HUMAN LANGUAGE TECHNOLOGIES AND AFFAN OROMO

Taye Girma*

Abstract: This paper deals with Afaan Oromo, a Cushitic languages family which cover approximately 34.49% of the Ethiopia's population, pertinent to the language processing and speech technology with an overview of the attempts made on Afaan Oromo with respect to natural language processing; which include part of speech tagging, grammar checker, Parser and speech technology such as speech recognition and text to speech synthesis with the performance evaluation for few of them. The paper specially deals with the current state of Text-to-Speech Synthesis through addressing the different Text-to-speech synthesis approaches existing before and after the '90s pinpointing their advantages and disadvantages. Although the main focus of the paper is Afaan Oromo, it also briefly reviews the current progress of the language processing and speech technology potentials for Ethiopian languages.

Keywords: Ethiopian languages, Cushitic languages, Afaan Oromo, language processing, speech technology, Text to Speech Synthesis

*Department of Computer Engineering, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia

I. INTRODUCTION

The advancement of tools and methods for language processing has so far focused on a small numbers of languages and mainly on the ones used in the developed world and some of them are English, Spanish, French, Germany from Europe and Chinese, Japanese from Asia. However, there is a potentially even larger need for investigating the application of computational linguistic methods to the languages of the developing countries like Ethiopia and others. The need for localization in these countries for different purposes is enormously growing as most of the people in those countries do not speak the European and East-Asian languages where the computational linguistic community and speech technology groups have so far mainly concentrated.

Thus, there is an obvious need to develop a wide range of applications using human language technologies (HLT). HLT includes: speech recognition, speech synthesis, text categorization, text summarization, text indexing, information extraction, data fusion and text data mining, question and answering, report generation, spoken dialogue systems, Translation technologies, spelling and grammar checkers, information retrieval and filtering, and so forth [20]. However, the concern of this paper is to show the level of research made on Afaan Oromo with respect to natural language processing and speech technology. The main target of the researchers of this paper is at showing the current status of Afaan Oromo and text-to-speech synthesis with the current states of the arts. It also shows the attempted research work on POS, grammar checker, stemmer and speech recognition with respect to the specified language as it is indicated in table1.

A Text-to-Speech system (TTS) is the process of transposing written text into sound. But, in order to create the sound, thorough linguistic specifications for the text processing part of the TTS system (known as front-end) are required as it is highly dependent on the language, and includes– a number of processes like the transliteration of non-standard words (numbers, abbreviations, etc), word-to-phoneme conversion, part-of-speech tagging, etc [15]. However, the waveform or speech generation part of a TTS system is mostly independent on language and the state-of-the-art technology extracts the acoustic information from the training data, which consists of speech recordings annotated with the linguistic specification [15].

The paper will specially deals with the current state of Text-to-Speech Synthesis through addressing the different Text-to-Speech synthesis approaches existing before and after the ‘90s pinpointing their advantages and disadvantages. Although the main focus of the paper is Afaan Oromo, it also briefly reviews the current progress of the language processing and speech technology potentials for Ethiopian languages.

II. SPEECH SYNTHESIS METHODS

Speech synthesis systems have, over the years, been developed for various languages all over the world [5]. To realize speech synthesis systems, many synthesis methods have been proposed. Before the ’90s, methods based on ruled-based, formant synthesis had often been studied. These methods use phonetic units based on rules and those units used in the rules are hand-crafted.

Then after the ‘90s, various corpus-based methods, like concatenative synthesis were proposed and these methods generate speech by concatenating speech units from a database. This approach, simply stores the pre-recorded entire speech corpus itself for some selected parts of it; for example, from the given set of the limited size of the corpus, one instance of speech sounds. And then indexing the stored form of speech with the linguistic specification where it is also called ‘labeling the stored speech data’ such that appropriate parts of it can be found, extracted then concatenated during the synthesis phase. The index is used like the index at the back of the books to map all the occurrences of a particular linguistic specification. In a typical unit selection system, the labeling will comprise both aligned phonetic and prosodic information.

