報告番号	※甲　　第　　号

主　論　文　の　要　旨

論文題目　Acoustic Feature Transformation Based on Generalized Criteria for Speech Recognition (音声認識における音響特徴変換の最適化基準の一般化に関する研究)

氏　名　坂井 誠

論　文　内　容　の　要　旨

音声認識の基本認識性能向上のために、本論文では音声認識の音響特徴変換の研究を行った。音声認識では入力音声を短区間（フレーム）に分割して認識処理を行っていく。このとき、隣接する複数のフレームを結合して新たな特徴量とすると、音声の動的な時間変動を特徴に組み込めるため認識性能向上が期待できる。一方、フレーム数を大きくすると特徴量の次元数が高くなるため、音響モデルの推定精度が劣化し音声認識性能も劣化することが考えられる。そこで通常、複数フレームを結合した高次元の音響特徴を低次元空間へ埋め込む音響特徴変換が広く用いられている。音響特徴変換は、高次元の特徴量に対し音素識別情報を失わないように特徴量の次元を削減することが重要になる。音響特徴変換には次の二つのアプローチがある。

1. 音素クラス間変動とクラス内変動の比を最大化する
2. 音素クラス間の分類誤差を最小化する

本研究では両者を研究対象とした。また、音響特徴変換はさらに線形変換と非線形変換とに分けられるが、本研究では線形変換を研究対象とした。

はじめに、音素クラス間変動を大きくしながらクラス内変動を小さくするように音響特徴を変換する手法に注目した。このアプローチの代表的な従来手法として線形判別分析 (LDA), 異分散線形判別分析 (HLDA) や異分散判別分析 (HDA) があげられる。LDA, HLDA や HDA は音声認識の特徴変換として広く使われている。これらは独立に提案された手法であるため、これまではそれらの手法の関係についてはあまり議論されてこなかった。本研究では、これら従来手法の間に密接な関係があることを説明し、従来手法を包含した新しい枠組みを提案する。LDA は、クラス間変動としてクラス間の平均距離または全データの分散を用い、クラス内変動として各クラス分散の算術平均を用いている。HDA と HLDA は、クラス間変動と

してそれぞれクラス間の平均距離と全データの分散を用い、クラス内変動としてどちらも各クラス分散の幾何平均を用いている。つまり、LDA と HDA、または、LDA と HLDA は、クラス間変動の定義は同じで、クラス内変動としてクラス分散の算術平均をとるか幾何平均をとるかの違いであると見ることができる。以上より、従来手法はクラス間変動の定義の違いと、クラス内変動の定義の違いで説明づけられる。そして、クラス内変動は各クラス分散の一般化した平均を用いることにより統一した枠組みで見ることができることに注目し、従来手法を一般化した手法 (PLDA) を提案する。PLDA は平均の取り方を決定するための制御パラメータを変更することで、さまざまな特徴変換を行うことができる。PLDA は従来手法である LDA, HLDA と HDA を特殊事例として包含している。また、与えられた音声データに準最適な PLDA の制御パラメータを自動的に決定するために、制御パラメータ選択手法を提案した。さらに、これらの音響特徴変換と、近年音響モデルの学習に広く用いられている識別学習法との組み合わせ効果を調査した。認識実験結果より、PLDA は従来手法を上回る認識性能を達成し、さらに、特徴変換と識別学習を組み合わせることで相乗効果が得られることがわかった。

上記手法はいずれも各クラス間の平均距離などを大きくしながらクラス内変動を小さくするように特徴を変換する手法である。これらは各クラスのデータがひとかたまりになっているとき、すなわち単峰性の時に、各クラスのデータをうまく分離することができる。しかしながら、同一クラスのデータが単峰性ではなく、複数の離れたかたまりとなっているとき、すなわち多峰性の時、従来手法は複数のかたまりを一つの大きなかたまりとして扱うため、各クラス間の平均距離やクラス内変動を正しく評価することができなかった。一般に、音声データは多数の話者の音声が様々な騒音環境下で収録さており、複数のかたまりで構成されていると考えられる。つまり、音声は単峰性ではなく多峰性であると考えられる。そのため、従来手法では必ずしも適切な音響特徴変換となっていなかったと考えられる。このような多峰性データを適切に取り扱うために、いくつかの手法が提案されている。局所性保存射影法（LPP）では、高次元空間上で近くにあるデータは低次元空間でも近くになるように特徴を変換することを特徴とする。これにより、多峰性データの局所性を保存しながら特徴を低次元に変換することができる。LPP は教師なしの手法であったが、LPP と教師あり手法である LDA を組み合わせた局所 LDA も提案されている。本研究では、LPP と HDA を組み合わせた手法と LPP と PLDA を組み合わせた手法を提案する。これらの拡張手法は、クラスが多峰性であってもクラス間変動とクラス内変動を適切に評価できる。その効果を実験により確認した。

次に、音素間の重なりを分類誤差として定義し、特徴変換後の低次元空間で音素間の分類誤差が小さくなるように音響特徴を変換する手法に注目する。音素間の分類誤差が小さくなるように音響特徴変換を行えば、音素の識別性能が保持されるこ

ととなり、その音響特徴を使って音響モデルを構成すれば高い音声認識性能が期待できる。分類誤差を小さくする際は、次の二つを満たすことが重要である。

- クラス間の分類誤差の総和を小さくする
- 極端に分類誤差が大きくなるクラスが生じることを避ける

分類誤差の総和が大きければ複数の音素の識別性能が低くなり音声認識の性能劣化につながる。一方、分類誤差の総和が小さくても、特定の音素の識別性能が極端に悪くなれば、その音素を含む認識対象語彙の性能劣化につながることになる。このアプローチの従来手法は、低次元空間上での全てのクラスペアの分類誤差を測定し、その平均分類誤差を最小にするように音響特徴変換しようとしてきた。この方法では、誤差の総和は小さくできるが、特定音素の識別性能を極端に低下させることを防ぐ枠組みがない。そのため、音響特徴変換により識別性能が大きく劣化する音素があると、その音素を含む認識対象語彙の性能劣化につながるおそれがある。そこで、特定の音素の識別性能が極端に劣化することを抑えるために、クラスペアの中で最大となる誤差を最小にする基準を提案する。この基準では特定音素の誤差が大きくならないことが保証される。さらに、平均誤差を小さくすることと最大誤差を小さくすることは、どちらも一般化した平均を使って説明できることに注目し、これら2つの基準を一般化した。この一般化した基準を使うことで平均誤差を小さくしながら最大誤差を同時に小さくするような特徴変換を行うことができる。認識実験より、提案法は従来手法に比べ高い認識性能を達成できることが示された。