

科学研究費助成事業 研究成果報告書

平成 30 年 6 月 28 日現在

機関番号：82505

研究種目：基盤研究(C) (一般)

研究期間：2014～2017

課題番号：26350466

研究課題名(和文) 確率モデルを用いた地域性情報による話者分類

研究課題名(英文) Speaker classification using residence information based probabilistic model

研究代表者

鎌田 敏明 (KAMADA, Toshiaki)

科学警察研究所・法科学第四部・室長

研究者番号：10356173

交付決定額(研究期間全体)：(直接経費) 2,700,000円

研究成果の概要(和文)：本研究では、声道共鳴特性に由来する音響特徴量を用いた話者の地域性情報の抽出及び話者の地域分類について検討を行った。ケプストラムによって得られた話者間距離を地域性情報とした特徴量抽出を行い、話者分布によりその有効性及び話者の地域属性による分類の可能性を示した。確率モデルを基盤とした地域モデルを構築し、検証用データによる話者分類の性能評価を行った。

研究成果の概要(英文)：This study aimed to extract of residence information of speakers using speech features originating in vocal tract resonance, and local classification of speakers. We extracted the features which made residence information distance between speakers acquired by cepstrum, and speaker distribution showed the possibility of the classification by residence information of speakers.

研究分野：人間情報学

キーワード：音声情報処理 話者分類 法科学

1. 研究開始当初の背景

本研究において話者の地域性情報の観点に着目した背景は、音声認識や話者認識を行う上での話者モデルを、外国語話者と日本語話者を区別して構築することは一般的に行われているにもかかわらず、日本語話者モデルを構築する上では、男女の性別や大人と子供の区別を行うことはあっても、話者の居住地ごとに話者モデル構築するような研究はほとんど行われていないという点である。日本語話者の母音「ア」と米語話者の母音「a」を区別するために異なった話者モデルを構築するというのであれば、大阪居住者の母音「ア」と福島居住者の母音「ア」を区別するための異なる話者モデルを構築すべきであると考えられる。ある居住地の母音「イ」は「エ」に似ているが、他の居住地の母音「イ」は「ウ」に似ているといったような、居住地による日本語5母音の母音図のずれというものには存在し、それは方言を構成する要素の1つである。

これまでこのような信号処理技術による居住者ごとの話者モデルや地域性情報について十分な検討が行われてこなかったのは、研究で利用している音声コーパス(データベース)にあると考えられる。一般的に行われている研究で利用されている日本語音声コーパスにおいて、47都道府県を網羅した地域性情報が付加された大規模音声コーパスは市販されていない。一方我々が利用している音声コーパスは、47都道府県の居住者それぞれ約60名から構成される3000人のデータベースであり、地域性情報抽出の実験を行う環境が整備されている。

音声学の分野では方言研究が行われているものの、音声音響工学の分野では方言分離などの研究が行われていることは少ない。そこで本研究では、信号処理技術による方言の一側面であると考えられる話者の地域性情報について検討を行う。

2. 研究の目的

音声による話者認識では、音声中に含まれる個人性情報は声道の伝達特性に由来するスペクトル包絡から得られるケプストラムに基づく研究が行われてきた。また、このような信号処理手法による話者認識において、日本語を母語とする日本人の話者認識を行う研究では、日本人モデルを構築した上での理論及び実験的検討が行われてきた。

しかし本来我々が話者、特に知らない話者の音声を聞いて話者を判断しようとする場合、その話者が自分の知っている誰かに似ていると判断する前に、その話者の居住地を判断する試みをすることがある。例えば、東京居住者が大阪居住者の音声を聞いて大阪弁だと思ふことは頻繁に起こりえることである。そこで本研究では、話者認識において、ケプストラムの特徴量から話者の居住地として

地域性情報を抽出すること、及び得られた地域性情報による話者のクラスタリングを目的とする。

3. 研究の方法

音声コーパスからの特徴量の抽出として、電話を通して録音された約3000名の男性話者音声コーパスを利用した実験を行い、地域性情報の抽出に関する実験を行った。利用するケプストラムはLPCケプストラム係数、メル周波数ケプストラム係数などやそのデルタケプストラムである。得られたケプストラム空間から、2話者における話者間距離を算出し、これらを実験で用いる特徴量とした。

得られた話者間距離から地域属性ごとの話者分布を求め、地域性情報抽出の有効性及び話者分類の可能性について検討した。

4. 研究成果

(1) 音声データと特徴量抽出

照合実験に使用した音声データは、3000人程度の音声データベースから抽出された音声である。3000人の音声データベースは47都道府県全ての居住者を網羅しており、各都道府県でそれぞれ約60名ずつから構成されている。実験で使用した話者は、複数の発声内容において1人当たり9回の発声が全て存在するデータを抽出している。実験に使用した音声データを表1に示す。電話音声であることから、11.025 kHz から 8 kHz にダウンサンプリングを行った。

