

論文内容の要旨

博士論文題目

Self-adaptive and Incremental Machine Speech Chain
自己適応型インクリメンタルスピーチチェーンの研究

氏名 Sashi Novitasari

(論文内容の要旨)

In human spoken communication, speech production and perception are inseparable. It is reflected in the human speech chain mechanism, showing that humans speak while listening. This mechanism allows them to monitor and improve their speech performance in various situations. It is also important for language acquisition.

Inspired by the human speech chain mechanism, a machine speech chain framework based on deep learning was recently proposed for a semi-supervised development of a text-to-speech (TTS) system and an automatic speech recognition (ASR) system. However, the basic framework was aimed only for non-incremental TTS and ASR training, in which the systems require a long delay when encountering a long input sequence. Moreover, the TTS and ASR still perform separately during inference. They could not do self-adaptation or change the speech by considering environmental situations. By contrast, humans can listen to what they speak in real-time and enhance the intelligibility of their speech,

氏 名	Sashi Novitasari
-----	------------------

(論文審査結果の要旨)

This thesis proposes self-adaptive and incremental machine speech chain frameworks for training and inference by mimicking the human speech chain closely. To achieve this, first, the thesis proposes a method to reduce the latency of the basic machine speech chain by replacing the components with an Incremental Text-To-Speech system (ITTS) and an Incremental Automatic Speech Recognition (ISR). During speech chain training, the proposed system improved together through a short-term loop. Second, a self-adaptation framework focusing on speech synthesis in noisy environments through a speech chain mechanism is applied. It synthesizes the speech not only by taking text input but also by the auditory feedback representing the current system performance and the environmental situation. This mechanism allows the TTS to speak in a Lombard effect automatically according to real situations, where the Lombard effect controls speech intensity and quality to enhance speech intelligibility. Finally, it is confirmed that the proposed self-adaptive incremental speech synthesis with a low-latency adaptation is critical in noisy environments.

The thesis research aims to construct the self-adaptive speech processing system together with ASR and TTS. The candidate successfully built the low-latency self-adaptive incremental TTS system based on the speech chain and showed the system can produce Lombard speech in noisy environments. This is the first research in this direction. The proposed research provides a general framework for a self-adaptive and incremental speech chain system. A series of her research resulted in two high-quality peer-reviewed international and domestic English journal papers, and four peer-reviewed international conference papers. As a result, the thesis is sufficiently qualified as a Doctoral thesis of Engineering.