J.Sci.Univ.Kelaniya **3** (2007) : 63-72

PRACTICAL ISSUES IN THE DEVELOPMENT OF TTS AND SR FOR THE SINHALA LANGUAGE

R.I.P. WICKRAMASINGHE[1], K. H. KUMARA[2] AND N.G.J. DIAS[1*]

[1]**Department of Statistics & Computer Science, University of Kelaniya, Kelaniya, Sri Lanka**
[2]**Department of Mathematical Science, University of Wayamba, Kuliyapitiya, Sri Lanka**

*****Corresponding author (E-mail: ngjdias@kln.ac.lk)

ABSTRACT

In this study, discussed some issues that are arising in the development of Text-to-Speech (TTS) and Speech Recognition (SR) systems for the Sinhala language are discussed. There are numerous benefits of TTS and SR to Sinhala community. In the context of Sinhala language, the Natural Language Processing (NLP) technologies are in the primary stage in comparison with other languages such as English, French, German, Japanese etc. Since current TTS and SR development environments are far from being language independent, the existing systems cannot be adopted directly to the Sinhala language. This contribution outlines problematic areas that come across in the development of TTS and SR systems for Sinhala language. While discussing those problems, some of the possible solution strategies for both TTS and SR development have been formulated.

Keywords: Text-to-Speech, Speech Recognition, Sinhala Language, Natural Language Processing.

INTRODUCTION

TTS can be considered as the automatic production of speech, through a grapheme-to-phoneme transcription of the sentences to utter (Dutoit, 1993). TTS software can "read" text from a document, Web page or e-Book, generating synthesized speech through a computer's speakers. There are variety of applications of TTS software such as proofreading, Voice Announcements and Readouts, Dual Party Telephone Relay, Reverse Directory Assistance, Spoken E-mail and Fax, Talking Books, Internet Access. Also TTS can be used to convert text files into audio files that can then be transferred to portable forms.

SR is the process of converting an acoustic signal, captured by a microphone or a telephone, to a set of words (Ronald, 1996). Similar to the TTS applications there are numerous usages of SR systems in education**,** in data entry for people with physical or learning difficulties, in automated telephone-based information retrieval systems, in voice response devices industry and security considerations.

Sinhala is a language of around 13 million people living in Sri Lanka. It is not spoken in any other country, except of course by enclaves of migrants. Being a descendant of a spoken form (Pali) of the root Indic language, Sanskrit, it can be argued that it belongs to the large family of Indo-Aryan languages (Weerasinghe, 2004).

To benefit from above described TTS and SR applications to the Sinhala community, high quality TTS and SR systems have to be developed.

ISSUES

Some basic issues that arise at the development of TTS and SR systems for the Sinhala language are discussed here. Among the major issues that we are going to discuss are speech and text corpora, code switching of text and utterance, lexicalized foreign origin names or terms, prosody, dealing with dates and numbers.

Speech and text corpora

The problem with the development of TTS and SR for Sinhala language is the lack (or very small amount) of speech and text corpora. It is one of the reasons why its linguistic structure is very poorly investigated in the aspects of Computational Linguistic and Mathematical Linguistics disciplines. In Sri Lanka Computational Linguistic and Mathematical Linguistics disciplines have not been developed even in University level though there are many specialists in Computing, Mathematics and Linguistics separately. This may be one of the major barriers to over come the above issues.

Speech corpora for TTS and SR are usually recorded by speakers with normative pronunciation. In Sri Lanka there are different pronunciation patterns depending on the geographical regions. As an example the way upcountry people speak is different from that of down south people. Therefore it is not straightforward to define the term “normative speech” for Sinhala language with various dialects. This is one of another issue that we have to consider when speech corpora are developed for Sinhala. Some of the possible suggestions to overcome above issue are considering the speech of broadcast readers as the normative, carrying out Socio-linguistic studies and the speech of a particular person (professor, writer, actor etc.). As a start point the speech of a broadcast reader can be taken nevertheless the most reliable method is carrying out Socio-linguistic studies that requires a lot of effort.

Code switching of text and utterance

As Sri Lanka has a multilingual community, code switching is very common in communication. This can be easily seen among people as well as among the media, especially in FM Radio channels and news paper advertisements. Code-switching can range from a single word to a complete shift in the primary language of the interaction. Switching abruptly for short regions, however, may make the overall utterance difficult to understand in the context of a SR system.

Both in TTS and SR, code switching can create various problems. For an example when the input is given as “**මම** call **කරන්නම**“ to the SR system, it finds very difficult to handle the wave pattern generate by word, “call”. Though system tries to recognize the wave pattern for the word “call” by comparing Sinhala corpus, the system fails to do so. The reason is Sinhala corpus does not contain such a word. This type of code-switching is another issue that we need to handle in a SR system of

Sinhala language. In these cases identifying the regions where code-switching takes place is, of course, one problem for SR. Suitable methods should be introduced to overcome this type of issues.

