Progress Report of TTS project during the period April 2012 to December 2012
IIT Guwahati

1. Title:
Development of Text to speech synthesis systems in Assamese and Manipuri Languages.

2. Work Mandated:
1) Identification of text for children story telling application and news bulletin in Assamese language.
2) Completion of recording for children story and news bulletin for 10 hrs (male 5 hrs and female 5 hrs) and recording of English Arctic database for 1132 Sentences for building bilingual voice.
3) Transliteration of the recorded data of 10 hrs.
4) Labeling of the transliterated data.
5) Building of the Text to speech Synthesis systems for Assamese and Manipuri Language using Festival and HTS synthesis, including the small footprint of TTS ported to Android based systems using HTS synthesis.
6) Integration of TTS with NVDA (a Windows based screen reader) and ORCA (a Linux based screen reader)

3. Work Done:

1) Text Identification:
   • For Assamese story telling application we have identified stories from the book “Shoadloga shadhu” and “Buhri aair shadhu”.
   • Text is chosen such that it gives new syllables as compared to the already existing syllables.

2) Recording: Recording is done in studio environment at 48000 Hz sampling rate using mono channel. Recording is done for
   • Assamese: Female voice (Children story telling, 85 mins, 625 sentences)
     Male voice (Newsbulletin 1180 sentences 3-1/2 hours)
   • English: Female voice (English Arctic Database -1132 sentences)

3) Transliteration: After the completion of the recording process the data obtained is to be transcribed. We have done the Transliteration of the recorded data for Assamese Language by using IT3 chart. During this process the signal is carefully listened and looked into so as to minimize transcription error as much as possible. Transliteration is done for
   • 625 sentences for Assamese story telling
   • 1180 sentences for Assamese News bulletin

4) Chipping:
   Chipping of the data into smaller parts has been done comparable to the length of a general sentence and the junk part is removed. Chipping is done using wavesurfer removing the junk part from the wav files and resampled them to 16 kHz as this is the frequency of wavfile labeling tool accepts.
Chipping done for

- Approx 85 mins for Assamese story telling.
- Approx 3-1/2 hours for Assamese news bulletin.

5) Parser script for syllablification:

- A Parser script which generates a syllabic word structure for labeling tool and CAR Tree building is developed
- Basic phones used in the script are different compared to other languages since the pronunciation of /tx/, /txh/, /dx/, /dxh/ is similar to /t/, /th/, /d/, /dh/
- Script is made robust by making it compatible for the english words written in assamese graphemes.
- Script accepts text in IT3, so syllabification for both labeling and CAR Tree building is done in IT3

6) Labelling:

Assamese Labelling is done using DON Label for below given data

- 236 sentences for Assamese Story telling with 538 unique syllables.
- 500 sentences for Assamese News bulletin with 1000 unique syllables.

1132 sentences of English data is labeled using EHMM (automatic labeler) with phoneme word structure.

7) Voice building:

Voice building is done for 236 sentence of Story telling (Female voice) and 300 sentences of News bulletin (Male voice). Created the parser script, unique syllables and phoneset while building the CAR tree

1. Installations done:
   1. Installation of “Festival” package.
   2. Installation of “speech_tools”.
   3. Installation of “festvox”.

During testing the expected input is in UTF8. So Data is converted from UTF8 to IT3. A generic script which is developed for all indian language UTF8 to IT3 conversion is used. Since the Bengali alphabets are mostly closer to IT3, we utilised Bengali phoneset and made compatible to Assamese. The differences are showed in Table 1.

<table>
<thead>
<tr>
<th>IT3</th>
<th>Bengali</th>
<th>Assamese</th>
</tr>
</thead>
<tbody>
<tr>
<td>ra/ri/ru/r</td>
<td>र/रि/रू/र</td>
<td>र/रि/रू/र</td>
</tr>
<tr>
<td>va/vi/vu/v</td>
<td>व/वि/वू/व</td>
<td>व/वि/वू/व</td>
</tr>
</tbody>
</table>
9) Bilingual TTS system: Single TTS system which can synthesise text having both Assamese (UTF8) and English is developed. Basic differences for Assamese and English are shown below.

<table>
<thead>
<tr>
<th>Basic Unit</th>
<th>Assamese</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Structure</td>
<td>Parser Script</td>
<td>CMU Dictionary</td>
</tr>
<tr>
<td>Labeling</td>
<td>DON Label(Semi automatic)</td>
<td>EHMM Labeling</td>
</tr>
</tbody>
</table>

Assamese system is built by the normal procedure. English is accompanied with the basic system by building CAR Tree for basic phonemes using 1132 English sentences. During testing if the word is in UTF8 it is converted to IT3 and passed to parser script to get the word structure in syllables and if it is an English word, phoneme word structure is extracted using the CMU Dictionary. So a complete bilingual system is developed and the quality is satisfactory for Assamese and English.

10) Modification of Transliteration tool is done to suit Assamese language.

11) HTS voice is developed for 1 hour of Assamese News bulletin.

12) Issues faced:
- Parser script is unable to handle some English words which are written in Assamese language. Eg: card (kaard).
- During testing if a new syllable comes, system fails to synthesize.
- For English if the word is not in dictionary it cannot get the phoneme structure and there by synthesis fails.