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２５年　６月　３日現在

機関番号：13904
研究種目：挑戦的萌芽研究
研究期間：2010～2012
課題番号：22650034
研究課題名（和文）　音声言語情報に含まれるプライバシ情報の保護に関する研究

研究課題名（英文）　Study on privacy protection in spoken language

研究代表者
中川　聖一（NAKAGAWA　SEIICHI）
豊橋技術科学大学・大学院工学研究科・教授
研究者番号：20115893

研究成果の概要（和文）：音声には、発声者を特定できる声質情報と発声内容に含まれる個人名などのプライバシ情報が含まれている。本研究では、まず音声と背景音を分離するためにベクトル量子化(VQ)に基づく手法と非負値行列因子分解法(NMF)に基づく手法を提案し、背景音との重畳音声から、音声だけを抽出し除去する技術を開発した。また、抽出された音声を声質変換する技術、および音声認識により、音声に含まれている人名を抽出し除去する技術を開発した。

研究成果の概要（英文）：Spoken language contains speaker characteristics and contents related to privacy information such as personal names. In this study, we developed mixed sound separation techniques based on Vector Quantization and Non-negative Matrix Factorization (NMF), elimination technique of only speech in mixed sound and conversion technique of speaker characteristics. Finally, we developed a personal name detection/elimination method based on the modification of a language model.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2010 年度	1,000,000	0	1,000,000
2011 年度	1,200,000	360,000	1,560,000
2012 年度	600,000	180,000	780,000
総　計	2,800,000	540,000	3,340,000

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：　音声情報処理、プライバシ保護、声質変換、音源分離、音声認識、人名抽出

１．研究開始当初の背景

ごく近い将来に社会の至る所で様々なユビキタスセンサネットワークが構築されるようになると予測されるが、次のステップとして、オープンで汎用的なセンサデータ資源として WWW のように社会全体で共有化する“センシングウェブ”へと進化すると予想される。このようなオープンなセンサデータに含まれる情報利活用の一部として、音センサ（マイクロホン）を通して収集される音声言語情報に含まれるプライバシ情報の保護（除去）を行うための技術開発が望まれている。

２．研究の目的

ごく近い将来において様々なユビキタスセンサネットワークが構築されるようになると考えられるが、そのネットワーク上ではプライバシ保護が重要になってくると考えられる。本研究では、世界的に見て他に例のない音声言語情報に対するプライバシ保護技術として、背景音だけを通信するために「混合音からの音声除去技術」の開発、発声者が特定できないように「実環境下での声質

変換技術」の開発、発話内容に含まれている固有名などの除去を目的とした「実環境下音声認識と認識結果に対するプライバシ保護言語処理技術」の開発を行う。これらは、現状の音声認識技術から見て、非常にチャレンジングな課題であると同時に、今後のユビキタス環境下における基盤技術になっていくと考えられる重要な研究課題である。

３．研究の方法

(1) 実環境下での声質変換法

現在の声質変換技術は、クリーン音声をベースとした研究であり、背景音が存在する環境ではその変換性能が低下する。また、背景音が存在する環境では基本周波数の推定が難しい問題があり、これも声質変換の性能を落とす原因となっている。これに対して、これまでに実環境下音声認識で培った背景音低減処理技術を応用して、背景音が存在する実環境での高性能且つリアルタイム動作可能な声質変換技術を開発する。

(2) 音声除去法の研究

本手法では、予め背景音を重畳した音声と、それに対応するクリーン音声（背景音がない音声）のスペクトルをペアとしたものを特徴量としてベクトル量子化（VQ）コードブックを作成しておき、入力音声（背景音＋音声）と VQ コードブックの背景音を重畳した音声部分のスペクトルとの距離を計算、最も距離が近い VQ コードからクリーン音声部分のスペクトルを取り出し、そのスペクトルを入力音声スペクトルから差し引くことで、背景音を復元する。

