TDIL Roadmap for Speech Technology Development in India

Speech technology is the field of *Interactive Technologies*. There is ongoing *shift from Speech component* research *to* research on integrated *Speech Systems*. Together with Speech, are the modalities that constitute full natural human - human communication (e.g.. Gesture, lip movements, facial expression, gaze, bodily posture) leading towards multi-modal interactive systems

TDIL Roadmap:

2007: Basic Technologies:

- *Speech corpora* for all Indian Languages annotated with phoneme levels, Stress emphasis tags & phrase break tags(10Mn Words)
- Software to compute *prosodic pattern* (phoneme duration & pitch contour), to retrieve acoustics units (diphones)
- Design of *CSR* (continuous Speech Recognition) engine, Speech Synthesis engine, Data sets (Acoustics & language models, vocabularies). Language model requires text data of over 50 Million words.
- Translation Memory based Translation model
- CSR in mobile devices (Voc: 10,000w) in Hindi
- High Quality (with concatenated format) speech synthesis
- Generally usable cross-lingual text retrieval
- Morph analysis, disambiguation of homographs, letters to phonemes, assigning syllable stress, pitch contours, pitch accents & durations.
- Usable ontological lexicons for limited domains (20,000W)
- *Concept - to- speech* synthesis
- Task oriented animated character dialogue for the web (Voc: 2000 w)

Systems:

- Hindi dictation system (Voc: 75, 000 w)
- Useful broadcast transcription systems for information extraction
- Useful special-purpose spoken sentence translation systems (Portable, web)

2008 : Basic Technologies

- Task oriented fully natural animated characters (speech, lips facial expression, gesture) output
- Context sensitive summarization (responsive to user's specific needs)
- Answering questions by making logical inferences from database content
- Speech synthesis with several styles and emotions in major ILs
- Continuous speech understanding in workstations with standard dictionaries (5000 w) in major ILs
- Controlled languages with syntactic and semantic verification for specific domains
- General speaker identification, robust speech recognition in hard-to-model noise conditions and real speaker-independent recognition
- Large coverage grammars with automatic acquisition for syntactic and semantic processing for limited application
- Task oriented fully natural speech, lips, facial expression, gesture input understanding and output generation

Systems:

- Useful speech summarization systems in major Indian languages
- Useful text summarization systems (100:10:1)
- Useful multiple - speaker meeting transcription systems
- Medium-size vocabulary (3,000 W) conversational systems
- Speech driven personal assistant systems

Tools, Platforms, Infrastructure

- Platform for generating intelligent multimedia presentation systems with spoken interaction
- Infrastructure for general portability of spoken dialogue systems across domains and tasks

2010 : Target:

- Unlimited-vocabulary spoken multilingual conversation
- Unlimited-vocabulary spoken translation systems
- Unlimited on-line understanding & generation of integrated natural speech, lips, facial expression and gesture communication
- Fully natural interactive communication