Digits, Numbers, Fractions and Dates

Digits, Numbers and dates must be expanded into full words. For example in Sinhala, numerals 1973 would be expressed in following ways.

If 1973 is number would be expanded as **එක් දහස් නවසිය හැත්ත තුන**

If 1973 has been expressed in years, then the expansion should be **අවුරුදු එක් දහස් නවසිය හැත්ත තුනයි**

If 1973 is about students, then the expansion should be **සිසුන් එක් දහස් නවසිය හැත්ත තුනයි**

The expressing fractions and dates are also problematic. When a fraction is expressed that can be wrongly interpreted by the system as a date or vise versa. When both numbers are above 31 and that can be easily understood by the system as a fraction. The real problem occurs when dealing with numbers that are below 31 as given in the following.

Fraction 1/2 can be expanded **භාගය** or **එක බෙදීම දෙක** (if 1/2 is a fraction).

If 1/2 express as a date the expansion should be ජනවාරි දෙක.

Therefore these types of problematic situations should be handled when developing TTS and SR systems. Expansion of ordinal numbers has been found also problematic.

.

Handling Acronyms

Abbreviations may be expanded into full words, pronounced as written or pronounced letter by letter (Macon, 1996). There are also some contextual problems. For example, ලි.ප.තැ and සැ.යු are usually pronounced with expansions and ස.ණස or සතොස letter-by-letter.

When there is one acronym for two completely different things it would be a very difficult problem to handle. For an example the English acronym ASP stands for

ඇක්ටිව් සර්වර් පේජස් in computer science point of view though the general public refers to *සහකාර පොලිස් අධිකාරි.*

Special characters and symbols

Special characters and symbols such as 'Rs', '. ','%', etc, also cause special kinds of problems. Consider the following situation.

123.45 should be expressed as එකසිය විසි තුනයි දශම හතරයි පහයි. but the same value is expressed with 'Rs' symbol, it would be රුපියල් එකසිය විසි තුනයි සත හතලිස් පහයි.. This creates a serious problem. Similar to above when '-' is used it is very difficult to understand the system whether '-' represent minus sign or a range between two numbers. For an example consider the following.

X-Y should be expressed as X අඩු කිරීම Y if the '-'sign represents a minus symbol.
X-Y should be expressed as X සිට Y if the '-'denotes all the numbers from X to Y.

Same Pronunciation with different Spelling

The same pronunciation with completely different meanings as well as different spellings creates difficult problems in SR systems for Sinhala language.

The meaning of ණය is *loan* and the meaning of නය is *theory*. But the problem is both words have same pronunciation. This is very difficult to tackle by the SR system. Consideration of the context of the sentence would be a possible solution to over come this problem.

Consider the following set of words. They have almost the same pronunciations with different meanings.

අදෙස් :- ආදේශය, ආදර්ශය
අදෙස් :- නිර්දෙෂ

අබියෙස් :- සමීපය
අබිසෙස් :- අභිෂේකය

අනුභවය :- කෑම
අනූභාවය :- බලය

These types of words are common in Sinhala. Therefore a high quality Sinhala SR system should be able to identify them correctly.

Phrasing

Phrasing is one of the most important factors of speech recognition. Proper phrasing will give the correct massage. Basically, Sinhala words can be categorized into four groups as “namapada”, “kriyapada”, “nipatha” and “upasarga”. “upasarga” should be appeared before “namapada” or “kriyapada”. “upasarga” like ප, පර, අව, ස, අනු, නි, දු, වී … etc., will combine with “namapada” or “kriyapada” and display as a single word. අනුනායක, දුදන, පසිදු, අවවාදය are examples for such instances.

“nipataha” can be occurred before/after a word or in between two words as ළමයාට ම, නගරයේහි දී, ගමක් වී ය. “nipatha” should also be separated from the root word. But the “nipatha” ව should not be separated as කියනුව, බලනුව, ලියනුව …etc. Punctuation marks also play a major role in written Sinhala language with respect to the meaning of the sentence. Using correct punctuations in the correct position will give the correct message. Consider the following two sentences:

මා කපා තිබෙන ගස් ගෙන යන්න

and මා කපා, තිබෙන ගස් ගෙන යන්න

Even both sentences have same words they will give different meaning because of the comma in the second sentence. Consider the Sinhala sentence:

ළමයා මළා

In colloquial form, we can give three different meanings by changing the prosody of the sentence. If we wrote it without any punctuation marks then it will not be given the correct meaning of the sentence. Because it can be an ordinary expression as:

ළමයා මළා

or can be a question as:

ළමයා මළා?

or can be an exclamation as:

ළමයා මළා!

