A New Prosodic Phrasing Model for Indian Language Telugu

N. Sridhar Krishna, Hema A. Murthy

Department of Computer Science and Engineering, Indian Institute of Technology, Madras, Chennai - 600036
Email: \{sridhar,hema\}@lantana.iitm.ernet.in

Abstract

Prosodic phrasing is an important and more difficult a problem for Indian languages, as the Indian language scripts use very little or no punctuation. This paper reports a preliminary attempt on data-driven modeling of prosodic phrase boundary prediction for the Indian language Telugu. In an effort to identify meaningful features that affect the prosodic phrasing, a new feature, namely morpheme tag, is defined. A Classification and Regression Tree (CART) based data-driven phrasing model is developed for the prosodic phrase boundary prediction and the usefulness of the morpheme tag feature is further demonstrated in an evaluation process. The phrasing model developed has been implemented in an Indian language Text-to-Speech synthesis system [1] being developed within Festival framework [2].

1. Introduction

In natural speech, humans tend to group words together with noticeable breaks or disjunctions between them. These groups can be identified as prosodic phrases. Prosodic phrasing plays an important role in structuring utterances by dividing them into meaningful chunks of information. Text-to-Speech systems should be able to identify these prosodic phrases to produce intelligible and natural sounding speech. Good models for identifying the prosodic phrases are crucial. They not only enhance intelligibility and naturalness but also the structure these models impose serves as input for several other modules in text to speech conversion and thus has a major influence on their performance. For example,

- Pronunciation generation module which gives the phoneme sequence needed to synthesize an utterance introduces pause at phrase boundaries and might even introduce some allophonic variation of the same phoneme at the phrase boundaries.

- The duration module lengthens the segments which occur immediately prior to a phrase boundary.

- In fundamental frequency contour generation, phrase boundaries delimit intonation phrases which help in identifying the boundary tones.

A number of different models have been proposed ranging from simple deterministic rules (rules based on punctuation) to most complex models which require full syntactic parse of the sentence to be synthesized. Traditional research on the location of phrase boundaries has focused primarily on the relationship between prosodic structure and syntactic structure, and uses some sort of syntactic information to predict prosodic boundaries often in the form of heuristic rules [3, 4]. However, such hand-crafted rules systems are very difficult to write, maintain and adapt to new domains and languages. To avoid these problems, recent efforts have focused on acquiring phrasing rules automatically by data-driven methods that require large prosodically labeled corpora. These data-driven methods differ mainly in two aspects: the learning model and the feature set. The learning techniques currently used include classification and regression trees [5, 6, 7], hidden markov models [8], neural networks [9] and transformational rule-based learning (TRBL) [10]. All these learning techniques primarily use a feature set that can be derived automatically from the raw text. For example, features such as punctuation, Part-of-Speech information, morphological information etc.

In highly inflective languages like Telugu, most words in running texts occur in inflected forms, occurring in the bare-stem or dictionary form very infrequently. And also, as there are very few function words exist, even the deterministic rules based in content/function word distinction are not applicable. In an effort to identify meaningful features that affect prosodic phrasing, a new feature, namely morpheme tag, is defined. For identifying prosodic phrase boundaries, a Classification and Regression Tree based data-driven method which makes use of the new feature, morpheme tag, is developed. Classification and Regression Trees are models based on self learning procedures that sort the instances in the learning data by binary questions about the attributes that the instances have. It starts at the root node and continues to ask questions about the attribute of the instance down the tree until a leaf node is reached [11]. The decision tree algorithm selects the best attribute and question to be asked about that attribute at each node. The selection is based on what attribute and question about it divide the
learning data so that it gives the best predictive value for
the classification. CART modeling is particularly useful
in the case of less researched languages like Indian lan-
guages, for which the most relevant features that affect
the various prosody events and the way they are inter-
related with each other are not studied in detail.

This paper is organised as follows. In Section 2, background for the work presented in this paper is given. In Section 3, need for new features for prosodic phrase boundary prediction and the definition of a new feature 
morpheme tag
d is given. In Section 4, details about the corpus that is used for the present study is given. In Section 5, stepwise construction of CART based phrasing model for analysis on contribution and relative important of various features, is described. In Section 6, objective evaluation of the phrasing model is presented.

2. Background

Research on Text-to-Speech conversion for Indian lan-
guages is a much younger enterprise in comparison with
Text-to-Speech research for English and other European
languages. The major obstacle for speech synthesis
research in Indian languages is, we neither have the
databases annotated with prosodic and linguistic informa-
tion nor the tools required to generate the appropriate lin-
guistic information (for example, the syntactic, morpho-
logical and lexical information) that is essential to predict
various prosody events from the text. Further, a Text-to-
Speech system, in general, is targeted for one particular
language. In India, there are 18 officially recognised lan-
guages each with its own set of dialects. It is very difficult
to have one speech synthesizer for each language (and for
each dialect!).

