4.10 MBROLA Formatted Diphone Database for Sinhala Language

K.H.Kumara, N.G.J.Dias and R.I.P.Wickramasinghe
Department of Statistics & Computer Science, University of Kelaniya

ABSTRACT

Diphone synthesis is one of the most popular methods used for creating a synthetic voice from recordings or samples of a particular person. Diphones are speech units that begin in the middle of the stable state of a phone and end in the middle of the following phone. The main interest in diphone synthesis is that they minimize the concatenation problems.

The aim of the MBROLA project, recently initiated by the Faculte Polytechnique de Mons (Belgium), is to obtain a set of speech synthesizers for as many voices, languages and dialects as possible, free of use for non-commercial and non-military applications. Central to the MBROLA project is MBROLA 2.00, a speech synthesizer based on the concatenation of diphones, takes a list of allophones associated with prosodic information as input and outputs 16 bit linear speech samples. Diphone databases tailored to the MBROLA format are necessary to run the synthesizer.

Therefore we put forward a Diphone database, tailored to the MBROLA format, to generate synthetic voice for Sinhala language through MBROLA .pho reader. The first step of building the diphone database was the fixing a list of all the phones (acoustic instances of phonemes) of Sinhala language. Creating the diphone database was achieved in three steps: Creating a text corpus, Recording the corpus and Segmenting the speech corpus. For the text corpus, we used few selected chapters of two Sinhala novels. The corpus was then red by two (Male and Female) native Sinhala speakers, digitally recorded and stored. Then all diphones were spotted manually with the help of SpeechViewer of CSLU toolkit which was developed by the Center for Spoken Language Understanding, Oregon Graduate Institute of Science and Technology, USA. A diphone database was finally created with 1004 diphone segments, which summarizes the results in the form of: the name of diphones, the related waveforms, their duration, and internal sub-splittings.

Since we did not consider allophone variations in all instances, it may reduce the naturalness of the resulting synthetic speech. It is also possible that the number of diphone segments may higher than the above number (1004). However, most of the common occurrences of diphones were included in the database that we have developed.