

Prosodic Pause Prediction Using Reduced Tag Set and Sub Word Units

by

E. Veera Raghavendra (200650023)

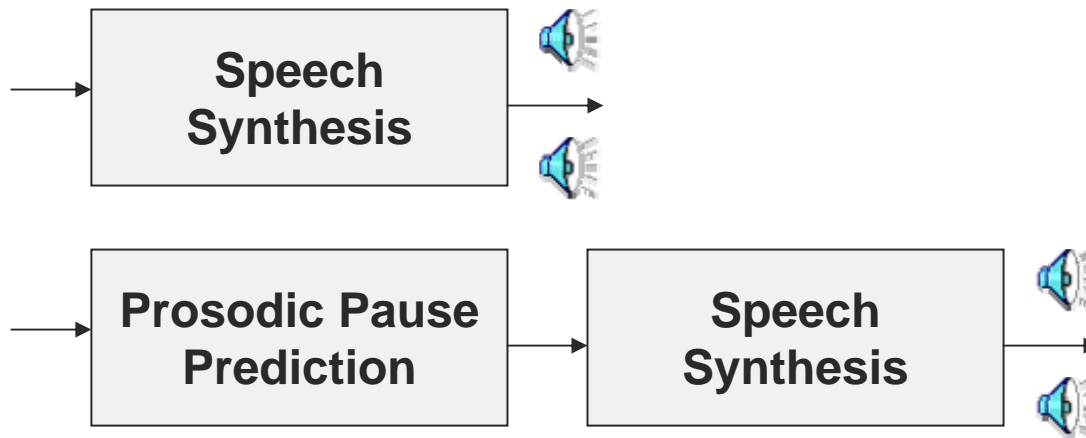
Srinivas Desai (200650024)

[Overview]

- What is Prosody?
- Why Reduced Tag Set?
- Reason for Sub Words usage.
- Approach Used.
- Results

[Demo]

- Hindi:
unkei paasa, raajdhanaanii mein: doo phle-t: hain: .
- Telugu:
amitaab, jayaabachchan, vaari pillalu abhisheik, shveita,
amitaab kut:un:ba mitrud:u, amar sin:g an:daruu kalisi kaashi
vachchaaru



[Prosody]

Prosody is a term used to refer to speech elements as below.

- Pause
For example [pau] speak this sentence
- Pitch
The highness or lowness of a tone
- Duration
Slow speaking vs. Fast speaking
- Loudness
Relative energy and volume.

For our project we have worked on the element – Pause.

[Reduced Tags]

- Parts of Speech enhances the performance of pause predictions, the results of which are shown later.
- 26 Tags are used at LTRC, IIIT.
- To determine the pause in a sentence we have reduced the number of tags to 10.
- The reduced tags are shown in the next slide.

[Reduced Tags Mapping]

	<u>Original Tags</u>	<u>Reduced Tags</u>
■	NN	NN
■	NNP	NN
■	PRP	NN
■	NLOC	NN
■	NNC	NN
■	NNPC	NN
■	NVB	NN
■	VFM	VRB
■	VAUX	VRB
■	VJJ	VRB
■	VJJN	VRB
■	VRB	VRB
■	VRBN	VRB
■	VNN	VRB
■	JVB	ADJ
■	JJ	ADJ
■	QF	ADJ
■	INTF	ADJ
■	RB	ADV
■	RBVB	ADV
■	NEG	NEG
■	QW	QW
■	PREP	PREP
■	RP	CC
■	CC	CC
■	QFNUM	QFNUM
■	QFN	QFNUM
■	SYM	SYM
■	UH	UH

[What is Syllable ?]

- A unit of pronunciation having one vowel sound, with or without surrounding consonants, and forming all or part of a word. (C*VC*)
- C is a consonant.
- V is a vowel.
- Examples: V, CV, CCV, CVC, CCCV.

Hindi

Maataa – maa/taa
pitaa – pi/taa

Telugu

amma – a/mma
naanna – naa/nna

Why do we need this?

- Pause in a sentence is required to bring naturalness in a synthesized speech.
- A sentence without a pause will make it difficult to understand the meaning of the sentence.
- A pause in a sentence is based on the language syntax and also on the lexical structure of the sentence.
- A lexical structure approach uses information such as syllable sequence to predict the pauses.
- Language based approach predicts the pauses on the basis of parts of speech.

[Different Approaches]

- Lexical lookup

The is a very old technique used, which needs a large database to cover all the possible words and, assign pauses manually.

- Rule Based

The models cannot be generalized, since syntax of language varies vastly from one language to another.

- Statistical methods

- Classification and Regression Tree – CART
- Artificial Neural Network – ANN
- Conditional Random Fields – CRF

Unit Selection

- We experiment the use of sub-words such as syllables as units for POS tagging.
- The motivation for the sub word approach is from the fact that the POS can be determined from the syllable structure.
- The sequence of syllables is very important.
- Example:
 - Hindi: Ka/raa/naa - VRB
 - Ka/raa/nei - VRB

 - Telugu: Ve/l:l:aa/li - VRB
 - Ve/l:l:aa/nu - VRB
- If we use the words as a unit, we will need a larger database when compared to the database with sub-words as units.

[Model Used for Training]

- As part of tools for statistical modeling we use Classification and Regression Tree for the training and testing.
- CART is powerful because it can deal with incomplete data, multiple types of features and the trees it produces, often contain rules which are humanly readable.
- Decision trees contain a binary question (yes/no answer) about some feature at each node in the tree. The leaves of the tree contain the best prediction based on the training data.
- The basic CART building algorithm is a *greedy algorithm* in that it chooses the locally best discriminatory feature at each stage in the process.

Features Used

POS Features:

- F1: <predictee> <Prev first 2> <Current first 2> <Next first 2>
- F2: <predictee> <Prev last 2> <Current first 2> <Current last 2> <Next first 2>
- F3: <predictee> <Prev last 2> <Current first 2> <Current last 1> <Next first 2>
- F4: <predictee> <Prev first 2> <Current first 2> <Current last 2> <Next first 2>
- F5: <predictee> <Prev last 2> <Current first 1> <Current last 2> <Next first 2>

Pause Features:

- F1: <predictee>< Prev first 2 > <Next first 2>
- F2: <predictee><Prev Prev-tag> <Prev-tag> <Next-tag> <Next Next tag>
- F3: <predictee><Prev first 2><Next first 2><Prev-Prev-tag><Prev-tag> <Next-tag> <Next-Next-tag> <Next-tag>
- F4: <predictee><Prev first 2><Next first 2><Prev-tag><Next-tag>
- F5: <predictee><Prev-tag><Next-tag>
- F6: <predictee><Prev first 2><Next first 2><Prev-Prev-tag><Prev-tag> <Next-tag> <Next-Next-tag>
- F7: <predictee><prev first 2><Next first 2><Prev-tag><Next-tag><Next Next-tag> <Next-tag>

[Results (Hindi)]

POS

Feature	Performance
F1	82.859
F2	83.415
F3	84.316
F4	83.906
F5	83.128

Pause

Feature	Performance
F1	87.190
F2	87.574
F3	88.941
F4	89.581
F5	88.129
F6	89.026
F7	88.684

[Results (Telugu)]

POS

Feature	Performance
F1	84.138
F2	88.967
F3	87.586
F4	88.276
F5	88.210

Pause

Feature	Performance
F1	66.159
F2	71.356
F3	71.482
F4	71.482
F5	71.482
F6	70.722
F7	71.863

[Issues]

- Need to improve POS tagging, and Pause prediction performance
- Need to explore more features.
- Typographical mistakes.
- Improper Data.



ThankYou