氏名	中井（なかい） 満（みつる）
授与学位	博士（工学）
学位授与年月日	平成8年3月26日
学位授与の根拠法規	学位規則第4条第1項
研究科，専攻の名称	東北大学大学院工学研究科 （博士課程）電気及通信工学専攻
学位論文題目	**韻律構造を利用した連続音声認識に関する研究**
指導教官	東北大学教授 阿曽 弘具
論文審査委員	東北大学教授 阿曽 弘具　東北大学教授 曽根 敏夫 東北大学教授 阿部 健一　東北大学教授 牧野 正三（大計センター） 北陸先端科学技術大学院大学教授 木村 正行

論文内容要旨

第1章 序論

連続音声認識は，より知的なマン・マシン・インターフェースを実現するための必須技術として期待されている。しかし，孤立単語音声と比較とする連続音声は構文構造や発話様式などにおいて複雑なため，認識・理解に膨大な処理時間とメモリを費しているのが現状である。それゆえに，認識精度や処理効率の向上のためにはスペクトルなどの音響的情報に加えて，文脈情報や韻律情報等の支援が必要であり，単語や文節等の句境界推定は重要な課題となっている。

本論文では基本周波数パターンなどの韻律特徴量を利用し，統語構造の文節句に類似したアクセント句を単位とした韻律構造について，

1) 連続音声認識の前処理・後処理に有効である句境界位置，
2) 実時間性を要する連続音声認識に利用可能な時間同期出力型の句境界尤度

の2点について，その推定法を提案し，これらの韻律構造を用いた連続音声認識システムを与え，韻律情報の有効性を明らかにしている。

第2章 韻律構造と韻律特徴量抽出

韻律構造の基本単位であるアクセント句は，基本周波数（F_0）パターン上の一つの山型に起伏するパターンとして明瞭に現れるため，韻律構造推定の 特徴量としてF_0パターンを用いることが多い。本章ではアクセント句境界検出に使用する韻律特徴量として，F_0パターン，ΔF_0パターン，F_0信頼度を定義し，既存の手法を用いた特徴抽出についてまとめている。

また，藤崎らによって定式化されているF_0パターンの生成モデル（3章で用いる）について概要を説明している。このモデルでは，F_0パターンは話者特有のバイアス成分（F_{min}）に文頭から文末へと降下するフレーズ成分と，局所的な起伏を形成するアクセント成分を重畳したものであり，発声区間を連続的に変化するパターンとして生成される。

F_0分析にはラグ窓法を用い，パターンの連続性を重視して有声・無声の区別なしに抽出を行う。分析の信頼度は各時刻毎の自己相関で得られる最大値で与えられ，これをF_0信頼度とよぶ。また，不特定話者を対象とする認識システムではF_{min}の予測は困難となるので，対数F_0パターンの線形回帰分析によりバイアス成分を除去した回帰係数パター

ンをΔF_0パターンとする。

第3章　アクセント句境界検出（非フレーム同期出力型）

本章では，F_0パターン連続整合法による句境界検出について，観測F_0パターンに基づく手法，およびF_0生成モデルに基づく手法を提案している。

F_0パターン連続整合法の基本的な概念は下平と嵯峨山によって基礎付けられており，

1)　アクセント句のF_0パターンの形状は多様であるがランダムではなく，少数個のクラスタに分類できる，

2)　一つの発話によるF_0パターンはクラスタの代表パターン（テンプレート）を接続した連続パターンである。

という性質を利用したものである。まず第1の性質より，アクセント句のF_0テンプレートは，視察によって切り出されたアクセント句で観測されるF_0パターンをクラスタ分類することによって学習することができる。次に第2の性質より，未知入力音声のアクセント句境界は，入力音声のF_0パターンとF_0テンプレートの連続整合（One-Stage DP）により得られる最適整合テンプレート系列の接続境界として検出することができる。この最適整合基準として，入力パターンとテンプレートの各時刻の値の二乗誤差にF_0信頼度を乗じた累積歪みを使用する。

観測F_0パターンに基づいた手法では，この先行研究に対し，テンプレートの整合自由度を上げることにより整合歪みを減らし，さらに上位N位の候補を探索するN-best法やテンプレートbigram情報による接続制御等を適用して句境界検出精度を高めている。この手法は韻律のモデルを全く必要としないため学習が容易であり，異なるデータセットに対して汎用性がある。しかし，句境界の自動検出においては処理時間の問題，多すぎる句境界挿入誤りの問題という短所もあり，それらはモデルを仮定しないことによるテンプレートの自由度の高さに起因する。

