Syllabification Design and TTS System for Afaan Oromo Using Unit Selection

Argaw Korssa
Pact Ethiopia
argaw100@gmail.com

Sebsibe Hailemariam
Department of Computer Science, Addis Ababa University, Ethiopia
sebsibe2004@yahoo.com

Abstract

Though Afaan Oromo is widely spoken in Ethiopia, there is no previous work done that generates quality TTS synthesis using unit selection speech synthesis as well as no syllabification algorithm is developed for it. To address the research problem, review of documents and consulting experts on the linguistic structure of the language were done which are vital to construct new phone set, pronunciation lexica, G2P conversion and syllabification algorithm design. Also recording of speech using praat and C++ code for G2P converter, manual labeling using wavesurf and implementation of syllabification algorithm was used. Furthermore, modules in Festival framework and Festvox where one can build speech synthesis with their own voice are used to develop a TTS prototype.

Finally, the syllabification algorithm is implemented and tested using selected words. The result showed 100% word accuracy rate which depicts the rule-based approach works very well but so far no comparison is made with the other approaches. In TTS synthesizer, new phone set, pronunciation lexica and rules for G2P conversion are constructed. Subjective tests have been carried out for evaluating the synthesized speech quality and the ultimate speech synthesis system got 4.24 MOS.

Keywords: Text-to-speech synthesis; Unit selection speech synthesis; Syllabification; Rule-based techniques; Grapheme-to-phoneme, CV

1. Introduction

TTS synthesis is an automatic conversion of unrestricted natural language sentences into speech [1, 3, 4]. TTS system must be developed for every language to extend the applicability and use of computer technology. It is also arguable that speech synthesis developed for one language is not applicable for others due to differences in language structure, intonation, duration, accent, tone, stress and other contexts. And many of the fully functional TTS products are based on the languages of developed countries. It is with this reality that Morka [6] and Sebsibie [8] tried to develop a prototype TTS for Afaan Oromo using diphone and Amharic language, respectively. However, in the previous work the quality of synthetic speech in Afaan Oromo language was limited and appears now to be insufficient in modern applications.

This study tried to address how to select speech unit and Afaan Oromo phone set for suitable TTS synthesis, how can unit selection be used to represent naturalness and intelligent synthesis and how to design and develop a rule-based syllabification algorithm.

To do so methodologies like data collection that helped to prepare appropriate training and testing dataset (corpus) for the desired system, tools such as speaker for playback record, C++ programming language to develop G2P converter and syllabification rule and praat for manual labeling were used. In addition, a new TTS system with a unit selection compatible voice was built from a new designed and developed large speech corpus within Festival/Festvox framework. Festvox offers general tools for building unit selection voices in new languages.

2. Analysis of Phonology of Afaan Oromo Language

As pointed out by Beekamaa [2] and Thaha [10], the script of Afaan Oromo language is commonly known as phonetic language. In Afaan Oromo there are five short vowels, five long vowels, seven digraph including foreign ones and 28 consonants. With the exception of vowels which stand by their own, all consonants have the combination of both short and long vowels to form a meaning. Afaan Oromo has another glottalized phone that is more unusual, an implosive retroflex stop, "dh" in Oromo orthography, a sound that is like an English "d" produced with the tongue curled back slightly and with the air drawn in so that a glottal stop is heard before the following vowel begins [2, 5].

Afaan Oromo words use both consonantal and vowel roots with vowel variation expressing difference in interpretation. Each syllable pattern comes in different orders, reflecting the 10 vowel sounds with the exclusion of the vowels themselves. There are 28 basic forms, giving 10*28 syllable patterns (syllographs) [2, 5, 10].

There are about 38 phonemes of Oromiffa alphabets (Table 1). These phonemes have been used to contain the acoustics units of the unit selection and speech file which are then used to get the synthesized speech corresponding to the given word. For this study, we have used the eight main templates of syllabic synthesis units (V, VV, VC, CV, CVV, VVC, CVC, and CVVC clusters).

