様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成２４年　６月２０日現在

機関番号：８２５０５
研究種目：若手研究（Ｂ）
研究期間：2009～2011
課題番号：２１７１０１７４
研究課題名（和文）　音声の大局的な分類を用いた法科学的話者認識

研究課題名（英文）　Forensic speaker recognition with global classification of speech

研究代表者
蒔苗　久則　（MAKINAE HISANORI）
警察庁科学警察研究所・法科学第四部・研究員
研究者番号：20415441

研究成果の概要（和文）：本研究では、法科学分野への応用のため、音声の大局的な分類を利用した話者認識手法の開発を目指した。上記の提案手法の実現に向け整備を行った音声データベースに収録された音声データから、フォルマント周波数を音響特徴量としてARX音声分析法により抽出し、抽出した音響特徴量により話者を小数クラスに分類可能かを確認するため、クラスター分析を行った。さらに、効率的な情報表現が可能となる音響特徴量の抽出手法の開発を行った。

研究成果の概要（英文）：This study aimed to develop a method for forensic speaker recognition, using global classification of speech. To confirm whether speakers were classified to a few classes, we conducted cluster analysis in which speakers were classified according to their formant frequencies. Formant frequencies were extracted from vowels of 319 female and 313 male speakers by ARX-based speech analysis method. In addition, a novel model for predicting formant frequency from speech signals was developed, which can efficiently represent speech information.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
２００９年度	1,100,000	0	1.100,000
２０１０年度	1,000,000	0	1.000,000
２０１１年度	1,100,000	0	1.100,000
総　計	3,200,000	0	3,200,200

研究分野：音声科学
科研費の分科・細目：社会・安全システム科学、社会システム工学・安全システム
キーワード：話者認識・法科学

１．研究開始当初の背景

近年、犯行時の音声が、防犯カメラや多様な録音機能を備えた携帯電話等により録音される機会が増加している。録音された犯人の音声は、犯人特定のための有力な手がかりであり、犯人と直結する唯一の手がかりとなる場合もある。こうした際、犯人と特定の人物の同一性を音声から判定するために利用される認証技術が話者認識（話者照合）技術である。

話者認識による個人認証は、以前から研究がなされ、現在では一部実用化が進められているものの、依然として認識誤りは避けられない。話者認識において生じる認識誤りは、一般に２種類に分類される。第一は、犯人であるにもかかわらず棄却してしまう（本人棄却）誤りであり、第二は、無実であるにもかかわらず犯人であるとする（詐称者受理）誤

りである。これらの誤りのうち、後者の誤りは冤罪に直結するため、話者認識技術の法科学的な応用の際には、特に低減・排除する必要がある。

現在の代表的な話者認識手法では、話者に関する情報を音響特徴量として音声データから抽出し、抽出した音響特徴量から話者を表現する確率モデルを統計的手法によって話者毎に構築する。そして、話者モデル間の類似度から話者の認識を行う。そのため高い認識性能を得るには、適切な音響特徴量と多量の音声データを使用した完成度が高い話者モデルの構築が必須となる。しかしながら法科学分野では、犯罪に関わる音声を対象とするため、得られるデータ量には限界がある。そのため、統計的手法による話者モデルの構築自体が難しく、現在主流の手法の適用は困難である。したがって法科学分野での話者認識には、話者モデルの構築が不要な手法の利用が望ましい。

２．研究の目的

本研究では、話者認識技術の法科学分野への応用のため、詐称者受理による認識誤りを極力排除し、話者モデルの構築が不要である話者認識手法を提案する。提案する手法の原理は、血液型のように音声を大局的な観点のもとで少数に分類可能であれば、同一の分類に属した場合は同一人・別人の判定は困難であるもの、異なる分類に属した場合、別人の判定は容易となることを利用した手法である。

３．研究の方法

上記の提案手法の実現に向け、研究に利用する(1)音声データベース(DB)の整備を行った後、整備したDBを使用し(2)音響特徴量を抽出し、抽出した音響特徴量により話者を小数のクラスに分類可能かを確認するため、(3)クラスター分析を行った。さらに、(4)音響特徴量の抽出手法の開発を行った。