一方，F_0生成モデルに基づく手法では，学習用のアクセント句を生成モデルパラメータで表現するためパラメータ推定等の熟練を要するが，個々のアクセント句の成分を前後のアクセント成分から分離することが可能であり，またテンプレートの学習過程におけるF_0抽出誤差の問題が解消される。句境界の自動検出においては生成過程に則した拘束条件として，モデル化テンプレートから生成するF_0パターン上の任意の値はすべて時間の関数によって一意に決定するという条件を与え，生成F_0テンプレートによる線形整合を行う。但し，観測パターンに基づく手法のようにテンプレート整合の終端を固定するとアクセント句の時間長の自由度が著しく小さくなるので，クラスタ内の統計によりテンプレートの終端可能な区間を設定する。この線形整合により，N位候補の選択を含めた整合処理が高速になる。また，非線形な整合と異なり，F_0抽出誤差による入力パターンの不連続区間に対しても，比較的短時間であれば句境界挿入誤りを回避することができる。

ATR日本語音声データベース（連続音声データ編）を用いた句境界検出実験を行い，F_0生成モデルに基づく手法において，挿入誤り率を50%以下に抑制しながら10位までの累積句境界検出率90%以上を達成した。

第4章　アクセント句境界尤度推定（フレーム同期出力型）

アクセント句境界検出は当該時刻における句境界の有無を二値で表現する。しかし，人間のアクセントの知覚には，話者および聴者による個人差があるので，同じ発声内容に対しても韻律構造が一意に定まるというものではない。前章では句構造の曖昧さをN-best複数候補によって表現していたが，本章では各時刻の句境界らしさを数値化した句境界尤度を定義し，この尤度を出力する手法を提案している。この手法では入力音声の全区間に対するテンプレート系列の最適解を探索する必要が無くなるので，入力の終了まで数値化処理を待機する必要が無く，時間に同期した出力が可能になる。この特徴は電話通信の音声自動翻訳等の実時間性を要する音声認識において重要である。

アクセント句境界尤度推定の基本アルゴリズムは句境界検出法であるF_0パターン連続整合法と同様であり，使用するテンプレートはF_0生成モデルに基づくものである。このテンプレートは自由終端区間が設定可能なため，分析当該時刻において直前の時刻の同一統合候補から生じる句境界仮説と非句境界仮説の2つの仮説を定義することができる。まず，各時刻における句境界仮説は，当該時刻においてテンプレートの始端と整合する候補である。入力の最終時刻まで最適整合経路上で，当該時刻が句境界となる場合にはこの句境界仮説から始まる整合歪みが極めて小さいことが予測される。一方，非句境界仮説は，句境界仮説への接続元となる候補がテンプレートの終端とはならずに，先行テンプレートとの整合を継続するという仮説を指す。この2つの仮説が単位時間経過したときの歪みの増加分を比較して，句境界

スコアを定義することができる。差や比を用いた幾通りかの数値化方法が考えられるが，本章では（非句境界仮説の対数歪み増加分）－（句境界仮説の対数歪み増加分）と定義することにより，句境界らしい区間のスコアを正の値に，逆の場合を負の値にしている。但し，スコアの値の最大・最小が特徴量や発声環境に依存して異なるので，スコアの正規化（－1～＋1）を行い，それを句境界尤度とよぶ。なお，正規化には時間同期性を考慮して単位時間あたりのスコア分布から推定する最大・最小値を用いて行う。

この句境界尤度推定法の有効性を確認する実験では，第3章において高検出精度を達成した設定条件を与え，F_0 分析，句境界スコア推定窓幅，正規化窓幅による尤度出力の遅延は0.5秒に設定した。この結果，発声区間の約半分の区間の句境界尤度が負の値となり，この区間にはほとんど句境界が存在せず，約90%の句境界については正の尤度が得られることを確認した。

第5章　フレーム同期型連続音声認識における韻律情報の利用

第3章から得られる句境界位置情報は音声認識の前処理である単語検出や音声認識の後処理である構文検証などに有効であり，これらは従来の研究において報告されている。一方，時間同期形連続音声認識システムの認識過程においても句境界情報の利用が可能であり，生成される文仮説候補の構文を句境界情報を用いて制御すれば，不要な仮説の展開を刈ることによって，処理効率，認識精度が向上することは容易に推測できる。しかしながら，従来の韻律情報推定の研究では時間同期性をあまり重視していないため，このような利用が不可能であった。本章では，第4章で提案した句境界尤度情報が時間同期形の音声認識システムにおいて有効であることを検証している。

対象には時間同期形SSS-LR連続音声認識システムを使用する。このシステムはOne-Pass Viterbiサーチに基づいてHMnetの音響モデルと照合を行う探索部と，拡張LR構文解析法を用いて統語解析を行う統語解析部とから構成された，時間同期形の認識手法である。したがって，各時刻毎に異なる音素系列を持つ複数の仮説を逐次展開し，最終時刻で文法的に受理された候補のうち最も音響尤度の高いものを認識結果として出力する。ただし，同一の音素列に対して構文の曖昧性があり，従来では最終認識結果を韻律構造的に評価するという手法を用いていた。

