論文内容の要旨

博士論文題目

Japanese Incremental Text-to-speech Synthesis based on Accent Phrase Unit
アクセント句単位に基づく日本語インクリメンタル音声合成

氏　　名　　　　Tomoya Yanagita

（論文内容の要旨）

Incremental Text-to-Speech (iTTS) synthesizes a speech incrementally from a synthesis chunk smaller than sentence units. The iTTS is a key component for simultaneous speech-to-speech translation and spoken dialogue systems. The challenge of iTTS is to maintain high speech quality with low latency by optimizing synthesis chunks. Existing iTTS systems use a word unit as a synthesis chunk. The Japanese language has a mora-timed rhythm and a tonal aspect accent, where an accent phrase is a critical unit for representing accents and meaning. This dissertation proposes a high-quality Japanese iTTS system with low latency by using accent phrase units. We first propose HMM-based Japanese iTTS and investigate the speech quality and latency. Experimental results show that (1): an accent phrase unit and features are necessary to improve speech quality for Japanese iTTS, (2): using the following one accent phrase unit effectively improves speech quality. Second, we investigate neural iTTS for Japanese. The neural iTTS systems proposed for English use a prefix-to-prefix neural iTTS framework with 1-2 word units a look-ahead. Since the Japanese language is based on accent phrase units, using a prefix-to-prefix neural iTTS with a look-ahead approach increases latency. We propose an alternative approach to the neural iTTS that does not use look-ahead. We propose a method that uses an accent phrase unit and exploits information embedded in the previous synthesizing

氏　名	Tomoya Yanagita

（論文審査結果の要旨）

This thesis proposes an incremental machine text-to-speech (iTTS) system for the Japanese language. The iTTS synthesizes a speech incrementally from a synthesis chunk smaller than sentence units. The iTTS is a key component for simultaneous speech-to-speech translation and spoken dialogue systems. The challenge of iTTS is maintaining high speech quality with low latency by optimizing synthesis chunks. Existing iTTS systems use a word unit as a synthesis chunk. The Japanese language has a mora-timed rhythm and a tonal aspect accent, where an accent phrase is a critical unit for representing accents and meaning. This dissertation proposes a high-quality Japanese iTTS system with low latency by using accent phrase units. The thesis proposes the following

① The thesis proposed an HMM-based Japanese iTTS. Experimental results show that (1): an accent phrase unit and features are necessary to improve speech quality for Japanese iTTS, (2): using the following one accent phrase unit effectively improves speech quality.

② Second, the thesis proposed a neural iTTS for Japanese. The neural iTTS systems proposed for English use a prefix-to-prefix neural iTTS framework with 1-2 word units a look-ahead. We propose an alternative approach to the neural iTTS that does not use look-ahead. We propose a method that uses an accent phrase unit and exploits information embedded in the previous synthesizing step.

The thesis research succeeded to build the Japanese iTTS system with high quality and low latency, considering the Japanese language structure. This is the first research in this direction. The proposed study provides a general framework for a Japanese iTTS system. A series of his research resulted in two high-quality peer-reviewed international and domestic English journal papers and two peer-reviewed international conference papers. As a result, the thesis is sufficiently qualified as a Doctoral thesis of Engineering.