

森勢 将雅

明治大学総合数理学部
専任准教授

Human-in-the-loop 型歌唱デザインの開発

§ 1. 研究成果の概要

2019 年度は、すでに多くの知見があるテキスト音声合成を対象に、イントネーションやタイミングなどのパラメータを人間が調整し、その調整結果から自然な歌声を生成する技術を開発した。従来の技術では、人間が調整する必要が無く自然な音声を出力することは可能であるが、調整することで音質が劣化する課題があった。この課題に対し本技術では、人間が調整した結果から自然な音声を生成するように音声生成モデルを学習させることで解決を試みた。

加えて、この技術を組み込んだインタフェースのプロトタイプを実装した。図 1 がプロトタイプの実行画面であり、図中の線が声の高さの時系列に相当する。ユーザは高さや音素のタイミングを調整可能であり、

調整に伴う劣化を抑えた音声生成が可能である。有効性を検証するため、生成された音声の品質、およびインタフェースのユーザビリティに関する評価を実施した。実験の結果、提案する生成モデルは、モデルを用いずに高さやタイミングを調整した音声と比べて音質劣化を抑えられることが示された。ユーザビリティに関しても、同様の尺度を用いて評価されたシステムの平均的なスコアよりも高いことが示された。



図 1: 実装したプロトタイプの実行画面