様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２８ 年　 ６ 月　 ９ 日現在

機関番号：１２６０１
研究種目：基盤研究(A)（一般）
研究期間：2011～2014
課題番号：２３２４００２３

研究課題名（和文）高次統計量追跡による自律カスタムメイド音コミュニケーション拡張システムの研究

研究課題名（英文）A study on custom-made augmented speech communication system based on higher-order statistics pursuit

研究代表者
猿渡　洋（Saruwatari, Hiroshi）

東京大学・情報理工学（系）研究科・教授

研究者番号：３０３２４９７４

交付決定額（研究期間全体）：（直接経費）　　36,800,000円

研究成果の概要（和文）：本研究では、高次統計量追跡による自律カスタムメイド音声コミュニケーション拡張システムに関して研究を行った。具体的なシステムとして、ブラインド音源分離に基づく両耳補聴システムや声質変換に基づく発声補助システムを開発し、以下の成果が得られた。（１）両耳補聴システムに関しては、高精度かつ高速なブラインド音源分離及び統計的音声強調アルゴリズムを提案し、聴覚印象の不動点を活用した高品質な音声強調システムが実現できた。（２）発声補助システムに関しては、データベース間における発話のミスマッチを許容する声質変換処理を開発した。実環境模擬データベースを用いてその評価を行い、有効性を確認することが出来た。

研究成果の概要（英文）：In this study, we address an unsupervised custom-made augmented speech communication system based on the higher-order statistics pursuit. This system consists of two parts, namely, a binaural hearing aid using blind source separation and a speaking aid via speech conversion. The following results are obtained. (1) As the binaural hearing-aid system, we propose new algorithms for an accurate and fast blind source separation and statistical speech conversion, yielding a high quality speech enhancement system utilizing a fixed point of auditory perception. (2) As the speaking-aid system, a new robust speech conversion algorithm against a mismatch between speech database is proposed. The evaluation using real-world sound database shows the efficacy of the proposed method.

研究分野：知能情報処理

キーワード：音声情報処理　統計的学習理論　音響信号処理

１．研究開始当初の背景

近年、音声・音声信号処理においては、従来のような決定論的な定常・線形信号処理のみならず、統計的かつ非定常・非線形信号処理に基づくものが提案されている。これは線形信号処理以上のパフォーマンスを求めた結果によるものであるが、その一方で、処理アルゴリズムの挙動解析（数理面での最適性の保証）が困難になるばかりでなく、未だに人間が聞いて心地よいと感じる信号処理の確立（聴覚的な意味での最適性）には至っていない。逆に、動的かつ非線形な信号処理は、しばしば、人間の聴覚上において不自然と感じるアーチファクト（人工的な副作用）を発生することが報告されている。つまり、将来的に非線形信号処理をも統一的に扱うには、処理のアルゴリズムをただ考えるだけではなく、「信号がどのような統計的な変形を受けたのか？」という観点から数学的にその処理結果を予測・推定することが必要となってくる。

また、全ての信号処理（線形・非線形信号処理を含む）を統一的かつ数学的に記述するには、従来のように二次統計量のみを扱いかつ単一ガウス分布のみで信号分布を近似する従来法では限界がある。例えば、2000 年に入って研究開発が活性化したブラインド音源分離では、ガウス分布近似から開放された理論が構築されたゆえに、あらゆる統計的性質の信号を分解解析することが可能となった経緯がある。音メディア信号はその発生要因が多岐に渡るため、多様な統計的な性質を考慮することが肝要である。例えば、音声信号は平均的にはラプラス分布に従うことが確認されており、更にミクロな視点で眺めれば各音韻ごとに最適な分布形状は異なる。また、突発的な音を含む実環境騒音もガウス雑音というよりもスーパーガウス分布に近いことがしばしば確認されている。

近年、本研究代表者（猿渡）らによって、高次統計量、特に４次の統計量に基づくカートシス（尖度）の値の変化が、人間の聴覚的印象と強い相関があることが明らかになった。端的に言うと、４次統計量を大きく変化させるような信号処理は、概して瞬間的に音のトーン性（音程感）を付与することに等しく、それゆえに雑音の中から音声成分を取り出すブラインド音源分離にて用いることが出来るし、また過度にトーン性を強調するような非線形信号処理はしばしば自然界の聞きなれた雑音を人工的な雑音に変えてしまい、結果として人が聞いた時に不快感や了解性の低下を引き起こす。そこで我々は、「さまざまな音響信号処理を行った際に信号の高次統計量がどのように変形を受けたか」を追跡すれば、全ての世の中の音響信号に対する線形・非線形信号処理結果を予測・評価できるという着想を得た。また、前記の流れとは独立に、本研究分担者の戸田らは、音声合成における最近の研究において、音声を統計的なパラメトリックモデルで表現し、動的な混合ガウス分布の統計パラメータ推定や変形を行うことによって、高精度に声質やイントネーションを制御可能であることを示した。また同時に、人間の声質表現語（例えば男／女らしい声、太い／細い声など）をも統計的なパラメータとして導入することにより、聴覚的な印象を保持したまま任意の声質を実現できることも示した。

