

研究種目：基盤研究（B）
 研究期間：2005～2008
 課題番号：17300065
 研究課題名（和文） 時期差・時間差・気分による音声変動の解明と分散型話者認識への応用に関する研究
 研究課題名（英文） Analysis of Intra-Speaker Variation and Development of Distributed Speaker Recognition System
 研究代表者
 黒岩 眞吾（KUROIWA SHINGO）
 千葉大学・大学院融合科学研究科・教授
 研究者番号：20333510

研究成果の概要：

本研究では、音声による個人認証技術である話者認識に関する基礎研究、及び携帯電話での利用を想定した応用研究を行った。基礎研究としては、同一の話者が、毎週1日、朝・昼・夕に2年～4年かけて発声した音声を収集し、時期差・時間差・気分等の様々な要因で個々人の音声はどのように変動するか等の解明を試みた。応用研究としては、音声に含まれる個人性の抽出手法、個人音声のモデル化法、モデルと音声の比較手法および判別手法を考案し、従来手法に比べ頑健で高精度な話者認識システムを実現した。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2005年度	7,100,000	0	7,100,000
2006年度	3,300,000	0	3,300,000
2007年度	2,400,000	720,000	3,120,000
2008年度	2,400,000	720,000	3,120,000
年度			
総計	15,200,000	1,440,000	16,640,000

研究分野：総合領域

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：音声学、話者認識、話者識別、話者照合、音声データベース、分散型音声認識

1. 研究開始当初の背景

研究当初、音声認識の分野では分散型音声認識(DSR; Distributed Speech Recognition)の標準化がほぼ終了し、DSRは携帯電話での実用化の段階に進みつつあった。これに伴いITU-T(国際電気通信連合 電気通信標準化部門: International Telecommunication Union, Telecommunication Standardization Sector)では分散型話者照合(DSV; Distributed Speaker Verification)の標準化の検討が行われていた。これに対し我々はDSRと共通の

フロントエンドを使用した高精度な話者認識手法を研究開発することが急務であると考へ、本研究を実施するに至った。一方で、音声科学における未解決の問題として話者内の音声変動の問題があった。この問題は実用的な局面では、原因不明の音声認識誤りや話者認識誤りにつながるもので、これらの技術に対するユーザの不信感を増大させる原因となっていた。特に話者認識においては、同一話者でも、体調、微妙な気分の違い、環境のちょっとした違い等により予想外の認

識誤りが生じており、日本において話者認識技術が普及しない原因のひとつとなっていた。これに対し我々は、この問題を検討するためのデータベースの収集・整備を進めると共に、基礎・応用の両側面からこの問題の解決を図ることとした。

2. 研究の目的

本研究では、個々人の音声、時期（経時変化、加齢による変化、季節変化等）、時間（朝・昼・晩等）、気分・体調等によってどのように変化するかを調査するためのデータベースを構築すると共に、それらの変化の調査及びその仕組みを解明することを基礎研究的側面からの目的とした。また、同基礎研究の結果得られた知見に基づき、時期差や環境の変化に性能を左右されない話者認識手法を確立することを応用研究的側面からの目的とした。応用研究においては、音声の利用が最も期待される携帯電話分野での利用を念頭に、既に標準化されている分散型音声認識フロントエンドに適用可能な話者認識手法（特に話者照手法）を研究開発することを目的とした。

3. 研究の方法

(1) 長・短期間音声データの収集

音声の話者内変動の調査・研究を行うためのデータとして、同一の話者が、毎週1日、朝・昼・夕の3回、音素バランス文50文等（延べ11分程度）を防音室で発声した音声を長期間に渡って収集する。

(2) 時期差・時間差等による音声変動の分析

(1)で収録した音声データ及び科学警察研究所において収集された『多人数話者骨導音声データベース』を用い、音響特徴空間での統計的分析、音声認識・話者認識手法を用いた話者内音声変動の分析を行う。