In our work, a multilingual Text-to-Speech system
[1] for Indian languages is being built within the Festival
framework [2]. As a starting point, a common multi-
linguial diphone database is prepared and linguistic/prosodic
processing modules are being developed for two Indian
languages Hindi and Telugu. The two languages are
chosen so as to represent one from each of the Aryan and
Dravidian family of languages. The languages Hindi and
Telugu are also the first and second largest spoken lan-
guages within India respectively.

This paper reports our work on prosodic phrase
boundary prediction for the Indian language Telugu. The
phrasing model developed in this paper has been imple-
mented in the Text-to-Speech synthesis system described
above.

3. New Feature - Morpheme tag

Traditionally syntactic and morphological information
has proven to be very useful in the prediction of prosodic
phrase boundaries [3, 4, 5, 6, 7, 8, 9, 10]. But for In-
dian languages, no complete morphological analyzers or
Part-of-Speech taggers are available. And also the syn-
tactic and morphological information and its relation to
the prediction of prosodic phrase boundaries is not stud-
ied in detail.

In an effort to identify linguistically meaningful fea-
tures that affect prosodic phrasing, a new feature, namely
morpheme tag, is defined. A set of 19 ‘morpheme tags’
are identified that occur at word boundaries(word endings).
Morpheme is a meaningful linguistic unit consisting
of a word or a word element that cannot be divided
into smaller meaningful parts. However, not all the iden-
tified ‘morpheme tags’ confirm to meaningful linguistic
units. They are proposed and established purely from en-
gineering viewpoint. In an analysis done on 10,022 word
Doordarshan news bulletin database [12], it is observed
that more than 50% of words have one of the identified
‘morpheme tags’. Table 1 gives a list of the identified
morpheme tags and their distribution in the 10,022 word
database. The next section details about the speech cor-
pus used for the phrase break analysis, various features
considered to derive from the text and generation of fea-
ture vectors from which the CART is trained.

<table>
<thead>
<tr>
<th>Morpheme name</th>
<th>Example word</th>
<th>Number of occurrences</th>
<th>Percentage %</th>
</tr>
</thead>
<tbody>
<tr>
<td>iO</td>
<td>dhE samiO</td>
<td>591</td>
<td>5.89</td>
</tr>
<tr>
<td>thO</td>
<td>pattudhalathO</td>
<td>147</td>
<td>1.46</td>
</tr>
<tr>
<td>Aru</td>
<td>annAru</td>
<td>484</td>
<td>4.83</td>
</tr>
<tr>
<td>ndhi</td>
<td>cheppindhi</td>
<td>144</td>
<td>1.43</td>
</tr>
<tr>
<td>ani</td>
<td>chEyAlani</td>
<td>337</td>
<td>3.36</td>
</tr>
<tr>
<td>lu</td>
<td>vi sEshAlu</td>
<td>467</td>
<td>4.66</td>
</tr>
<tr>
<td>nni</td>
<td>prabhuthvAnni</td>
<td>118</td>
<td>1.18</td>
</tr>
<tr>
<td>nna</td>
<td>chErukunna</td>
<td>138</td>
<td>1.37</td>
</tr>
<tr>
<td>Oni</td>
<td>rAshtramlOni</td>
<td>89</td>
<td>0.89</td>
</tr>
<tr>
<td>chi</td>
<td>nu.nchi</td>
<td>97</td>
<td>0.96</td>
</tr>
<tr>
<td>na</td>
<td>jarigina</td>
<td>482</td>
<td>4.81</td>
</tr>
<tr>
<td>ki</td>
<td>adhupulOkI</td>
<td>151</td>
<td>1.50</td>
</tr>
<tr>
<td>ini</td>
<td>purOgathini</td>
<td>37</td>
<td>0.37</td>
</tr>
<tr>
<td>gA</td>
<td>sandharbh.nG</td>
<td>211</td>
<td>2.10</td>
</tr>
<tr>
<td>ku</td>
<td>prAnthAlaku</td>
<td>284</td>
<td>2.83</td>
</tr>
<tr>
<td>nu</td>
<td>lakshyAlunu</td>
<td>246</td>
<td>2.45</td>
</tr>
<tr>
<td>pai</td>
<td>charyapai</td>
<td>77</td>
<td>0.77</td>
</tr>
<tr>
<td>la</td>
<td>charyala</td>
<td>423</td>
<td>4.22</td>
</tr>
<tr>
<td>.n</td>
<td>prabhuthv.n</td>
<td>598</td>
<td>5.96</td>
</tr>
</tbody>
</table>

Table 1: List of morpheme tags and their distribution in a
10,022 word database.

4. Corpus and feature vector generation

Corpus used for the study include 156 sentences of news
reading style, taken from Doordarshan news bulletin
database [12]. The corpus contained 1821 word bound-
aries, of which 447 are marked as breaks. The corpus
is divided randomly into training data (125 sentences, having 1418 word boundaries with 352 breaks) and test data (31 sentences, having 403 word boundaries and 95 breaks).

