

# 平成30年度 情報工学コース卒業研究報告要旨

戸田 研究室	氏 名	多 賀 遥 香
卒業研究題目	ユーザの協力的動作を活用したリアルタイム声質変換	

音声合成の技術を応用した VOCALOID や、仮想空間における YouTuber であるバーチャル YouTuber が、個性的なキャラクタにより人気を集めている。キャラクタへの自己投影を高めるために、服装や髪型などの見た目をキャラクタに似せる行為が盛んに行われる一方で、音声に関しては、身体的制約により個々の発話者が生成できる声は限定される。結果、キャラクタへの自己投影に違和感を残す原因となる。

身体的制約を超えた発声を可能とする技術として、声質変換 (Voice Conversion: VC) の研究が行われている。統計的手法に基づく枠組みが主流であり、入力話者と目標話者の同一発話データから変換モデルを事前に学習することで、入力話者の声質を目標話者への声質へと変換する処理が実現される。これまでの研究から、リアルタイムで動作する声質変換処理は実現されてはいるものの、1) ボコーダと呼ばれる数理モデルに基づく音声波形生成処理を用いることで、声の高さや声色といった様々な特徴の変換が可能となる一方で、変換音声の品質は自然音声と比べて劣化すること、2) ボコーダを用いない波形加工に基づく手法を用いることで、声の高さの変換が不要な際には、変換音声の品質劣化を防ぐ一方で、変換可能な特徴は声色に限定され、変換音声の話者類似度が劣化すること、が明らかにされている。

本研究では、キャラクタ性の強い発話音声への変換に着目し、入力話者であるユーザの協力的動作を活用する枠組みを提案する。声の高さや話し方など、ユーザが声色以外の特徴を真似て発声した音声を入力として利用することで、波形加工に基づく VC の高い品質を保持したまま、話者類似度の高い変換音声の実現を試みる。また、男性ユーザが女性キャラクタを真似る際には、裏声などの特殊な発声を取り扱う。これらのユーザの協力的動作による特殊な発声に対して VC を適用し、実験的評価により変換性能を明らかにする。

女性ユーザ1名および男性ユーザ1名から女性キャラクタ4名への変換音声に対して、音質および話者類似度に関する主観評価実験を行った。女性ユーザからの変換結果 (図1) から、1) 真似発声を用いることで、変換音声の音質を保持したまま話者類似度を改善できること、2) 波形生成 (VC-gen) と比べて波形加工 (VC-mod) を用いても音質改善効果は得られないことが分かった。また、男性ユーザからの変換結果 (図2) から、3) 真似発声に裏声を用いる際には音質および類似度改善効果は得られないこと、4) 波形生成と比べて波形加工を用いることで音質が劣化することが分かった。このことから、真似発声の利用は、一定条件の下では有効であるが、波形加工の音質改善効果を抑圧することが明らかとなった。

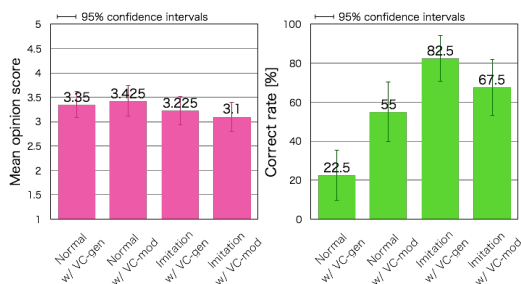


図1：通常の発話を用いた評価結果 (左図：音質, 右図：類似度の正答率)

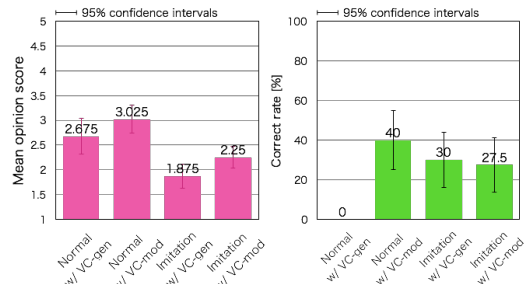


図2：特殊な発話を用いた評価結果 (左図：音質, 右図：類似度の正答率)