Table 1: Phonetic representation of Afaan Oromo consonants and its IPA Equivalence

Manner		*Bilabial*	*Labiodental*	*Alveolar*	*Palatal*	*Velar*	*Glottal*
Stops	Voiced Voiceless Ejective Implosive	B (p) ph /p□/		D T x /t□/ dh /□/		G K q /k□/	 ' /□/
Fricatives	Voiced Voiceless	(v)	 F	(z) S	(zy) sh /s'/		 H
Affricative	Voiced Voiceless Ejective				J Ch (c') C		
Nasals		M		N	ny /□/		
Lateral				L			
Rhotic				R			
Semi-vowel		W			Y		

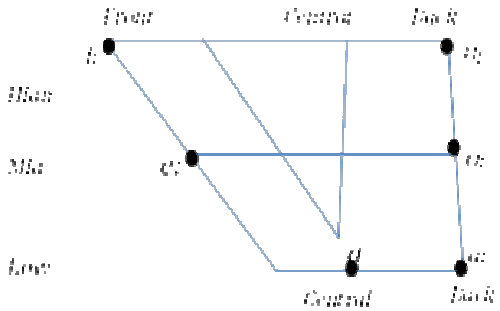

Figure 1: Diagram of Afaan Oromo vowels (Habte and Adugna: 6, 10)

3. System Architecture

3.1 Syllabification Model Design and Architecture

Having germination handling, syllabification and deletion rules and syllable templates of the language, we developed a rule-based syllabification model. Syllabification of any words with the exception of abbreviations and acronyms were considered in our system development. In order to rigger the module, Afaan Oromo text is used. Subsequently

normalization is followed. In the normalization module grammatically wrong characters that appear in the word or text will be deleted using devised rule. Mapping all phonemes and digraphs is carried out, and if digraphs exist re-mapping is carried out. The normalized text then goes to the germination module. Following the output from germination, syllabification is carried out by the syllabifier module. This is done using an algorithm to syllabify any input words in their legal sequence. The final output will be syllable boundary marked Afaan Oromo text as indicated in Figure 2.

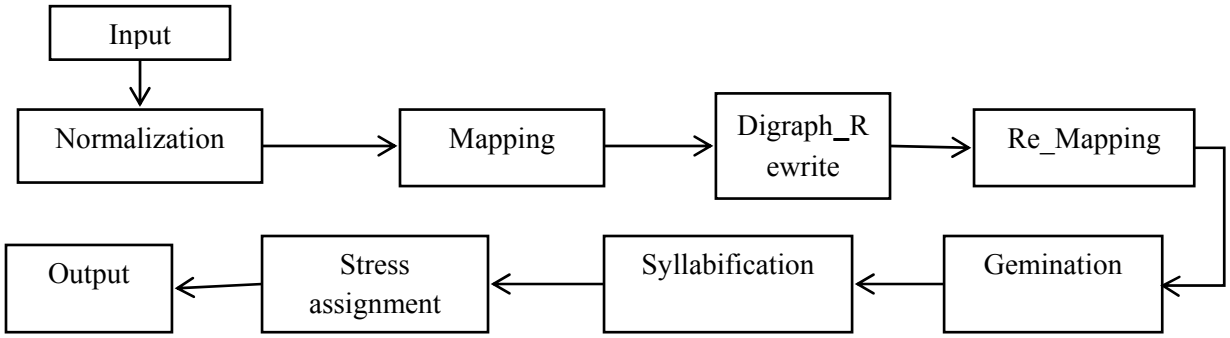

Figure 2: General Automatic Syllabification Model for Afaan Oromo Word

A word may be syllabified in different syllable structure but the algorithm selects the legal syllable structure sequence for the input word. Here, the well-known linguistic syllabification implementation principles namely the maximum-onset principle and sonority hierarchy principle are implemented. The expected output is an identical phonetic representation as input but including modified if incorrect input provided and syllable boundary markers.