２．研究の目的

前記２つの研究の流れ（猿渡による高品質音源分離と戸田による高品質音声合成）は、共に、統計量の追跡により非線形・動的信号処理の最適化を行うことや、人間の聴覚印象までも制御することを目標としている点において共通である。また、両者を組み合わせることにより、自然性が特に重要視される受聴および発声障害補助システムが実現できる。よって、我々は、統計量追跡による音質制御理論の確立、およびそれを応用した「誰がどのような音環境で使用しても安定かつ高品質なパフォーマンスを提供出来る、自律カスタムメイド音コミュニケーション補助システム」の実現可能性について検討を行うこととした。

本研究の目的としては、まず、「統計量追跡による声質制御理論」の数理的な拡張やその実証を、音補助システムを使用するユーザが存在する様々な音環境を想定して行う。次に、最終的なアプリケーションを想定した実証システムの開発として、以下の双方向システム構築を目標とする。

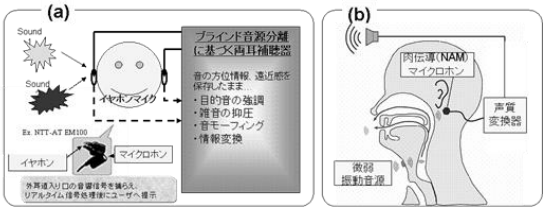

図 1(a) 受聴補助システム、(b) 発話補助システムの概形

（１）受聴補助・拡張システム（図 1(a)参照）

両耳に取り付けられた補聴器を用いて、マルチチャネル信号集音および信号処理を施し、再度ユーザに適切な加工音情報を提示する音拡張現実感（注：音は画像と異なりオクルージョンが生じないので観測音の分解・再合成が必須である）の開発を行う。特に、ユーザを取り巻く複雑な音環境を統計的な独立成分へリアルタイムに分解し、必要な音情報のみを常に一定の品質で選択的に提示することの出来る「高次統計量追跡による高品質ブラインド音源分離に基づく両耳補聴器」を実装する。さらに、分解抽出された独立な音情報を、ユーザの聴覚補正特性・好みに合わせて常に一定の聴覚印象になるよう加工・拡張再現したり、抽出された音響イベン

トを音声認識・理解によって言語化し、情報検索等を通じて必要な情報提供を行う「音コンシェルジュ機能」を付与し、総合的な音拡張現実感システムを実現する。

（２）発話補助システム（図 1(b)参照）

人工喉頭によって生成された口腔内の微弱音を肉伝道型 Non-Audible Murmur マイク（NAM：2004 年鹿野らによって発明）によって取得し、統計的信号分離及び統計的声質変換に基づいて自然な音声を任意の音質で再合成できるシステムを開発する。また、その究極拡張として、発話障害者の非常に少数の元音声が残存していた場合に、それを元に統計的声質変換を駆使して元音声を復元する「失われた声を取り戻す」機能も実現する。

３．研究の方法

本研究おいては、大きく分けて（１）高次統計量追跡による高品質ブラインド音源分離に基づく両耳補聴システム、及び（２）統計量追跡に基づく声質制御可能な発話補助システムの開発、の二項目に従って研究を遂行した。項目（１）に関しては、教師無し適応理論に基づく個人性に依存しない補聴方式、高次統計量追跡による高精度なリアルタイム音源分解、主観的に快適な声質自律制御、及び音情報拡張によるコミュニケーション補助処理の確立を研究遂行課題とした。項目（２）に関しては、NAM 劣化音声の高精度変換に基づく品質保証型音声復元、及び発話障害者自身の元音声復元処理の確立を行った。上記に関して、猿渡の研究統括のもと、主に猿渡・鹿野・川波、小野・宮部、牧野らが項目（１）を担当し、戸田・猿渡・鹿野が項目（２）を担当することで、受聴・発話補助の両面から網羅的に研究を進める。以下に詳細を述べる。