The process of retrieval is not entirely trivial, since the exact specification required at synthesis time may not be available in the corpus, so a selection must be performed to choose, from amongst the many slightly mismatched units, the best available sequence of units to concatenate. The speech may be stored as waveforms or in some other representation more suitable for concatenation (and small amounts of signal modification) such as residual-excited linear production coefficients (LPC) [15]. In terms of naturalness, the unit selection methods have proved to produce very natural sounding speech; though at the cost of very large speech training corpus [9]-[13].

After the mid-’90s, statistical model was proposed. In contrast to concatenative approach, the statistical approach does not store any speech. Instead, the model will be aligned to the speech corpus during the training phase and the model will be stored. The model which is stored after the training will be 'constructed in terms of individual speech units, such as context-dependent phonemes: the model is thus indexed by a linguistic specification'. At synthesis time, an appropriate sequence of context-dependent models like in [16] is retrieved and used to generate speech. 'Again, this may not be trivial because some models will be missing, due to the finite amount of training data available'. It is therefore important to prepare just like a 'look up table' to store model that holds linguistic specification to be used at a time of need. This is achieved by sharing parameters with sufficiently similar models: a process analogous to the selection of slightly mis-matched units in a concatenative synthesis approach [15].

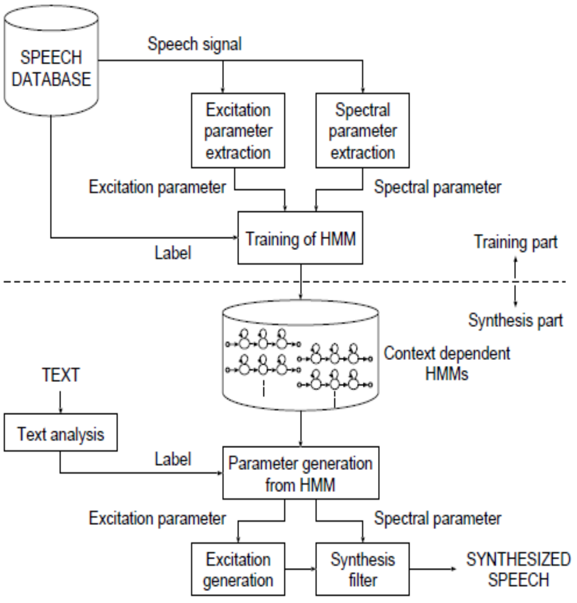

Fig. 1. HMM-based Speech synthesis system [8]

The HMM-based speech synthesis system (HTS) being statistical model-based system, it becomes popular after the mid-'00s [5]-[7].In this approach, spectrum, excitation, and duration of speech are simultaneously modeled by context-dependent HMMs, and speech waveforms are generated from the HMMs themselves [7]. Figure 1 shows the overview of HMM-based speech synthesis system where, the training part is similar to those used in HMM-based speech recognition systems. However, the essential difference between them is that the state output vector includes not only spectrum parameters, e.g., mel-cepstrum, but also source excitation parameters, F0 parameters. On the other hand, the synthesis part does the inverse operation of speech recognition: phoneme HMMs are concatenated according to the text to be synthesized. Then a sequence of speech parameters is determined in such a way that its output probability for the HMM is maximized. Finally, speech signal is synthesized by a speech synthesis filter.

HMM-based speech synthesis is a context-dependent approach and, in [16], an example of context-dependent label format for HMM-based speech synthesis in English has been presented. In contrast to the concatenative approach, the parametric methods, mainly based on Hidden Markov Models (HMM) require less speech data to train the system. In addition to this, it is statistical because it describes the parameters using statistics (e.g., means and variances of probability density functions) which capture the distribution of parameter values found in the training data.

In addition to a smaller training corpus, HTS also requires very little memory for the synthesis engine at runtime. As a result, TTS systems based on this approach can easily be integrated into handheld devices [1]-[2]. The other quality of HMM-based speech synthesis system is the possibility to generate various voice characteristics [5]. Therefore, the statistical method is the best and current state-of-art in the area. In summary, the advantages of Statistical Parametric Synthesis are: generates average speech units which are smooth and stable, stores statistics rather than waveforms, it is context dependent rather than language and it is easy to change style and emotions [5].

