

電子・情報 工学専攻	学籍番号	099308	指導 教員	中川 聖一
申請者 氏名	藤井 康寿			秋葉 友良

論文要旨 (博士)

論文題目	音声ドキュメントの音声認識、整形、要約に関する研究
------	---------------------------

(要旨 1,200字程度)

近年、動画共有サービスの発展により、Web上の音声ドキュメントが飛躍的に増加している。これらの内、大学で収録される講義音声ドキュメントは、最近になって利用できるコンテンツが増加してきたこともあり、初段の音声認識から後段の要約などの後処理にいたるまで、依然として研究段階にある。本研究では、これら講義音声ドキュメントのための音声認識、整形、要約に関する研究を行った。

講義音声は自然発話のため認識が非常に難しく、最先端のシステムを用いても認識率は50%程度である。このような状況において音声認識性能を改善するためには、従来の枠組みを超えた手法の構築が必須であると考えられる。本研究では、従来の枠組みを超えた手法として、非線形な特徴抽出機能を有した系列識別モデルである隠れニューラル確率場 (Hidden Conditional Neural Fields ; HCNF) を用いた音声認識手法の提案を行った。連続音素認識実験により HCNF が従来の識別モデル及び生成モデル (隠れマルコフモデル ; HMM) を用いた手法を上回ることが示され、音声認識性能改善のための基盤を与えることができた。

音声認識性能の改善に加え、認識結果の整形処理に関しても研究を行った。講義音声は自然発話であるため、発話の忠実な書き起こしは、認識誤りが存在しない場合でも読み難いものとなる。そのため、ユーザに書き起こしを提示する前に、読み易さを向上するための整形処理を施すことが望ましい。本研究では、認識が非常に難しい講義音声の認識誤りに対処可能な方法として、音声認識結果の複数仮説を用いる手法を実現した。さらに、複数仮説と同時に文レベル情報を効率的に利用できるように手法を拡張した。

音声認識結果から複数仮説を取得するために、本研究室において長年に亘り開発が続けられてきた大語彙連続音声認識システム SPOJUS の再実装を行った。再実装されたシステム—SPOJUS++は、音声認識結果の複数仮説を出力可能となっただけなく、探索誤りを低く抑えながらも現在の一般的な PC において実時間動作を実現した。

整形された認識結果の提示による「読む」という観点からの改善だけでなく、音声ドキュメントを「聞く」という観点からも改善を行うために、講義音声ドキュメントの要約に関する研究を行った。長時間の音声であり、明確な構造が存在せず、冗長性が高いという講義音声ドキュメント特有の現象に対処した要約手法を提案し、高精度な講義音声ドキュメント要約を実現した。

本研究を通して、音声ドキュメントの音声認識、整形、要約という3つの観点から音声ドキュメント処理技術を向上することができた。