また、非負値行列因子分解法(NMF)では、音声の基底ベクトルは VQ コードベクトル集合で与え、背景音の基底ベクトルも背景音の集合から VQ コードベクトル集合として求めておく。これらの基底ベクトルの重み付き和で入力の混合音を分解し、音声の基底ベクトルの重み和で音声を抽出する。この手法は、計算量が多いので、高速化を図る。

(3) 音声に現れる固有名除去法

音声認識システムが認識できない語彙外の単語の内、固有名詞が占める割合は多く、また固有名詞はプライバシ情報となりやすい。さらに、新しい固有名詞が時間の流れと共に生まれるが、これは当然認識辞書に含まれないため音声認識システムでは認識できない。これに対処する方法として新出語を音声認識辞書に登録する方法があるが、無差別に大量の単語を認識辞書に登録しても認識性能の低下を招くだけであり、効率的でない。本研究の枠組みでは、これらの未知語（新造語）がプライバシ情報たり得るか否かのみが重要であるため、この観点から未知語への対処（音声を提示する場合は音声除去技術により提示音声から除去、テキスト情報として提示する場合は認識結果の削除を行う）を検討する。

実際の音センサーは、発話者との距離が遠くに設置されており、数メートルの遠隔発話になりうる。遠隔発話の音声認識は、音量が小さくなり、残響や雑音の影響で、非常に難しい。本研究では、残響処理を中心に、遠隔発話の音声認識技術を開発する。

４．研究成果

音声中のプライバシ情報を保護するためには、発声者の隠蔽（声質変換）とプライバシに関する内容の除去が必要である。本研究では以下の成果を得た。

(1) 声質変換

音声に含まれる最も大きなプライバシ情報は、発声者が誰であるかが分かることである。そのため、テレビのインタビュー等では、主に、声の高さを変更する音声変換装置で声質を変えている。しかし、これだと不自然な音声になったり、抑揚などの韻律情報は保存されていて、話者情報が残ってしまう。また、背景雑音中の音声では、背景雑音も変形すると正確な場の雰囲気などの情報が伝わらなくなる。そこで、本研究は、背景音の重畳した音声から音声だけを抽出し、音声の声質を変換した後、元の背景音に加える方法を開発した。背景音の重畳した音声からの音声の抽出に関しては、次節で述べる。

音声の個人性を表すパラメータは、ピッチ、スペクトルピークの周波数、スペククトルピークの形状、スペクトルの傾きなどであるが、これらを個々に変形するのではなく、スペクトルを表すパラメータがガウス分布をしていると仮定して、この分布を変形した。スペクトルパラメータの抽出には、メル LPC 分析を使用し、分析合成には MLSA フィルターを使用した。センシング情報の実時間処理を目指しているので、計算量の少ない方法で声質変換を実現した。

レストランの背景雑音を重畳した音声(10dB と 0dB の信号対雑音比)に対する評価結果を図１に示す。図は、５点満点（１～５点）の被験者 10 人の平均値と標準偏差を示している。図より、分析合成音の質は低下しているが、発声者の個人性は除去されていることが分かる（スコア５が、元音声と大きく異なることを示す）。しかし、音声自体も変わっていて、性能はまだ不十分である。その第１の原因は、雑音中の音声のピッチ抽出が不正確なことによる。