Even though most of the papers presented in [1] were about automatic Speech Recognition (ASR), the conference was held for special sessions on under-resourced language in easing information access and awareness as there is a great interest to port speech technology to under-resourced languages. In [18], the HMM-based speech synthesis system was proposed by the researchers for the cross-lingual use of resources for one of the under-resourced language, Malay, with few resources including recorded speech and segmental labels. First, they produced the time-aligned phone transcriptions for Malay using Festival English speech synthesis system and constructed Malay grapheme-to-phoneme database and English CART. Then they validated the result by intelligibility and naturalness tests on the synthetic speech after training.

III. SPEECH TECHNOLOGY FOR ETHIOPIAN LANGUAGES

Ethiopia, as one of the multilingual and multicultural countries, has faced the critical problem of development and implementation of language use policy that could satisfy the needs of various societies in question and contribute to their socioeconomic and socio-cultural development. The various governments that ruled Ethiopia since the reign of Emperor Tewodros II followed various language use policies that suit their political orientation [22]. However, the Ethiopian constitution of 1994 allocated Ethiopia into nine independent regions, each with its own “nationality language”, but still with Amharic being the federal working language.

Until 1994, Amharic was also the principal language and medium of instruction in primary and secondary schools of the country, but higher education in Ethiopia is actually carried out in English [21]. As Ethiopia was the only African country which managed to avoid being colonized during the big European Power struggles over the continent during the 19th century, it would thus be reasonable to assume that the country would have been using its languages for the educational system and for day to day activities especially for those who are unable to read and write using their respective mother tongue and the disabled community. However, this is not the case. This is because of lack of professionals who had been working on the local languages to localize the existing technologies for the regions and the nations as a whole [21].

However, at present Afaan Oromo has different official functions in the Oromiya Regional State and also in Oromiya Zone of the Amhara Region. It is a regional official language, a medium of instructions in primary schools, teacher training institutions and colleges, and these days it is found to be a field of study in higher educational institutions such as Addis Ababa University, Jimma University, Ambo University and etc. Moreover the language also serves as a language of the courts, religions, mass-media and so on.

A. The Cushitic Languages

The Cushitic languages are a branch of the Afro-Asiatic language family spoken primarily in the Horn of Africa (Somalia, Eritrea, Djibouti, and Ethiopia), as well as the Nile Valley (Sudan and Egypt), and parts of the African Great Lakes region (Tanzania and Kenya). The branch is named after the Biblical character Cush, who was traditionally identified as an ancestor of the speakers of these specific languages as early as 947 CE (in Masudi's Arabic history Meadows of Gold) [19].

Afaan Oromo:

The Oromo (Ge'ez: ⍰⍰⍰, ’Oromo) is an ethnic group inhabiting Ethiopia, northern Kenya, and parts of Somalia [25]. With 30 million members, they constitute the single largest ethnicity in Ethiopia and the wider Horn of Africa, at approximately 34.49% of Ethiopia's population according to the 2007 census [26]-[28].

Oromos speak the Oromo language as their mother tongue (also called Afaan Oromo and Oromiffa), which is part of the Cushitic branch of the Afro-Asiatic family. According to Gragg (1976) and Kebede (2005), in Ethiopia, Afaan Oromo has five major dialects: Rayya (Northern), Boorana (Southern), Tulama (Central), Harar (Eastern) and Mecha (Western). Then, following Oromo, Somali is the next with about 18 million speakers, and Sidama with about 3 million speakers. Other Cushitic languages with more than one million speakers are Afar (1.5 million) and so on [19].

B. Natural Language Processing and Afaan Oromo

This section will review the attempt made on natural language processing (part of speech tagging, grammar checker, Afaan Oromo stemmer and etc) for Afaan Oromo and the speech technologies (Speech Recognition and Speech Synthesis) that were attempted for Afaan Oromo. Gragg (1976) has produced an article that deals with the phonology, morphology and syntax of the language and few more works are described as follows.