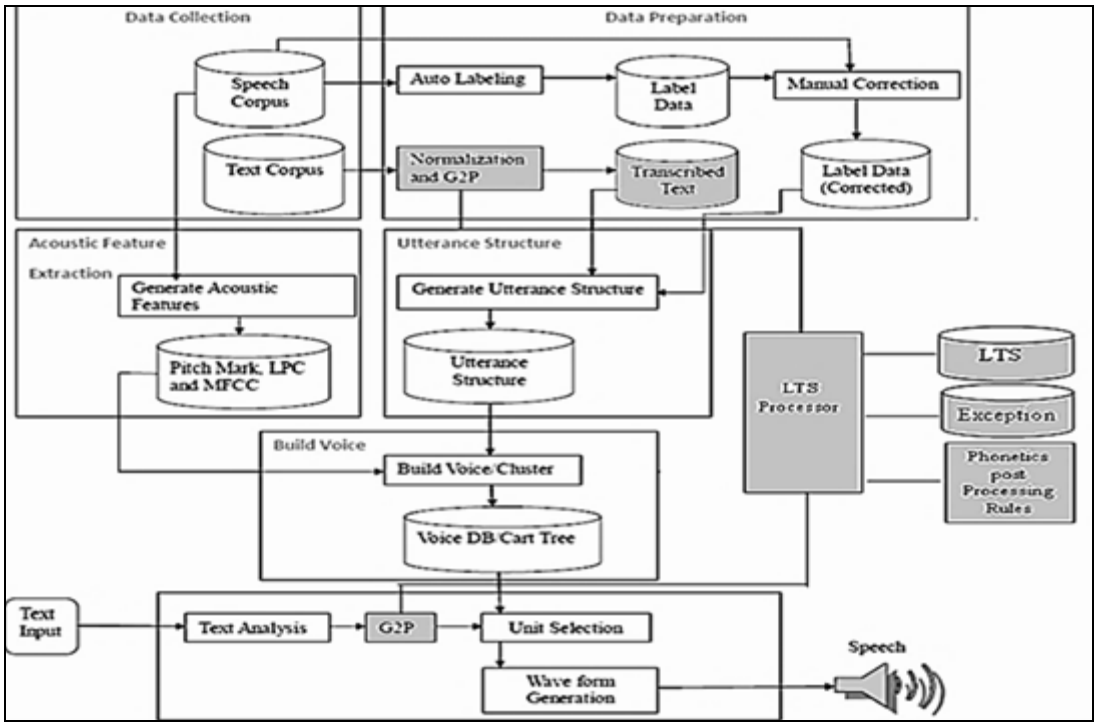

Figure 3: Afaan Oromo TTS System Implementation Model

3.2 Afaan Oromo TTS design and Architecture

Phonemes are the smallest units of speech sound in a language that can serve to distinguish one word from another [9]. Afaan Oromo alphabet has 38 letters classified as 10 vowels and 28 consonants including five adopted foreign phoneme sounds [7, 10]. In our system for phonetic transcriptions we

adopted a phoneme set based on the IPA SAMPA standard.

G2P/LTS conversion, developed using C++ programming language, is best suited for Afaan Oromo words due to close relationship between orthography and phonology of the language. However, some exceptional loanwords might exist. Afaan Oromo pronunciation lexicon is mainly used for determining the pronunciations of these words [10].

4. Implementation

4.1 The Syllable Structure and Syllabification of the Language

All the syllable type of Afaan Oromo can also be defined at the phonological representation. We can classify the eight templates into different classes based on the grammatical structure and/or weight distinction of the syllable templates, for instance, open or closed, long or short, geminated or not, etc. In a simple weight distinction, there are heavy and light syllables in Afaan Oromo. A heavy syllable is a syllable with branching nucleus or a branching rime. A branching nucleus generally means the syllable has a long vowel; this type of syllable is abbreviated as CVV. A syllable with a branching rime is a closed syllable, that is, one with coda; this type of syllable is abbreviated as CVC or CVVC. Light syllable with a short vowel as the nucleus and no coda (CV). A closed syllable is one that ends in a consonant and an open syllable is one that ends in a vowel. Short vowel open syllables are light, all others are heavy. However, heavy syllables attract stress, while light syllables only get stress if they happen to be in the right location in the string of syllables. The following example and Table 2 explain about heavy, light, open and closed syllables.

a. Humnaan hum-naan (By force)

b. Firfirse fir-fir-se (be distributed)

c. Jijijjiiramaa ji-jij-jii-ra-maa (Changeover)

In the words (a, b and c) above, *se, ji, ra, maa* are open syllable whereas *hum, naan, fir and jij* are closed syllable. Syllables such as *naan, jii and maa* are heavy syllable while *fir, se, ji and ra* are light syllables.

