Abstract

Nowadays, Text-To-Speech (TTS) has been rapidly adopted in many areas. One of them is navigation system where the machine will speak the instruction so that users no longer need to read it. In TTS, there is something called prosody wich is known as phrase break, duration, intonation, and emphasis on the word. Phrase break plays an important role in prosody. Determination of the break is affecting duration and intonation of utterance of a sentence that will certainly affect the quality of the utterance.

This research uses Hidden Markov Model (HMM) to determine phrase break in a sentence. The system uses a bigram HMM with observation symbols in the form of part-of-speech (POS) which is two consecutive words in a sentence. The state sequences represent the sequence of phrases break in a sentence and can be used as reference to assign duration and pitch in order to generate prosody. Quality of the utterance will be measured using the Mean Opinion Score (MOS).

The results showed that the average value of MOS (1 to 5 in scale) for affirmative sentence is 3.55 and the best is 3.84. The average value of MOS for questions is 3.73 and the best is 3.87. The average value of MOS for imperative sentence is 3.76 and the best is 3.82. The system's accuracy of phrase breaks is 79,23%.

Keywords: HMM, TTS, prosody