LDCIL – Speech Sub-group
Minutes of Meeting held on August 30 2012, CIIL, Mysore

1. L Ramamoorthy, CIIL, Mysore

2. Hema A Murthy, IIT Madras

3. K Samudra Vijaya, TIFR Mumbai

4. A G Ramakrishnan, IISc Bangalore

5. Arundhati Sengupta, CIIL Mysore

6. M Venkatesh, CIIL Mysore

7. Richa, CIIL Mysore

8. Atreyee Sharma, CIIL, Mysore

The meeting began with Dr. Ramamoorthy welcoming all the members.

1. The status of the current work was presented. Arundhati gave us the details of the work done:

   - Nineteen languages a minimum of 85 hours of data has been collected.
   - Further, six hours of each language have been labeled at the sentence level.
   - The guidelines used for labeling were as per the discussions at previous meetings.
   - The labels were verified by linguists from all the languages.

2. There are requests from CDAC, Pune to label the data at syllable and phoneme levels.

3. Hema A Murthy said that this can be done by building a bootstrapped speech recognition system and then using the back trace to generate the labels.

4. A person is to be deputed from IIT Madras to give a training on the HTK toolkit to the CIIL staff.

5. After a discussion with the staff, it was felt that another 28 hours of data can be labeled at the sentence level. Text will be corrected (wherever required) to match with the spoken utterance.

6. The next item on the agenda was to propose the collection of "telephonic corpora," for large vocabulary independent speech recognition.

   - This was discussed at length.

- It was felt unless there are targetted goals for collecting the data, it may not serve much purpose.

- In this context, it was felt that both DIT, MHRD should have a discussion and collect relevant data jointly.
  - DIT suggests various applications that can be developed in the context of the Indian market.
  - It becomes the responsibility of CIIL to collect the data, label them and make them available for research and development.
  - It was felt that a meeting between TDIL, DIT and CIIL should be conducted as soon as possible to decide the modus operandi.