VLSI Architecture of Robust Speech Recognition System

(ロバスト音声認識システムの VLSI アーキテクチャ)

学位論文内容の要旨

現在、高度情報処理システムにおいて、音声、画像処理技術に基づくメディア処理は不可欠なものとなりつつある。その中でも高度ヒューマンインターフェイスとして期待される分野として音声認識がある。現在の音声認識システムはコンピュータ上で動作し、比較的静かな室内環境で使用する場合が多いが、今後は室外環境へと使用範囲の拡大することが期待されている。しかしながら、室外環境の場合では、騒音(人の話し声、自動車走行など)、マイク性能の低下、話者とマイク間の距離の拡大など、音声認識システムの動作環境が悪いことが前提となり、そのような劣悪な環境でも認識精度を維持することが必要である。また、室外では認識装置は携帯できるように小型でかつ極低消費電力であることが求められる。携帯機器ではバッテリー容量の制限があるので、バッテリー消費を抑えるための低電力化が重要である。また、音声認識装置を音声入力インターフェイスとして利用する場合、例えば、音声入力による機器操作、対話によるロボットの操作では、認識精度とも応答速度が求められる。つまり、認識処理を短時間で終わることが重要となる。

小型化、低消費電力化、高速処理を実現する有力な方法として、音声認識処理を汎用プロセッサで実行するソフトウェア処理から認識処理に特化した専用回路を用いてハードウェア処理を移行する方法がある。本研究では、ハードウェア処理で行うことを前提とした音声認識処理 VLSI アーキテクチャを取り扱う。また、ハードウェア処理の観点から、これまでソフトウェア処理では計算時間の増大により困難であった高精度な雑音ロバストアルゴリズムを導入する。本研究では音声入力から認識結果出力までの全ての処理を実行する回路を1個のチップに実装することを目標として、以下の項目について研究を行った。

- ・高速処理及び低消費電力を実現する音声認識処理 VLSI アーキテクチャの設計
- ・多くの雑音環境に対して認識精度を維持することのできるロバストアルゴリズムとそのアルゴリズムの VLSI 実装
- ・音声認識、音声分析、ロバスト処理を統合したロバスト音声認識システムの VLSI 設計

これまでの多くの音声認識の研究はソフトウェア上での実装を前提としていたため、音声認識処理回路アーキテクチャの研究例はこれまでに数件程度しかない。その研究例も認識計算処理の一部のみをハードウェアで実装し、他の部分は汎用プロセッサで演算処理を行うことを想定している。先行研究と比較しての本研究の新規性は、(1) 認識処理全体を実行することのできる VLSI アーキテク

チャを提案し、提案アーキテクチャと汎用プロセッサの認識処理での処理速度、消費電力の評価比較を行い、定量的な効果を示していること。(2) 従来のソフトウェア処理の観点からは困難と考えられていた新しいロバストアルゴリズムを提案し、回路に実装したこと。(3) マイクからの音声入力から認識結果の出力までロバスト処理を含めたロバスト音声認識チップや検証用ボードの試作を報告したこと。以上の3点が挙げられる。上記の研究項目に対する成果は以下にまとめられる。

- ・音声認識処理 VLSI アーキテクチャでは隠れマルコフモデル法単語音声認識アルゴリズムのハードウェア化について検討を行った。隠れマルコフモデルの計算手順について、並行処理とパイプライン処理が可能でかつハードウェア効率の高い計算方法を提示し、上記の計算手順に基づいた回路アーキテクチャを提案した。提案アーキテクチャはメモリでのデータ記憶処理回数を削減し、並列処理とメモリ部に対するゲーテッドクロックの導入による。性能評価では汎用プロセッサで同じ演算を処理する場合と比べて、処理時間が約 1/50、消費エネルギー量が約 1/10 となることを示した。

- ・ロバスト音声認識アルゴリズムでは、ケプストラム(パワースペクトラムに対して対数変換、逆フーリエ変換を行う)領域の時間軌跡に着目した手法を提案した。ケプストラム時間軌跡における最大値と最小値間の距離を規格化することにより、加法性雑音の影響を低減できることを提示し、同じく時間軌跡に対するバンドパスフィルタをかける手法を組み合わせる。この提案手法が従来手法と比較して多くの種類での雑音環境下で高い認識精度をもつことを示し、また、元来の認識性能(雑音がない場合での認識精度)を劣化させないことを評価実験により示した。

- ・ロバスト音声認識システムの VLSI 設計では、上記の隠れマルコフモデル法認識処理回路とケプストラム領域ロバスト処理回路とメルケプストラム分析による音声分析回路を統合した音声認識システムを設計した。設計した回路はデジタル CMOS チップ上にて実装した。また、音声認識システムの動作検証とハードウェア設計を支援するものとして、設計支援 CAD の開発と FPGA ボードによる評価ボードの構築を行い、その実装例を報告した。

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

VLSI Architecture of Robust Speech Recognition System

(ロバスト音声認識システムの VLSI アーキテクチャ)

