

論文内容の要旨

博士論文題目

氏 名

(論文内容の要旨)

本論文では、音声インタフェースに対する強いニーズが存在する、車載機器の音声によるコマンド・コントロールのアプリケーションに焦点をあて、車室内における加法性雑音、乗法性歪みに頑健な音声認識を実現するための方法を提案する。

まず、話者毎に音声フレームと非音声フレームで別々にケプストラム平均を求めることにより、正規化音声スペクトルに対する乗法性歪み(話者の個人性、発話様式、空間伝達特性、マイク特性)を一括して補正できるアルゴリズム E-CMN(Exact Cepstrum Mean Normalization)を提案する。不特定話者用のHMM(Hidden Markov Model)を音源未知の加法性雑音に適応化させるHMM合成法と E-CMN 法を組合せ、繰り返し演算によることなく、HMMを乗法性歪みと音源未知の加法性雑音に適応化させる E-CMN/PMC 法を提案する。

次に、音声/非音声フレームを区別せず、連続的に雑音スペクトルを推定し、それを入力スペクトルから減じる連続スペクトル減算法(CSS:Continuous Spectral Subtraction)に着目し、従来法であるスペクトル減算法(SS)や、最小平均2乗誤差推定法(MMSE:Minimum Mean Square Error Estimation)と比べて、音源未知の加法性雑音に対し、低SNRにおいて、E-CMN/PMC法よりも頑健であることを示す。

さらに、音源未知の加法性雑音に埋もれた音声をフレーム毎に正確に検出する音声検出法(VAD:Voice Activity Detection)を利用して、学習同定法(NLMS)のフィルター係数の推定処理を制御するNLMS-VAD法を提案する。ロードノイズなどの音源未知の加法性雑音が存在する環境であっても、スピーカ音などの音源既知の加法性雑音が効率良く除去できることを示す。さらに、E-CMN/CSS法との組合せにより加法性雑音、乗法性歪みに対処できることを示す。

最後に、逐次的な話者適応方式を提案し、最大事後確率推定法やEMアルゴリズムによる方法と比較して、高い話者適応性能が得られることを示す。

(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。音声認識における一つの大きな問題は、実環境における雑音と歪みの問題である。本論文では、自動車内での音声認識に焦点を絞り、その環境における加法性雑音と乗法性の歪みに対処し、頑健な音声認識を実現するための基礎研究について述べられている。

本論文の成果は、次のように要約できる。

1. 実環境における種々の環境要因、話者性、音韻性などのモデル化を行ない、正規化スペクトルを用いることの提案を行なっている。さらに、話者毎に音声フレームと非音声フレームで別々にケプストラム平均を求めることにより、正規化音声スペクトルに対する乗法性歪み(話者の個人性、発話様式、空間伝達特性、マイク特性)を一括して補正できるアルゴリズム E-CMN(Exact Cepstrum Mean Normalization) を提案している。さらに、不特定話者用のHMM (Hidden Markov Model) を音源未知の加法性雑音に適応化させるHMM合成法と E-CMN法を組合せ、繰り返し演算によることなく、HMMを乗法性歪みと音源未知の加法性雑音に適応化させる E-CMN/PMC法を提案し、有効性を実験により確認している。
2. 音声/非音声フレームを区別せず、連続的に雑音スペクトルを推定し、それを入力スペクトルから減じる連続スペクトル減算法 (CSS:Continuous Spectral Subtraction) に着目し、従来法であるスペクトル減算法 (SS) や、最小平均2乗誤差推定法 (MMSE:Minimum Mean Square Error Estimation) と比べて、音源未知の加法性雑音に対し、低SNRにおいて、E-CMN/PMC法よりも頑健であることを示している。
3. 音源未知の加法性雑音に埋もれた音声をフレーム毎に正確に検出する音声検出法 (VAD:Voice Activity Detection) を利用して、学習同定法 (NLMS) のフィルター係数の推定処理を制御する NLMS-VAD法を提案している。ロードノイズなどの音源未知の加法性雑音が存在する環境であっても、スピーカ音などの音源既知の加法性雑音が効率良く除去できることを示し、さらに、E-CMN/CSS法との組合せにより加法性雑音、乗法性歪みに対処できることを示している。
4. 逐次的な話者適応方式を提案し、最大事後確率推定法や EM アルゴリズムによる方法と比較して、高い話者適応性能が得られることを示している。

以上のように、本論文は音声認識における環境に対する頑健性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。