４．研究成果

(1) 音声データベースの整備

我々が研究を進めている法科学分野で有効な話者認識手法の開発に用いる DB には、一般的な音声 DB と同様に多数の話者が発話した多様な発話内容の音声が収録されている必要があるのに加え、犯罪捜査や鑑識活動の実態に即した音声が収録されている必要がある。たとえば、犯行から容疑者の検挙までには数ヶ月が経過することもあるため、犯行時と検挙後に録音された音声資料間での分析を考慮すると、時期差を設けて録音された音声が収録されている必要がある。さらに一般的な DB では、低雑音環境下で録音されたクリーン音声が収録されていることが多いものの、法科学で用いる DB には、振り込め詐欺といった犯罪が示すように、電話を介して録音された音声が収録されている必要がある。

そこで法科学的な話者認識システムの開発や認識性能の検証のため音声 DB を整備した。DBには、同時録音されたクリーン音声や携帯電話を介した音声などが収録されている。音声の録音は数ヶ月を隔てた２時期に渡り行われている。1 時期目には男性 336 名と女性 328 名の合計 664 名から録音を行い、2 時期目の録音を 1 時期目に引き続き行った話者は、男性 313 名と女性 319 名の合計 632 名であった。また、録音の際、話者の身長、体重、年齢、主な居住履歴を話者属性として記録した。話者の年齢は 18 歳から 78 歳であり、２時期のいずれもから録音を行った話者の年齢構成を表１に示す。

年齢層	男性話者		女性話者	
	人数	割合[%]	人数	割合[%]
10 歳代	15	5	16	5
20 歳代	82	26	82	26
30 歳代	63	20	60	19
40 歳代	61	19	63	20
50 歳代	61	19	65	20
60 歳以上	31	10	33	10

*小数第一位を四捨五入した値を示している

表１　収録話者の年齢構成

発話内容は、単音節、単語、短文の 3 種類からなる計 230 発話である。単音節は、明瞭度試験などに広く用いられている 100 音節である。次に単語は、「警察」、「もしもし」、「爆発」などの 66 単語である。最後に短文については「携帯をもって車に乗れ」などの 14 個の短文に加え、ATR 音素バランス文セット 50 文を加えた 64 短文である。

録音した音声は、発話内容毎に切り出しを行った。さらに電話回線の遅延を補償し、同時に収録した音声データ間での遅延時間は最大 20msec 程度となっている。以上の切り出しは、録音を行った全ての話者と発話に対して行われ、そのうち２時期のいずれでも録音を行った話者から得られた音声ファイルの総数は、2,325,760（＝632 話者×230 発話×2 回×2 時期×4 チャンネル）である。

(2)音響特徴量の抽出

法科学的な応用では、電話音声などのように雑音の影響が避けられない。そこで話者認識に広く用いられているものの雑音に脆弱なケプストラム特徴量ではなく、音響特徴量としてフォルマント周波数に注目した。

①ARX 音声分析法によるフォルマント抽出

フォルマント周波数の抽出には線形予測

分析が広く用いられているものの、音源特性と声道特性の分離の困難さに起因する抽出誤差が指摘されている。そこで本研究では、線形予測分析よりも高精度な抽出が可能とされる大塚ら(2001)が提案したARX音声分析法を利用してフォルマント周波数の抽出を行った。

フォルマント周波数の抽出には、音声DBに収録されている音声のうち、2時期のいずれからも録音を行った成人男女632名が発話した5母音を用いた。最初に各母音信号s(t)について自己相関関数により定常時刻と基本周波数F0を算出し、次に、定常時刻でのフォルマント周波数F1～F3を求めた。これらの抽出には、分析窓長30 ms、サンプリング周波数11.025 kHz、全極型の声道伝達関数を使用し、極の次数は入力信号に合わせて10～18次の範囲で調整した。

