THE
DOCTORAL
RESEARCH ABSTRACTS

Volume: 14, October 2018

14th ISSUE

33

Name : MUHAMMAD IZZAD BIN RAMLI

Title : GRAMMAR-BASED PROSODY MODIFICATION FOR EXPLICIT CONTROL MALAY LANGUAGE STORYTELLING SPEECH SYNTHESIS

Supervisor : ASSOC. PROF. DR. NURSURIATI JAMIL **(MS)**

DR. NORAINI SEMAN **(CS)**

Storytelling speech synthesis is a process of converting written text to the spoken speech in storytelling speaking style. It has gained much interest in the area of digital storytelling and storytelling humanoid robot for children in learning environment. Reviews have shown that storytelling speech synthesis can be developed using implicit control, explicit control or playback approach. The literatures stated that each approach has its own drawbacks and needs to be tackled for a better quality synthesized speech. In this thesis, explicit control is selected because it is commonly used in the storytelling speech synthesis and has shown to produce good intelligibility and reasonably natural speech. However, modification of prosody in explicit control approach remains a problem as it may lead to speech quality degeneration due to extreme over-exaggeration of speech. Furthermore, perception evaluation showed that the similarity score between the natural and synthesized speech can also be improved for a more satisfactory result. Therefore, this research aims to introduce a new prosody modification technique to reduce over-exaggeration and simultaneously improve the similarity between the natural and synthesized speech. Three narrative children short stories in neutral and storytelling styles are recorded by nine storytellers. A total of 522 speech sentences, 5,238 words and 12,294 syllables are collected to be utilized as experimental datasets and prosody analysis. Based on the prosody analysis, a grammar-based prosody modification rules are proposed by integrating grammatical structure. Consequently, new rules and algorithm that is MustFront rule, limitation rule, and two-steps pitch contour algorithm are introduced to increase the synthesized speech quality. Using Harmonic Noise Model (HNM) as the synthesizer, the grammar-based prosody modification rules are used to produce the synthesized storytelling speech. The synthesized storytelling speech is then compared to baseline methods of synthesized storytelling speech that are global and local prosody modification rules. The evaluation of the synthesized storytelling speech was conducted using objective test (Perceptual Evaluation of Speech Quality (PESQ) test, and aspects or components test) and perceptive test (naturalness, intelligibility, similarity test, and recognition test). The result of PESQ test showed that grammar-based prosody modification with limitation rule produced the highest Mean Opinion Score (MOS) of 3.35 based on five-point scale. The prosody parameters test also demonstrated that the synthesized storytelling speech using grammar-based with limitation rule is much closer to the natural storytelling speech. As for the perception test evaluated by nine native speakers, results showed that grammar-based rule with limitation rule is able to outperform local and global rules by achieving the naturalness, intelligibility and similarity Mean Opinion Score (MOS) of 4.11, 4.47 and 4.06, respectively using a five-point scale. The proposed rule also managed a high accuracy rate of 92% for the recognition test. As conclusion, the performance of the synthesized storytelling speech using grammar-based with limitation rule is better than local and global rules.