論文内容の要旨

申請者氏名　　古川　慧

A significant gap remains between our understanding of human intonation mechanisms and efforts to replicate human-like intonation in speech synthesis. To bridge this gap, the research objectives in this dissertation can be categorized into two types: linguistic and engineering.

The linguistic objective is to validate the boundary-driven account of downstep in Japanese. Second, the engineering objective is to apply this boundary-driven theory to speech synthesis, aiming to reproduce various phonological phenomena in Japanese. Regarding the linguistic objectives, this dissertation delves into the fundamental factors that trigger downstep. Traditionally, downstep has been defined as a pitch range compression triggered by lexical pitch accents. The accent-driven account posits that downstep is directly caused by accents. In contrast, I propose the boundary-driven account, suggesting that downstep arises from the insertion of phonological boundaries.

In terms of the engineering objectives, this dissertation applies the proposed boundary-driven theory to neural sequence-to-sequence speech synthesis, aiming to reproduce phonological phenomena such as initial lowering and rhythmic boost. Previous studies on speech synthesis have primarily focused on subjective naturalness, often lacking objective assessments of their ability to replicate pitch patterns of phonological phenomena in Japanese that stem from syntactic structures and phonological constraints. In this study, particular attention is given to the phenomena of initial lowering and rhythmic boost, and novel criteria for objectively evaluating synthesized speech are introduced.

論文審査結果の要旨

申請者氏名　　古川　慧

This thesis examines both linguistic and engineering perspectives. From a linguistic standpoint, it investigates the boundary-driven account of downstep in Japanese. Traditionally, downstep has been characterized as a pitch range compression triggered by lexical pitch accents. However, this research proposes an alternative explanation, arguing that downstep results from the insertion of phonological boundaries. On the engineering side, the study applies this boundary-driven theory to speech synthesis, aiming to replicate various phonological phenomena in Japanese. To achieve this, the proposed theory is integrated into a neural sequence-to-sequence speech synthesis framework. The findings of this research have been published in two high-quality journals, *The Journal of the Acoustical Society of America* and *IEEE Access*, along with three papers presented at international conferences. These contributions fulfill the requirements for a doctorate degree in engineering.