抽出したF0の頻度分布を性別毎に図１に示す。F0の平均と標準偏差は、男性話者で138.9±25.8 Hz、女性話者で231.2±26.3 Hzであった。また表２に抽出したフォルマント周波数の平均値を示す。以前から知られているように、F0とフォルマント周波数のいずれにも性別による違いが認められた。

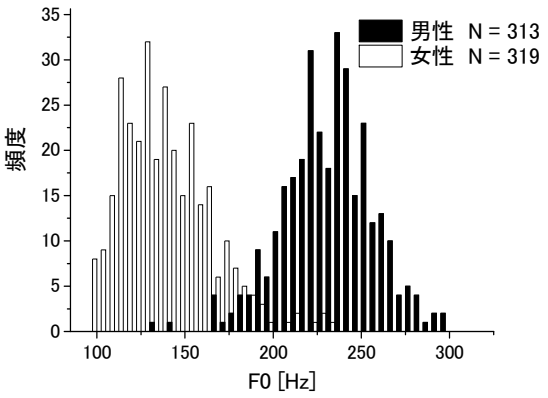

図１　基本周波数の頻度分布

	男性			女性		
	F1	F2	F3	F1	F2	F3
/a/	807	1217	2625	1014	1517	3107
/i/	292	2278	3101	304	2883	3597
/u/	329	1269	2318	351	1476	2872
/e/	477	1985	2641	542	2460	3205
/o/	468	817	2645	515	955	3097

表２　フォルマント周波数の平均［Hz］

(3)話者のクラスター分析

男女632名の５母音から抽出したF1～F3とF0を特徴量とした話者の分類をward法による階層的クラスター分析により行い、得られたたデンドログラムを図２に示す。

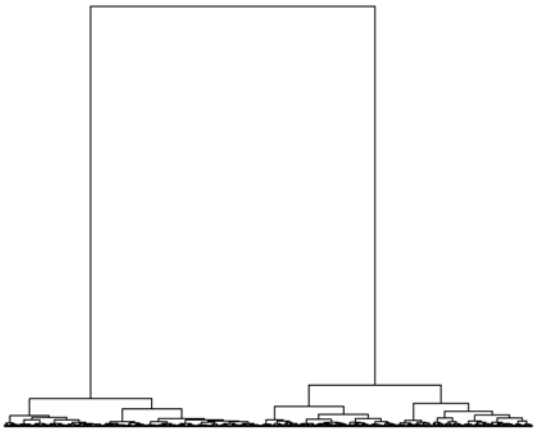

図２　クラスター分析結果：男性・女性話者

図２が示すように、明確に２群に分類されていた。これらの群はF0やフォルマント周波数の平均値が性別に応じて異なっていたことからも示唆されるように、主に性別に対応していた。以下では、性別毎に同様のクラスター分析を行った。

①男性話者のクラスター分析

性別毎のクラスター分析には、性別毎の平均値と分散を用いて規格化したフォルマント周波数やF0を用いた。313名の男性話者についてクラスター分析を行い、得られたデンドログラムを図３に示す。

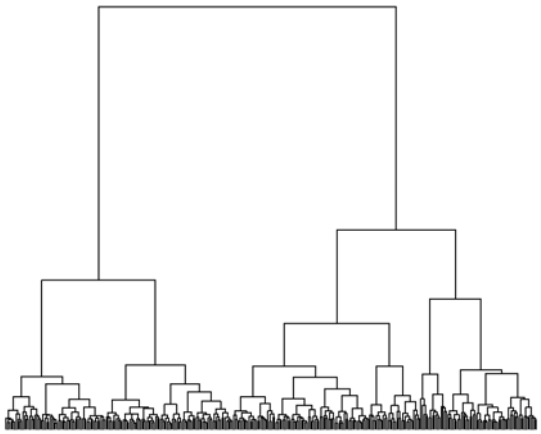

図３　クラスター分析結果：男性話者

図３が示すように、数個のクラスターが明確に生成されていた。生成されたクラスターの特徴を明らかにするため、各クラスターに属する話者属性についての解析を３個のクラスターに注目して行った。それぞれのクラスターに属する話者の年齢、身長、体重の平均値とSEM（標準誤差）を図４に示す。図が示すように、主に身長や体重に応じたクラスターが生成されていた。３個のクラスター（C1～C3）間で検定を行った所、年齢、身長、体重の平均値には、表４に示すような有意差が認められた。

