令和元年度　情報工学コース卒業研究報告要旨

戸田 研究室	氏　　名	中　島　健　斗
卒業研究題目	リアルタイム統計的声質変換における聴覚フィードバックに関する検討	

近年，VR コンテンツを利用して，好みのキャラクタを演じる行為が人気を集めている．しかし，仮想空間において，理想的な外見を得ることが可能である一方で，理想的な声を発することは身体的制約を受けるため困難である．

その問題の解決法の一つとなり得るのが，リアルタイム統計的声質変換である．まず，学習データを用いて，入力話者の音響特徴量から目標話者のものへと変換する関数を事前に推定する．得られた変換関数を入力話者の任意の音声に対して適用することで，目標話者の声質の音声への変換を行う．リアルタイム変換処理を用いることで，入力話者となるユーザは，所望の目標話者の声質による発声が可能となる．その一方で，慣れ親しんだ自身の音声による聴覚フィードバック (Auditory Feedback, AF) のみでなく，システムが出力する変換音声による AF も受ける状況になり得る．AF は発声に大きく影響を与えることが知られているが，リアルタイム統計的声質変換を用いた発声における AF の影響については，未知のままである．

本研究では，リアルタイム統計的声質変換使用時において，AF がユーザの発声に与える影響について調査を行う．システムから与える AF として，1) 何も与えない場合 (NF)，すなわち，自身の音声による AF のみを与える場合と，2) 目標話者への変換音声を与える場合 (TF) を検討する．TF は，より適切な変換音声が得られるように，ユーザが自ずと発声を動的に調節する可能性がある一方で，変換音声の声質につられて学習データ収録時と大きく異なる発声を行う可能性もあり，その場合は変換精度の低下を招くと予想される．そこで，学習データ収録時の発声に誘導させるという考え方から，3) 学習データの自身の声質への変換音声を与える場合 (IF) についても検討する．

各 AF を与えた際の変換音声の自然性および目標話者との類似性について，主観評価実験を行った．入力話者として男性話者 1 名，目標話者として女性話者 1 名，キャラクタ 1 名 (VOICEROID2 結月ゆかり) を据えた．また，女性話者を目標話者とした場合については，変換音声と目標音声との間のメルケプストラム歪みを用いて，変換精度に関する客観評価を行った．主観評価実験の結果（図 1）から，目標話者によっては，TF を用いることで，NF よりも高い自然性が得られることが示された．また，客観評価実験の結果（図 2）においても，同一の目標話者に対して，TF により変換精度の改善が得られることが示された．以上の結果により，目標話者によっては，変換音声を AF として与えることで，より自然性の高い変換音声が得られるように，ユーザが発声を自ずと調節できることが分かった．

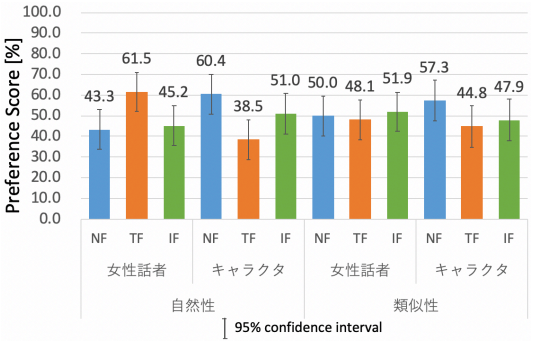

図 1:主観評価実験の結果

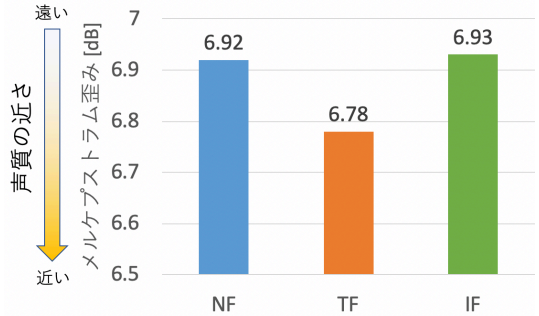

図 2:各変換音声と女性話者の音声との間のメルケプストラム歪みの比較結果