These types of ambiguity occurrences should be taken into consideration when developing SR and TTS systems for Sinhala.

Foreign Names and Terms

Finding correct pronunciation for proper names, especially when they are borrowed from other languages, is usually one of the most difficult tasks for any TTS system.

Many foreign words and loan words are common in Sinhala. Some of them have derived into Sinhala accent and they will not be problematic in phonemic transition process. කැමරාව, ට්‍රොලිය, කාරය, එන්ජිම …etc are examples for such instances. But some words such as Eiffel, Blare, TOEFL ..etc are in their original form. Therefore those words are ambiguous to both SR and TTS. To overcome this type of ambiguities, these kinds of words may be included in a specific exception dictionary with their pronunciations. It is clear that building a database with all the proper names in the world is an impossible task.

Identification of Phonemes

The identification of phoneme boundaries in continuous speech is an important problem in areas SR and TTS. In particular, speech synthesis requires accurate knowledge of phoneme transitions, in order to obtain a naturally sounding speech waveform from stored parameters. Although Sinhala is a phonetic language there are many ambiguities in the process of identifying phonemes. As an example if we take the Sinhala word අක්කා (/akka/), generally it is divided into two phonemes as /ak/ and /ka / or /ak/ and /a:/ . But the acoustic is as follows (Fig. 1).

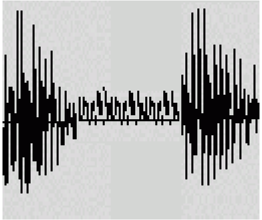

Figure 1. Acoustic graph for the Sinhala word /akka/

It is clear that according to the acoustic feature we cannot separate /akka/ in to two segments within “kk”. Usage of acoustic features instead of parametric features or vise versa sometimes may be problematic in the automatic segmentation.

Prosody

Though the written text usually contains very little information of prosody features and some of them change dynamically during the speech, for better pronunciation of the TTS this prosodic information may be given to a speech synthesizer. Mapping the appropriate prosody (intonation, stress, and duration) from written text is the most challenging problem for TTS. For an example when the phrase මොකද කරන්නේ? is expressed by changing the intonation it will give two different meanings. With the low intonation it means *what is to do*. But with high intonation it gives a sense of annoyance.

The duration also matters lot in SR as there is a high tendency to interpret following two words because of the changes of the duration. සමය and සාමය are two words with different meanings. But when it is pronounced by changing the duration of the utterance it will create an ambiguity.

Timing at sentence level or grouping of words into phrases correctly is difficult because prosodic phrasing is not always marked in text by punctuation, and phrasal accentuation is almost never marked (Santen et al., 1997). If there is no breath pauses in speech or if they are in wrong places, the speech may sound very unnatural or even the meaning of the sentence may be misunderstood. For example, the input string "මා මට පොතක් ගත්තා" can be spoken as two different ways giving two different meanings as " මා මට පොතක් ගත්තා " and මාමා ට පොතක් ගත්තා. In the first sentence " මා මට පොතක් ගත්තා " means *I myself bought a book*, and in the second it means *I bought a book for uncle*.

Although Sinhala language is not a tone language the prosody feature plays a major role in the communication.

CONCLUSIONS

Switching to Natural Language Processing in Sri Lanka is not an easy task because of the poor computational and mathematical linguistic infrastructure available. Clearly, still we have to do lot of things to learn how to develop high quality TTS and SR systems for Sinhala Language. Most improvements are now expected to arise in the field of Natural Language Processing in Sri Lanka. We should try to originate from a better modelization (Mathematical, Statistical or Rule based) of the human prosody rather then from more advanced text analyses.

It is obvious that in the development of both TTS and SR issues like Speech and Text corpora, Code switching of text and utterance, Digits, Numbers and Dates, Fractions and dates , handling Acronyms, Special characters and symbols, Same Pronunciation with different Spelling, Foreign Names and Terms and Prosody should be handled properly. In this paper we discussed some practical issues and solutions both as we view and as it seen in the rest of the literature.

REFERENCES

Cole R.A., J. Mariani, H. Uszkoreit & A. Z. V. Zue 1997. “Survey of the State of the Art in Human Language Technology”, Cambridge University Press and Giardini, Cambridge.

Dutoit T. 1993. High Quality text to Speech Synthesis of the French Language, PhD dissertation, Faculte Polytechnique de Mons, TCTS Lab, 31 bvd Dolez, B-7000 Mons (Belgium).

Jurafsky D. & J. H. Martin 2004. Speech and Language Processing, Pearson Education Series.

Weerasinghe R. 2004. "A Statistical Machine Translation Approach to Sinhala-Tamil Language Translation", SCALLA working conference, Katmandu.