Parts of Speech Tagging (POS):

Part of speech tagging is one of the linguistic knowledge required for speech synthesis and it is the act of assigning each word in sentences a tag that describes how that word is used in the sentences. [23] That means POS tagging assigns whether a given word is used as a noun, adjective, verb, etc.

According to Pla and Molina [23] notes cited by Getachew and Million, one of the most well-known disambiguation problems is POS tagging. A POS tagger attempts to assign the corresponding POS tag to each word in sentences, taking into account the context in which this word appears. According to Getachew and Million, the performance of the prototype for Afaan Oromo tagger is tested using tenfold cross validation mechanism and the result shows 87.58% and 91.97% accuracy both for unigram and bigram models, respectively.

Grammar Checker for Afaan Oromo:

Grammar checker determines the syntactical correctness of a sentence written in any human languages, which is mostly used in word processors and compilers [24]. Debela stated that for languages, like Afaan Oromo there is lack of advanced tools and the area is still in the early stages [30]. He developed the entire rules based on the morphology of Afaan Oromo in his paper and the evaluated result of the checker was very promising as the precision result shows 88.89%. But, still there is a need for further study to enhance the performance of the checker. Owens has made a study of the grammar of Afaan Oromo as indicated in [31] and described the phonology, morphology and syntax of the Harar dialect.

Afaan Oromo stemmer:

Most natural language processing systems use stemmer as a separate module in their implementation. Especially, it is very important to develop Afaan Oromo machine translator, Afaan Oromo speech recognizer and search engines for Afaan Oromo [30]. The developed stemmer is rule based and of course, the rules can't be complete because of the complexity of the language and hence this stemmer didn`t include any rule that handles compound words. However, the evaluation of the experiments for the developed rule gave an overall accuracy of about 94.84% and it is really an encouraging result. [30]

C. Speech Technology and Afaan Oromo

In this section, the speech technology will be reviewed in relation with Afaan Oromo. The speech technology includes Text to Speech Synthesis and Speech Recognition.

Speech Recognition for Afaan Oromo:

The ultimate goal of any automatic speech recognition is towards developing a model that converts speech utterance to texts words. Kassahun Gelana, in his study, tried to develop prototype for a continuous, speaker independent Afaan Oromo speech recognizer so as to check possibility and suitability of the tools and techniques selected from the various literatures [29]. A continuous, speaker independent Afaan Oromo speech recognizer's experiment is performed having similar objective of transforming Afaan Oromo continuous speech in to its text word formats for continuous Afaan Oromo speaker independent speech utterances using HMM and sphinx system (sphinx train for training and Sphinx4 for decoding). According to his performance evaluation using test data sets and the recognizer performance is found to be 68.514% with sentence accuracy of 28% for continuous Afaan Oromo speech and a phoneme based trigram performance of 89.459% with sentence accuracy of 42% achieved [29].

Text to Speech Synthesis and Afaan Oromo:

There were different attempts made to develop the TTS system for Afaan Oromo language by different researchers locally in Addis Ababa University at the master's level. However, the majority of them are on speech recognition.

Morka Mekonnen in his study, tried to address the issue of having textual information in speech forms using diphone based text-to-speech system for the language of Afaan Oromo [17].

He found that transcribing the orthography (writing system) into phonetic units for Afaan Oromo is well suited for developing rules and that made his transcription to be accurate. As a result, success in recognizing the utterance of the transcribed phonetic unit was 43.33% for naive listeners and 83.33% for listeners who heard the utterance at least three times in different day.

Samson Tadesse has also developed the dictionary based TTS for Afaan Oromo with good performance evaluation result but it needs a larger data base so as to include all possible utterance in the language [14]. Even though there were attempts on Afaan Oromo, the available Afaan Oromo TTS systems use diphone, dictionary-based and concatenative approach for synthesis.

V. CONTRIBUTION OF THE STUDY

This research on one hand aims to contribute to the speech technology domain of language technology studies in general and on the other hand to the field of TTS in particular by uncovering the Afaan Oromo and adding new insights to existing natural language processing literatures. The selected and used references to consult the TTS issues in this paper, are expected to significantly devise comprehensive concepts that can maximize the TTS mechanisms in a better way. The outcome of this research can also be used as a spring-board for future researchers who wanted to engage in detail TTS studies.