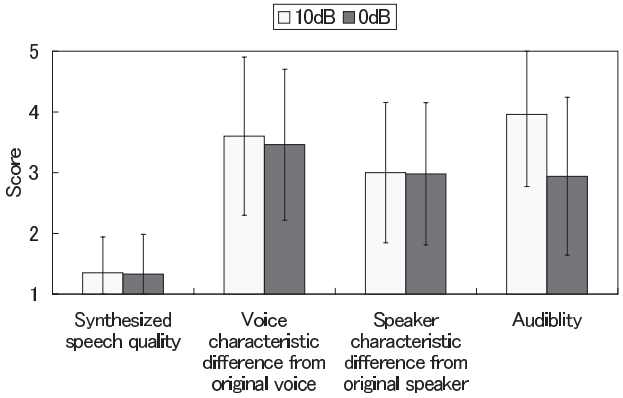

図１　背景雑音中の音声の声質変換の評価結果

(2) 背景雑音と音声の分離と音声除去

背景雑音中の音声から音声だけを除去する、いわゆる音源分離の研究に取り組んだ。まず、ベクトル量子化（VQ）手法と非負値行列因子分解（NMF）によって音楽重畳音声から音声を取り出す方法を考案した。主な点は、①両手法共に、単に音声部分だけを再構成するのではなく、原音と音声推定値からウィナーフィルター方式により推定を行った、②頑健な推定を行うために、ウィナーフィルターにスムーズイング法を導入した、③NMF 法を高速化するために、VQ 手法を併用した高速 NMF を開発した、④VQ コードブックから入力音に近いコードベクトルを選択する際に、音声認識の距離尺度に合致するケプストラム距離を使用した、ことである。これにより、音声の抽出と除去精度が向上した。

NMF 法の大きな欠点は大量の計算を必要とし、実時間処理が不可能なことである。これを高速化するために、NMF 法に VQ 手法を導入した。つまり、音声や音楽の基底ベクトルを、音声や音楽の VQ コードベクトルで定義し、しかも音楽重畳音声も VQ コードブック化し、このコードベクトルに関して、予め NMF で音声と音楽に分解しておく方法である。これにより、音楽重畳音声入力に対し、VQ 化するだけで、実時間で音声を抽出（除去）出来るようになった。図２に VQ 手法の模式図、図３に高速 NMF の模式図を示す。

本手法を評価するために、不特定話者が発声した単語音声にピアノ３重奏（ピアノ、チェロ、ギター）の背景音を-5dB，0dB，10dB，20dB で重畳し、音声認識実験を行った。音源分離を行わないと音声認識率は 20dB で 85.6%（10dB で 56.3%）であったのが、本手法を適用することにより 91.4%（74.5%）まで、向上させることができた。さらに、音楽重畳音声で学習したモデルを用いた場合では、それぞれ 97.8%（92.4%）から 98.4%（95.0%）まで改善することができた。

図４が、雑音重畳音声から音声スペクトルを抽出した例である。背景雑音と音声の分離が、出来ていることが分かる。図中、楕円で囲んだ範囲が、音楽（雑音）除去により、音声スペクトルが明確になった所を示している。

