New Prosodic Phrasing Model for Sinhala Language

W.M.C. Bandara, W.M.S. Lakmal, T.D. Liyanagama, S.V. Bulathsinghala, Gihan Dias and Sanath Jayasena
Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka.

Abstract **- This paper describes a new model of predicting prosodic phrase breaks in Sinhala language in order to improve the quality of the existing TTS Sinhala voices. In a Text To Speech (TTS) system, quality of the synthetic voice is mainly dependent on, how well its prosodic model is implemented. The prosodic model adjusts the phrasing and the pitch of the voice while applying suitable durations and tones for words and diphones. Out of these, phrasing and pitch of the voice carries much importance since appropriate phrase breaking helps to clearly understand the synthesis voice. In a real world scenario, when we speak a sentence, we automatically divide it to small segments and apply pauses at those breaks. Also the pitch of the voice gets lowered near a break and gets increased in the other segments automatically. But in a TTS system, we do not have that advantage and therefore need to be précised with the phrase breaks. Otherwise it will create wrong meanings as well as producing unnatural speech. Existing Sinhala TTS systems lacks proper prosody implementations and hence difficult to understand when it reads, especially long sentences. This issue can be overcome by applying a suitable phrase breaking technique.**

I. INTRODUCTION

The existing method [1] of identifying phrase breaks for Sinhala language is capable of identifying only the presence of two possible occurrences and applying phase breaks in those scenarios. First it can identify punctuation marks and applies long breaks marked. Also it can identify content words and it applies short breaks at those positions. There is a major issue with the current way of identifying and applying breaks on content words. If a sentence has consecutive function words the system stops at each and every word resulting unnatural phrase break output. Also it maintains a gap of five words between a short break and a long break. If the gap is less than five then it will not apply a break for that content word. This feature is common in both English and Sinhala phrase break modules even though the size of gap is different. [2]

In order to achieve more suitable phrase breaking technique, we need to understand how the phrase breaks are occurred in Sinhala language, since it is language dependant. After doing some background analysis [3] and getting expertise help [4][5][6] on how phrase breaks are applied in Sinhala, we found out that it is associated Part Of Speech (POS) words like adjectives, participles, pre-verbs and etc. In this paper we discuss how those associations can be implemented.

A. Background

Although phrase breaking plays a major role in the quality of the voice, there are very few phrase breaking models exists. One reason for that is these models are language dependant and need lot of research and knowledge about the language. Lack of knowledge and not having a well defined POS database leads to a failure in the model. However there are some models exists in English language which can be used for build a new model for a local language. Festival [2] has one such implementation for English language using a Classification And Regression Tree (CART) in Scheme language. In that tree, language specific phrase breaking rules which are developed according to the POS databases of that language are applied. There is also another phrase breaking method with much statistical approach, but it cannot capture useful information in long distance [7]. Nevertheless it requires less amount of linguistic knowledge compared to CART approach.

B. Problem

Festival phrase break CART uses a categorized word base to predict phrase breaks in a given input text. It has *Function Words (fn)* and *Punctuation (punc)* as Part Of Speech (POS) tags for allocating phrase breaks, yet when it comes to Sinhala language, the algorithm is to be improved for other categories as adjectives, pre-verbs, participles and etc... After going through many Sinhala voice clips, we were able to build the following rules which give proper phrase breaks.

Rule No. 1 **(When there is an 'adjective' and a 'noun', not followed by a preposition)**

When there is an adjective in the sentence, there should be a pause after the noun followed by that particular adjective.
e.g.
වියපත් යාචකයා අප සමඟින් සිනාසුනේය.
යාචකයා අප සමඟින් සිනාසුනේය

When there is a preposition appears after the following noun, the break will not apply after the noun
e.g.
මම වියපත් යාචකයා වෙත මුදල් දීඥ කලෙමි.

Rule No. 2 **(When there is a 'pre-verb' occurs in a sentence)**

When there is a pre-verb exists in the sentence, there should be a pause after the pre-verb.
e.g.
සුමනදාස කොළඹ ගොස් මාතර බසයකට ගොඩවුනේය.
සුමනදාස මාතර කොළඹ බසයකට ගොඩවුනේය.

When a pre-verb is followed by a participle, the pause will not be applied after the pre-verb.
e.g.
මල්ලී සෙල්ලම් කිරීමට ගොස් තිබුණේ මගේ බයිසිකලයත් රැගෙනය.

