Abstract

Speech Synthesizer for
Thai Text-to-speech (TTS) system on embedded device

May 2010

by

Konlakorn Wongpatikaseree

B.Eng, Sirindhorn International Institute of Technology
Thammasat University, 2007

This thesis presents the development of a speech synthesizer for Thai Text-To-Speech (TTS) system on embedded devices. Although the Thai TTS systems have been implemented on personal computers and the quality of most systems are acceptable, not so many systems have been developed on embedded or mobile devices, such as mobile phone. In this research, we focus on the practical aspects of the Thai TTS systems on a mobile device including application size and processing time. We aim at developing a Thai speech synthesizer that produce an output speech in real-time on the mobile device. The proposed synthesizer, Flite_Thai, is based on a unit concatenative synthesis technique, and the Flite software which is a speech synthesis engine designed for the resource-limited devices. To evaluate the proposed system, we conduct an experiment to compare the computational time, and the naturalness of the output speech. The experimental result shows that the system worked in a real-time manner. However, the quality of the generated speech is still lower than the standard level.

Hence, we propose an approach to improve the quality of the generated speech by designing a new speech corpus using a new speech unit. Non-sense carrier sentence technique is used in this design of the new speech corpus. Co-articulation affect is concerned in carrier sentence technique. Each of the carrier sentences contains a sequence of speech units without concerning the meaning. In addition, we propose a new speech unit called hybrid diphone, which is extracted from the new speech corpus. This new speech unit combines the advantage of diphone and demi-syllable. It aims to reduce the storage usage size and improve the quality. From the experiment, we found that the Mean Opinion Score (MOS) values of all the speech units that came from the non-sense carrier sentences are better than the MOS values that came from the natural sentences in the existing speech corpus, TSynC. However, among the three unit types in the news corpus, demi-syllable obtained the highest score. Although, the proposed hybrid diphone obtained higher MOS than the existing system and the diphone, it still suffers from a similar problem which is unsmooth joints between units. Some smoothing techniques are required in the future work to improve the quality of the speech generated using the hybrid diphone unit.