



Pronunciation Lexicon in W3C Framework

Dr. Shyamal Kumar Das Mandal

Assistant Professor

sdasmandal@cet.iitkgp.ernet.in

Indian Institute of Technology, Kharagpur

5th May 2011

WORKSHOP ON PRONUNCIATION LEXICONS FOR INDIC LANGUAGES

Jawaharlal Nehru University



What is Pronunciation Lexicon?

Representation of Pronunciation information of the Lexical items along with its Grapheme Representation

Why Pronunciation Lexicon ?

It is required for the development of Speech technology such as Text to Speech Synthesis and Automatic Speech Recognition



What is PLS of W3C?

The Pronunciation Lexicon Specification (PLS) is designed to enable interoperable specification of pronunciation information for both **ASR** and **TTS** engines within voice browsing applications

How it is used in TTS and ASR?

The PLS is the standard format of the documents referenced by the **<lexicon>** element of **SSML**. The PLS engine will load the external PLS document and transparently apply the pronunciations during the processing of the **SSML** document. An application may contain several distinct PLS documents to be used in different points of the application.

If a **pronunciation lexicon** is referenced by a **SRGS** grammar it can allow multiple pronunciations of the word in the grammar to accommodate different speaking styles

Present Pronunciation Lexicon Markup Language Definition



Elements	Attributes	Description
<u><lexicon></u>	version xml:base xmlns xml:lang alphabet	root element for PLS
<u><meta></u>	name http-equiv content	meta data container element
<u><metadata></u>		meta data container element
<u><lexeme></u>	xml:id	the container element for a single lexical entry
<u><grapheme></u>	orthography	contains orthographic information for a lexeme
<u><phoneme></u>	prefer alphabet	contains pronunciation information for a lexeme
<u><alias></u>	prefer	contains acronym expansions and words' substitutions
<u><example></u>		contains an example of the usage for a lexeme



```
<?xml version="1.0" encoding="UTF-8" ?>
<!-- This pronunciation lexicon is licensed under the GPL. -->
- <lexicon version="1.0" xmlns="http://www.w3.org/2005/01/pronunciation-lexicon"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.w3.org/2005/01/pronunciation-lexicon
  http://www.w3.org/TR/2007/CR-pronunciation-lexicon-20071212/pls.xsd" alphabet="ipa"
  xml:lang="de">
- <lexeme>
  <grapheme>আমি</grapheme>
  <phoneme>emi</phoneme>
</lexeme>
- <lexeme>
  <grapheme>ঘড়া</grapheme>
  <phoneme>gʱeɽe</phoneme>
</lexeme>
- <lexeme>
  <grapheme>দিত্তি</grapheme>
  <phoneme>ditijo</phoneme>
</lexeme>
- <lexeme>
  <grapheme>কেট</grapheme>
  <phoneme>keu</phoneme>
</lexeme>
</lexicon>
```

Multiple pronunciations for the Same Orthography



Problem no.1 → Homographs

Homographs are words with the **Same Orthography** and **Different Meanings** and **Different Pronunciations**.

Solution under the existing PLS specification

- ❖ Using “Role” attribute under the Lexeme element
- ❖ Using “Prefer” attribute under the Phoneme element

This solution is erroneous



Proposed Solution

Proposal -I: (POS as an attribute)

The "*pos*" (parts of speech) can be an optional attribute under the phoneme element which indicates the proper pronunciation

Advantage:

- ❖ **Reduction of Lexeme numbers.**
- ❖ **Removal of Disambiguity**

Example



```
<?xml version="1.0" encoding="UTF-8"?>
<lexicon version="1.0" xmlns="http://www.w3.org/2005/01/pronunciation-lexicon"
  alphabet="ipa" xml:lang="bn">
  <lexeme>
    <grapheme>সরল</grapheme>
    <phoneme pos= "adjective"> ʃɔrol </phoneme>
    <!-- IPA string is: " ʃɔrol" -->
    <!--Itrans is: "sarala" -->
    <!--Meaning is : "easy" -- >
    <phoneme pos= "verb" > ʃorlo </phoneme>
    <!-- IPA string is: " ʃorlo" -->
    <!--Itrans is: "sarala" -->
    <!--Meaning is: "moved" -- >
    <phoneme pos= "null"> ʃɔrol </phoneme>
    <!-- IPA string is: " ʃɔrol" -->
    <!--Itrans is: "sarala" -->
    <!--Meaning is: " easy" -- >
  </lexeme>
</lexicon>
```