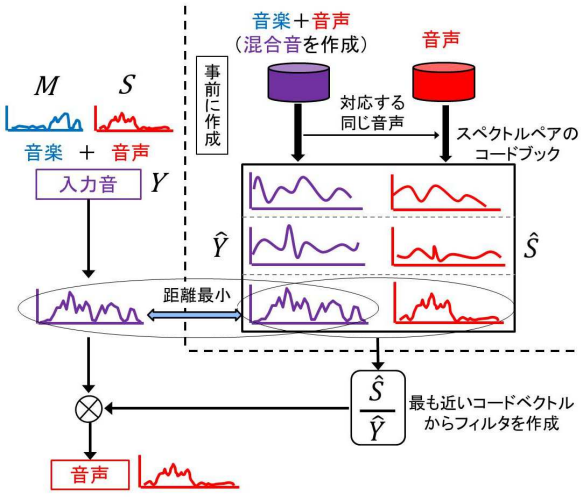

図２　ＶＱ手法による音源分離

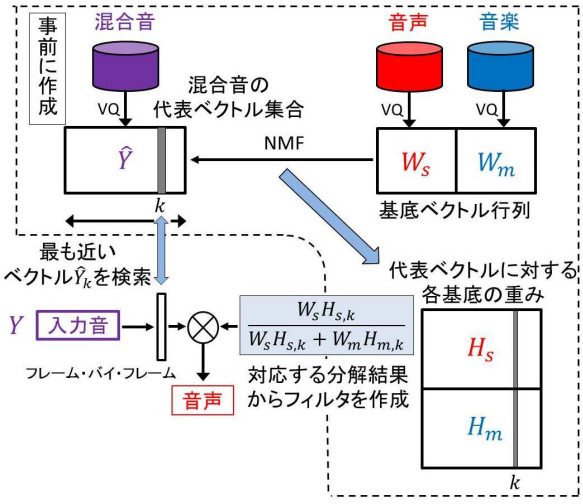

図３　高速ＮＭＦ法による音源分離

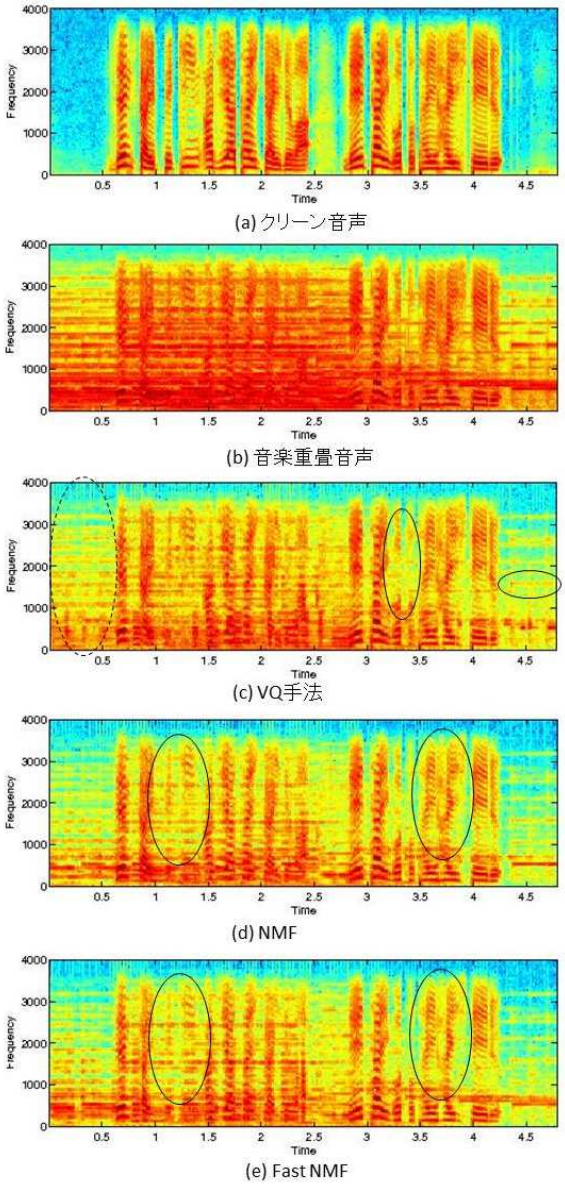

図４　雑音重畳音声から音声を抽出した結果

(3) 音声中の人名の抽出と除去

プライバシの代表的な情報である人名の抽出とその部分の音声除去に取り組んだ。これは、大語彙連続音声認識で人名と認識された区間を除去すれば、目的が達成できるわけではない。その理由は音声認識誤りと、音声認識用の辞書に含まれていない人名の認識は認識できないことによる。そこで、人名の抽出精度を上げ、抽出漏れを極力減らす方法を提案した。そのため、①大語彙連続音声認識における言語モデルで、人名の出現予測確率を大きくすることで人名の漏れを防ぐ方法、②類似した発音の人名を辞書に存在するように辞書に登録する人名を増加させ、人名は発音の類似した人名に誤認識されるようにする方法、③登録した人名の予測確率を与えるためにクラス言語モデルを導入する方法、④辞書に存在しない人名は任意の音節列として認識する方法、を開発した。