本章では認識システムに対して，以下の改良を行っている。

1) アクセント句構造を基にした文脈自由文法を作成する。但し，文節内文法は統語構造によって記述している。
2) 文仮説のスコアを（音響尤度）＋（韻律尤度）によって与え，同一の音素列に対しても構文の曖昧性に応じて文仮説候補を分割する。但し，韻律尤度はアクセント句構造の相違にのみ依存するので，候補の分割はLR解析によってアクセント句に還元する状態と還元しない状態の混在している仮説についてのみ行う。
3) 同一音素列を持つ異なる候補が同一時刻にアクセント句へ還元することによって共通のLR解析状態になった場合には，スコアの高い候補のみを残すことによって，効果的に仮説を削減する。

以上の認識システムについて，ATR日本語音声データベースを用いた評価実験を行った。音響モデルについては学習に使用した話者と使用しなかった話者の2名，言語モデルについてはパープレキシティの異なる2種類の文脈自由文法を作成し，組合せて4通りの実験を行った。なお，使用した句境界情報は句境界尤度の負の値のみであり，統語解説部における句境界挿入の抑制を目的とした。実験の結果，文認識結果についてほとんど改悪されることがなく，約3%の認識率の向上が見られ，これらは音響モデル，言語モデルに依らないことが確認できた。また，展開仮説数の異なる上限（ビーム幅）を用いた実験では，同程度の認識率を得るためには韻律情報を使用した方が狭いビーム幅で済み，処理効率の向上が確認できた。

第6章　結　　論

本論文では，高精度な韻律情報抽出法としてF_0パターン連続整合法によるアクセント句境界推定法を提案し，連続音声認識の前処理・後処理を目的とした句境界位置の検出，および時間同期性を重視した句境界尤度推定法について研究した成果をまとめた。また，後者の句境界尤度情報を時間同期型連続音声認識システムで活用する手法を提案し，その有効性を実証した。

審 査 結 果 の 要 旨

音声認識は，より知的なマン・マシンインタフェースの実現で必須となる対話処理のために必要不可欠であり，孤立単語認識から連続音声認識へ，朗読音声認識から自然発話認識へと研究対象が複雑化している。著者は，連続音声から単語や文節の境界を正確に検出することが認識精度の向上に有効であるという観点から，従来は困難とされた，音声自体がもつイントネーションやアクセントなどの韻律構造の抽出法を検討するとともに，実時間で韻律構造を連続音声認識に活用できる，認識タスクの複雑化に対応する認識技術を開発した。本論文はこれらの成果をとりまとめたもので，全編6章よりなる。

第1章は序論である。

第2章では，韻律構造推定処理系の入力となる，音声の基本周波数パターンであるF_0パターンなど3種類の韻律特徴量の定義とその抽出法を与えている。

第3章では，連続音声のアクセント句境界の検出法として，F_0パターン連続整合法を提案している。観測F_0パターンに基づくものとF_0パターン生成モデルに基づくものを提案し，実験的に比較検討し，F_0パターンの抽出エラーの影響を受けにくい後者が検出率90%以上となることを述べている。音声認識の精度限界となるアクセント句の検出精度を向上させたことは，実用上重要な成果である。

第4章では，人間でもアクセント句境界を正確に検出することができないという事実，および従来法では音声入力の終了時点でしかアクセント句境界を検出できないという問題点を解決するため，アクセント句境界のもっともらしさの尺度である句境界尤度を提案し，その妥当性を検証している。この尤度は各音声フレーム（周波数分析のための基本区間）ごとに決めることができ，特別なセグメンテーションをせずに音声認識過程で実時間の利用を可能にするもので，興味深い提案である。

第5章では，連続的な韻律情報である句境界尤度の計算を連続音声認識システムに組み込み，韻律情報を活用した認識手法を与えている。韻律情報は，文仮説候補の韻律表現に基づいて構文の曖昧性による仮説の分岐・統合の過程でその判断基準に用いられ，入力される音声フレームに同期した認識処理を可能にしている。この手法で連続音声認識実験を行い，尤度情報の利用で処理速度を向上させるとともに，文認識率で約3%の改善が得られることを示している。これは，句境界尤度の有効性を示すもので，連続音声認識への大きな貢献である。

第6章は結論である。

以上要するに本論文は，アクセント句境界検出のためのパターン連続整合法を確立するとともに，新たに考案した句境界尤度を活用する連続音声認識手法を開発しその有効性を明らかにしたもので，情報通信工学および音声工学の発展に寄与するところが少なくない。

よって，本論文は博士（工学）の学位論文として合格と認める。