

DATA ORIENTED PERCEPTUAL DISCONTINUITY SMOOTHING ALGORITHM

CHANDRA SEKHRAM BONDU¹ & RAMAKRISHNA S²

¹System Analyst, Department of Computer Science, Rashtriya Sanskrit Vidyapeetha, Tirupati, Andhra Pradesh, India

²Professor, Department of Computer Science, Sri Venkateswara University, Tirupati, Andhra Pradesh, India

ABSTRACT

The major drawback of concatenation based Text to Speech Synthesis is the presence of audible discontinuities in the continuous synthetic speech output. The speech units used to generate synthetic speech was extracted from continuous natural utterances. The way of segmenting and extracting the speech units from natural utterances for concatenative speech synthesis has significant effect on the naturalness and intelligibility of synthetic speech. Phoneme, Diphone and Syllable based units are three kinds of popular speech segmentation approaches. Each segmentation approach has its own merits and demerits. Diphone based units are segmented such a way that the transition effect that exists between two phonemes is not lost. When syllable like or phoneme based speech segments were used for generating synthetic speech there were audible disturbances relating to lacking of transitional effect between two phonemes. A Listener can subjectively distinguish an audible difference between the natural utterance and synthetic speech. A method called Perceptual Discontinuity smoothing to reduce such discontinuities using a command line software tool known as sox was experimented. This work majorly focused on subjectively measuring such discontinuities by conducting perceptual listener tests and developing a method to reduce such discontinuities. This work is a part of developing a Text to Speech Synthesis for Sanskrit.

KEYWORDS: Coarticulation Effect, LISS, RISS, TR, MOS, Prosodic Features, Sox, Abhinidhana