Proposal -II: (Pos as an element)

The <lexeme> element may contain optionally one or more <pos> element. Each <pos> element contains the pronunciation of the word depending on pos element information.

Advantage:

- ❖ **Reduction of Lexeme numbers.**
- ❖ **Removal of Disambiguity**

Example



```
<?xml version="1.0" encoding="UTF-8"?>
<lexicon version="1.0" xmlns="http://www.w3.org/2005/01/pronunciation-lexicon"
  alphabet="ipa" xml:lang="bn">
  <lexeme>
    <grapheme>সরল</grapheme>
    <pos> adjective </pos>
    <phoneme> ʃɔrol </phoneme>
    <!-- IPA string is: " ʃɔrol" -->
    <!--Itrans is: "sarala" -->
    <!--Meaning is : "easy" -- >
    <pos> verb </pos>
    <phoneme> ʃorlo </phoneme>
    <!-- IPA string is: " ʃorlo" -->
    <!--Itrans is: "sarala" -->
    <!--Meaning is: "moved" -- >
    <pos> null </pos>
    <phoneme> ʃɔrol </phoneme>
    <!-- IPA string is: " ʃɔrol" -->
    <!--Itrans is: "sarala" -->
    <!--Meaning is: "easy" -- >
  </lexeme>
</lexicon>
```

Problem no. 2 :

Ideolectic variation of the same orthographic information

Proposed Solution

```
<?xml version="1.0" encoding="UTF-8"?>
<lexicon version="1.0" xmlns=http://www.w3.org/2005/01/pronunciation-lexicon
alphabet="ipa" xml:lang="bn">
  <lexeme>
    <grapheme>উনত্রিশ</grapheme>
    <phoneme prefer="true">untriʃ</phoneme>
    <!-- IPA string is: "untriʃ" -->
    <phoneme>unotiriʃ</phoneme>
    <!-- IPA string is: "unotiriʃ" -->
  </lexeme>
</lexicon>
```

Multiple orthographic representation with same pronunciation



Homophones

Homophones are words with **Different Spellings** and **Different Meanings** but the **Same Pronunciation**

```
<?xml version="1.0" encoding="UTF-8"?>
<lexicon version="1.0"
xmlns="http://www.w3.org/2005/01/pronunciation-lexicon"
  alphabet="ipa" xml:lang="bn">
  <lexeme>
    <grapheme>কুল</grapheme>
    <phoneme>kul</phoneme>
  </lexeme>
  <lexeme>
    <grapheme>কুল</grapheme>
    <phoneme>kul</phoneme>
  </lexeme>
</lexicon>
```



Need of Morphological Information

In some of the languages like Bengali, English not only **POS** information but also **Morphological** information especially in case of verb **finiteness** and **honorificity** information are very crucial in determining the pronunciation of a homograph.

For Example:

Bengali Verb “kare” has two pronunciations depending on its finiteness

Similarly Bengali verb “dhara” has two pronunciations depending upon its honorificity

English verb “read” has two pronunciations depending on its tense information

POS.s1.s2.s3.s4.s5.s6.s7.s8

Example



```
<?xml version="1.0" encoding="UTF-8"?>
<lexicon version="1.0" xmlns="http://www.w3.org/2005/01/pronunciation-
lexicon"
  alphabet="ipa" xml:lang="bn">
  <lexeme>
    <grapheme>করে</grapheme>
    <phoneme: pos= "VM.s1.s2.s3.s4.fin.s6.s7.s8" >kɔre</phoneme>
    <!-- IPA string is: "kɔre" -->
    <!--Itrans is: "kare" -->
    <!--Meaning is: "do/does" -- >
    <phoneme: pos= "VM.s1.s2.s3.s4.nfin.s6.s7.s8" >kore</phoneme>
    <!-- IPA string is: "kore" -->
    <!--Itrans is: "kare" -->
    <!--Meaning is: "having done" -- >
  </lexeme>
</lexicon>
```

