Summary of Speech Research at SIT, IITKGP

The activity of speech processing has been started in School of Information Technology, since 2007. The basic objectives of our group are (1) Development of speech systems in Indian languages, (2) Characterization and incorporation of natural emotions in speech systems, (3) Development of speech systems for mobile devices and (4) Screen readers in Indian languages applicable to visually challenged people.

With the support from TDIL (Technology development for Indian languages) group of DIT, we have initiated the research activity for developing text to speech (TTS) synthesis systems in six Indian languages (Hindi, Bengali, Marathi, Tamil, Telugu and Malayalam). At IIT Kharagpur, we have developed Bengali TTS system, and it was finally integrated with screen reader applicable for visually challenged people. The developed screen reader was demonstrated at the workshop held in National Association for the Blind (NAB), Kolkata. We have also conducted a short term course about 5 weeks at NAB Kolkata with 30 visually challenged people on the utility of the developed screen reader to access the computer for different applications. Recently, on 22nd September 2011 TTS systems developed in six Indian languages were launched by Union minister Sachin Pilot, and soon they will be kept for public access.

Towards the development of sophisticated speech systems, natural emotions should be handled in an appropriate way. In this direction, we have proposed emotion specific features at source, system and prosodic levels for characterizing the emotions. For carrying out this research in Indian context, we have developed emotional speech corpora in Telugu and Hindi. We are also working on emotion synthesis using emotion speech corpus as well as by developing signal processing methods to incorporate the desired emotion specific information.

By observing the recent growth in usage of mobile devices, we have started working on addressing the issues in speech processing in mobile environments. We have proposed appropriate signal processing methods for accurate feature extraction in presence of variable background environments and different speech coders. In addition to feature extraction methods, we have also contributed in design and development of suitable models for speech and speaker recognition in the presence of speech coders and different background environments.

In addition to above mentioned areas, we also worked on other problems such as voice transformation and language recognition. For voice transformation, we have proposed mapping functions using neural networks for transforming speaker specific information between source and target speakers. We have also proposed pitch synchronous speaker specific features and developed prosody modification methods suitable for voice transformation application. For carrying out the language recognition task in Indian context, we have developed Indian language speech database consisting of 27 Indian languages. We have proposed language specific features at different levels and developed models at different hierarchy for achieving the better recognition performance.