

本論文は「Incorporation of Prosodic Modules for Large Vocabulary Continuous Speech Recognition (大語彙連続音声認識における韻律モジュールの導入)」と題し、韻律境界情報を用いて大語彙連続音声認識の性能向上を図ったものであって、全7章からなり、英文で記述されている。

第1章は「Introduction」であって、まず、現在の連続音声認識についてその問題点を指摘した上で、統語境界との関連が深い韻律境界情報を利用する可能性について言及している。次に、韻律境界検出の性能の低さにふれ、その要因を考察したうえで、性能向上が必ずしも容易でないことを指摘し、従来のような、入力音声の区分に利用する手法に限界があるとしている。さらに、連続音声認識で一般的に用いられている Cross-word Context Dependent (CCD) 音素モデルについて、韻律によって示される単語間の結びつきの強さに対応して、その利用を制御し得る可能性を示している。最後に、第2章以降の論文の構成を述べている。

第2章は「Large vocabulary continuous speech recognition」と題し、HMM 音響モデルと n-gram 言語モデルによる、統計論的な枠組みの連続音声認識アルゴリズムについて説明している。特に、本論文と関連の深い探索問題について、2段階の認識探索器を念頭におき、Viterbi ビーム探索、A*探索、木構造辞書等について詳説している。また、探索範囲(認識時間)と認識率の関係についても言及している。

第3章は「Prosodic information for automatic speech recognition」と題し、従来行われている音声認識への韻律の利用手法を整理した上で、論文との関連の高い有用なもの3つを特に紹介している。1つめは、韻律を組み込んだトータルな認識システムである Verbmobil プロジェクトにおける内容の紹介であり、これが単語候補の情報を利用し、韻律的特徴を用いて各言語境界の確率を推定した上で、それを認識における言語処理過程で利用するものであるとしている。2つめは統計的モデリングによる韻律語境界の抽出手法として提案された、モーラ基本周波数(F0)パターン遷移過程のモデルである。3つめは対話アクトの分類への利用であり、これによって適切な言語モデルを選択すれば認識性能の向上が期待される。最後に、韻律の利用形態について言及し、本論文で提案する手法の導入としている。

第4章は「Detection of prosodic-syntactic boundaries」と題し、まず、韻律の重要な特徴としてのF0パターンについて、その特徴とモデル化について説明している。次に、本論文で利用するF0パターンの大局的及び局所的特徴を併用した韻律境界(アクセント句境界)情報抽出手法について説明し、特に挿入誤りが多いとしている。

第5章は「Incorporation of prosodic modules for LVCSR」と題して、本論文で提案する2つの手法について述べている。1つめは、木構造辞書を利用した Viterbi ビーム探索で行われる言語尤度の factorization に起因する最適探索経路のスコアの変動を考慮した動的ビーム幅制御である。これは、単語あるいは文節の始めて最適探索経路のスコアが見かけ上悪化することに着目したものであり、韻律境界が文節境界に近似的に対応するとして、ビーム幅を境界付近で広げ、次の境界に向かって狭める手法である。2つめは2段階目の探索において、CCDモデル(tri-phoneモデル)の利用を制御する手法である。韻律境界では音響的な調音結合が弱まると仮定し、そこではbi-phoneモデルを用いる。

第6章は「Experimental results」と題し、実際に2段階の認識探索器に前章の手法を組み込み、大語彙連続音声認識を行った結果について述べている。テストデータは日本語新聞記事から選択した50文を10名のいずれかの話者が読み上げたものである(5文/1名)。音響モデル、言語モデルは日本語新聞記事とその音声から構築された公開のものを用いている。実験の結果、30%の認識時間の減少、14%の文正解率の向上を得、手法の有効性が示されたとしている。

第7章は「Conclusion」であって、本研究で得られた成果を要約し、統計的手法による韻律境界検出の性能向上など、将来の課題について述べている。

以上を要するに、本論文は、連続音声認識において韻律的特徴を利用する新しい方策として、韻律境界情報を用いた正解候補探索の際のビーム幅動的制御手法、CCDモデルの利用制御手法を提案し、その有効性を認識実験により実証したものである。これは、人間の音声知覚過程での重要性が指摘されているにもかかわらず、従来、認識にほとんど利用されていなかった音声の韻律情報を利用する一般的かつ有効な手法を提示した初めてのものと見える。今後の大語彙連続音声認識の性能向上の発展に大きく寄与するものであり、電子情報工学に貢献するところが少なくない。

よって、本論文は博士(工学)の学位請求論文として合格と認められる。