

学 位 論 文 題 名

変調スペクトルの特徴を考慮した音声認識における
雑音に頑健な特徴抽出に関する研究

学位論文内容の要旨

第 1 章では研究の背景と目的について述べる。近年のパーソナルコンピュータや PDA(Personal Digital Assistants), 携帯電話などの情報機器が急速に普及し, より身近なものとなってきた。しかしながら, その操作性はお年寄りや小さい子供にとっては必ずしも使いやすいとは限らない。また, カーナビゲーションシステムをはじめとする車載機器などは目や手が自由に使えない特殊な環境で操作を強いられる。これら現状のインターフェースが持つ課題を解決する方法の一つとして, 音声を用いることが挙げられる。音声インターフェースは他のインターフェースに比べ情報処理速度, 習得時間の面で優れており, 音声認識に寄せられる期待は大きい。このような背景において, 音声認識は確立モデルの導入による技術的進歩とコンピュータの処理能力の著しい向上により急速に発達し, 音声認識を組み込んだ製品も見かけるようになってきた。しかし, 依然として音声認識技術を用いたインターフェースは本格的な普及に至っていないのが現状である。その原因は認識性能の不十分さ, 特に雑音環境下での認識性能の不十分さにある。実際の製品でも, 静かな環境では高精度の認識が可能だが雑音の影響が大きい環境では頻繁に誤動作を起こし使いものにならないことも少なくない。

第 2 章では音声認識の概要に触れる。音声認識は音響分析部と音声認識部に分けて考えることができる。我々のシステムで用いた音響分析法である MFCC(Mel Frequency Cepstral Coefficients)と, 音声認識部で用いる HMM(Hidden Markov Model)について詳しく説明する。

第 3 章ではこれまで行われてきた耐雑音性向上のための手法を紹介する。音として入り音声に影響を与える雑音を加法性雑音, マイクロフォン特性の違いなど伝達特性に与える雑音を乗法性雑音と呼び, それぞれ様々な対策法が提案されてきた。中でも加法性雑音に対してはスペクトルサブトラクション法, 乗法性雑音に対してはケプストラム平均除去法が簡便な手法のわりに効果が大きい。そのため多くの認識システムで採用されてきた。認識実験によりその性能の評価を行う。

広く用いられているスペクトルサブトラクション法は雑音の種類や雑音の量により安定した認識ができない場合があり大きな問題となっている。そこで第 4 章では, 雑音の種類や雑音の量に依存しない新たな雑音ロバスト手法を提案した。具体的には変調スペクトルと呼ばれる特徴量などの時系列データを周波数領域で表現したものを効果的に加工するランニングスペクトルフィルタリング(RSF)法と雑音の影響により減少したケプストラムのダイナミックレンジを補正するダイナミック

レンジ調整(DRA)法を組み合わせた手法である。それぞれの手法について従来法に比べ認識性能が向上する理論的根拠を述べ、孤立単語音声認識実験によりその有効性を確認した。

RSF 法は孤立単語認識実験において高 SNR 環境下では認識精度の向上度が低く、無雑音環境下では逆に認識率を低下させることになった。これは学習と認識の分析アルゴリズムの違いによるミスマッチに起因している。第 5 章ではこのミスマッチを改善すべく、推定 SNR に基づき RSF 法の実行を調整する拡張 RSF 法を提案した。拡張 RSF 法は推定した SNR の値がある閾値を超える場合にのみフィルタリングを実行する手法で、学習・認識ともに同じアルゴリズムで分析を行える。認識実験により RSF 法の問題点であった高 SNR 環境下での認識精度がさらに向上し、無雑音環境下での認識率も低下させず高い性能を保持することを示した。