NHK のニュース音声を対象に、人名の抽出実験を行った。本研究の目的は、人名を正しく抽出することであり、他の人名に誤認識となっても正解とした。結果を図６と図７に示す。図５は、通常の大語彙連続音声認識で人名の認識を行った結果である。ただし、人名を含むトライグラムの確率を α 倍する手法で、α_1=1.0 のとき（図の左端）が、通常の認識手法のベースラインである。図６は、本研究で開発した手法である。ベースラインで辞書に存在しない未知語の人名の抽出で、再現率 14%、精度 2%であったものが、上述の４手法を併用することにより、再現率 87%、精度 12%に向上した。これは、ニュース音声中の人名の出現率は約 1%であることから、1000 単語中（人名が 10 単語)、人名として 80 単語抽出し、そのうち、9 単語が正しい人名であったということに対応する。音声の 10%程度が誤っても（欠如しても)、意味的にはほぼ正しく理解できることから、本手法は、初期の目標を達成したと言える。勿論、実用的には、再現率を 100%に近づける必要はあるが、今後の音声認識システムの向上により、可能になっていくと考えられる。

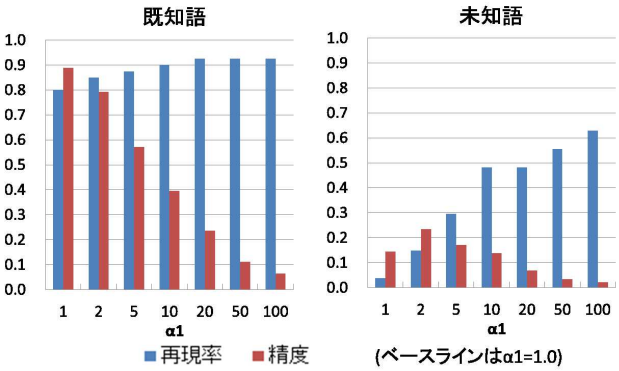

図５　ベースラインによる人名抽出結果

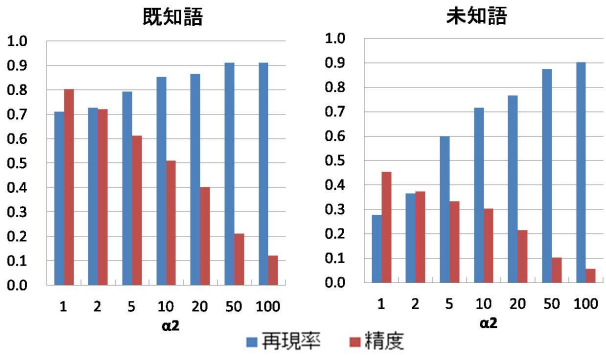

図６　提案手法による人名抽出結果

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　２件）

① W. Naptali, M. Tsuchiya, S. Nakagawa, “Class-based n-gram language Model for new words using out-of-vocabulary to in-vocabulary similarity”, IEICE Trans. Inf. & Syst., 査読有, Vol.E95-D, No.9, pp.2308-2316 (2012.9), 10.1587/transinf.E95.D.2308

② A. Y. Nakano, S. Nakagawa, K. Yamamoto “Distant speech recognition using a microphone array network”, IEICE Trans. Inf. & Syst. 査読有, Vol.E93-D, No.9, pp.2451-2462 (2010.9), 10.1587/transinf.E93.D.2451

〔学会発表〕（計　１５件）

① 仲野翔一, 山本一公, 中川聖一『ケプストラム距離に基づく NMF の高速化手法と VQ 手法による音楽重畳音声の認識』, 日本音響学会春季講演論文集, 1-Q-26b, (2013.3.13) 東京工科大学八王子キャンパス, 八王子市

② 川口亮, 土屋雅稔, 中川聖一『音声ドキュメント中の人名抽出』, 日本音響学会春季研究発表会, 3-P-37b, (2013.3.15) 東京工科大学八王子キャンパス, 八王子市

③ S. Nakano, K. Yamamoto, S. Nakagawa, “Fast NMF based approach and improved VQ based approach for speech recognition from mixed sound”, Proc. APSIPA, SLA. 7, 4pages, (2012.12.4) California, USA

④ A. A. Nugraha, S. Nakagawa, “Improving distant speaker identification robustness using a non-linear regression based dereverberation method in feature domain”, 日本音響学会秋季研究発表会, 3-P-17, (2012.9.21) 信州大学工学部, 長野市

