

平成 17 年 1 月 11 日

電子・情報工学専攻	学籍番号	983738
申請者氏名	福田 隆	

指導教員氏名	新田 恒雄 中川 聖一
--------	----------------

論文要旨(博士)

論文題目	A Study on Feature Extraction and Canonicalization for Robust Speech Recognition (音声認識のための頑健な特徴量と正準化に関する研究)
------	--

音声認識技術はこの10年間で飛躍的な進歩を遂げ、カーナビゲーションシステムや電話音声自動応答システムなど、様々な局面で実用化が始まった。音声認識の基本性能は着実に向上しているものの、いまだ多くの問題に直面する。現在の主流である統計的手法では、周囲の雑音や利用環境が、確率的分類器(HMM)の設計範囲を超えると、誤認識が急激に増える。そのため、不明瞭な発声や発話スタイル、周囲の雑音などによる音声パターン変動を吸収する何らかの枠組みが必要である。本論文では、クラス間の尤度差を減少させる要因を特徴抽出の段階で取り除く正準化処理を提案し、それに通ずる頑健な特徴抽出器について述べる。

一般的な音声認識システムは特徴抽出器とそれに続く確率的分類器から構成されている。特徴抽出段階では、MFCCとその時間的変化を表現した動的特徴のセットが多用されている。動的特徴は、時間-スペクトル(TS)パターン上の時間軸に沿った変動を表現しているとなすことができる。実験音声学の成果から、TSパターン上には音韻の種類に対応して、様々な幾何学構造が現れることが示されているが、これまで種々の変動を表現する特徴は利用されてこなかった。本論文では、はじめに、TSパターン上で認識に有効な特徴を抽出し、特徴パラメータに組み込む方法を提案する。

一方、音韻論の分野では、調音様式(母音性、子音性、連続性、...)や調音位置(高舌性、前方性、舌端性、...)を表す音素弁別特徴(DPF)による音素分類が古くから提案されている。音声認識においても古くからDPFの利用が研究されていたが、近年、再びDPFの利用が検討され始めた。しかし、DPF単独で高い性能を得ることが難しいため、MFCCに代わる特徴として広く利用されるに至っていない。本論文では、次に、単一の多層ニューラルネットワーク(MLN)からDPFを抽出すると共に、DPFパラメータ単独で、雑音に頑健な音声認識を実現することを目指す。MLNにより出力されるDPFは多くが対数正規分布に近い分布を示し、また各次元間で相関を持つ。そのため、対角共分散で表現されるHMM分類器の入力としてDPFを利用するには問題がある。そこで、KL変換とGram-Schmidtの直交化法を併用して、DPFパラメータを無相関化する方法を提案する。対角共分散型のHMMを利用する場合、直交化前後の性能差は大きい。他方、DPFを無相関正規分布に変換するのではなく、DPF分布が持つ非対称性をHMMで直接表現する方法も検討する。具体的には、標準的なHMMで用いられる正規分布の代わりに、対数正規分布を利用することを試みる。

確率的分類器(HMM)に基づく認識システムは、性別、話速、音響環境等に起因するHMMの隠れ変数のバイアスにより性能が低下する。この影響はDPFパラメータを利用する際にも現れる。より頑健な認識システムを構成するには、隠れ変数の影響を特徴抽出の段階で吸収することが望ましい。本論文では、最後に、雑音強度と男女の声質を対象とした特徴パラメータ正準化方式について述べる。正準化は音響特徴空間と音響モデルの間に、中間表現である音素弁別特徴(DPF)空間を導入することではじめて実現する。正準化器は、音響的変動要因に対応させた複数のDPF抽出器と一つのDPF選択器から成る。雑音に頑健なDPFパラメータを介する正準化処理は、隠れ変数の影響を極めて少なく抑えることができる。