

論文内容の要旨

博士論文題目

氏 名

(論文内容の要旨)

音声認識システムの性能は音響モデルの精度と頑健性に大きく依存する。本論文では、音響モデルの性能向上のため、ゆう度を用いたコンテキスト依存音響モデルの構築手法を提案する。また、従来の音声認識には利用されていない韻律情報を用いた新たな認識手法を提案する。

まず、日本語音声の継続時間をモーラ単位でモデル化し、CYKを一般化した枠組を用い音響ゆう度の計算と構文解析を統一的に扱う手法を提案する。更に、ピッチとスペクトル情報間の相関を用いた正規化により認識クラス内の音声の音響的特徴の変動を削減し、クラス間の識別性能の向上を図る。

不特定話者音声データを基に、音素コンテキストと時間方向への変動を考慮した音響モデルを構築するため、ゆう度を基準としたHMMの状態分割と決定木に基づくHMMの状態分割を統合した、最ゆう推定逐次状態分割法(ML-SSS)を提案する。提案手法を自由発話音声認識に適用し、ATR自由発話データベースを用いた音声認識実験によりその有効性を示す。

最後に、研究用ソフトウェアの実システムへの導入を容易にするため、2つの異なる統合手法を提案する。この手法に基づき2つの音声認識システム(MIKOSHI、ATRSPREC)を構築した。MIKOSHIはソフトウェアのモジュール化を進めると共に認識の実時間処理を可能としている。また、ATRSPRECに於いては更に各モジュールをイベントループに組み込むことで、新規機能の追加を容易にしており、本システムはATR日英音声翻訳システム(ATR-MATRIX)のフロントエンドとして機能している。

氏 名	Harald Singer
-----	---------------

(論文審査結果の要旨)

音声による人間と機械のインタフェースは、人間にとって最も自然なインタフェースである。近年、音声による人間と機械とのインタフェース、特に、音声の認識の研究については、極めて高い関心を持って研究が推進されている。本論文では、自由に発話された音声を入力とする他言語間翻訳通信システムを実現するための音声認識部に関して行なわれた基礎研究について述べられている。具体的には、「音声認識のための音響モデル」に関する研究に焦点が当てられている。

本論文の成果は、次のように要約できる。

1. これまで音声認識には利用されていなかった韻律情報を用いた新たな認識手法を提案している。まず、日本語音声の継続時間をモーラ単位でモデル化し、CYKを一般化した枠組を用い音響尤度の計算と構文解析を統一的に扱う手法を提案している。更に、ピッチとスペクトル情報間の相関を用いた正規化により認識クラス内の音声の音響的特徴の変動を削減し、クラス間の識別性能の向上を図る方法について提案し有効性を確認している。
2. 音声認識システムの性能は音響モデルの精度と頑健性に大きく依存する。本論文では、音響モデルの性能向上のため、尤度を用いた前後音素のコンテキスト依存の音響モデルの構築手法を新規に提案している。不特定話者音声データを基に、音素コンテキストと時間方向への変動を考慮した音響モデルを構築するため、ゆう度を基準としたHMMの状態分割と決定木に基づくHMMの状態分割を統合した、最ゆう推定逐次状態分割法(ML-SSS)を提案している。さらに、提案手法を自由発話音声認識に適用し、ATR自由発話データベースを用いた音声認識実験によりその有効性を示している。
3. 最後に、研究用ソフトウェアの実システムへの導入を容易にするため、2つの異なる統合手法を提案している。この手法に基づき2つの音声認識システム(MIKOSHI、ATRSPREC)を構築した。MIKOSHIはソフトウェアのモジュール化を進めると共に認識の実時間処理を可能としている。また、ATRSPRECに於いては更に各モジュールをイベントループに組み込むことで、新規機能の追加を容易にしており、本システムはATR日英音声翻訳システム(ATR-MATRIX)のフロントエンドとして機能している。

以上のように、本論文は音声認識における音響モデルの高度化を目的として、上記の点について新たな提案を行ない、有効性を示した。これらの研究成果は、情報科学、音情報処理学の分野における学術的な貢献として高く評価できる。平成9年12月25日開催の公聴会の結果を参考に、本博士論文の審査を行なった結果、本論文は博士(工学)の学位論文として十分に価値のあるものと判断する。