***Rule No. 3*_(When there is a 'preposition' appear in a sentence)**

When there is a preposition exists in the sentence, it should be followed by a pause and when there are two consecutive prepositions in a sentence the pause should be applied after the last preposition.
e.g.
සාලිය අමරා සමඟ පාසැලට පැමිණියේය.
සාලිය අමරා සමඟ සිට අප දෙසට ඇවිද ආවේය.

When there is a pre-verb after the preposition, the pause will not be applied after the preposition.
e.g.
සාලිය අමරා සමඟ ගොස් පන්සිල් ගත්තේය.

***Rule No. 4*_(When there is a 'participle' occurs in a sentence)**

When there is a participle in the sentence, it should be followed by a pause.
e.g.
මඟ දෙපස අරලියා මල් පිපී තිබුණේ අප සැම පිළිගන්නට මෙනි.
අප එන මග දෙපස අරලියා මල් පිපී තිබුණි.

When there is a preposition appears after the participle, the break will not come after the participle.
e.g.
මම සොරුන්ට කැගැසූ නමුත් කිසිදු පලක් නොවිනි.

***Rule No. 5* (In a word where a 'participle' is joined with a 'preposition')**
Participle + preposition-ත් - බැලීමත්, නැරඹීමත්, කිරීමත්, පිරිනැමීමත්...

When an above sort of word found in the sentence, it should be immediately followed by a pause.
e.g.
පොත්පත් බැලීමත් රුපවාහිනිය නැරඹීමත් ඔහුගේ දින චර්යාවේ අංගයන් විය.
පොත්පත් හා රුපවාහිනිය ඔහුගේ දින චර්යාවේ අංගයන් විය.

II. Methodology

In order to cope up with above complex rules, we defined three new POS tags, *Adjectives (adj)*, *Participles (pp)* and *Pre-verbs (pv)* and added them to the word base. Currently these three categories consist of words which are considered as most used words in their respective categories. This word base is then mapped with data driven CART which implements the algorithm for phrase breaks.

When the input text utterances are given to the CART, it starts from root and traverse through different intermediate nodes. All the nodes except the leaf nodes have conditions which will check for particular features in the given text such as whether the previous word is an adjective, whether the current word is a pre-verb, whether two or more consecutive fn words are in the text… etc. An example of a partial Phrase CART is shown in Figure 1. Finally at a leaf node we assign either no break (*NB)*, small break (*B)* or large break (*BB)* depending on the path that traversed. Evaluation, detailed analysis and the contribution from the proposed phrase breaking model is given in Section 3.

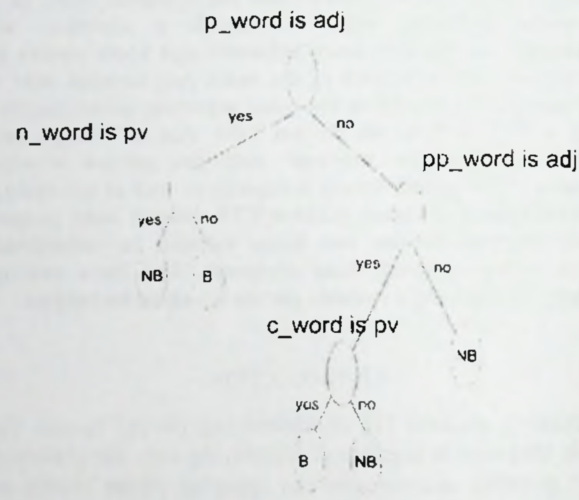

Figure 1

In order to overcome the consecutive phrase breaks issue in the existing systems, previous breaks lookup approach is incorporated to the phrase break tree. This function check where the last phrase break occur and if the distance to the last phrase break is greater than some pre defined value, then only it will apply a break. If the distance is too small, the system will not apply breaks on the current word hence the quality and the naturalness of the voice is preserved.

III. Results And Discussions

When it comes to the testing of the new phrase breaking algorithm we had two main concerns. First one was to evaluate the accuracy of the phrase breaking algorithm against the existing one. And the second one was to test for the effect of this algorithm, on the intelligibility of the overall system.

For the evaluation of new phrase breaking model in contrast to the existing one, we selected 10 people in age from 20 years to 30 years and asked them to predict when should be the phrase breaks appear in the given sentences. Then we synthesized the selected text from the existing phrase breaking model and our phrase breaking model and noted the accuracy of the phrase breaks and non-phrase breaks as follows.