(3) 話者認識に有効な特徴量の検討

(2)の分析結果に基づき、時期差・時間差とうによる音声変動の影響を受けにくい音響特徴量を検討する。さらに、それらの特徴量を用いた頑健な話者認識手法を検討する。

(4) 分散型話者認識手法

DSR フロントエンドで離散化されたデータに対しても頑健かつ精度の高い話者認識手法を研究開発する。また分散型話者照合として実用化するための閾値設定法を検討する。

(5) 感性情報抽出法

話者の気分を抽出しモデル化する手法を検討する。本研究期間では言語情報からの感性情報抽出及びモデル化を目標とする。

4. 研究成果

(1) 長・短期間音声データの収集

[成果] 週1日、朝（10時前後）・昼（13時前後）・夕（16時前後）、各11分程度、女性話者1名が2年間、男性話者2名が3年間、男性話者1名が4年間に渡り発声した音声データを収集した（延べ400時間程度の音声）。発声内容は、ATR音素バランス文Aセット50文、単語発声、4桁数字発声、単母音発声及び本人の氏名を含んでいる。収集した音声データのうち、男女各1名の2年分のデータをデータベース化（1文、1単語発声毎にファイル化）した。これらは、国際的にも唯一無二のデータであることから、話者認識研究だけでなく音声学一般の研究に広く利用されるものと期待される。

[今後の展望] 男性話者1名（研究代表者）については、今後も同データの収録を続けていく。データベース化については、一部自動化したが、人手による複数回の聞き直し作業が必要なため、収録時間の10倍以上の時間を要している。そのため、今後もデータベース化を続ける必要があり、ボランティアによる作業に加え、別途予算の獲得も目指す。

(2) 時期差・時間差等による音声変動の分析

[成果] 特定話者の音響モデルを用いた場合でも収録毎で音声認識率のばらつきが大きい（音素認識率50%～75%）こと、及び、3ヶ月程度の時期差により話者認識誤り率が増大（時期差なし：1.5%⇒時期差あり：20%）することが明らかとなった。しかしながら、音声認識率と、収録時の体温、血圧、脈拍、自己申告による気分、部屋の湿度との相関は見られなかった。唯一、温度のとの若干の相関が見られたが、話者内音声変動の物理的要因を特定するには至らなかった。なお、収録毎に発声速度の変動があり、音声認識率にかなりの影響を与えることも明らかとなり、その後の収録では発声速度が同じになるよう注意した。

時期差については、学習時と認識時の期間差が大きくなるにつれ精度は低下しており、朝・昼・夕の違いよりも1週間の期間差の影響がより大きいことが分かった。これに対し、切り出し作業からは、「朝・昼・夕の音声の違いは概ね判定でき、その違いの方が1週間の違いよりも大きいように思える」との報告があり、音声認識・話者認識に用いていないプロソディ等に関わる情報が人間の音声知覚に影響を与えていることが明らかとなった。また、COSMOS法により音響特徴量の時期差による変化を可視化して観測した。その結果、音響特徴量は時期の変化とともに徐々に変化しているものの、その軌跡には規則性は無かった。但し、風邪等による明らかな音声変動については、その違いを検出できた。

【今後の展望】 話者内音声変動に関するいくつかの知見は得られたものの、科学的な結論は得られなかったため、下記(3)の研究成果を参考に話者内変動が現れる部分空間を特定し、話者内変動のモデル化および特徴量と生体の物理的な関連性を明らかにする。

(3) 話者認識に有効な特徴量の検討

【成果】 対数パワースペクトル空間において、音韻性・時期差を表す部分空間を構成し、それらの部分空間と直交する部分空間で特徴量を算出する方法を提案した。時期差1ヶ月～2年のデータ(『NTT 話者認識データベース』、男性話者23名)を用いた話者識別実験において、この特徴量が MFCC 等の一般的な音響特徴量に比べ、高い話者認識性能を示すことを確認した(MFCC:96.7%、提案特徴量:97.4%)。
【今後の展望】 話者認識では用いられることの少なかった声帯振動に起因する音源情報も考慮し、話者性が現れる部分空間や話者内変動が現れる部分空間を特定する方法を検討すると共に、非線形手法を導入しより分離度の高い部分空間構成法を開発する。