（１）両耳補聴システムの開発

これは、ステレオ型イヤホンの外部にマイクが取り付けられたものを想定しており、外耳道入口付近での音を両耳マイクで収録し、それを前記の独立成分分解や情報拡張系にて適切な処理を加え、イヤホンにて両耳再生する、という音拡張現実感システムである。これにより、軽度難聴者（例えば初～中期の老人性・騒音性難聴など；周波数帯域幅やダイナミックレンジの大きな縮小が生じている重度難聴者は医療的対応が必要なので対象に含まない）の身体能力補助のみならず、健聴者におけるコミュニケーション能力拡大が実現できる。具体的な実施項目を以下に記す。

① 頭部形状の個人性に依存しない補聴方式の確立：　一般に、従来の両耳補聴器の問題として、頭部形状の個人性の問題があった。つまり、ユーザ毎に両耳間の時間差や音量差、頭部による回折効果などが異なるため、従来の固定型の信号処理では音を適切に分解することは不可能であった。一方、高次統計量追跡に基づく独立成分分析では、音源間の独立性のみに基づいて音分解処理を行うため事前にユーザ頭部の情報が一切不要であり、頭部特性への自動適応化処理を内包していると考えられる。これを検証するため、複数ユーザから得られる複数頭部形状を用いて音源分離シミュレーションを行い、個人非依存性を実証する。本音源分離処理系では、広く音声以外の音響イベントも取り扱うことより、マクロな統計モデルとして一般化ガンマ分布等を採用し、そのパラメータの変形度合いを追跡することによって音源分離を実現する。

② 高精度なリアルタイムブラインド音源分離の確立：　一般に、独立成分分析によるブラインド分離処理は、その数理的な自由度と相反して、高次統計量推定における演算量増加および信頼度が問題となり、リアルタイム性に欠けている。これを解決するため、低次補助関数にて反復最適化上限を追跡する補助関数型 ICA(2010 年に小野が発見)を検討する。また、多数の混合状態を分解するため、非線形分解処理である非負行列分解やカーネル ICA（宮部が提案）等も検討する。

③ 主観的に快適な補聴処理の確立：　前記の独立成分分析による分離性能の不十分さを低減するため、高速な非線形ポスト処理を接続する。これにより不要音の抑圧性能は大きく向上するが、非線形処理特有の歪みが発生する可能性がある。本研究では、この問題に関して、主観的に最適な雑音抑圧量と歪みの関係を統計量追跡によって導き出す。また、異なる雑音環境において自動的に最適な雑音抑圧・歪み制御パラメータが設定されるアルゴリズムを開発する。ここでは、マクロな統計モデルとして一般化ガンマ分布、更に（特に音声イベントに関しては）ミクロなモデルとして動的混合ガウス分布を考え、両者の統一的な最適化に取り組む。

④ 音情報拡張による音コミュニケーション補助処理の確立：　抽出された統計的独立音単位での音拡張・変換処理を確立する。具体的には、ユーザがより聞き取りやすい音声を柔軟に生成するために、声質表現語などに基づく直感的な声質・イントネーション制御機能を備えた音声変換技術を構築する。ここでは、音声のミクロな統計量を扱うため、動的な混合ガウス分布によるパラメトリックなモデリングを考え、そのパラメータ変換に関する統計量追跡を行うことで処理の最適化を行う。本技術をさらに拡張し、どのような対象音声も常に所望の一定声質を持つ音声へと変換するシステムも構築する。次に、分解取得された音響イベントを、話者・環境適応（信号処理への適応も含む）されたハンズフリー音声認識器によって言語化し、情報検索等を通じて必要な情報提供を行う機能を付与する。これにより、総合的な音拡張現実感システムを実現する。

（２）発話補助システムの開発

従来の発話障害者（特に喉頭摘出者）は、外部バイブレータ等によって音声を直接生成するが、その品質は通常のコミュニケーションに支障をきたすほど低い。よって本研究では、生成された口腔内の微弱音を首元に装着したNAMマイクによって取得し、統計的な声質変換に基づいて高品質な音声を再合成するという新しい発話補助システムを提案する。具体的な研究課題を記す。

① NAM 劣化音声の高精度変換に基づく聞きやすい音声復元の確立：　一般にNAMマイクで収録された口腔内微弱音声は、その周波数帯域幅が大変狭く、また基本周波数の欠落および SN 比の低下などの問題を抱えている。これを解決するため、まず複数NAMマイクによるブラインド音源分離や、（1）④で述べたような統計的声質変換技術を利用し、高品質かつ聞きやすい音声の復元を図る。また同時に、変換音声に対して、発話障害者の好みを反映させた聴覚印象を定常的に付与できるシステムを構築する。