Existing model

	Occurred Non-break	Occurred break	Accuracy
Predicted Non-break	457	4	99.13
Predicted break	23	57	71.25

New model

	Occurred Non-break	Occurred break	Accuracy
Predicted Non-break	453	8	98.26
Predicted break	2	78	97.5

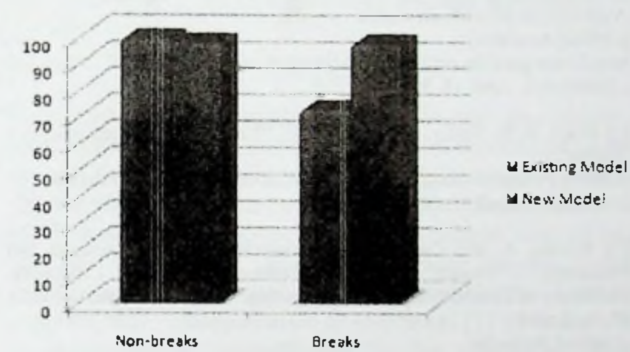

Figure 2 Breaks and Non-breaks in contrast with existing and new models

When testing for the intelligibility, we used some of the standard methods of evaluating TTS Systems, in this testing process. Mainly we used couple of Rhyme Tests namely; Diagnostic Rhyme Test (DRT) and Modified Rhyme Test (MRT) for intelligibility testing. We conduct these tests in three phases.

In the first phase we used twenty five pairs of words with different level of confusability and we played only a single word from each pair. Then we asked the observers to mark the word that he heard on the answer sheet.

In the second phase we repeated the same test we done before using five pair of sentences. In the final phase we played ten synthesized sentences using the TTS system and we asked the listeners to write down what they heard.

The first test was to determine the word level intelligibility while the second and third tests to evaluate the sentence level intelligibility.
After performing these tests we took average error rate as a measure. But we can also get individual error rates for different consonants to determine how they confused with each other as well.

The test group consists of 20 students from Computer Science and Engineering Department of University of Moratuwa. When selecting the group we decided to take people who have Sinhala language as their mother tongue in order to maintain a test group with excellent command in Sinhala language. Since all the participants were Computer Science and Engineering students, the group had a good understanding about the TTS systems and their real world applications. Also the age group was 23-25 and out of 20 participants we choose 8 to be female students.

We took one participant at a time and performed these tests individually. Also we did not let them to take all the tests in a single run. Because we believed that it would help to avoid the situations like the user get familiarized to the voice.

IV Test Results For Intelligibility Testing

A. Perception of the words (intelligibility in word level)

This test contained twenty five pairs of words with different levels of confusability. Those pairs differ from the initial consonant, the middle consonant or from the last consonant. The listener hears the word and then marks the word that he thinks correct on the answer sheet.

According to the received error rates most of the words were recognized correctly by the listeners. Only three words were misunderstood by the listeners. By going with the results we can say that 98.6% of the words were accurately recognized by the intelligibility test performed at the word level.

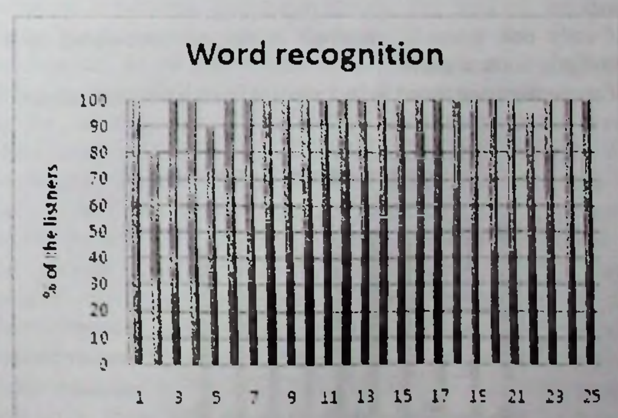

Figure 3Test results for word recognition

B. Perception of the sentences (intelligibility in sentence level) - Part one

Here we play five pairs of sentences with different levels of confusability and listener has to mark the correct sentence from the two sentences in the answer sheet.

Since the last two sentences are somewhat confusing, some of the listeners got wrong in those instances. But in the first three sentences, especially in the second one, where the comparison between the question and the declarative sentence listeners identified correct one with on effort at all.

