

多様な方言に対応した音声認識システムの開発 (152302005)

A Study of a Speech Recognition System for Multi Dialect

研究代表者

樽松理樹 岩手県立大学

Masaki KUREMATSU Iwate Prefectural University

研究分担者

関 義則[†]

Yoshinori SEKI[†]

[†]株式会社日立ソリューションズ東日本

[†]Hitachi Solutions East Japan, Ltd.

研究期間 平成 27 年度～平成 29 年度

概要

現在、音声認識技術は、精度や処理速度の向上とともに、様々な場面で活用されている。しかし、現在の音声認識技術は標準語を中心としており、方言への対応がまだ不十分である。そのため、方言を話す人が多い地域においては、誤認識を起し、発言者の意図を正確に変換できない事が予想される。今後、音声認識技術を利用する場面が増え、様々な人の発話を認識する事を考えれば、方言による誤認識は解決すべき課題であるといえる。

以上の背景から、本研究課題では、多様な方言に対応した音声認識システムの開発に取り組む。本システムにおける音声認識とは、従来の音声認識と異なり、方言音声を変換列に正しく変換することを意味し、その実現を目的とする。そのため、従来の音声認識手法で活用されている音韻モデル、言語モデルは用いず、異音モデルを用い、音声を変換列への変換を試みる。異音モデル構築手法としては、従来手法をベースに本目的に適合するように機械学習手法を用いることで実現する。さらにこのモデルを用いた異音列認識手法を構築する。また、複数の異音モデルを利用する枠組みも検討し、認識率(変換精度)の向上を試みる。

本手法を研究代表者がいる地域の方言音声に適用し、その有用度を評価する。

1. まえがき

音声認識の研究分野において、方言認識は大きな課題の1つであり、多様な取組みが行われている。しかし、方言認識を行うためには、多様な方言に対する汎用性、少ない方言言語資源での動作、方言変換・言語理解との接続容易性といった点を考慮しなくてはならず、困難な課題である。現在、方言に対応しているとされる商用的なシステムやサービスも開発されているが、対応できる方言は限定されているものが多く、精度の面からも、まだ不十分と言える。以上の背景から、本研究課題では、多様な方言に対応した音声認識システムの開発に取り組む。また先に述べた方言音声の特徴を鑑み、本研究課題では、方言音声を変換列に変換する手法の開発に取り組む。基本方針としては、従来の音声認識手法と同様の枠組みを用いる。

本研究では、はじめに音韻モデル構築に必要な対象となる方言音素を収集する。次に方言音声から方言音素用の音韻モデルを構築する。最後に、構築した音韻モデルを用いて音声認識を試みる。ただし、方言音声の収集と音韻モデル構築は並行して行う。

2. 研究開発内容及び成果

本研究開発においては、大きく2つのことに取り組んだ。1つ目は、専門家の知見に基づく方言音声資源の収集であり、もう1つは、機械学習手法に基づく音韻モデル構築手法の開発である。以下、それぞれについて述べる。

2.1 専門家の知見に基づく的確な方言音声資源の収集

専門家として、岩手県立大学宮古短期大学部・田中宣廣教授に協力をいただき、対象とする方言音声については、同教授の助言のもと、岩手県の宮古地域及び秋田県の能代地域とした。また、方言音声認識においては従来の音韻ではなく、異音に着目することが重要であるという助言のもと、

収集の対象となる異音 158 パターン(152 種類)を選定し、それらの表記方法およびそれらを集めるのに適した単語について整理した。表 1 にその一部を示す。表 1 において、a や i などの列見出しは母音を、k や s などの行見出しは子音を示す。これらのクロスしているのが目標とする異音であり、カナ表記の下線部である。