表1 音声資料

話者数	3000人
発声時期差, 時期	3~4ヶ月, 3時期
1時期の発声回数	3回
発声内容	6単語(はい, 車, 連絡, 爆弾, 銀行, 警察)
サンプリング周波数	8 kHz
量子化精度	16 bit

表2 分析条件

分析窓	ハミング窓
フレーム長	32 ms
フレーム周期	16 ms
高域強調	1次の適応型
分析方法	LPC分析
分析次数	12次
特徴量	LPCケプストラム係数 LPCケプストラム係数

表2に示す分析を行い、DPマッチングから得られたDP距離による話者内、話者間距離をもって、地域性情報が包括されている特徴量とした。

(2) 居住地ごとの距離分布

音声データの話者の居住地を限定した話者間距離の分布についての検討を行った。使用した音声データは3000人のデータベースから東京都、大阪府、青森県及び福島県在住

の話者それぞれ 50 人を抽出した。なお音声データには話者に対する居住地情報の他に、自己申告ではあるが出身地情報も存在する。今回の実験では話者の出身地については考慮していない。また使用した単語は「はい」である。

DP マッチングによる話者照合過程から、同一人及び別人の距離分布を求めた。同一人の話者間距離（話者内距離）は、異なる時期の発声全ての組み合わせから得た。別人の話者間距離は居住地の話者ごとに分布を求めた。

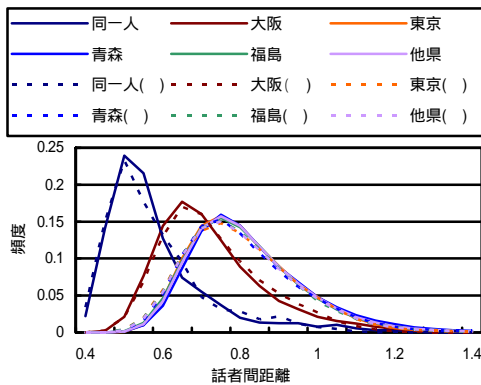


図 1 居住地ごとの話者間距離分布

大阪府在住の話者から見た、LPCC 及び LPCC を特徴量とした同一人及び居住地ごとの距離分布を図 1 に示す。明らかにわかるのは、LPCC、LPCC の場合も、同一地域居住の別人の分布と別地域居住の別人の分布が大きく異なっている点である。LPCC 及び LPCC の分布の差はわずかであるが、その違いは確認できる。

(3) 大規模データベースによる距離分布

前述と実験条件を同一とし、東京及び大阪居住という特定の地域における、居住地を除くすべての都道府県ごとの話者間距離を求め、距離分布による実験を行った。

大阪在住者からみた距離分布を図 2、東京在住者からみた距離分布を図 3 に示す。同一地域及び異なる地域での別人による話者分布には大きく差があることが示された。既存のデータベースと比較することによって居住地、架電地域の推定、分類ができる可能性があることが示された。

