論文審査の結果の要旨

氏名　　朝川　智

本論文は「音声の構造的表象に基づく単語音声認識に関する研究」と題し，全九章から成る。音声は発声者の声道形状・サイズ，収録機器・環境によって音響的には様々に変化する。従来の技術ではこれら不可避的な変動に対処するために，1)集めて統計的にモデル化すること，2)各々の場面で対象に合わせること，で対処して来た。本論文ではこれらの方法論とは全く異なる観点からこの問題の解決を試みている。対象とする話者や環境による変動は一般的に時不変であり，静的なバイアス項となる。このバイアス項を数学的にモデル化し，任意のバイアス項に不変な音声の表象を提唱し，それに基づいた音声認識手法を提案，実験的にその妥当性を示している。

第一章は序論であり，本論文の背景・目的・構成について述べている。続く第二章では従来の音声認識システムについて概説している。本論文は特に音声の音響的特徴及びそれを用いた照合方式についての新提案であるので，音声認識システムの中でも音響的特徴抽出部，音響照合部について概説している。更に第三章にて，音声に不可避的に混入する非言語的特徴（話者の年齢や性別，収録機器・環境の違いによる音響的な変化）について説明し，その数学的なモデル化を行なっている。更に，これらの変動に対して従来どのように対処してきたのか，その先行研究についてまとめている。

第四章で，非言語的要因に不変な音声表象として音声の構造的表象について解説している。第三章で行なった非言語的特徴の数学的モデル化に従って，その不変量を導出し，不変量のみを使って音声を表象している。ここでは，微分可能かつ可逆な全ての写像関数に対して不変なる量を導出している。本論文の直接の先行研究となる研究において，孤立単語列に対する音声認識が試みられており，それについて解説している。なお，この構造的表象は古典的な言語学の一分野である，構造音韻論を物理的実装として解釈可能であること，即ち提案法の言語学的妥当性についても触れている。

第五章以降，本論文で提案する新しい技術について詳細に検討している。音声の構造表象を用いた音声認識を考える場合，大きな問題が二つ生じる。一つは「強すぎる不変性問題」と呼ばれるものであり，他方は「高すぎる次元数問題」と呼ばれるものである。本章では前者の解決を図る。非言語的要因に対する高い不変性は，例えば，異なる単語を同一視してしまう問題を引き起こす。これは１）「あ」と「え」の違いも，話者Ａの「あ」と話者 B の「あ」の違いも，スペクトル包絡と呼ばれる同一の物理量によって表現されること，２）提案する不変量が微分可能かつ可逆な全ての写像関数において不変となるため，不変性が強すぎること，が原因である。話者の違いにだけに不変な音響照合方式の開発が必要となるが，ここでは，話者変換行列の一実装例に着眼し（帯行列として変換行列を実装），その変換行列に対してのみ不変性を有する音響照合方式を提案した。具体的には特徴量を次元分割し，部分空間へと射影した上で構造表象を構成する。

第六章では，提案手法の有効性を，日本語五母音を並び替えた母音単語音声認識，及び音韻バランス単語音声認識という二つのタスクに対して検討している。両タスクにお

いて，特徴量の次元分割は非常に有効に働くことが実験的に示された。しかし，母音単語の場合は従来の方法論と同等の性能が得られたが，音韻バランス単語の場合はまだ性能的には従来の方法論とは開きが大きい結果となった。

第七章にて，第二の問題「高すぎる次元数問題」を，判別分析を二段階に分けて適用する方法を導入することで解決した。更には動的特徴による構造表象，複数の特徴ストリームによる構造間の距離の導入など，種々の改善を図ることで，母音単語の場合は従来の方法論よりも高い性能を，音韻バランス単語の場合も，それに匹敵する精度を示すことができた。即ち，学習条件と評価条件間のミスマッチが少ない場合は，従来法と同等の性能を示すことを実験的に確認することができた。

第八章にて，多様な話者性に対する本手法の頑健性について実験的に検討している。提案手法は話者やマイクなどの非言語的要因に対する不変性を唱っており，ここでは身長を人工的に制御した音声を作成し，それを入力することで提案手法の頑健性について検討した。母音単語，音韻バランス単語ともに従来の方法論ではおよそ不可能と考えられる「高い頑健性」を示すことに成功した。特に母音単語の場合は，ある環境下で構築された提案手法の音響モデルが，学習／評価の条件を揃えて何種類も用意した（従来法による）音響モデルと，常に，同等の性能を示すことができ，話者適応や環境適応という技術を使わずに，常にそれらを駆使した従来法の精度を出すことに成功した。

第九章にて，本研究を総括している。提案する構造表象による音声認識は，従来の音声認識研究が対象としてきた音声の音響特徴とは全く異なる音響特徴を捉え，それに基づく超頑健な音響照合方式を提案している。話者の違いなどの非言語的な音響変動はそもそも時不変な静的バイアス項であるにも拘らず，従来の方法論では，それら静的変動を大量の音声データを集める事で統計的に対処し，個々の話者性をサンプリングに伴うランダム雑音と見なして来た。これは，話者性は時間軸に沿って変わりうるもの，という前提を置いており，事実とは大きくかけ離れている。本論文ではこれらの問題を，音声の音響的相対量を駆使することで解決する方法を提案し，その実用性を示した。と同時に，提案手法が抱える未解決問題についても，それを明確にすることができた。

以上要するに，本手法は従来音声工学が採択して来た方法論，しかも，ほぼ常識となりかけた方法論に対して敢えて疑問を投げかけ，その疑問を解くための一手法を提案しており，情報学の基盤に貢献するところが少なくない。

よって，本論文は博士（科学）の学位請求論文として合格と認められる。