In general, the study is expected to shed a light on its implications for policy makers and practitioners as the output of this research can be used as important input to decision makers and other concerned bodies.

VI. CONCLUSIONS

The speech technologies (speech synthesis, speech recognition, and etc) and the natural language processing (POS tagging, grammar checker and etc) are at a very infant stage for Ethiopian languages and particularly for Afaan Oromo.

Few attempts are made on TTS for Afaan Oromo at the master's level with a promising performance evaluation result. However, the types of synthesis approaches employed by those researchers have difficulties when trying to improve the naturalness or intelligibility of the synthetic speech. On top of that, the existing locally developed TTS for Afaan Oromo were not using large high-quality speech corpus, which can then be used for statistical training of the HMM model, the current state of the arts. As far as the researchers' knowledge and other researchers who are working in the area it is well understood that there were no any local attempts made to develop TTS for Afaan Oromo using the statistical parametric approach and hence need further research on the languages [3]-[4].

In a nutshell, the author collected and presented the summary of all the NLP (Natural Language Processing) and ST (Speech Technology) attempts made by different researchers along with their area and their results as given in table 1 below.

Table1. Summary of various researchers' attempts for Afaan Oromo

Types	**Research area**	**Result of the study (%)**	**Researchers**	Ref
NLP	Parts of Speech Tagging (POS)	Unigram:87.58	Getachew and Million	[23]
		Bigram: 91.97		
	Grammar checker	88.89	Debela Tesfaye	[24]
	Afaan Oromo stemmer	94.84	Debela Tesfaye & Ermias Abebe	[30]
ST	Speech Recognition	Sent. Acc of 28: 68.51	Kassahun Gelana	[29]
		Sent. Acc of 42: 89.45		
	Text to Speech Synthesis	Naive listener: 43.33	Morka Mekonnen & Samson Tadesse	[17]
		Repeated listener (3 times): 83.33		

Therefore, since the area is at its infant stage for this language the author recommends the researchers to work on Afaan Oromo with large sized population so as to help the society benefited from the existing technologies.

REFERENCES

[1]ISCA.Interspeech:http://www.interspeech2011.org/specialsessions/ss-7.html, 2011 [Accessed on: 13-03-2014]

[2] A. Stan, P. Bell, and S. King, *A grapheme-based method for automatic alignment of speech and text data,*" In Proc. IEEE Workshop on Spoken Language Technology, Miami, Florida, USA, 2012

[3] B. Toth and G. NemethHidden Markov Model based speech synthesis in Hungarian, Information Communication Journal, vol. LXIII, no. 7, 2008, pp.30-34.

[4] A. Stan, Romanian hmm-based text-to-speech synthesis with interactive intonation optimization," Ph.D. dissertation, Technical University of Cluj-Napoca, 2011

[5] Junichi Yamagishi, Average-Voice-Based Speech Synthesis, PhD thesis, Tokyo Institute of Technology, 2006

[6] K. Tokuda, H. Zen, J. Yamagishi, T. Masuko, S. Sako, A. Black, and T. Nose, The HMM-based speech synthesis system (HTS) Version 2.1, http://hts.sp.nitech.ac.jp/, [Accessed on: March 15, 2014].

[7] H. Zen, K. Oura, T. Nose, J. Yamagishi, S. Sako, T. Toda, T. Masuko, A. W. Black, and K. Tokuda, Tutorial on Recent development of the HMM-based speech synthesis system (HTS), Nagota Institute of Technology, Nara Institute of Science and Technology, University of Edinburgh, Carnegie Mellon University.

[8] A. Black, H. Zen, and K. Tokuda*, Statistical parametric speech synthesis*, in Proc. ICASSP 2007, Apr. 2007, pp. 1229-1232.