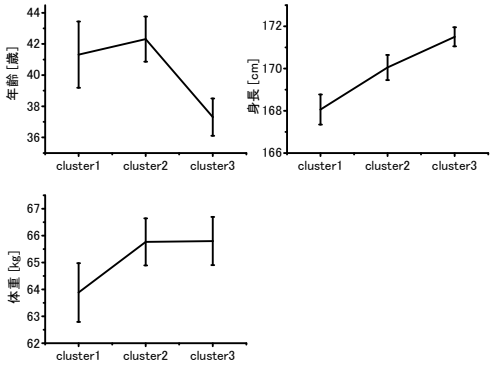

図４　各クラスターに属する話者の
年齢、身長、体重：男性話者

	年齢		
	C1	C2	C3
C1			
C2	ns		
C3	ns	**	

	身長		
	C1	C2	C3
C1			
C2	**		
C3	**	*	

	体重		
	C1	C2	C3
C1			
C2	ns		
C3	ns	ns	

表３　統計解析結果：男性話者
*:P<0.1，**:P<0.05，ns:非有意

②女性話者のクラスター分析
319 名の女性話者に対して行ったクラスター分析の結果を図５に示す。

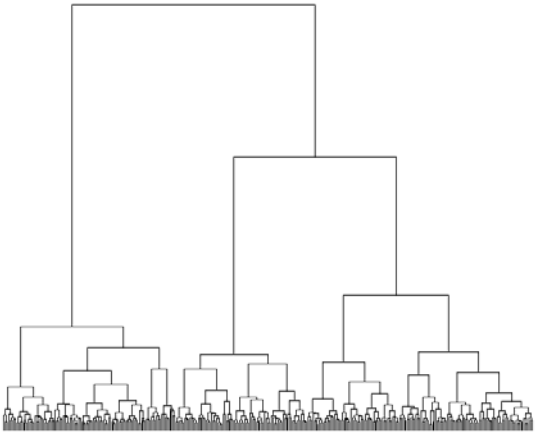

図５　クラスター分析結果：女性話者

図５が示すように、女性話者でも数群のクラスターが明確に生成されていた。男性と同様に３クラスターに注目し、各クラスターに属する話者の年齢、身長、体重の平均値と SEM を図６に示す。図６が示すように、生成されたクラスターには男性と同様に身長や体重と対応関係が認められるものの、表４が示すように統計的な有意差は認められなかった。

このように女性では、身長や体重との有意な対応関係が認められないものの、全ての母音やフォルマント周波数ではなく、特定の母音やフォルマント周波数のみを用いた分析では、有意な対応関係が認められた。例えば、母音/u/のみを使用したクラスター分析を行った結果を図７と表５に示す。これらの図と表が示すように、分類に有効な母音や、また、クラスター分析に用いるフォルマント周波数の組み合わせにも依存していた（data not shown)。

