Abstract

Speech synthesis systems which involve concatenation of recorded speech units are currently very popular. These systems are known for producing high quality, natural-sounding speech as they generate speech by joining together waveforms of different speech units. This method of speech generation is quite practical. However the speech units that are being concatenated may have different spectra on either side of the concatenation points. Such mismatches are spectral in nature and give rise to spectral discontinuity in concatenated speech waveforms. The presence of such discontinuities can be very distracting to the listener and degrade the overall quality of output speech. This paper proposes a speech signal processing technique that deals with the problem of spectral discontinuity in the context of concatenated waveform synthesis. It involves the post-processing of the synthesized speech waveform in time domain. This technique is implemented on different single channel Punjabi wave audio files which were created by concatenating different Punjabi syllables. A listening test was conducted to evaluate the proposed technique, and it was observed that the spectral discontinuity is reduced to a large extent and the output speech sounds more natural with the reduction of audible noise.
Removal of Spectral Discontinuity in Concatenated Speech Waveform


Index Terms
Keywords
Speech waveform  Concatenative speech synthesis  Spectral discontinuity