POS → Part of Speech Marker; s1 → person marker; s2 → tense marker; s3 → aspect marker; s4 → mood marker; s5 → finite/nonfinite; s6 → emphatic/non emphatic; s7 → negative/ non negative; s8 → honorific/ non honorific

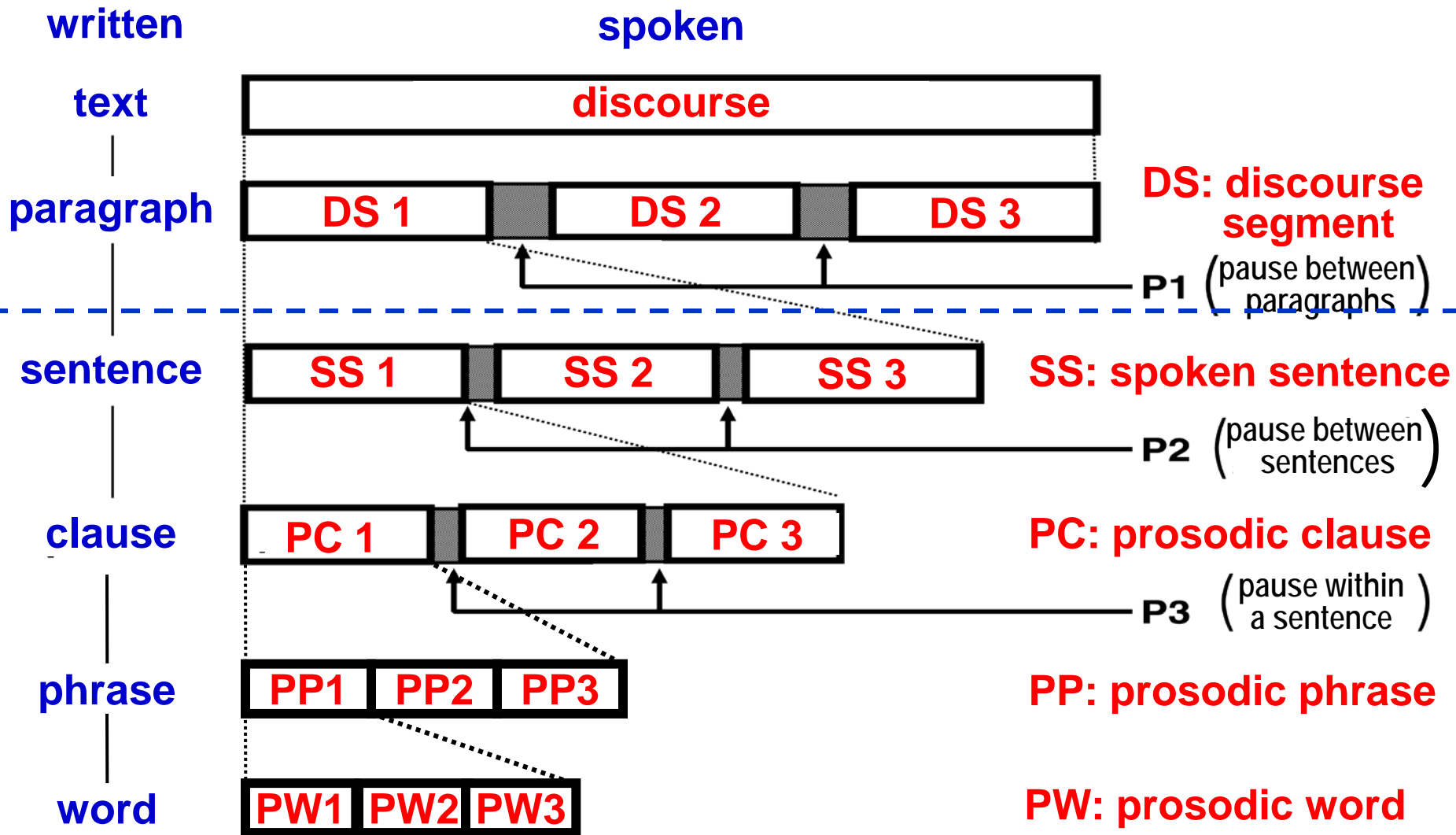


Morphological Analysis for reducing the size of the Lexicon

Morphological Analysis is also very helpful in reducing the size of the Lexicon. Most of the