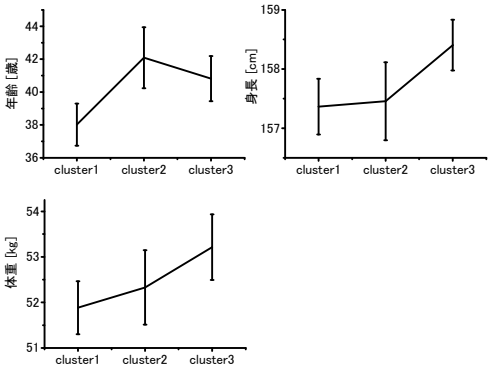

図６　各クラスターに属する話者の
年齢、身長、体重：女性話者

	年齢		
	C1	C2	C3
C1			
C2	*		
C3	ns	ns	

	身長		
	C1	C2	C3
C1			
C2	ns		
C3	ns	ns	

	体重		
	C1	C2	C3
C1			
C2	ns		
C3	ns	ns	

表４　統計解析結果：女性話者
*:P<0.1，**:P<0.05，ns:非有意

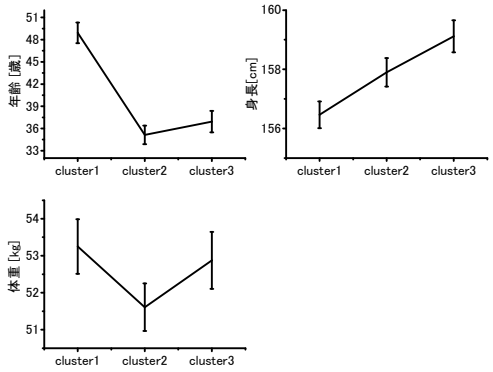

図７　各クラスターに属する話者の
年齢、身長、体重：女性話者、/u/

	年齢		
	C1	C2	C3
C1			
C2	**		
C3	**	ns	

	身長		
	C1	C2	C3
C1			
C2	**		
C3	**	**	

	体重		
	C1	C2	C3
C1			
C2	*		
C3	ns	ns	

表５　統計解析結果：女性話者，/u/
*:P<0.1，**:P<0.05，ns:非有意

以上のように適切な母音や音響特徴量の利用により、音声が身長・体重などの身体的な特徴に応じて音声が分類されうることが明らかとなった。

(4)フォルマント抽出手法の開発
以上のように話者の分類に有効な特徴量となりうることが明らかとなったフォルマ

ント周波数を、ARX 分析法よりもシンプルかつ低計算コストで抽出し、効率的な情報表現が可能な手法を開発した。

開発手法でのフォルマント抽出では、まず音声信号 s(t)の対数パワースペクトルを cos 展開し、12 次元特徴ベクトルによりスペクトル形状を表現する。そして、得られたスペクトル形状を表現する特徴ベクトルの線型または非線形結合によりフォルマント周波数の抽出を行う。結合係数は、ARX 分析法により抽出したフォルマント周波数と抽出値 の二乗平均誤差を最小化することにより算出する。こうして得られたモデルを用いて推定したフォルマント周波数と、ARX 分析法により得られたフォルマント周波数との関係を図８に示す。

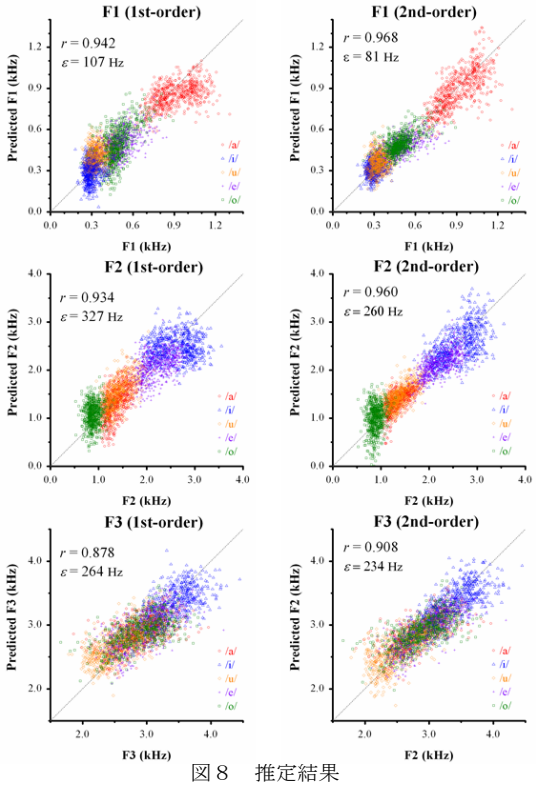

図８　推定結果

左：線型、右：非線形モデルでの結果

図８の横軸は ARX 分析法により得られたフォルマント周波数であり、縦軸はモデルにより推定したフォルマント周波数である。図に示したように対角線上に分布していることから、線型・非線形のいずれのモデルであっても、高精度でフォルマント周波数が推定可能であった。特に、相関係数 r が示すように、線型モデルよりも非線形モデルの方が、高精度な推定が可能であった。つまり、母音のスペクトル形状によるフォルマント周波数の推定には、非線形の利用が有効であることを示している。