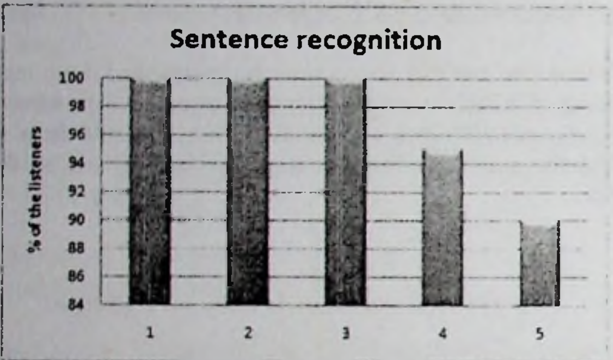

Figure 4 Test results for sentence recognition -1

C. Perception of the sentences (intelligibility in sentence level) - Part two

Here we played ten sentences once at a time and asked the observer to write down the sentence that he heard in the space given in the answer sheet. A sentence would be played only once and these sentences were short in length and hence we expected the memorizing capacity would not have affected to the results of the test.

In this test we have taken these results based on the following assumption.

If the sentence is totally correct it is considered as a correct answer.

If only one word is incorrect it can be considered as a partially correct answer.

If more than one word is incorrect it is an incorrect answer.

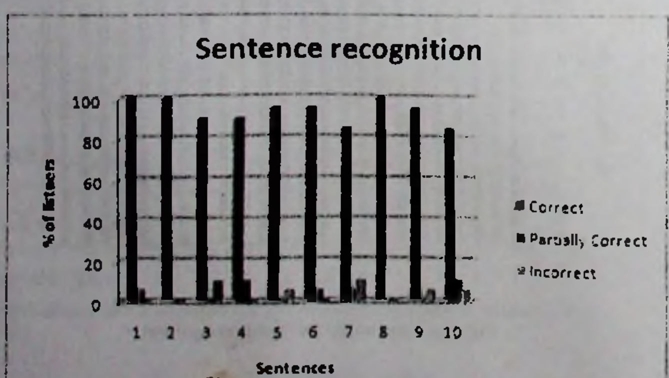

Figure 5 Test results for sentence recognition -2

From the results, the new model provides a significant advantage over the old model on defining phrase breaks as appropriate.

Accuracy of defining of non-breaks from new model is slightly less than the old model. Still the new model is acceptable due to its very high accuracy.

Analyzing the results obtained from the intelligibility testing, we can clearly see that the new algorithm does not impose a negative effect on the intelligibility of the overall system.

Taking all these facts into the consideration, we can say that this new phrase breaking algorithm improves the overall performance of a Sinhala TTS system.

REFERENCES

[1] Wasala, A. Weerasinghe, R. Welgama, V. and Gamage, K. "Festival-si: A Sinhala Text-to-Speech System", Language Technology Research Laboratory, University of Colombo School of Computing, 35, Reid Avenue, Colombo 07, Sri Lanka.
[Online].Available:
http://www.panl10n.net/english/outputs/SriLanka/Sinhala%20TTS/Reports/. [Accessed: April 18, 2010]

[2] Black, A.W. Taylor, P. and Caley, R. "The Festival Speech Synthesis System",
[Online].Available: http://www.cstr.ed.ac.uk/projects/festival/manual/ [Accessed: March 15, 2010]

[3] Wasala, A. and Gamage, K., "Research Report on Phonetics and Phonology of Sinhala" , Language Technology Research Laboratory, University of Colombo School of Computing, 35, Reid Avenue, Colombo 07, Sri Lanka .
[Online].Available:
http://www.panl10n.net/english/outputs/Working%20Papers/Sri%20Lanka/. [Accessed: April 18, 2010]

[4] Karunatillake, W.S. "An Introduction to Spoken Sinhala", 3rd edition, M.D. Gunasena & Co. ltd, 217, Olcott Mawatha, Colombo 11, 2004.

[5] Alwis, P. 2006, "Niwaeradi Wahara", Suriya Prakashakayo, 109, Colombo 10. 2006.

[6] Alwis, P. "Niwaeradi Wahara -2", Suriya Prakashakayo, 109, Colombo 10. 2007.

[7] Tesprasit, V. Charoenpornsawat, P. and Sornlertlamvanich, V. " Learning Phrase Break Detection in Thai Text-to-Speech" Information Research and Development Division National Electronics and Computer Technology Center, Thailand; Thai Computational Linguistics Laboratory CRL Asia Research Center, Thailand.[Online].Available: http://www.tcllab.org/virach/paper/virach/.
[Accessed: April 20, 2010]