

ABSTRACT

In the last decade there has been a shift towards development of speech synthesizer using concatenative synthesis technique instead of parametric synthesis. There are a number of different methodologies for concatenative synthesis like Time Domain Pitch Synchronous Over-Lap Add method (TDPSOLA), Pitch Synchronous Overlap and Add (PSOLA), and MultiBand Resynthesis Overlap Add (MBROLA). This project, describes a concatenative speech synthesis system based on Epoch Synchronous Non-Over Lapp Add (ESNOLA) technique, for standard colloquial Bengali, which uses the partnemes i.e. part of the phonemes as the smallest signal units for concatenation. The system provided full control for prosody and intonation.

This project presents a new Concatenative text-to-speech (TTS) system for Standard Colloquial Bengali (SCB) using a new set of signal units in sub-phonemic level, namely, partnemes. The Epoch Synchronous Non Overlap Add (ESNOLA) algorithm is developed for concatenation, vowel generation or deletion as a part of duration modification, pause generation as well as for pitch modification. It may be noted that the prosody of the stored units is often not consistent with that of the target utterance and must be altered at the time of synthesis. Furthermore, several types of mismatches can occur at unit boundaries of the synthesized signal, which have to be properly truncated and matched. ESNOLA technique provides the complete control on implementation of intonation and prosody. It allows judicious selection of signal segments so that smaller fundamental parts of the phonemes may be used as units reducing both number and size of the signal elements in the dictionary. Further the methodology of concatenation provides adequate processing for proper matching between different segments during concatenation. The use of special type of basic signal segments makes the size of signal dictionary very small so there is a possibility of its implementation in low-cost general-purpose electronic devices.