表 1 音声収集時に用いた単語票の一部

	a	i	u	e	ε	o
r	アサマ 〔朝間〕	イチゴ 〔苺〕	ウサギ 〔兎〕	エンズー 〔違和感〕	エアーダッコ 〔間〕	オメサン 〔アナタ〕
k	カモシカ 〔カモ鹿〕	キモノ 〔着物〕	クジラ 〔鯨〕	ケバライチ 〔花原市〕	ケアール 〔帰ル〕	コメ 〔米〕
s	サドー 〔砂糖〕	シマ 〔島〕	スマッコ 〔隅〕	ゼトク 〔千徳〕	セアータ 〔咲イタ〕	ソゴ 〔其処〕
t	タナバタ 〔七夕〕	チカナイ 〔近内〕	ツーダ 〔着イタ〕	テッテ 〔手〕	テアーク 〔太鼓〕	トツコ 〔絡マリ〕
n	ナカグズ 〔長靴〕	ニシマチ 〔西町〕	ヌマ 〔沼〕	ネイチ 〔根市〕	ネア 〔ナイ〕	ノサッテ 〔乗ッテ〕
...						
r	クラ△ 〔葎〕	クリ△ 〔栗〕	クル△スー 〔苦シイ〕	クレ△ネア 〔クレナイ〕	クレ△デー 〔暗イ、黒イ〕	クロ△ 〔黒〕

作成した単語票を用い、田中教授の協力のもと、方言話者(成人男性数名、成人女性数名)が単語票中の単語を発音した音声、3年間で合計約9時間分の収集を行った。能代地域の方言音声も同様に、収録対象とする異音を含む単語票を作成し、その発話音声を9時間分収集した。これらの作業工程全体をまとめ、マニュアル化を進めた。また、既存研究の手法を参考に開発した音声解析ツールを用いて、音声へのラベル付けを行った。本ツールでは、作

業者によって結果に差異が生じる、作業者の錬度によって作業時間が変動するといった課題を解決し、作業効率・作業精度の向上を図るために、音声データから音素の区切り位置の検出を行う。具体的には、ボリンジャーバンド分析による音声振幅特徴量抽出と現行の音声認識で用いられる Δ MFCC 特徴量算出の 2 つの手法を用いている。本提案手法の有用性を評価するために行った評価実験においては、一定時間以上の作業経験者と未経験者との作業において、作業結果が 60% 以上一致した割合を最大 30% 向上でき、また作業時間を最大 29% 短縮できた。

2.2 機械学習手法に基づく音韻モデル構築手法の開発
予備実験として、既存の音声認識ソフトを用い、収集した音声認識を試みた結果、十分な精度を得ることができなかった。また、既存のツールを用いて構築した音韻モデルを利用し、方言認識を行った結果、モデル生成に利用した音声の発話者の音声に対しては、最初の音素に対しては 90% 近くの正答率を得た一方、別の話者においては 10% を下回る結果となった。このことから、方言音声に個人差が大きいことが再認識された。

これらを踏まえ、本研究課題においては、新たな音韻モデル構築手法として、① IPA (国際発音記号) 音素、② 品質工学で利用される RT 法、③ K-NN 法、④ リカレンスプロットおよび ANN、⑤ SOM (自己組織化マップ)、⑥ AutoEncoder を、それぞれ利用する 6 手法を提案し、収集した方言音声による評価実験を行った。

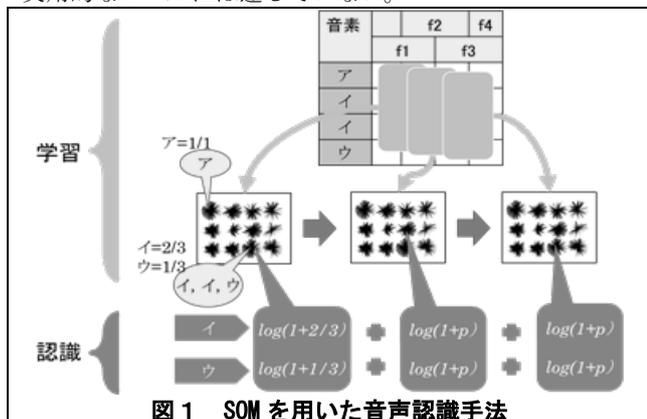
手法①: 本手法では、IPA の各音素から構築した特徴ベクトルとの類似度の集合によって音韻モデルを構築する。評価実験の結果としては 10% 以下と低く、有用とは言えなかった。

