deBas: A Sinhala Interactive Voice Response (IVR) System

J D Nallathamby, K K R Kariyawasam, H D Pullaperuma, D C Vithana and S Jayasena

Abstract

Although there are widely used Interactive Voice Response (IVR) systems in many languages today, there is no Sinhala language IVR system yet. This paper talks about deBas IVR: a complete Sinhala IVR with automatic speech recognition (ASR) and text-to-speech (TTS) synthesis modules that work in compliance with Media Resource Control Protocol (MRCP). It discusses some background literature, the process taken, the overall design and implementation aspects and the future work that can be carried out in this area. In the ASR component, training the acoustic model is done with SphinxTrain, and decoding with PocketSphinx, which are based on Hidden Markov Models (HMM). In the TTS component, AMoRA Sinhala TTS knowledge base is used, which uses Festival speech synthesis engine and a female diphonic voice, built using Festvox voice building tools. Asterisk is used as the IVR gateway and dial-plan interpreter. MRCPv2 protocol has been followed in developing the speech resources, which uses Session Initiation Protocol (SIP) for establishing controlled connections to external media streaming devices and Real-time Transport Protocol (RTP) for media delivery. The language model of the ASR component has been restricted to digits from 0-9 that are commonly used in IVR systems and the set of words used for our demo application. The Word-Error-Rate and the Sentence-Error-Rate of the ASR component are reported to be 31.4% and 54% respectively, as observed in our experiments. In addition to these, we also introduce a new intonation model that can be applied to any existing Sinhala diphonic voices.