dictionary used a standard format for describing the morphological information of a particular word.

This will reduce the multiple entry of the same word with different suffixes and word compounding.

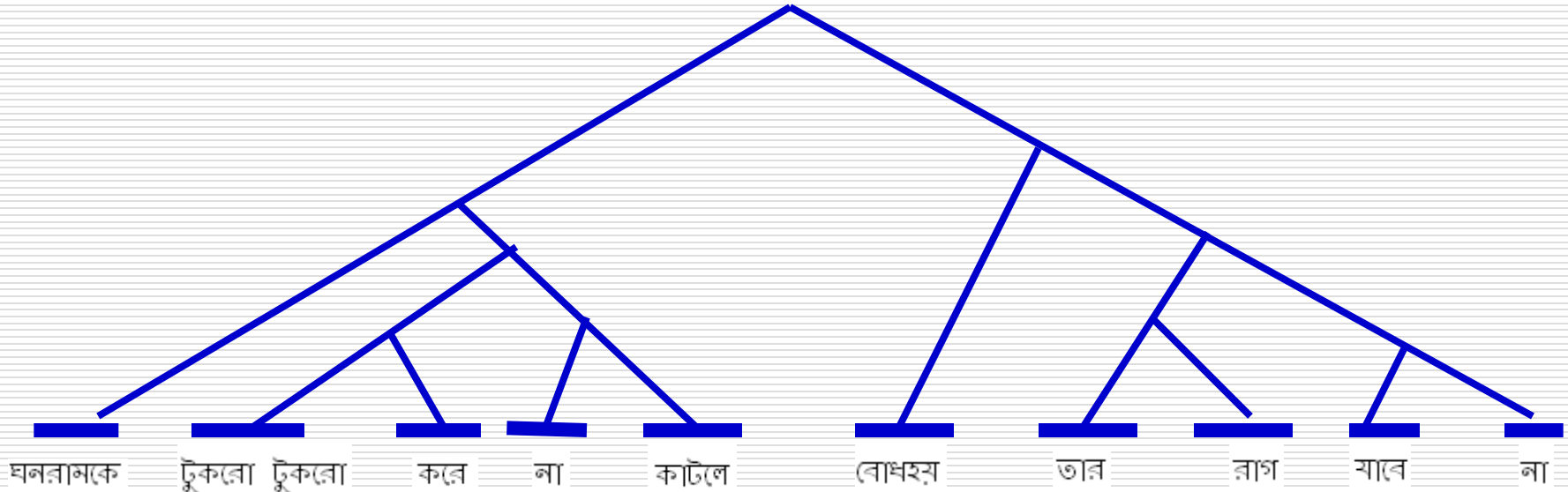
Prosodic Structure of Bangla Text Reading



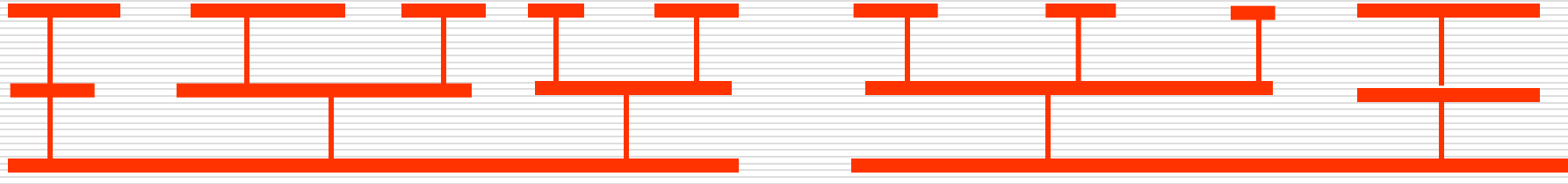
An Example of Disparity Between the Syntactic Structure and the Prosodic Structure