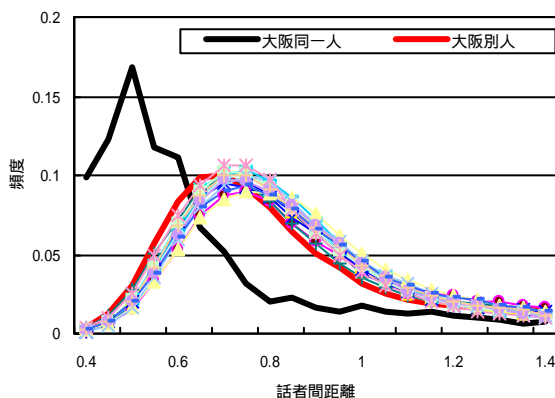


図 2 大阪居住者からみた距離分布

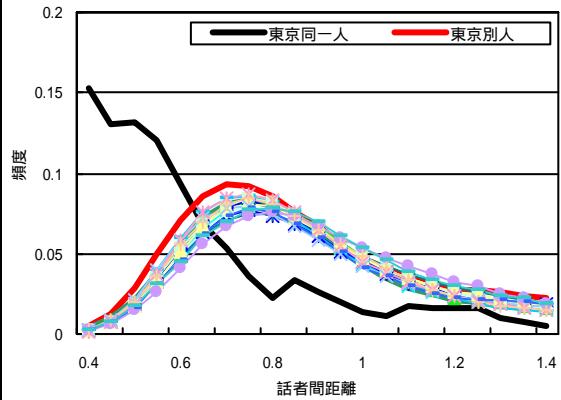


図 3 東京居住者からみた距離分布

(4) 母音間距離による地域分類

音声データベースがありから抽出した関東（東京、神奈川、埼玉、千葉各 30 名）、関西（大阪、兵庫、奈良、京都各 30 名）、東北（福島、山形、秋田 30 名）在住の男性話者による電話音声により、地域クラスタリングのための地域モデルを作成した。地域モデルの特徴量はケプストラム係数による同一時期の母音間距離を利用し、日本語の母音図からフォルマント周波数の近い母音を主たる特徴量とした。次にオープンデータによるクラスタリングの性能評価を行うため、地域モデルに使用しない関東（東京、神奈川、埼玉、千葉 40 名）、関西（大阪、兵庫、奈良、京都 40 名）、東北（福島、山形、秋田 30 名）在住の男性話者を評価データとしてクラスタリングの検証を行った。1 回の発話による同一時期母音間距離を特徴量として、話者モデルとのユークリッド距離を算出し、最小値となったものを属性地域として判定させ、どの地域の話者がどの地域に判定されたかについて確率を算出した。その結果を図 4 に示す。

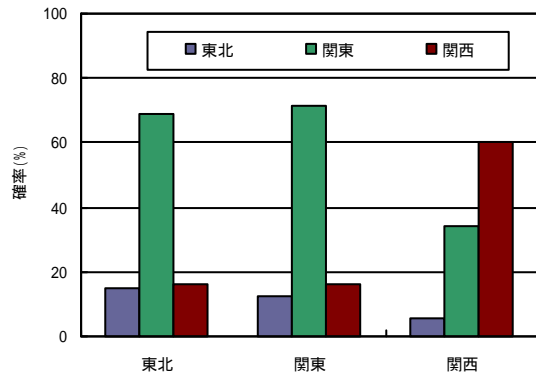


図 4 確率モデルによる地域分類評価

この結果から、関東は 7 割程度正しく判定、関西は 6 割程度正しく判定されているが、東北地方は 7 割が関東と誤認識してことが分かった。よってこの結果を考慮すると、提案し手法によるクラスタリングでは、関東/関西の 2 大地域の地域分類の可能性は示されたが、東北地方を分類するためには、新たな特徴量をモデルに追加するなどによる他の手法で行う必要が分かった。

5. 主な発表論文等

〔学会発表〕(計 12 件)

蒔苗久則、網野加苗、鎌田敏明、長内隆、
正弦波モデルを用いた非定常雑音の抑圧、
日本法科学技術学会第 20 回学術集会、2014

長内隆、網野加苗、鎌田敏明、蒔苗久則、
音声データベースの違いによる話者照合性能の比較、日本法科学技術学会第 20 回学術集会、2014

鎌田敏明、蒔苗久則、網野加苗、長内隆、
母音間距離を利用した話者の地域性、日本法科学技術学会第 20 回学術集会、2014

網野加苗、蒔苗久則、鎌田敏明、長内隆、
言語形態を用いた出身地推定法の提案、日本法科学技術学会第 20 回学術集会、2014

蒔苗久則、網野加苗、鎌田敏明、長内隆、
非定常雑音の抑圧性能の評価に関する研究、
日本法科学技術学会第 21 回学術集会、2015

網野加苗、蒔苗久則、鎌田敏明、長内隆、
聴取による合成音声と自然音声の識別、日本法科学技術学会第 21 回学術集会、2015

長内隆、網野加苗、蒔苗久則、鎌田敏明、
収録環境の異なる音声を用いた話者照合における標準化・正規化変換の効果、日本法科学技術学会第 21 回学術集会、2015

長内隆、網野加苗、蒔苗久則、鎌田敏明、
異なる環境下の単語発話を用いた話者照合における標準化・正規化変換の効果、日本法科学技術学会第 22 回学術集会、2016

鎌田敏明、蒔苗久則、網野加苗、長内隆、
話者認識における静的特徴量と動的特徴量の比較、日本法科学技術学会第 22 回学術集会、2016

蒔苗久則、網野加苗、鎌田敏明、長内隆、
正弦波モデルを用いたブラインド雑音抑圧、
日本法科学技術学会第 22 回学術集会、2016

遠藤泰徳、平田達也、矢崎智仁、網野加苗、蒔苗久則、鎌田敏明、
発話様式の異なる音声の音響分析、日本法科学技術学会第 22 回学術集会、2016

長内隆、網野加苗、蒔苗久則、鎌田敏明、
話者照合における発話様式の影響に関する予備的検討、日本法科学技術学会第 23 回学術集会、2017

6. 研究組織

(1) 研究代表者

鎌田 敏明 (KAMADA, Toshiaki)
科学警察研究所・法科学第四部・室長
研究者番号：10356173

(2) 研究分担者

なし

(3) 連携研究者

なし

(4) 研究協力者

なし