⑤ 仲野翔一, 山本一公, 中川聖一『音楽重量音声の音声認識のためのＮＭＦによる音楽除去の高速化およびＶＱ手法の改善』, 日本音響学会春季研究発表会, 1-P-18, (2012.3.13) 神奈川大学横浜キャンパス, 横浜市

⑥ 嶋田晃太, 山本一公, 中川聖一『残響に頑健な遠隔発話の話者認識の検討』, 日本音響学会春季研究発表会, 1-P-29, (2012.3.13) 神奈川大学横浜キャンパス, 横浜市

⑦ S. Nakano, K. Yamamoto, S. Nakagawa “Speech recognition in mixed sound of speech and music based on vector quantization and non-negative matrix factorization”, Proc. Interspeech, pp.1781-1784 (2011.8.29) Florence, Italy

⑧ 中川聖一, W. Naptali, 岩見圭祐『音声認識・検索のための未知語の扱い』, 情報処理学会, 第 87 回音声言語情報処理研究会, SLP87-6, (2011.7.21) 札幌市

⑨ 仲野翔一, 山本一公, 中川聖一『ＮＭＦとＶＱ手法による音楽重畳音声の音声認識』, 電子情報通信学会, 音声研究会, SP2011-34, (2011.6.23) 名古屋大学, 名古屋市

⑩ W. Naptali, M.Tsuchiya, S. Nakagawa “Multi class-based n-gram language model for new words using Web data”, Proc. 11th Wseas International Conf. MUSP-11, pp.125-131 (2011.3.7-9) Venice, Italy

⑪ 南和江, 藤井康寿, 土屋雅稔, 中川聖一『大規模コーパスを用いた固有表現抽出手法の検討』, 言語処理学会, 第 17 回年次大会, pp.2-9 (2011.3.8) 豊橋技術科学大学, 豊橋市

⑫ 仲野翔一, 山本一公, 中川聖一『ＮＭＦとＶＱ手法による音楽重量音声の音楽除去と音声認識』, 日本音響学会春季研究発表会, 2-P-14, (2011.3.10) 早稲田大学, 東京都

⑬ W. Naptali, M. Tsuchiya, S. Nakagawa “Class-based n-gram language model for a out-of-vocabulary words”, 日本音響学会春季研究発表会, 2-P-37, (2011.3.10) 早稲田大学, 東京都

⑭ W. Naptali, M. Tsuchiya, S. Nakagawa “Modeling out-of-vocabulary words using multi class-based n-gram language model for automatic speech recognition”, 第 5 回音声ドキュメント処理ワークショップ, 5-1 (2011.3.7) 豊橋技術科学大学, 豊橋市

⑮ K. Yamamoto, S. Nakagawa, “Evaluation of privacy protection techniques for speech signals”, Proc.Int. Conf. Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU-2010, pp.653-662 (2010.6.28-7.2) Dortmund, Germany

〔図書〕（計　２件）

① 中川聖一, 小林聡, 峯松信明, 宇津呂武仁, 秋葉友良, 北岡教英, 山本幹雄, 甲斐充彦, 山本一公, 土屋雅稔『音声言語

処理と自然言語処理』，コロナ社，(2013.3)，264 ページ
② 中川聖一『情報理論―基礎から応用まで―』，近代科学社（2010.6），242 ページ

〔その他〕
ホームページ等
http://www.slp.cs.tut.ac.jp

６．研究組織
(1)研究代表者
中川　聖一（NAKAGAWA SEIICHI）
豊橋技術科学大学・大学院工学研究科・教授
研究者番号：20115893

(2)研究分担者
山本　一公（YAMAMOTO KAZUMASA）
豊橋技術科学大学・大学院工学研究科・准教授
研究者番号：40324230
土屋　雅稔（YAMAMOTO KAZUMASA）
豊橋技術科学大学・情報メディア基盤センター・助教
研究者番号：70378256