

論文の内容の要旨

論文題目 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

氏 名 柏木 陽佑

音声認識システムは、スマートフォンの普及や計算機パワーの発展などに伴い、非常に身近なものとなりつつある。単純な入力インタフェースとしての性能は実用に耐えうるものとなって来たが、その認識性能は人間のそれと比べた場合未だ充分とは言えない。今後、自動音声認識技術のさらなる応用を考えた場合、認識性能の向上は必要不可欠である。

我々が普段何気なく聞いている音声には様々な情報が内在する。これらは大きく言語情報、パラ言語情報、非言語情報の3つに分けることができ、音声認識とは、この内の言語情報を抽出するタスクであると考えることができる。なお、本研究において前提とする音声認識システムでは、パラ言語情報による影響は音声分析の過程で無くなるものと考えられる。しかし、非言語情報は音響特徴量に影響を与えるためノイズであり、認識性能の低下の原因となる。そのため、これをいかに制御するかが音声認識において長年の課題であった。

さて、近年、統計的機械学習はニューラルネットワークの台頭という大きな転換点を迎えた。音声認識技術においてもこの影響は大きく、ニューラルネットワークベースの識別的な音声認識システムが高い性能を示すことも報告され、もはやニューラルネットワークベースのシステムが主流となったと言える。しかし、ニューラルネットワークは従来のガウス分布、ガウス混合分布に基づく生成的なモデルとは性質が大きく異なる。そのため、従来の非言語情報に関する要素技術をそのまま利用することが困難であり、現在のニューラルネットワークをベースとする音声認識における非言語情報の制御に関する研究の潮流は、手探りな状況であることが否めない。これを打開するためには、ニューラルネットワーク音声認識に適した理論的背景に重点を置いた非言語情報の制御技術の研究が重要である。

そこで、本論文では、非言語情報の違いに頑健なニューラルネットワーク音声認識システムの実現を目指す。非言語情報の制御として大きなウェイトを占める特徴量と音響モデルにおいて、従来のガウス混合分布に基づく非言語情報の制御に関する要素技術を基にした、ニューラルネットワークベースの非言語情報の制御手法を提案する。先に述べた通り、ニューラルネットワークとガウス混合分布は性質が大きく異なるため、単純な応用は困難である。そのため、ニューラルネットワークとガウス混合分布の融合により、両者の性質を組み合わせたアプローチを提案する。これにより、従来の生成的アプローチの持つパラメータの意味づけ、制御に関する要素技術を

ニューラルネットワーク音声認識に取り入れることが可能となり, 認識性能の向上が可能となる. また, モデルの性質のみを考慮したニューラルネットワークとガウス混合分布の融合では, 音声認識システムに対する非言語情報の扱いに対する本質的な解決策とはならない. そこで, 本論文の後半では非言語情報の違いに頑健な特徴量表現である音声の構造的表象, そしてそれを構成する分布間距離の計算に対してニューラルネットワークを用いたアプローチを導入する. これにより, ニューラルネットワークの高い識別性能とそれを支える特徴量空間の表現能力を, 従来のガウス分布をベースとする音声学的知見に基づいた手法との融合が可能となり, ニューラルネットワークを用いた新しい非言語情報の違いに頑健な特徴量表現が実現できる.