学位論文題名

A Study on a Robust Speech Recognition System against Various Additive Noises

(種々の雑音に対する雑音ロバスト音声認識に関する研究)

学位論文内容の要旨

本論文の目的は、種々の雑音に対するハンズフリーなロバスト音声認識システムについて研究することにある。近年音声認識は、手を使わずに機械を操作することを可能とする次世代のインタフェースとして注目を集めており、カーナビゲーションシステムやビデオゲーム、ペット型ロボットなどの製品において実用化が始まっているが、現在そのような製品の多くでは、十分な認識性能が得られていないことが問題となっている。その原因として、実環境における騒音が挙げられる。音声認識とは基本的にパターン認識であり、ここでは入力された音声データと事前に学習されたデータの比較が行われるが、実環境においては、車の走行音や人の話し声などの種々の雑音が入力音声に付加されることが予想され、これが学習データとの間に差異を生じ、認識性能を劣化させることが指摘されている。そのため、実用的な音声認識システムにおいて、システムの耐雑音性は極めて重要な要素である。

雑音の影響を低減する分析手法には二つのアプローチがある。一つはマイクロホンアレーと呼ばれる、複数の音声入力を録音し、それらの差異を用いて話者方向の音声成分を強調する手法である。しかしながら、本手法は複数のマイクロホンを使用するため装置が大掛かりになり、携帯端末などの小型の機器に搭載するのが難しくなってしまう。もう一つのアプローチは、単一の音声入力から音声成分や雑音成分を予測し、フィルタリングやサブトラクションによりそれらの成分の強調あるいは低減を行うものである。本アプローチにおいて最も代表的な手法が、非音声部分の入力信号より雑音成分のスペクトルを予測し、全体の入力音声スペクトルより減算を行うスペクトルサブトラクション(Spectral Subtraction: SS)であるが、雑音成分の予測の正確さが雑音の種類や大きさに依存し、スペクトルの引き過ぎや引き残しによりミュージカルノイズと呼ばれる新たな雑音を生じる問題が指摘されている。また従来手法では、ドアの開閉音やアラームなどの突発的な雑音を予測して、それに対応した分析を行うのは難しい。そのため提案手法は、様々な雑音に柔軟に対応することのできるアルゴリズムであることが求められる。

また、本論文の目的を達成する上で、頑健な音声区間検出の実現は避けることのできない重要な課題である。それはもちろん音声認識の利点であるハンズフリーなインタフェースを実現するため

であるが、同時に音声区間検出による頑健な音声認識を実現するためでもある。現在実用化がなされている製品の中には、発話の前にボタンを押すなどの操作を必要としているものも多く含まれているが、この方式ではタイミングのずれにより発話の一部が欠損してしまい、認識性能が損なわれる恐れがある。さらに録音された区間に含まれる余分な雑音区間による影響も考えられるため、音声認識システムの実装を考えたとき、雑音に頑健な音声区間検出の必要性は非常に大きいものがある。

これらの点を踏まえた研究の成果を以下にまとめることができる。

- ・メルケプストラム(Mel Frequency Cepstral Coefficients: MFCC)の差分データと MFCC 分析時のフィルタバンク出力を基準値に用いることにより、必要な音声特徴成分を強調し、通常のスペクトルなどを用いた手法よりもより雑音に頑健な音声区間検出を実現した。さらに本手法は、音声認識そのものの認識正解率の改善にも貢献した。

- ・突発的な雑音の影響を低減するため、そのような雑音の区間を飛ばすバースト雑音スキップアルゴリズムを提案した。雑音区間の推定には自己回帰移動平均(Auto Regressive Moving Average: ARMA)モデルを用いた原音声信号予測における予測誤差を用い、原音声信号が区間検出に影響を与えないようにした。これにより本手法は、突発的な雑音が音声信号と同程度の大きさであっても、効果的にその区間を検出し、影響を低減することを可能とした。

- ・様々な定常的な雑音に対する頑健な音声分析手法として、変調スペクトルにおいて重み付けを行うことで音声成分を抽出する変調スペクトルコントロール(Modulation Spectral Control: MSC)と、雑音によるゲインの減少を補正するケプストラム振幅正規化(Cepstral Amplitude Range Normalization: CARN)を組み合わせた手法を提案した。本手法は SS 法に比べてさまざまな環境で有効に働くことが音声認識実験により確認された。さらに本手法と SS 法の組み合わせについても調査を行い、この組み合わせにおいて SS 法が本手法の雑音環境に対する柔軟性を損なうことなく、さらに認識性能を向上させることを確認した。最終的に、この SS と提案手法の組み合わせ法は、SN 比 10dB 環境において 90%以上の認識正解率を示し、耐雑音音声分析を適用しない音声認識システムに比べて、37%余りの認識正解率の改善をもたらした。