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計１件）

蒔苗久則、鎌田敏明、長内隆(2012)、「法科学分野における話者認識のための大規模音声データベースの構築」、科学警察研究所報告、(採択決定)、査読有

〔学会発表〕（計１１件）

① 伊藤仁、蒔苗久則、「ケプストラム係数を用いた母音のフォルマント分析」、日本音響学会春季研究発表会、２０１２年３月１５日、神奈川大学（神奈川県）

② 蒔苗久則、鎌田敏明、長内隆、「適応処理や SS 法を用いた 非定常雑音の明瞭化処理手法の検討」、日本法科学技術学会、２０１１年１１月１８日、ホテルフロラシオン青山（東京都）

③ 蒔苗久則、鎌田敏明、長内隆、「音声明瞭化ソフトウェアの開発」、日本法科学技術学会、２０１０年１１月１２日、ホテルフロラシオン青山（東京都）

④ 長内隆、 鎌田敏明、 蒔苗久則、 網野加苗、「テキスト独立型話者照合における発話内容の共通性の評価」、日本法科学技術学会、２０１０年１１月１２日、ホテルフロラシオン青山（東京都）

⑤ 鎌田敏明、 長内隆、 蒔苗久則、 網野加苗、「母音間距離を利用した話者の地域性情報に関する検討」、日本法科学技術学会、２０１０年１１月１２日、ホテルフロラシオン青山（東京都）

⑥ 長内隆、 鎌田敏明、 蒔苗久則、 網野加苗、「テキスト独立型話者照合における発話内容の共通性に関する検討」、日本音響学会秋季研究発表会、２０１０年９月１５日、関西大学（大阪府）

⑦ 蒔苗久則、鎌田敏明、長内隆、「周波数帯域強調 MFCC による話者認識実験」、日本法科学技術学会、２００９年１１月１２日、ホテルフロラシオン青山（東京都）

⑧ 四宮康治、 蒔苗久則、 鎌田敏明、 長内隆、「尤度比を用いた話者照合法」、日本法科学技術学会、２００９年１１月１２日、ホテルフロラシオン青山（東京都）

⑨ 鎌田敏明、長内 隆、蒔苗久則、「話者認識における話者の地域性情報の抽出に関する検討」、日本法科学技術学会、２００９年１１月１２日、ホテルフロラシオン青山（東京都）

⑩ 長内隆、鎌田敏明、蒔苗久則、「テキスト独立型話者照合における話者モデルとその学習に用いるデータ量に関する考察」、日本法科学技術学会、２００９年１１月１２日、ホテルフロラシオン青山（東京都）

⑪ 四宮康治、 蒔苗久則、 鎌田敏明、 長内

隆、「テキスト依存型話者照合を用いた法科学的検査法の検討」、日本音響学会秋季研究発表会、２００９年９月１７日、日本大学(福島県)

〔図書〕(計２件)
① Kanae Amino, Takashi Osanai, Toshiaki Kamada, Hisanori Makinae, and Takayuki Arai, “Historical and procedural overview of forensic speaker recognition.”, in Forensic Speaker Recognition: Law Enforcement and Counter-Terrorism, pp. 3-20, Springer-Verlag, 2011
② Kanae Amino, Takashi Osanai, Toshiaki Kamada, Hisanori Makinae, and Takayuki Arai, ”Effects of the phonological contents and transmission channels on forensic speaker recognition.”, in Forensic Speaker Recognition: Law Enforcement and Counter-Terrorism, pp. 275-308, Springer-Verlag, 2011.

６．研究組織
(1)研究代表者
蒔苗　久則 (MAKINAE HISANORI)
警察庁科学警察研究所・法科学第四部・研究員
研究者番号：20415441

(2)研究分担者
なし

(3)連携研究者
なし