Considering the nature of the Text-to-Speech problem, only those features that can be automatically derived from text are considered. Each word in the corpus is annotated with the following features together with the break index information (B for break, NB for no break):

- Distance from the previous break (in terms of number of words).
- morpheme tag of the words included in a five word window centered at the word under study. If the word doesn’t have any of the identified morpheme tags, it is given the tag ‘NT’ which stands for No-Tag.
- Word length of the words included in a five word window centered at the word under study (in terms of the number of syllables).

5. Generation of CART phrasing model

A Classification and Regression Tree based data-driven phrasing model is generated using the feature data described in Section 4. Since there is no previous knowledge about the usefulness of the features and their relative importance, the CART is built in a step-wise fashion to establish the usefulness and relative importance of the features. In this approach each single feature is taken in turn and a tree consisting of nodes containing only the conditions imposed by that feature is built. The single best tree is then kept and each remaining feature is taken in turn and added to the tree to find the best tree possible with just two features. The procedure is then repeated for a third, fourth, fifth feature and so on. This process continues until no significant gain in accuracy is obtained by adding more features. For running the CART building process, ‘Wagon’ classification and regression tree tool [13] is used.

Each word in the utterance is given input to the phrasing model to check whether it is at the end of a prosodic phrase. A decision is made by traversing the tree starting from the root node, taking various paths satisfying the conditions at intermediate nodes, till the leaf node is reached. The path taken depends on various features like, morpheme tag of the word, length of the word in terms of the number of syllables, distance from previous phrase break etc. The leaf node contains the predictive value for the decision. An example partial decision tree for phrase break prediction is shown in Figure 1. Objective evaluation of the phrasing model and detailed analysis on the usefulness and relative contribution of the proposed features is given in Section 6.

![Figure 1: An example partial decision tree (CART) for phrase break prediction. The triangles depict omitted parts.](image)

6. Evaluation and discussion

Evaluation of a phrasing model is very difficult as some of the errors might reflect the freedom of choice in placing prosodic phrase boundaries while other parts of the error are unacceptable errors. Two different types of errors can occur when predicting the placement of prosodic phrase boundaries: Insertion of a boundary when it is not present in the reference data, and deletion of a boundary present in the reference data. A missing phrase boundary makes the speech sound rushed and is not as bad as an extra phrase boundary, which can be distracting and confusing [14].

<table>
<thead>
<tr>
<th></th>
<th>Predicted Non-Breaks</th>
<th>Predicted Breaks</th>
<th>Percentage Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Breaks</td>
<td>274</td>
<td>34</td>
<td>88.96</td>
</tr>
<tr>
<td>Breaks</td>
<td>26</td>
<td>69</td>
<td>72.63</td>
</tr>
</tbody>
</table>

Table 2: Performance of the phrasing model.

For objective evaluation, the phrasing model is trained with training data (125 sentences, having 1418 word boundaries with 352 breaks) and evaluated with test data (31 sentences, having 403 word boundaries and 95 breaks). Table 2 shows the evaluation results of the phrasing model. In a total of 308 non-breaks, 274 are identified correctly giving an 88.96% accuracy while making 10.04% of false insertions. In a total of 95 breaks, 69 are identified correctly giving a 72.63% accuracy while making 23.37% of false deletions. The results obtained are inferior but comparable with the phrasing models for the much researched languages like English.

To assess the effectiveness of the features considered CART’s are built in a step-wise fashion as described in
Section 5. and the results are shown in Table 3. The First column gives the feature names, and the second column gives the correlation obtained between actual and predicted decisions (Break or No-Break) by the addition of the successive features in the CART modeling process. It can be seen that the most important feature contributing to the better performance (or correlation) is the *morpheme tag* information which justifies the use of *morpheme tag* as a feature for the prediction of phrase breaks. The other important feature is the word length (in terms of number of syllables).

<table>
<thead>
<tr>
<th>Feature used</th>
<th>Correlation between actual and predicted breaks (cumulative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word-MorphemeTag</td>
<td>0.7426</td>
</tr>
<tr>
<td>Next.Word-NumSyls</td>
<td>0.8015</td>
</tr>
<tr>
<td>Prev.Word-MorphemeTag</td>
<td>0.8235</td>
</tr>
<tr>
<td>Prev.Prev.Word-NumSyls</td>
<td>0.8309</td>
</tr>
<tr>
<td>Word-NumSyls</td>
<td>0.8456</td>
</tr>
</tbody>
</table>

Table 3: Analysis on usefulness of features in phrase break prediction.

7. Conclusions

Prosodic phrasing is an important and more difficult a problem for Indian languages, as the Indian language scripts use very little or no punctuation. In an effort to identify meaningful features that affect prosodic phrasing in the Indian language Telugu, a new feature, namely *morpheme tag*, is defined. A Classification and Regression Tree based data-driven phrasing model is developed for phrase break prediction and the usefulness of the *morpheme tag* feature is further demonstrated in an evaluation process. The same procedure could be extended to other inflective languages. Objective evaluation of the new phrasing model has given inferior but comparable results with the phrasing models for the much researched languages like English.

8. Criticism

- The feature *morpheme tag* is not linguistically motivated, instead, it is proposed and established purely from engineering viewpoint.
- Modeling and analysis is done on smaller data set.

9. References