現在、高度情報処理システムにおいて、音声、画像処理技術に基づくメディア処理は不可欠なものとなりつつある。その中でも高度ヒューマンインターフェイスとして期待される分野として音声認識がある。現在の音声認識システムはコンピュータ上で動作し、比較的静かな室内環境で使用するが多いが、今後は室外環境へと使用範囲の拡大することが期待されている。しかしながら、室外環境の場合では、騒音(人の話し声、自動車走行など)、マイク性能の低下、話者とマイク間の距離の拡大など、音声認識システムの動作環境が悪いことが前提となり、そのような劣悪な環境でも認識精度を維持することが必要である。また、室外では認識装置は携帯できるように小型でかつ極低消費電力であることが求められる。携帯機器ではバッテリー容量の制限があるので、バッテリー消費を抑えるための低電力化が重要である。また、音声認識装置を音声入力インターフェイスとして利用する場合、例えば、音声入力による機器操作、対話によるロボットの操作では、認識精度とも応答速度が求められる。つまり、認識処理を短時間で終えることが重要となる。

小型化、低消費電力化、高速処理を実現する有力な方法として、音声認識処理を汎用プロセッサで実行するソフトウェア処理から認識処理に特化した専用回路を用いてハードウェア処理を移行する方法がある。本研究では、ハードウェア処理で行うことを前提とした音声認識処理 VLSI アーキテクチャを取り扱う。また、ハードウェア処理の観点から、これまでソフトウェア処理では計算時間の増大により困難であった高精度な雑音ロバストアルゴリズムを導入する。本研究では音声入力から認識結果出力までの全ての処理を実行する回路を1個のチップに実装することを目標として、以下の項目について研究を行った。

- ・高速処理及び低消費電力を実現する音声認識処理 VLSI アーキテクチャの設計
- ・多くの雑音環境に対して認識精度を維持することのできるロバストアルゴリズムとそのアルゴリズムの VLSI 実装

・音声認識，音声分析，ロバスト処理を統合したロバスト音声認識システムの VLSI 設計

これまでの多くの音声認識の研究はソフトウェア上での実装を前提としていたため，音声認識処理回路アーキテクチャの研究例はこれまでに数件程度しかない．その研究例も認識計算処理の一部のみをハードウェアで実装し，他の部分は汎用プロセッサで演算処理を行うことを想定している．先行研究と比較しての本研究の新規性は，(1) 認識処理全体を実行することのできる VLSI アーキテクチャを提案し，提案アーキテクチャと汎用プロセッサの認識処理での処理速度，消費電力の評価比較を行い，定量的な効果を示していること．(2) 従来のソフトウェア処理の観点からは困難と考えられていた新しいロバストアルゴリズムを提案し，回路に実装したこと．(3) マイクからの音声入力から認識結果の出力までロバスト処理を含めたロバスト音声認識チップや検証用ボードの試作を報告したこと．以上の3点が挙げられる．上記の研究項目に対する成果は以下にまとめられる．

ロバスト音声認識アルゴリズムでは，ケプストラム(パワースペクトラムに対して対数変換，逆フーリエ変換を行う)領域の時間軌跡に着目した手法を提案した．ケプストラム時間軌跡における最大値と最小値間の距離を規格化することにより，加法性雑音の影響を低減できることを提示し，同じく時間軌跡に対するバンドパスフィルタをかける手法を組み合わせる．この提案手法が従来手法と比較して多くの種類での雑音環境下で高い認識精度をもつことを示し，また，元来の認識性能(雑音がない場合での認識精度)を劣化させないことを評価実験により示した．(第3章)

音声認識処理 VLSI アーキテクチャでは隠れマルコフモデル法単語音声認識アルゴリズムのハードウェア化について検討を行った．隠れマルコフモデルの計算手順について，並行処理とパイプライン処理が可能でかつハードウェア効率の高い計算方法を提示し，上記の計算手順に基づいた回路アーキテクチャを提案した．提案アーキテクチャはメモリでのデータ記憶処理回数を削減し，並列処理とメモリ部に対するゲーテッドクロックの導入による．性能評価では汎用プロセッサで同じ演算を処理する場合と比べて，処理時間が約 1/50，消費エネルギー量が約 1/10 となることを示した．(第4章)

ロバスト音声認識システムの VLSI 設計では，上記の隠れマルコフモデル法認識処理回路とケプストラム領域ロバスト処理回路とメルケプストラム分析による音声分析回路を統合した音声認識システムを設計した．設計した回路はデジタル CMOS チップ上にて実装した．また，音声認識システムの動作検証とハードウェア設計を支援するものとして，設計支援 CAD の開発と FPGA ボードによる評価ボードの構築を行い，その実装例を報告した．(第5章)

これを要するに，筆者は，新たなロバスト音声認識手法とその LSI 実現の検討を行い，種々の雑音に有効なロバスト性を持つ新しいフィルタを求め，その有効性を示した．さらに，そのハードウェア実現を行い，実時間音声認識システムの開発を行った．これにより，音声情報処理・実時間システム LSI の設計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．