SYNTACTIC STRUCTURE S

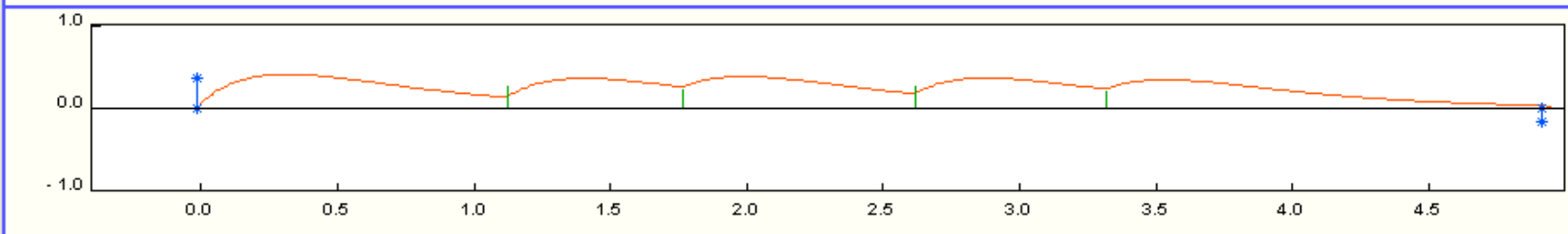
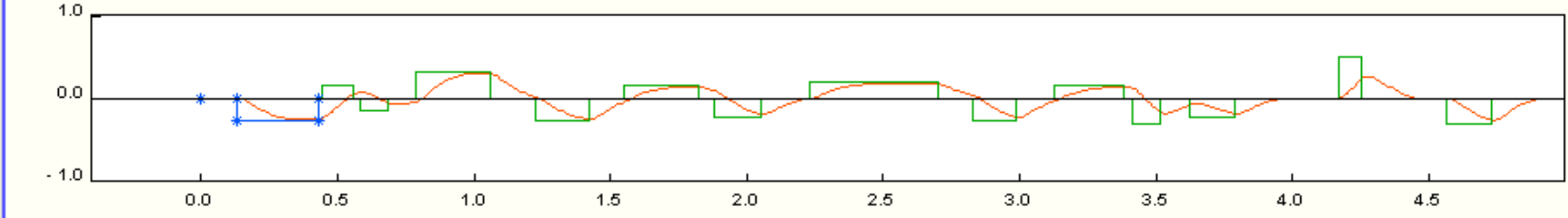
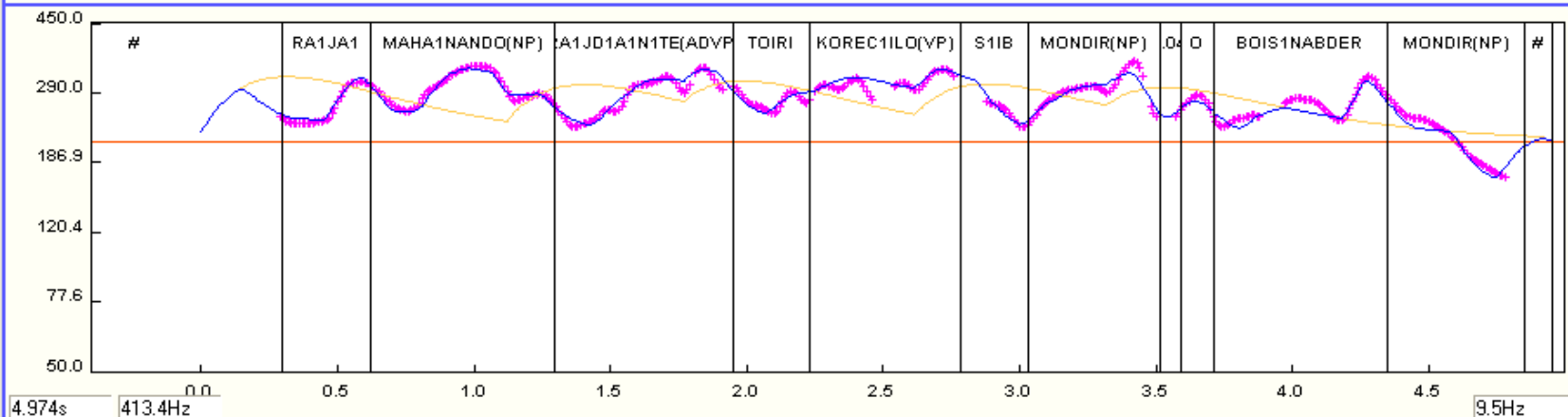
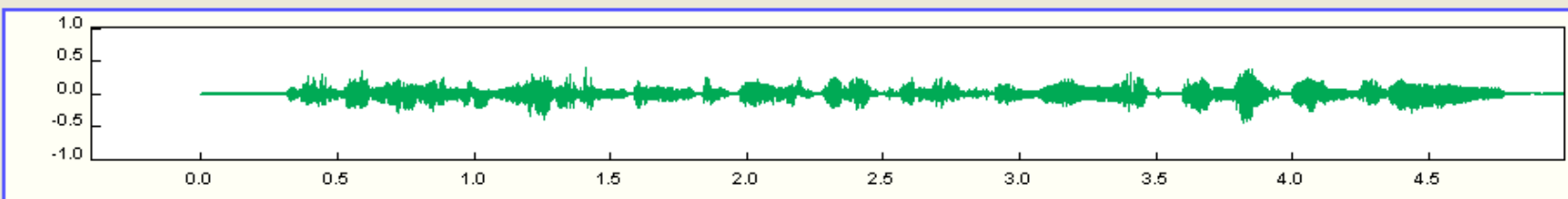