(4) 分散型話者認識手法

① 情報圧縮に頑健な話者認識手法

【成果】 情報圧縮により離散化された特徴量に対しても頑健な話者認識手法として、離散分布を用いた話者モデル構成法、及び、Earth Mover's Distance を用いた話者認識手法を提案した。同手法は、2006年11月に開催された International Symposium on Chinese Spoken Language Processing の話者認識コンテストの話者識別部門(時期差のない電話音声データ、登録時と識別時の回線種別は同一、評価データには登録されていない話者の音声データを含む)において1位の成績(話者識別率99.3%)を収めた。一方で、同手法は計算量を要するという問題(当初10GFLOPSの計算機で2秒の音声の処理に30分程度を要していた)があったため、高速化手法の検討を進めた。いくつかの手法を提案し比較検討した結果、離散分布のクラスタを話者毎に再構成する手法により、2秒の音声に対し2分程度で処理できるまでに至った。また、認識精度を若干犠牲にすることにより、実時間処理を可能とする手法の開発も行った。

本提案手法は、欧州電気通信標準化機構(ETSI)により2003年までにはほぼ標準化が終了し、2005年ごろから順次日本の携帯電話端末に搭載されている分散型音声認識フロントエンドに対応した話者認識手法である。そのため、現行の携帯電話回線においても端末の変更なしに本提案法を導入可能であることから、社会的なインパクトも大きいと考えられる。

【今後の展望】 現在、携帯電話通信会社等と実用化に向けた検討を開始している。一方で、計算量にとられない高精度な話者認識手法の研究も進めている。具体的には、GMM-supervector と Earth Mover's Distance をカーネル関数として用いたサポートベクターマシンによる話者照合手法の研究を行っている。また GMM の学習に識別的学習手法を導入する予定である。

② 詐称者受率率を設定可能な話者照合手法

【成果】 話者照合においては、本人であるか他人であるかを判定するための閾値処理が必要となる。本研究では、詐称者誤受率を事前に、かつ高精度に設定できる手法として、順序統計量を用いた閾値設定手法を提案した。提案手法は、話者識別の概念を話者照合に応用したものであり、入力音声と多数の話者モデルのスコアを各々計算し、本人モデルのスコアがN位以内に入るか否かにより照合を行う。提案手法は話者空間での話者分布の粗密が一定になるように話者分布を非線形変換していると解釈できる。実験の結果、T-norm 等の正規化手法に比べ30%程度の誤り削減率を達成した。また、時期差があっても詐称者誤受率にほとんど変化がないことを確認した(ただし、時期差により本人誤排除率は増加する)。このことは、セキュリティ分野への応用においては、大きな利点となる。さらに、提案手法の高速化を図るため、予め話者クラスタリングを行うことで、照合時に使用する多数話者のモデル数を1/8に削減することにも成功した。同手法は①の話者認識手法だけではなく、GMM、VQ等の従来手法に対しても有効な手法であり、現在実用化されている多くの話者照合システムに導入可能であり、工学的価値の高い方法であると考えられる。

(5) テキストからの感性情報抽出法

【成果】 200名のアンケート調査に基づき、喜び(happy)、悲しみ(sad)、驚き(surprise)、怒り(angry)、恐れ(fear)、嫌悪(disgust)、平静(quiet)に関する心的状態遷移ネットワークを構築した。同ネットワークは、外部からの刺激により心的状態がどのように変化するかを確率モデルとして表現したものであり、音声からの心的状態の推定に活用可能であると期待される。また、感情生起事象文型パターンや感情語辞書に基づき、会話文に含まれる感情の推定手法を提案した。今後、音声認識と同手法を組み合わせることで、音声に含まれる感情を推定する手がかりが得られるものと期待される。

