list of common words (with gemination) which it does not attempt to analyze. For example, given the unambiguous form ይሰበራሉ, it outputs /yɨssebberallu/, and given the ambiguous form ገና, it outputs both possible pronunciations /gəna/ and gənna/. Words like ገና can only be inferred by analyzing the context and finding out the parts-of-speech (POS) of the word. But this is beyond the scope of the current work.

6 Conclusion

The paper discussed orthographic problems in Amharic writing system and presented preliminary results on a method for automatic assignment of geminates and epenthetic vowel in GTP conversion for Amharic TTS system. Our method is the first attempt for Amharic language and can be easily customized for other Ethio-Semitic languages.

However, the work described in this paper is still on progress and for more accurate GTP conversion, parts-of-speech (POS) tagger and phrase break predictor needs to be implemented or addressed. For example, the word ገና, which can be pronounced as both / gəna/ meaning 'still/yet' and gənna meaning 'Christmas', can only be inferred by analyzing the context and finding out the POS of the word.

In the future, we have a plan to disambiguate words like ገና using syntax and use the analyzer to extract more features and finally customize our TTS system for other languages.

Acknowledgments
The authors would like to thank Daniel Yacob for providing the Armbruster verb collections where the geminatation is marked manually with the "Tebek" symbol (U+135F).

References

A. Chalamandaris, S. Raptis,, and P. Tsiakoulis. 2005. *Rule-based grapheme-to-phoneme method for the Greek.* In: Proc. of INTERSPEECH-2005, pp. 2937-2940, Lisbon, Portugal.

Amtrup, J. 2003. *Morphology in machine translation systems: efficient integration of finite state transducers and feature structure descriptions*. Machine Translation, 18, 213-235.

Armbruster, C. H. 1908. *Initia Amharica: an introduction to spoken Amharic*, Cambridge: Cambridge University Press.

Baye Yimam. 2008. *የአማርኛ ሰዋሰው (Amharic Grammar)*, Addis Ababa. (in Amharic).

Beesley, K.R., and Karttunen, L. 2003. *Finite state morphology*. Stanford, California: CSLI Publications.

Ethnologue. 2004: *Languages of the World*, http://www.ethnologue.com/

Gasser, M. (2011). HornMorpho: a system for morphological analysis and generation of Amharic, Oromo, and Tigrinya words. *Conference on Human Language Technology for Development*, Alexandria, Egypt.

Leslau, Wolf. 1995. *Reference Grammar of Amharic*, Wiesbaden: Harrassowitz.

M.L Bender, J.D.Bowen, R.L. Cooper and C.A. Ferguson. 1976. *Language in Ethiopia*, London, Oxford University Press.

M. Lionel Bender and Hailu Fulass. 1978. *Amharic Verb Morphology: A Generative Approach*, Carbondale.

Paul Taylor, 2005. *Hidden Markov Model for Grapheme to Phoneme Conversion*. In: Proc. of INTERSPEECH-2005, pp. 1973-1976.

R.I. Damper, Y. Marchand, M. J. Adamson, and K. Gustafson. 1998. *A comparison of letter-to-sound conversion techniques for English text-to-speech synthesis*", Proceedings of the Institute of Acoustics, 20 (6). pp. 245-254.

Saba Amsalu and Girma A. Demeke. 2006. *Non-concatinative Finite-State Morphotactics of Amharic Simple Verbs*, ELRC Working Papers Vol. 2; number 3.

T. Anberbir and T. Takara. 2006. *Amharic Speech Synthesis Using Cepstral Method with Stress Generation Rule*, INTERSPEECH 2006 ICSLP, Pittsburgh, Pennsylvania, pp. 1340-1343.

T. Takara and T. Kochi. 2000. *General speech synthesis system for Japanese Ryukyu dialect*, Proc. of the 7th WestPRAC, pp. 173-176.