PW
PP
PC



PROSODIC STRUCTURE

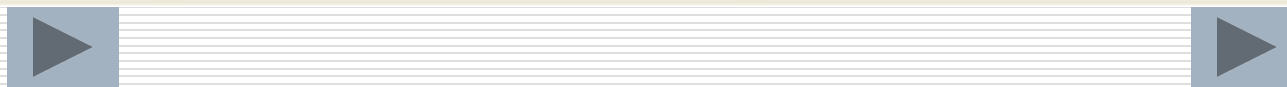
PW: Prosodic Word
PP : Prosodic Phrase
PC : Prosodic Clause



Open Wav
Save Wav
Open PAC
Save PAC
Auto Pac
Fb
214.0532
ReSynthesis

Delete
< >
Delete
< >

Tone command: Aa -0.2857 T1 0.1342 T2 0.4331
Phrase command: Ap 0.3673 T0 -0.0184 T0e 0
Parameters: alpha 3 beta 20 gamma 0.9
Exit



Pitch contour

```
<?xml version="1.0"?> <speak version="1.1"
xmlns="http://www.w3.org/2001/10/synthesis"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.w3.org/2001/10/synthesis
http://www.w3.org/TR/speech-synthesis11/synthesis.xsd"
xml:lang="en-US">
<prosody contour="(0%,+20Hz) (10%,+30%) (40%,+10Hz)">
good morning
</prosody>
</speak>
```

Emphasis Element

```
<?xml version="1.0"?> <speech version="1.1"
xmlns="http://www.w3.org/2001/10/synthesis"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.w3.org/2001/10/synthesis
http://www.w3.org/TR/speech-synthesis11/synthesis.xsd"
xml:lang="en-US">
```

That is a <emphasis> big </emphasis> car!

That is a <emphasis level="strong"> huge </emphasis>
bank account! </speech>

धन्यवाद