【今後の展望】 音声分野で感情を扱うためには、研究者が広く利用できる、心的状態をラベル付けした音声データベースの構築が

必要である。一方、感情のように曖昧な事象を扱うための客観的評価手法の確立も大きな課題である。

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 25 件)

- ① Shingo Kuroiwa, Satoru Tsuge, Fuji Ren, “Distributed Speaker Identification and Verification for Mobile Terminals,” International Journal of Biomedical Soft Computing and Human Sciences, Vol.14, pp.3-9, 2008, 査読有
- ② Junko Minato, Kazuyuki Matsumoto, Fuji Ren, Seiji Tsuchiya, Shingo Kuroiwa, “Evaluation of Emotion Estimation Methods Based on Statistic Features of Emotion Tagged Corpus,” International Journal of Innovative Computing, Information and Control, Vol. 4, No.8, pp.1931-1941, 2008, 査読有
- ③ Shingo Kuroiwa, Satoru Tsuge, Masahiko Kita, Fuji Ren, “Speaker Identification Method Using Earth Mover’s Distance for CCC Speaker Recognition Evaluation 2006,” International Journal of Computational Linguistics & Chinese Language Processing, Vol. 12, pp.239-254, 2007, 査読有
- ④ 松本和幸, 三品賢一, 任福継, 黒岩眞吾, “感情生起事象文型パターンに基づいた会話文からの感情推定手法,” 自然言語処理, Vol. 14, 2007, 査読有
- ⑤ 柘植覚, 黒岩眞吾, 獅々堀正幹, 任福継, 北研二, “分散音声認識における実時間周波数特性正規化手法,” 情報処理学会論文誌, Vol.48, pp.900-908, 2007, 査読有
- ⑥ Mohamed Abdel Fattah, Fuji Ren, Shingo Kuroiwa, “Effects of Phoneme Type and Frequency on Distributed Speaker Identification and Verification,” IEICE Transactions on Information and Systems, Vol. E89-D, pp.1712-1719, 2006, 査読有
- ⑦ Shingo Kuroiwa, Yoshiyuki Umeda, Satoru Tsuge, Fuji Ren, “Nonparametric Speaker Recognition Method using Earth Mover’s Distance,” IEICE Transactions on Information and Systems, Vol. E89-D, pp.1074-1081, 2006, 査読有
- ⑧ Satoru Tsuge, Masami Shishibori, Fuji Ren, Kenji Kita, Shingo Kuroiwa, “Specific Speaker’s Japanese Speech Corpus over Long and Short Time Periods,” Journal of Research on Computing Science, Vol.17, pp.115-124, 2006, 査読有
- ⑨ 黒岩眞吾, 小林邦嘉, 柘植覚, 任福継, “音声認識・音声合成を用いた音声途切れ補間手法,” 日本音響学会誌, Vol.62, pp.3-9, 2006.

査読有

[学会発表] (計 45 件)