以上の点より、本論文の目的である、種々の雑音に対するハンズフリーなロボスト音声認識システムの研究において、十分な成果を挙げる事ができた。

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

A Study on a Robust Speech Recognition System against Various Additive Noises

(種々の雑音に対する雑音ロバスト音声認識に関する研究)

本論文の目的は、種々の雑音に対するハンズフリーなロバスト音声認識システムについて研究することにある。近年音声認識は、手を使わずに機械を操作することを可能とする次世代のインタフェースとして注目を集めており、カーナビゲーションシステムやビデオゲーム、ペット型ロボットなどの製品において実用化が始まっているが、現在そのような製品の多くでは、十分な認識性能が得られていないことが問題となっている。その原因として、実環境における騒音が挙げられる。音声認識とは基本的にパターン認識であり、そこでは入力された音声データと事前に学習されたデータの比較が行われるが、実環境においては、車の走行音や人の話し声などの種々の雑音が入力音声に付加されることが予想され、これが学習データとの間に差異を生じ、認識性能を劣化させることが指摘されている。そのため、実用的な音声認識システムにおいて、システムの耐雑音性は極めて重要な要素である。

雑音の影響を低減する分析手法には二つのアプローチがある。一つはマイクロホンアレーと呼ばれる、複数の音声入力を録音し、それらの差異を用いて話者方向の音声成分を強調する手法である。しかしながら、本手法は複数のマイクロホンを使用するため装置が大掛かりになり、携帯端末などの小型の機器に搭載するのが難しくなってしまう。もう一つのアプローチは、単一の音声入力から音声成分や雑音成分を予測し、フィルタリングやサブトラクションによりそれらの成分の強調あるいは低減を行うものである。本アプローチにおいて最も代表的な手法が、非音声部分の入力信号より雑音成分のスペクトルを予測し、全体の入力音声スペクトルより減算を行うスペクトルサブトラクション(Spectral Subtraction: SS)であるが、雑音成分の予測の正確さが雑音の種類や大きさに依存し、スペクトルの引き過ぎや引き残しによりミュージカルノイズと呼ばれる新たな雑音を生じる問題が指摘されている。また従来手法では、ドアの開閉音やアラームなどの突発的な雑音を予測して、それに対応した分析を行うのは難しい。そのため提案手法は、様々な雑音に柔軟に対応することの

できるアルゴリズムであることが求められる。

また、本論文の目的を達成する上で、頑健な音声区間検出の実現は避けることのできない重要な課題である。それはもちろん音声認識の利点であるハンズフリーなインタフェースを実現するためであるが、同時に音声区間検出による頑健な音声認識を実現するためでもある。現在実用化がなされている製品の中には、発話の前にボタンを押すなどの操作を必要としているものも多く含まれているが、この方式ではタイミングのずれにより発話の一部が欠損してしまい、認識性能が損なわれる恐れがある。さらに録音された区間に含まれる余分な雑音区間による影響も考えられるため、音声認識システムの実装を考えたとき、雑音に頑健な音声区間検出の必要性は非常に大きいものがある。

これらの点を踏まえ、本論文での研究成果を以下にまとめることができる。

- ・メルケプストラム(Mel Frequency Cepstral Coefficients: MFCC)の差分データと MFCC 分析時のフィルタバンク出力を基準値に用いることにより、必要な音声特徴成分を強調し、通常のスペクトルなどを用いた手法よりもより雑音に頑健な音声区間検出を実現した。さらに本手法は、音声認識そのものの認識正解率の改善にも貢献した。

- ・突発的な雑音の影響を低減するため、そのような雑音の区間を飛ばすバースト雑音スキップアルゴリズムを提案した。雑音区間の推定には自己回帰移動平均(Auto Regressive Moving Average: ARMA)モデルを用いた原音声信号予測における予測誤差を用い、原音声信号が区間検出に影響を与えないようにした。これにより本手法は、突発的な雑音が音声信号と同程度の大きさであっても、効果的にその区間を検出し、影響を低減することを可能とした。

- ・様々な定常的な雑音に対する頑健な音声分析手法として、変調スペクトルにおいて重み付けを行うことで音声成分を抽出する変調スペクトルコントロール(Modulation Spectral Control: MSC)と、雑音によるゲインの減少を補正するケプストラム振幅正規化(Cepstral Amplitude Range Normalization: CARN)を組み合わせた手法を提案した。本手法は SS 法に比べてさまざまな環境で有効に働くことが音声認識実験により確認された。さらに本手法と SS 法の組み合わせについても調査を行い、この組み合わせにおいて SS 法が本手法の雑音環境に対する柔軟性を損なうことなく、さらに認識性能を向上させることを確認した。最終的に、この SS と提案手法の組み合わせ法は、SN 比 10dB 環境において 90%以上の認識正解率を示し、耐雑音音声分析を適用しない音声認識システムに比べて、37%余りの認識正解率の改善をもたらした。

以上の点より、本論文の目的である、種々の雑音に対するハンズフリーな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、電子情報工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(工学)の学位を授与される資格あるものと認める。