CS630: Speech Technology

LAB-1: Speech Signal to Symbol Transformation

OBJECTIVE:

Manually performing speech signal to symbol transformation

SEQUENCE OF STEPS:

(a) Collect speech files for one of the Indian languages.

(b) Display a speech file using the utility "wavesurfer".

(c) Transcribe it in ITRANS code.

(d) Convert your transcription into syllable-like units.

(e) Label the speech file with syllable-like units using "wavesurfer".

(f) Repeat the same procedure for other speech files.

(g) Write a brief report on your observation.

1 Displaying the speech data

Display the given speech data file using the utility "wavesurfer".
wavesurfer samplefile.wav

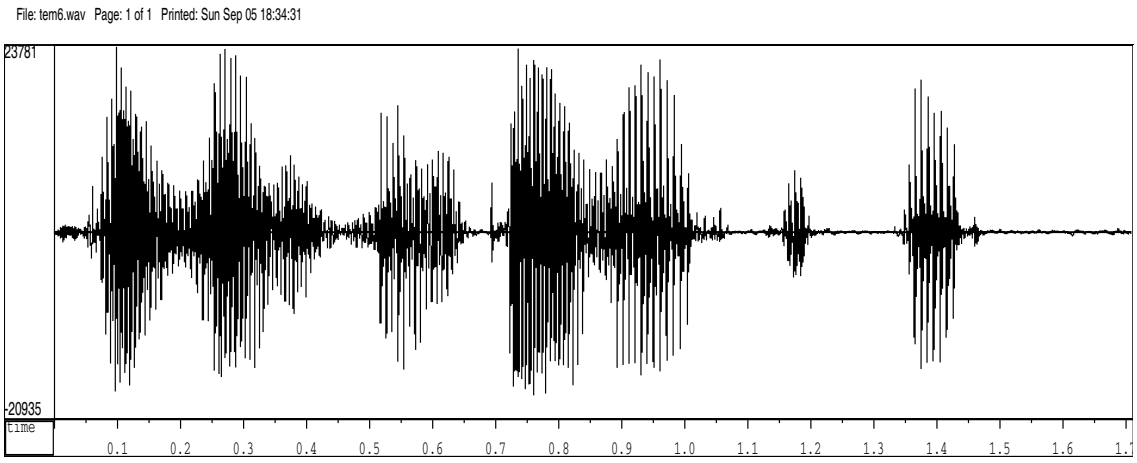

Figure 1: Speech waveform for the utterance "Sri M. Venkaiahnayudu chepperu"

2 Transcription into known language

Play the speech file using "wavesurfer"

By listening to the speech, write the utterance into the known (native) language.

3 Transcription into English using ITRANS code tables

Transcribe the utterance into English using ITRANS code tables.
ITRANS code: sri em ve.nkayyanayudu chepperu
(ITRANS code is the common transliteration code for Indian languages. Using these ITRANS code tables, any Indian language script can be transcribed into English. For each of the Indian language a separate ITRANS code table is used to transcribe the script from the given language into English)

4 Deriving the syllable-like units from English transcription

Split the utterance (text in ITRANS code) into syllable-like units. That is the ITRANS code text representation of the utterance is divided into subword units such as syllable-like units (text to symbol transformation, i.e., sentence to subword units)

Here syllable-like units are the symbols corresponds to the segments of the speech signal.

Reasons for choosing the syllable-like units as symbols against to phonemes (consonants or vowels (C or V)):

(1) It is very difficult to segment the speech data into phonemes.

(2) Due to coarticulation the effect of consonants is observed in the adjacent vowels.

(3) A character in an Indian language scripts is close to a syllable, and is typically one of the forms: V, CV, CCV, CVC, CCVC and CVCC, where C is a consonant and V is a vowel.

Utterance in ITRANS code: sri ve.nkayyanayudu chepperu
Syllable-like units of the utterance : sri em ve.n ka yya na yu du che ppe ru

5 Identifying the boundaries of the syllable-like units in speech waveform

Manually mark the boundaries of the syllable-like units by listening to the speech file segment by segment. The marked syllable boundaries and the associated waveform are shown in Figure 2. In the derived syllable boundaries some of the syllables present in the transcription are missing and some new syllables are marked. The expected syllable boundaries for a given speech file as per the transcription are shown in Figure 3.