[9] N. Baloyi and M. Manamela, An HMM-based Speech Synthesis System for Xitsonga, Master's thesis, University of Limpopo (Turfloop Campus), 2009

[10] A. W. Black, *Perfect synthesis for all of the people all of the time*," in IEEE 2002 Workshop on Speech Synthesis, 2002

[11]E. Moulines and F. Charpentier, Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," Speech Communication, vol. 9, no. 5-6, pp. 453-467, 1990

[12] A. Black and N. Campbell, *Optimizing selection of units from speech database for concatenative synthesis*," in Proc. EUROSPEECH-95, Sept. 1995, pp. 581-584.

[13]A. Syrdal, C. Wightman, A. Conkie, Y. Stylianou, M. Beutnagel, J. Schroeter, V. Storm, K. Lee, and M. Makashay, *Corpus-based techniques in the AT&T NEXTGEN synthesis system,* in Proc. ICSLP 2000, Oct. 2000, pp. 411-416.

[14] S. Tadesse, Speech Synthesis for Afaan Oromo, Master's thesis, Addis Ababa University, 2010.

[15] S.King , An introduction to statistical parametric speech synthesis, The Center for Speech- Technology Research , University of Edinburgh, http://www.ias.ac.in/sadhana/Pdf2011Oct/837.pdf , [Accessed on: 15-03-2014]

[16]Heiga Zen, Example of Context-dependent label format for HMM-based speech synthesis in English,https://wiki.inf.ed.ac.uk/twiki/pub/CSTR/F0parametrisation/hts_lab_format.pdf, [Accessed on: 15-03-2014]

[17] M. Mikonnen, Text-to-Speech for Afaan Oromo, Master's thesis, Addis Ababa University, 2001

[18] Mumtaz Begum Mustafa, Ainon Raja Noor, Roziati Zainuddin, Zuraidah M. Don, Gerry Knowles, A cross-lingual approach to the development of an HMM-based speech synthesis system for Malay, ISCA. Interspeech, 2011

[19]E.Wikipedia Cushitic Languages: http://en.wikipedia.org/wiki/Cushitic_languages[Accessed on: 8-01-2014]

[20] B.Gambäck and G. Eriksson, *Natural Language processing at the School of Information Studies for Africa Proceedings of the Second ACL Workshop on Effective Tools and Methodologies for Teaching NLP and CL*, Ann Arbor, Association for Computational Linguistics, 2005, pages 49–56. 2005.

[21] T. Bloor and W. Tamrat, Issues in Ethiopian language policy and education, Journal of Multilingual and Multicultural Development, 1996, 17(5):321–337.

[22]Getachew Anteneh and Derib Ado, Ethiop. J. Educ. & Sc. Vol. 2 No. 1, September 2006, pp 37-62.

[23]Getachew and Million, Part of Speech Tagging for Afaan Oromo, IJACSA, Special Issue on Artificial Intelligence, 2011, pp 1-5.

[24] Debela Tesfaye, A rule-based Afaan Oromo Grammar Checker, Vol. 2, No. 8, 2011, pp126-130.

[25]Merriam-Webster Inc, Frederick C. Mish, Merriam-Webster's Collegiate Dictionary, (Merriam-Webster: 2003), p.876

[26] Census Report. “Ethiopia’s population now 76 million”. (2008) available at: http://ethiopolitics.com/news, [Accessed on: March 21, 2014]

[27]The CSA estimates a population growth of 7.6% between the time the census was conducted and the date of its approval: "Ethiopia population soars to near 77 million: census". Google News. AFP, 4 December 2008. [Accessed on: March 29, 2014]

[28]Tesema Ta'a, The Political Economy of an African Society in Transformation (2006), books.google.com/books?isbn=3447054190, pp. 17

[29] K. Gelana, A continuous, Speaker Independent Speech Recognizer for Afaan Oromo, Master's thesis, Addis Ababa University, 2010.

[30]Debela Tesfaye, Ermias Abebe, Designing a Rule Based Stemmer for Afaan Oromo Text, International Journal of Computational Linguistics (IJCL), Volume (1): Issue (2), pp1-11, 2010

[31]Owens, Jonathan, A Grammar of Harar Oromo (Northeastern Ethiopia). Hamburg: Helmut Buske Verlag, 1985