

Notice for the PhD Viva Voce Examination

Ms Balaji V (Reg. No. 1345001), PhD scholar at CHRIST (Deemed to be University), will defend her PhD thesis at the public viva-voce examination on Wednesday, 19 May 2021 at 11.00 am on the WebEx Meeting platform.

- Title of the Thesis** : **Implementation of a Speech Recognizer and Synthesizer for the Physically Challenged**
- Discipline** : **Computer Science**
- External Examiner**
(Outside Karnataka) : **P Arul Mozavirman**
Professor
School of Electronics and Communication Engineering
VIT University, Vellore
Tamilnadu
- External Examiner**
(Within Karnataka) : **Dr M N Sree Ranga Raju**
Professor
Department of Electronics & Communication
Engineering
Bangalore Institute of Technology
V V Puram,
Bengaluru - 560004
Karnataka
- Supervisor** : **Dr G Sadashivappa**
Professor
Department of Telecommunication
Engineering
R V College of Engineering
Bengaluru
Karnataka

The faculty members of the Department and the School, interested experts and research scholars of all the branches of research are cordially invited to attend this open viva.



Registrar

Place: Bengaluru
Date: 15 May 2021

ABSTRACT

Speech Recognition and Speech Synthesis are two complementary technologies that are used in systems to which the human voice serves as input or output. People with physical, motor disabilities prefer systems that can be driven by their voice than using the strenuous, usual and standard input-output devices such as keyboard, mouse and monitor. Solutions under the umbrella of “Assistive Technology” are designed to support people with disabilities to overcome the difficulties in handling their diurnal chores. Present-day commercial speech processing systems have received wider customer acceptance, yet not suitable for people with speech disabilities. The fact is that present-day speech recognizers fail to recognize voices with distortions, misrepresentations and deformations. The unintelligibility of the input voice limits the use of off-the-shelf speech processing products by the speech-impaired user community. In such scenarios, the speech processing systems require alterations to become suitable for the specialised user group. Techniques of adaptation are popular in the field of speaker recognition, which can be applied in the domain of Augmentative and Alternative Communication (AAC). The main aim of this research is to model a speaker adaptive system for the speech-disabled users with articulation disorders and neurologically-based disorders due to illnesses like cerebral palsy.

The problem context for this research work is two-fold: accepting the incomprehensible speech input and transforming the same into a more understandable speech. The first portion is to adapt a speech recognizer and verify the recognition accuracy and the second portion is to substitute the recognized words with a better-comprehensible voice. Due to the medical requirements of the research subjects, collecting and using live speech data of individuals is a tedious task. Also, the collection and storage of patients’ data are restricted by ethical procedures. Hence, the data created by various Universities, following the standard procedures in a noise-free environment are used for this research work. Experiments are conducted on the voice data sets in order to improve the recognition accuracy for speakers uttering individual words. The Speech Recognizer is implemented using Hidden Markov Models and Speech Synthesizer is implemented using a pattern-searching algorithm on a database with text input and voice output. The adaptation techniques, viz., Maximum Likelihood Linear Regression (MLLR) and Maximum A Posteriori (MAP) are applied in a pipeline with adjusted language model and pronunciation dictionary. This has reduced the Word Error Rates (WER) of recognizing the incoherent speech. In the first phase, the parameters of the acoustic model of a generic speech recognizer are altered using the feature vectors generated from the training data set applying maximum likelihood linear regression. In the second phase, parameters of the updated model are used as informative priors to MAP adaptation. Speech Synthesizer, i.e., the Text-to-Speech system then translates the recognized text into a more-intelligible voice which is clearer to the listeners.

The simulation with test data sets measured the effectiveness of the combined algorithm proposed here; it produced improvements in recognition accuracy from 43% (for a speaker with 93% speech intelligibility) to 90% (for a speaker with 2% speech intelligibility). An analysis of the improvement in recognition accuracy and speed of recognition for each speaker reveals that the proposed methodology is more effective for severely dysarthric speakers than those with less speech impairments.

Keywords: Automatic Speech Recognition (ASR), Speech Synthesis and Text-to-Speech Systems (TTS), Speech Disabilities, Articulation Disorders, Neurological-based Disorders, Dysarthric Speech, Speaker Adaptation, Hidden Markov Model (HMM), Maximum Likelihood Linear Regression (MLLR), Maximum A Posteriori (MAP).