学位論文審査の要旨

主 査 教 授 宮 永 喜 一

副 査 教 授 野 島 俊 雄

副 査 教 授 小 柴 正 則

副 査 教 授 小 川 恭 孝

学 位 論 文 題 名

変調スペクトルの特徴を考慮した音声認識における 雑音に頑健な特徴抽出に関する研究

近年のパーソナルコンピュータやPDA(Personal Digital Assistants), 携帯電話などの情報機器が急速に普及し, より身近なものとなってきた。しかしながら, その操作性はお年寄りや小さい子供にとっては必ずしも使いやすいとは限らない。また, カーナビゲーションシステムをはじめとする車載機器などは目や手が自由に使えない特殊な環境で操作を強いられる。これら現状のインターフェースが持つ課題を解決する方法の一つとして, 音声を用いることが挙げられる。音声インターフェースは他のインターフェースに比べ情報処理速度, 習得時間の面で優れており, 音声認識に寄せられる期待は大きい。このような背景において, 音声認識は確立モデルの導入による技術的進歩とコンピュータの処理能力の著しい向上により急速に発達し, 音声認識を組み込んだ製品も見かけるようになってきた。しかし, 依然として音声認識技術を用いたインターフェースは本格的な普及に至っていないのが現状である。その原因は認識性能の不十分さ, 特に雑音環境下での認識性能の不十分さにある。実際の製品でも, 静かな環境では高精度の認識が可能だが雑音の影響が大きい環境では頻繁に誤動作を起こし使いものにならないことも少なくない。

第1章では本論文の研究の背景と目的について述べている。

第2章では音声認識の概要に触れている。音声認識は音響分析部と音声認識部に分けて考えることができる。我々のシステムで用いた音響分析法である MFCC(Mel Frequency Cepstral Coefficients)と, 音声認識部で用いる HMM(Hidden Markov Model)について詳しく説明している。

第3章ではこれまで行われてきた耐雑音性向上のための手法を紹介している。音として入り音声に影響を与える雑音を加法性雑音, マイクロフォン特性の違いなど伝達特性に与える雑音を乗法性雑音と呼び, それぞれ様々な対策法が提案されてきた。中でも加法性雑音に対してはスペクトルサブトラクション法, 乗法性雑音に対してはケプストラム平均除去法が簡便な手法のわりに効果が大いいため多くの認識システムで採用されてきた。認識実験によりその性能の評価を行った。

広く用いられているスペクトルサブトラクション法は雑音の種類や雑音の量により安定した認識

ができない場合があり大きな問題となっている。そこで第4章では、雑音の種類や雑音の量に依存しない新たな雑音ロバスト手法を提案している。具体的には変調スペクトルと呼ばれる特徴量などの時系列データを周波数領域で表現したものを効果的に加工するランニングスペクトルフィルタリング(RSF)法と雑音の影響により減少したケプストラムのダイナミックレンジを補正するダイナミックレンジ調整(DRA)法を組み合わせた手法である。それぞれの手法について従来法に比べ認識性能が向上する理論的根拠を述べ、孤立単語音声認識実験によりその有効性を確認している。

RSF法は孤立単語認識実験において高SNR環境下では認識精度の向上度が低く、無雑音環境下では逆に認識率を低下させることになった。これは学習と認識の分析アルゴリズムの違いによるミスマッチに起因している。第5章ではこのミスマッチを改善すべく、推定SNRに基づきRSF法の実行を調整する拡張RSF法を提案している。拡張RSF法は推定したSNRの値がある閾値を超える場合にのみフィルタリングを実行する手法で、学習・認識ともに同じアルゴリズムで分析を行える。認識実験によりRSF法の問題点であった高SNR環境下での認識精度がさらに向上し、無雑音環境下での認識率も低下させず高い性能を保持することを示した。

以上の点より、本論文の目的である、種々の雑音に対するロバスト音声認識システムの研究において、十分な成果を挙げている。

これを要するに、筆者は、新たなロバスト音声認識手法の提案とその開発を行い、種々の雑音に有効なロバスト性を持つ新しい信号処理技術を実現し、その有効性を示した。これにより、音声情報処理・音声認識システムの開発・実現に関する多くの有益な知見を得ており、情報科学の分野に貢献するところ大なるものがある。

よって筆者は、北海道大学博士(情報科学)の学位を授与される資格あるものと認める。