Table 2: Summary of Different Kinds of Afaan Oromo Syllable Structure

Kind	*Description*	*Example*
Heavy	Has a branching rhyme. All syllables with a branching nucleus (long vowels) are considered heavy.	CVVC, CVV, VVC CVC
Light	Has a non-branching rhyme (short vowel).	CV
Closed	Ends with a consonant coda.	CVC, VC,
Open	Has no final consonant	CV, V, CVV

4.2 Consonants cluster and their syllable structure

In Afaan Oromo, the maximum number of allowable same or different consonant sequences in a cluster is two. In case if below conditions (a-g) violated, there must be deletion of the phoneme/s. Accordingly, Afaan Oromo word structure allows the following consonant sequence.

a) No two consonant consecutively appear at the beginning or end of the word except ch, dh, ny, ph, ts, sh and zy.
b) No three consonant clusters consecutively appear anywhere in the text.
c) Single quotation (‘) is used to separate between two different vowels and vowels that have different sound and/or length either same or different, e.g., re’ee (goat), bu’aa (result), etc.
d) No three cluster vowels consecutively appear anywhere in a word without separator.
e) No two different vowels come together, if appeared separator is used.
f) No two same short or long sound vowels consecutively appear in a word, if appeared separator is used.
g) Clustering and germination of consonants is possible only word medially

4.3 Epenthesis in Afaan Oromo words

Though the process of epenthesis is not as such significant, we developed a rule for those foreign words. In case there are such words, epenthesis can occur at word-initially and final frequently and word medial sometimes. For instance the English word 'sport' is rewritten as 'ispoortii'. In this letter 'i' is inserted at beginning and end of the word. In the other case, the word 'president' is rewritten as 'piresedaantii' and others. This insertion is to obey the grammatical writing principle of the language unless it has no meaning and difficult to understand.

4.4 Rules and Algorithms

The syllables in Afaan Oromo are formed with the combination of consonants and vowels. Syllables are generally formed from 2-4 characters and contain a vowel and consonants. The most frequently used syllable for words which belong to the Afaan Oromo language are shown in Table 3. To dos so, a minimal set of rules were developed.

Table 3: The structure of Afaan Oromo Syllables-Most Frequently Used Syllables

Syllable structure	*Sample Syllables*
V	Eda (last night) -> e-da. Others start with single a, o, u, i
VV	Aanaa (woreda) -> aa-naa and others
CV	Ba, ci, fi, to… (Bakakkaa-thunder, cimaa-strong, fili-select, etc.)
VC	Ol, ej, ul … (e.g. Ol-up, ejjuu-prostitute, ullaa-, allaattii-birds, etc.)
CVV	Baa, dee, fii… (e.g. baala-leaf, deega-poorness, fiiguu-to run, etc.)
CVC	Bor, gad, bar… (e.g. barruu-inside thumb, barreessuu-write, etc.)
CVVC	Foon, baal, kaan…(e.g. baachuu-carry, foon-meat,)
VVC	Oof, aal… (e.g. aalbee-knife, eengadda-before yesterday, etc.)

In relation with deletion of insupportable phoneme/s, study was not undergone. Thus, from our empirical observation deletion of vowel and consonant has great role in syllabification of Afaan Oromo words. It can occur at word initial, word medial and word final position. Before we design the model of syllabification algorithm we first model the deletion of the inappropriate vowel and consonant in separate module. Both the syllabification and deletion of vowel and consonant algorithm reads input from left-to-right. The language normalization, mapping, germination, deletion and syllabification algorithm identified are presented in the form of a formal algorithm implemented in C++ Programming language.

Deletion of vowel and consonant procedure:

1. Accept input word and scan from left to right.
2. If consonant cluster occurs at word initial or final position, delete the preceding one.
   a) *Exception*: If the first phoneme or consonant is digraph. (Rule #1)
3. If 3 consonants are appeared in sequence word medially, delete the preceding one. (Rule #2)
4. If 3 vowels are appeared in sequence word initial, medially or final, delete the preceding one. (Rule #2)
5. If a cluster of consonants contains the geminate and singleton in sequence, delete after the geminated consonants. (Rule #3)
6. If a cluster of consonants contains the singleton and geminate in sequence, delete one after the geminated consonants. (Rule #3)
7. If a cluster of consonants contains two different geminates in sequence, delete either of the two geminate consonants. (Rule #4)
8. Repeat 2 up to 7 until all the phonemes are parsed in the phonemes list.

