

審査の結果の要旨

氏名 橋本 哲弥

本論文は「話者空間基底を用いた特徴量分解とそれに基づくパラレルデータフリー DNN 型声質変換」と題し、全 8 章より成る。入力音声はその言語的内容を変えずに、非言語的特性（主として話者特性）のみを変換し出力する技術は声質変換と呼ばれる。入出力の統計的写像を推定するために、通常、言語内容が揃った複数話者のパラレルコーパスを必要とするが、本論文ではこの技術的制約を解消する試みを提案している。また、従来 GMM (Gaussian Mixture Model) を通して写像推定が行われてきたが、近年 DNN (Deep Neural Network) による写像推定が GMM の精度を超えるようになってきた。本論文では、両者の長所を生かし、GMM に基づいて話者空間基底を構成し、入力話者から各基底話者への変換及び最終的な出力話者への変換を DNN に担当させる方式を提案している。従来方法との比較実験の結果、パラレルフリーな多対多声質変換タスクにおいて、その有効性が示された。

第一章は「序論」であり、本論文の背景、目的、及び、構成を述べている。

第二章は「一対一声質変換及びその拡張」と題し、まず、音声の基本的な音響特徴量を説明し、更に、一対一の統計的声質変換技術を紹介すると共に、その話者適応についても述べている。また、本研究で提案する手法の契機となった、多言語音声認識における言語非依存サブネットワークの自動適応についても触れている。

第三章は「従来の多対多声質変換手法」と題し、先行研究にて検討された多対多の声質変換技術について説明している。多対多で声質変換を実現する場合、話者性をどのような表現で捉えるのかが一つの焦点であり、話者表象について説明し、また、声質変換をパラレルデータフリーで行なった先行研究例についても触れている。これらは、本論文の直接の先行研究に対応する。

第四章は「提案手法：複数のサブネットワークを有する DNN に基づく多対一声質変換」と題し、第二章で紹介した DNN サブネットワークの枠組みを声質変換に適用する手法を提案し、幾つかの着眼点に基づいて、提案手法の変換精度を

検証している。実験の結果、従来手法と比較して精度の高い変換が実現された。

第五章は「提案手法：EVGMM に基づく話者空間基底を用いた DNN 声質変換手法」と題し、GMM と DNN の統合に基づく声質変換を検討している。GMM を用いて話者空間を設計し、一旦入力話者特性を基底ベクトルの重み付き和として表現し、各基底話者の成分を DNN によって変換し、最終的にそれらを足し合わせることで所望の話者性を表出する枠組みである。DNN のみを用いた変換と比較して、技術的構成・設計が理解しやすい特徴を持つ。比較実験の結果、従来手法よりも優れた変換精度を示した。

第六章は「提案手法：EVGMM に基づく話者空間基底を用いたパラレルデータフリー DNN 声質変換」と題し、EVGMM+DNN の枠組みをパラレルデータフリー実装する手法を提案し、実験的に先行研究例に対する優位性を示している。技術的には、欠損 EM アルゴリズムを用いて、共分散項の推定を実現したことが大きい。

第七章は「提案手法：FHL を用いた声質変換」と題し、前節で用いた Average Voice Model を拡張し、DNN ネットワークの各層が話者変換にどのように寄与しているのかを分析的に検証するため、Factorized Hidden Layers を用いた声質変換を検討した。

第八章は「結論」であり、本研究のまとめと今後の展望を述べている。

以上要するに本論文は、統計的声質変換において、従来検討されてきた GMM による変換と、近年、より高精度な変換を実現した DNN による変換の長所を生かし、機能的にモジュール分割可能な形で両者の統合を行い、更には、パラレルデータフリーの変換という、技術的難易度の高い問題に対して、一つの技術的解答を与えており、音声工学及び情報工学に貢献するところが少なくない。よって本論文は博士（工学）の学位請求論文として合格と認められる。