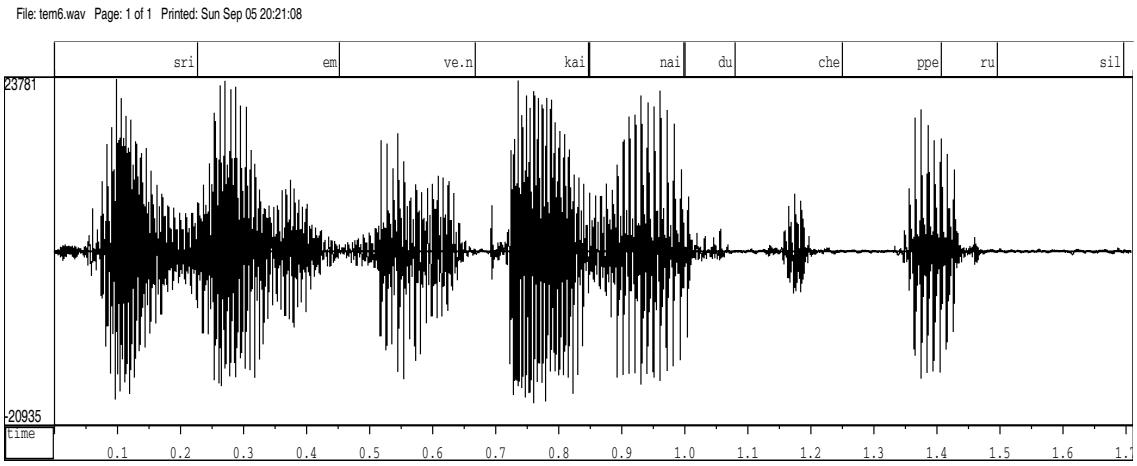

Figure 2: Speech waveform with the labels marked by listening to the segments of speech waveform

6 Observation

(1) While listening to the entire sentence, we can easily make out the message.

(2) Even the word boundaries can also be marked easily most of the times.

(3) While marking the boundaries of the syllables in a polysyllabic word a great difficulty is observed.

(4) Due to coarticulation effect, it is difficult to listen the particular syllable without the influence of the adjacent syllables in a polysyllabic word.

(5) In particular marking the boundaries of a syllable with semivowels found to be difficult.

(6) While listening to the segments of the speech data file, we are not able to perceive (completely missing) some syllables. But at the same time we can able to perceive a word clearly, where the syllables (missing) are present.

This observation shows the complexity involved in speech signal to symbol transformation. Even for human beings endowed with the natural speech

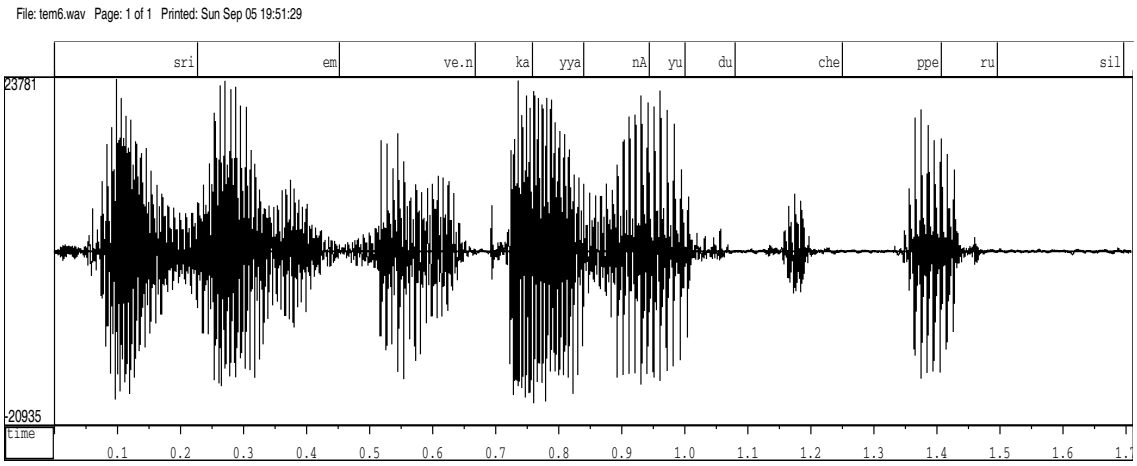

Figure 3: Speech waveform with expected labels

production and perception mechanism, found to be difficult to segment (label) the speech data into syllable-like units.