手法②: 本手法は、各異音の特徴を元に構築した特徴空間を音韻モデルとみなす。その特徴空間との距離により、判定を行う (RT 方法)。プロトタイプシステムによる評価では、一定の異音への偏りなどから認識精度は不十分であった。

手法③: 本手法は、手法②における RT 法を K-NN 法に置き換えたものである。手法②より偏りは減少したが、精度の向上は得られなかった。

手法④: 本手法では、フォルマント周波数の遷移を、リカレンスプロットで表現、それと音素との関係を ANN で学習することで音韻モデルを構築した。実験結果、精度は 28.8% から 35.6% と低いものであった。

手法⑤: 本手法では、音声特徴から構築した SOM を音韻モデルとし、新たな音声に対し、K-NN 法で認識を行う。図 1 にシステムの概要を示す。評価実験の結果、約 45% の正答率を得た。手法①~④に比べ、高い値ではあるが、実用的なレベルには達していない。



手法⑥: 本手法は、手法⑤において SOM の代わりに、音

素ごとに学習した AutoEncoder を判別器として用いる。評価実験の結果、再現性は約 50% を得たが、正答率は 5% 以下と低い結果となった。

6 つの手法と既存ツールとで方言音声認識を行った結果、初音に限れば、既存ツールを用いた場合、学習と認識で同一話者の場合のみ高い正答率を得たことから、方言の特殊性が見出された。一方、提案手法では、SOM を用いた手法が最も高い正答率であったが、実用レベルには到達していない。

3. 今後の研究開発成果の展開及び波及効果創出への取り組み

研究期間内においては、目標の十分な達成にいたらなかったが、いくつかの知見を得るとともに、一定量の音声データを収集できた。今後の展開として、これらの成果を利用し、音韻モデル生成手法の開発を継続実施する。具体的な手法としては、手法⑤および手法⑥をベースとし、IPA 音声の活用も視野に入れた改善を進める。

また、音声認識に用いた手法は、振動データなどの波形処理にも援用可能である。その点に着目し、異常検知などの波形データの異常検知に活用する。異常検知は音素よりもパターンが少なく明確であることから、有用に働くことが期待できる。

波及効果創出としては、目標を一定以上達した時点において、フリーウェアや API の形で外部公開し、利用してもらうことで、その有用性を評価することともに、評価をフィードバックすることで、精度の向上を図る。

4. むすび

本研究開発では、方言音声認識システムの構築を試みた。しかし、上述したとおり研究期間内においては、思うような研究開発成果は得られず、目標の十分な達成にいたらなかった。一方で、新たな音声データの収集、複数の手法の提案と検証を行った。この点では一定の前進はあり、今後につながるものと捉えている。

今後は、本研究開発期間で得られた成果をベースとし、これまでの知見およびデータを活用し、研究自体は継続実施することで、今後の前述の面への寄与につながることを考える。また今回研究を行った音素モデル構築の枠組みは、別の課題、例えば周波数データの異常検知に援用することも検討する。

【誌上発表リスト】

- [1] 吉田裕範、松本和芳、関義則、樽松理樹、“方言の音声認識モデル構築に向けた音声データ収集の効率化検討”、情報処理学会第 167 回 ヒューマンコンピュータインタラクション研究会 (2016 年 3 月 9 日)
- [2] 吉田裕範・関義則・樽松理樹、“方言の音声認識モデル構築に向けた音素データの音素分割点の検出手法の提案”、平成 28 年度情報処理学会東北支部研究会 (2017 年 3 月 7 日)
- [3] 樽松理樹、吉田裕範、関義則、“SOM と AutoEncoder を用いた方言音声認識手法の提案”、平成 29 年度情報処理学会東北支部研究会 (2018 年 3 月 9 日)

【報道掲載リスト】

- [1] “方言音声を理解する機械の研究”、北羽新報、2017 年 4 月 2 日