- ① 陸昊澤, 西田昌史, 堀内靖雄, 黒岩眞吾, “音韻変動を抑制する特徴変換に基づく話者認識,” 日本音響学会春季研究発表会, 1-P-26, pp.199-200, 東京, Mar. 2009.
- ② 黒岩眞吾, 柘植覚, “話者認識技術の紹介と最近の研究動向,” 電子情報通信学会・バイオメトリックシステムセキュリティ研究会, 第 17 回研究発表会, pp.19-25, 東京, 2009 年 3 月.
- ③ Satoru Tsuge, Minoru Fukumi, Shingo Kuroiwa, “Specific Speakers’ Speech Corpus Over Long and Short Time Periods,” Oriental COCOSA, 4pages, Kyoto, Nov. 2008.
- ④ Haoze Lu, Haruka Okamoto, Masafumi Nishida, Yasuo Horiuchi, Shingo Kuroiwa, “Text-Independent Speaker Identification Based on Feature Transformation to Phoneme-Independent Subspace,” 11th IEEE International Conference on Communication Technologies, pp.692-695, Hangzhou, China, Nov. 2008.
- ⑤ Satoru Tsuge, Osanaï Takashi, Makinae Hisanori, Kamada Toshiaki, Minoru Fukumi, Shingo Kuroiwa, “Combination method of Bone-conduction Speech and Air-conduction Speech for Speaker Recognition,” Interspeech, Brisbane, Australia, pp.1929-1932, Sep. 2008.
- ⑥ 喜多雅彦, 黒岩眞吾, 柘植覚, 蒔苗久則, 長内隆, 鎌田敏明, 谷本益巳, 土屋誠司, 福見稔, 任福継, “大規模話者骨導音声データベースを用いたテキスト独立型話者照合実験,” 第 9 回音声言語シンポジウム, 2007-SLP-69-1, pp.183-188, 京都, 2007 年 12 月.
- ⑦ Satoru Tsuge, Keiji SEIDA, Masami Shishibori, Kenji Kita, Fuji Ren, Minoru Fukumi, Shingo Kuroiwa, “Analysis of Variation on Intra-Speakers Speech Recognition Performances,” IEEE NLP-KE2007, pp.387-392, Beijing, China, Aug. 2007.
- ⑧ 喜多雅彦, 柘植覚, 黒岩眞吾, 任福継, “多数の話者モデル内での順位情報を用いた話者照合,” 人工知能学会言語・音声理解と対話処理研究会, SIG-SLUD-A701-2, pp.7-12, 徳島, 2007 年 7 月.
- ⑨ Satoru Tsuge, Minoru Fukumi, Masami Shishibori, Fuji Ren, Kenji Kit, Shingo Kuroiwa, “Study of Relationships Between Intra-Speaker’s Speech Variability and Speech Recognition Performance,” 2006 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 2006), pp.41-44, Tottori, Dec. 2006
- ⑩ Shingo Kuroiwa, Satoru Tsuge, Masahiko Kita and Fuji Ren, “Evaluation of EMD-based Speaker Recognition using

ISCSLP2006 Chinese Speaker Recognition Evaluation Corpus,” 5th International Symposium on Chinese Spoken Language Processing, pp.539-548, Kent Ridge, Singapore, Nov. 2006.

⑪ Satoru Tsuge, Masami Shishibori, Kenji Kita, Fuji Ren, Shingo Kuroiwa, “Study of Intra-Speaker’s Speech Variability over Long and Short Time Periods for Speech Recognition,” IEEE International Conference on Acoustic, Speech, and Signal Processing, pp.397-400, Toulouse, France, May 2006.

⑫ Satoru Tsuge, Masami ShiShibori, Kenji Kita, Shingo Kuroiwa, “Data Collection for Investigating Speech Variability in a Specific Speaker Over Long and Short Time Periods,” IEEE International Conference on Natural Language Processing and Knowledge Engineering, pp.152-157, Wuhan, China, Oct.-Nov., 2005.

⑬ Shingo Kuroiwa, Yoshiyuki Umeda, Satoru Tsuge, Fuji Ren, “Distributed Speaker Recognition using Speaker-Dependent VQ Codebook and Earth Mover’s Distance”, INTERSPEECH 2005, pp.3085-3088, Lisbon, Sep. 2005.

6. 研究組織

(1) 研究代表者

黒岩 眞吾 (KUROIWA SHINGO)
千葉大学・大学院融合科学研究科・教授
研究者番号：20333510

(2) 研究分担者

北 研二 (KITA KENJI)
徳島大学・高度情報化基盤センター・教授
研究者番号：10243734
任 福継 (REN FUJI)
徳島大学・大学院ソフテックサイエンス研究部・教授
研究者番号：20264947
柘植 覚 (TSUGE SATORU)
徳島大学・大学院ソフテックサイエンス研究部・講師
研究者番号：00325250
西田 昌史 (NISHIDA MASAFUMI)
千葉大学・大学院融合科学研究科・助教
研究者番号：80361442

(3) 連携研究者

堀内 靖雄 (HORIUCHI YASUO)
千葉大学・大学院融合科学研究科・准教授
研究者番号：30272347