

## 審査の結果の要旨

氏名 鈴木 雅之

本論文は「背景雑音と話者の違いに頑健な音声認識」と題し、全6章から成る。音声信号には背景雑音、部屋の音響的残響、マイク・チャンネルの音響特性、話者の体格・年齢差など、発声されたメッセージとは無関係の要因により様々な音響歪みが混入する。本論文はこれらの要因に対して「背景雑音」と「話者の違い」という二大要因を取り上げ、それらに頑健に動作する技術を個別に構築し、その有効性を実験的に検証している。

第一章は「音声認識の基本技術」と題し、音声認識システムを構築する際に必要となる4つの機能・モジュールと、音声認識問題の数理統計的な定式化を説明している。

第二章は「近年の音声認識技術とその問題点」と題し、第一章で述べた4機能の実装技術に関する近年の動向について非常に幅広いサーベイを行っている。本論文では、特に音響モデリングに着眼した検討を行っている。近年の音響モデリング技術は、識別モデルに基づくモジュール開発が行なわれるようになっており、本章でも複数の実装方式を紹介している。本章の後半では、実社会で利用されている大規模音声認識システム二つを取り上げ、**state-of-the-art** な認識システムについて記載すると共に、本章の最後では、現在でも未解決な問題である「背景雑音」と「話者の違い」による影響について触れ、その打開策について指針を述べている。

第三章は「非定常雑音に頑健なステレオベース特徴量強調」と題し、背景雑音に対する頑健性向上を試みている。**SPLICE** と呼ばれるステレオベース特徴量強調（雑音音声と対応する無雑音音声の平行ループスを用い、雑音音声から無雑音音声への写像を推定し、これをシステム動作中に利用することで雑音音声に対する精度向上を図る）に着眼し、これを様々な観点から改良する方法について検討している。**SPLICE** では従来入力音声側の音響空間を領域分割し、入力特徴量がどの領域に所属するのかに基づいて変換関数を適応的に求めている。提案手法では、1) 出力側のクリーンな音声特徴量空間で空間分割を行って写像関数を推定し、2) その空間における識別性を上げる目的で事前の識別的な特徴量変換を行なう技術を提案している。これらを検討する中で、本

研究が先行研究をより一般化したモデルであることが数理的に示されている。この一般化モデルを最適化できれば、従来研究に比べてより高い精度を示すことが理論的に予想され、雑音化連続音声認識実験を通して、実験的にその妥当性を示すことができた。提案手法は声質変換、帯域拡張などにも応用し、こちらでも良好な結果が得られることが実験的に明らかとなった。

第四章は「ミスマッチがない場合にも頑健な特徴量強調」と題し、ミスマッチが無い場合に特徴量強調がもたらす副作用を軽減する手法を提案しており、実験的にその性能を確認した。

第五章は「識別的リランキングにおける構造的表象の利用」と題し、話者性の違いに対する頑健性向上を試みている。ここでは先行研究で提案されている話者の違いに不変な特徴量である構造的表象を、連続音声認識へ適用することを試みている。従来、構造的表象に基づく音声認識は、特徴量が離れた二時刻の（特徴量の）コントラスト量として定義されるため、特徴量時系列が入力されることを仮定した従来のデコーディングアルゴリズムとの相性が悪かった。本研究では、従来の音声認識技術と構造表象の利点を同時に生かす手法として、リランキングに着目し、複数の仮説に対する再評価プロセスにおいて、構造的表象を利用する手法を検討した。リランキングでは、近年、音声認識のデコーディングには直接活用しにくい特徴／情報も使われており、そこに構造的特徴を導入した。実験の結果、多様な話者を含む無雑音下の音声入力に対して、従来の音声認識よりも高い精度を示すことが示され、提案手法の有効性を実験的に確認することができた。

第六章は「まとめ」と題し「背景雑音」と「話者の違い」に対して音声認識システムを頑健に動作させる技術的解決を示した本論文を総括し、と同時に、今後の展望について述べている。

以上要するに本論文は、音声認識精度を下落させる音響的な主要因である「背景雑音」と「話者の多様性」に着眼し、これを解決する二つの技術を提案し、音声認識実験を通してその有効性を実証しており、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。