

審査の結果の要旨

氏名 柏木陽佑

本論文は「非言語情報の違いに頑健な特徴量表現に着目したニューラルネットワーク音声認識に関する研究」と題し、6章より成る。音声には背景雑音、チャンネル特性、話者の体格・年齢差など、発声されたメッセージとは無関係な音響歪みが非言語的情報として混入する。近年深層学習に基づいたニューラルネットワーク (Deep Neural Network, DNN) 音声認識が脚光を浴びているが、本論文では DNN の一部を特徴抽出器と捉え、非言語的情報の違いに頑健な特徴量表現に着目し、DNN 音声認識の高度化を検討し、その有効性を実験的に検証した。

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「音声認識に関する基礎技術とニューラルネットワーク」と題し、従来の音声認識における主流の音響モデリング技術であった HMM や、それが対象としていた音声特徴量について解説している。その後、DNN とその音響モデリングへの利用について述べている。非言語的情報の違いに関する対策である、特徴量正規化やモデル適応についても説明している。更には、識別モデルであるが故に DNN が持つ制御の難しさについても HMM と比較している。

第三章は「雑音環境下音声認識のためのニューラルネットワークを用いた識別的区分線形変換」と題し、DNN を音声強調の枠組みで利用する手法を提案し、その有効性を実験的に示している。DNN の高い識別能力を利用し、音声強調の時に必要となる入力音声の音響空間インデックスを、入力空間 (雑音音声空間) ではなく、出力空間 (クリーン音声空間) で求め、それに基づく区分線形変換を通して音声を強調する手法を提案した。実験の結果、従来の音声強調技術よりも高い認識精度を実現できた。

第四章は「話者コードによるパラメータ制御を用いたニューラルネットワーク音響モデルの正規化学習」と題し、DNN による話者正規化学習において、話者依存層と話者非依存層をより明確に区分して DNN を学習する方法を提案し、その有効性を実験的に示している。音声認識時には話者情報は未知であるため、これを推定しつつ、ネットワークの適応を行なっている。実験の結果、話者の違い

に対する頑健性を向上させることに成功した。

第五章は「識別的アプローチによる分布間距離計算とその利用」と題し、DNNを用いた不変特徴の計算方法を提案し、その有効性を実験的に示している。**f-divergence** は分布間距離の一種であるが、これは変換不変性を有する。音声特徴をこの **f-divergence** で記述できれば、様々な変化に頑健な特徴となるが、その推定が難しい。本章では、**f-divergence** の計算式を等価的に変形し、DNNに基づく音素事後確率を効果的に利用し、**f-divergence** を推定する方法を提案した。タスクとしては言語識別に適用し、識別率を高めることに成功した。

第六章は「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、識別モデルであるが故にその制御や解釈が困難であった DNN に対し、特徴量強調、話者正規化学習、不変特徴推定という、三つの音声特徴量に関するタスクを対象として、効果的な DNN の利用方法を提案し、非言語的情報の違いにより頑健な DNN 音声認識を実現しており、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。