The following is the proposed syllabification procedure in Afaan Oromo:

1. Accept the input from normalization algorithm and scan from left to right.
2. At word initial position if one vowel phoneme (V) and (CV) occurs in sequence, mark syllable boundary between first V and C, which results V-CV.

3. At word initial position if two vowels phonemes (VV) occurs in sequence, and the next is (CV), mark syllable boundary between VV-CV but if (VVCCV) mark as VVC-CV.
4. If the initial phoneme is vowel and the next two phonemes are consonant and vowels respectively; mark the syllable boundary just at the second (VC-CV)
5. If (VCCV) pattern occurs at any position, mark syllable boundary between the two consonant clusters.
6. If (VCVC) pattern occurs at word initial position, mark syllable boundary before the second vowel.
7. If (CVV) type sequence occurs at any position, mark syllable boundary at the end of the second vowel.
8. If (CVCCV) phoneme sequence occurs at word initial position mark syllable boundary between the middle consonant clusters (CVC-CV).
9. If (CVVC) pattern occurs at word final position and if there is phoneme before the first consonant mark syllable boundary before the initial consonant in this pattern.
10. If (CVCV) pattern occurs at any position, syllable boundary becomes CV - CV pattern.
11. If (CVCCVC) pattern occurs in a word mark syllable boundary between the geminated/clustered consonants. (CVC- CVC).
12. Repeat 2 up to11 until all phonemes are parsed.

4.5 Voice Building

4.5.1 Creation of Speech Database

To build Afaan Oromo speech database, the text corpus represents different sections of Afaan Oromo texts like political essays, literary works, sports sections, science, newspaper, and business pages in which some exist in hardcopy and others in softcopy. A single male speaker reads carefully the designed script. The selection of the speaker was based on subjective aspects. Signal was recorded using 16 bits and 16 kHz of sampling rate. To this effect, our database can include single or double letter sounds as the smallest phoneme in the current system. The minimum number of phonemes kept in the database is calculated as 598 and the rest of the syllables are formed from the concatenation of these sounds.

4.5.2 Feature Extraction and Clustering

The labeled speech database was processed by applying simple power normalization on each utterance. The maximum and minimum pitch value of the speaker was determined using the Wavesurfer Pitch Marker Tool. The Festvox pitch extraction parameters were adjusted accordingly to obtain pitch features for the utterances. MFCCs were also extracted. The Unit selection was built by applying unit clustering algorithm on the units of the database.

5. Experimental Results and Evaluation

5.1 Performance of Syllabification Algorithm

The test corpus selected and used for syllabification performance measurement and evaluation is 1000 words which satisfies all syllabification rule and regulation of the language. Each syllabified word contains one to seven syllables. The expert compares the input phoneme sequence and the output of the algorithm, and gives their remark on each output. Finally, the expert agreed 100% word accuracy rate.

Table 5: Distribution of syllable patterns over the test set

Syllable Pattern	*Frequencies*			*Total*	*Percentage (%)*
	Word Initial	*Word medial*	*Word final*		
V	70	–	–	70	2.5%
VV	13	–	–	13	0.5%
CV	288	404	392	1084	39.0%
VC	63	–	–	63	2.3%
CVV	167	224	290	681	24.5%
CVC	234	155	48	437	15.7%
CVVC	84	142	193	419	15.1%
VVC	10	–	–	10	0.4%
Total	*929*	*925*	*923*	*2777*	*100%*

5.2 Intelligence and Naturalness of TTS

5.2.1 Evaluation and Methods

To evaluate the quality of the synthetic voice produced by the developed system, we carried out formal listening tests. In such a way, we decided the listeners to rank the voice quality using a Mean Opinion Score (MOS) like scoring as a suitable method for overall evaluation of synthetic speech. The MOSs for speech synthesis are generally given in three categories: intelligibility, naturalness, and pleasantness; and has a five level scale from excellent to bad and it is easy to estimate the listening quality using a 5-point scale: 1-bad, 2-poor, 3-fair, 4-good, and 5-excellent.

In order to validate speech quality, participants that speak different dialects were selected from different parts of Oromia. Accordingly, 12 participants (3 females) were used and they listened the sentences using headphones. The subjects were familiarized with the speech synthesis by listening some example utterances of varying quality.

Of the total recorded and collected corpus, 48 complete sentences in Afaan Oromo had been selected randomly from the text corpus. Each of these sentences contains on average 10 to 32 words. The participants are allowed to listen to the recorded male voice samples before they check the developed TTS. Subsequently, each participant plays the sample voice to check the quality of the voice. However, in order to make the test effort easy and understandable by the participants, a questionnaire was delivered beforehand to familiarize with it. After listening audios, the participants were requested to fill the questionnaire properly.

The means and standard deviations were calculated as per the respondent's opinion. The participants are satisfied with pleasantness and the naturalness of the TTS voices with the average value of the mean being 4.24; and the acceptability of the developed system found to be very encouraging.

6. Conclusion and Recommendation

6.1 Conclusion

In this paper, corpus-based concatenative unit selection speech synthesis system architecture for Afaan Oromoo has been designed and implemented. We have done a literature survey on corpus-based concatenative speech synthesis research. Different techniques have been investigated and applied for optimal speech corpus design.

A unit selection algorithm based on dynamic programming has been implemented. Target and concatenation costs to be used in unit selection have been extracted. A speech representation based on harmonic coding has been implemented. As a result a unit selection based concatenative speech

synthesis system capable of generating highly intelligible and natural synthetic speech for Afaan Oromo has been developed. Subjective tests have been carried out to assess the speech quality generated by the system in terms of intelligence, naturalness and pleasantness. Finally, the system got 4.24 MOS. The variation emerged due to prosody issues, and concatenation of phones which will need further research.

An effort was also made in modeling syllabification algorithm and implementation for Afaan Oromo language. A set of formal rule was defined for the syllabification and epenthesis of its words. The algorithm achieves an overall 100% word accuracy which is well acceptable as rated with the experts. The experts are satisfied with the result. Thus, the rule-based syllabification algorithm developed for Afaan Oromo words is impressive given the linguistic rules and syllabification principles.

6.2 Recommendation

The following are the recommendations to further improve the quality of the Afaan Oromo synthesizer and syllabification:

- Handling prosodic issues (intonation, stress, and duration modeling)
- Developing germination handling algorithm.
- Application of HMM-based speech synthesis method

Similarly, it is also possible to make better and work the syllabification in a different approach:

- Stress and syllabification have their own relationship. Therefore, by studying their relationship thoroughly, we can have stress assignment algorithm.
- A comparison study using data-driven approaches.

References

[1] Samuel Thomas, “Natural Sounding Text-To-Speech Synthesis, based on Syllable-Like Units”, Master of Science, India, May 2007.

[2] Beekamaa Likkaasaa, “Seer Lugaa Afaan Oromoo”, Finfinne, 2004.

[3] Per Olav Heggtveit “An overview of Text-to-Speech Synthesis” Perolavhe Ggtveit, 2003.

[4] Hasim Sak, “A corpus-based concatenative speech synthesis system for Turkish”, 2000.

[5] Habte Bulti, “Analysis of Tone in Oromo”, Addis Ababa University, June 2003.

[6] Morka Mekonen, “TTS Synthesis for Afaan Oromo Language using Diphone Method”, 2001.

[7] Adugna Barkessa, “Concept of Afaan Oromoo Grammar”, Finfine, 2012.

[8] Sebsibe H/Mariam, S P Kishore, Alan W Black, Rohit Kumar, and Rajeev Sangal, “Unit Selection Voice for Amharic using Festvox”, 5th ISCA Speech Synthesis Workshop, Pittsburgh.

[9] Claire-A. Forel and Genoveva Puskás, “Phonetics and Phonology”, Geneva, March 2005.

[10] Thaha M. Roba, “Modern Afaan Oromo Grammer”, USA, 2004.