② 発話障害者自身の元声の復元：　前述の（2）②を更に進め、発話障害が後天的かつそのユーザの少量の元声が存在する場合に、そのサンプルより発話障害者個人の声を復元する統計的声質変換・生成システムを構築する。

４．研究成果

本研究で得られた具体的な研究成果及び開発システムに関して、以下にその概要を述べる。

（1）実環境における聴覚障害者の音響環境を模擬するため、両耳補聴器に関する基礎データベースの収録を行った。ここでは、主に実環境騒音の収集に注力し、最終的に 20 名分の頭部伝達関数と併せ、騒音化での両耳受聴が模擬できるシミュレータシステムを構築した。

（2）両耳補聴器システムを確立するため、統計量追跡による非線形信号処理の最適化問題を数理的に整理した。特に、4 次統計量不動点に基づく聴覚印象不動処理を独立成分分析アルゴリズムに導入し、実環境模擬データに対する分離評価を行い有効性を確認した。

（3）ブラインド音源分離アルゴリズム部の高精度処理に向けて、補助関数型ベクトルICA や高次統計量型方向推定の数理を統計量追跡の観点から理論整備した。実環境模擬データベースを用いて評価を行い、その有効性を確認することが出来た。

（4）発話補助システムの実現に関し、データベース間における発話のミスマッチを許容する声質変換処理を開発した。実環境模擬データベースを用いてその評価を行い、有効性を確認することが出来た。

（5）事前に予測できなかった特筆すべき成果として、ベイズ型音声振幅スペクトル推定における4次統計量不動点を世界で初めて発見し、それを応用したミュージカルノイズフリー音声強調法を開発した。図２に従来音声強調法との音声歪み性能（ケプストラム歪み）の比較を示す。本図より、提案する高次統計量制御型ミュージカルノイズフリー音声強調法は歪みが少なく、聴感上も優れた雑音抑圧性能を示すことが分かる。特に、一般によく使用される音声事前分布（音声振幅スペクトルに特定のレイリー分布を仮定するもの）だけでなく、他のより一般化されたもの（分布形状に自由度を有するカイ分布を仮定するもの）を用いた場合において、事後 SN 及び事前 SN に関するパラメータにバイアスを付加したところ、聴覚品質を不変に保つ 4 次統計量の不動点が確認された。また、非常に興味深いことに、この不動点は全ての事前分布（任意の形状を持つカイ分布）に関して生じるわけではなく、スパースな分布形状を仮定する場合には不動点が消失する現象も確認された。結論として、音声を良く表すスパースな事前分布を利用する際のトレードオフ現象（良い音声事前分布は必ずしも聴覚品質不変にならない）が発見された。

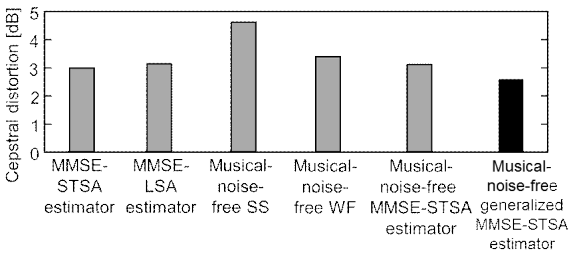

図２　高次統計量制御型音声強調法の比較
（右の 4 提案手法はカートシスが不動）

５．主な発表論文等

〔雑誌論文〕（計 20 件）

① Ryoichi Miyazaki, Hiroshi Saruwatari, Satoshi Nakamura, Kiyohiro Shikano, Kazunobu Kondo, Jonathan Blanchette, Martin Bouchard, “Musical-noise-free blind speech extraction integrating microphone array and iterative spectral subtraction,” 査読有, Signal Processing (Elsevier) vol.102, pp.226-239, 2014.
DOI:10.1016/j.sigpro.2014.03.010

② Hironori Doi, Tomoki Toda, Keigo Nakamura, Hiroshi Saruwatari, Kiyohiro Shikano, “Alaryngeal speech enhancement based on one-to-many eigenvoice conversion,” 査読有, IEEE Transactions on Audio, Speech and Language Processing, vol.22, no.1, pp.172-183, 2014.
DOI:10.1109/TASLP.2013.2286917

③ Ryoichi Miyazaki, Hiroshi Saruwatari, Takayuki Inoue, Yu Takahashi, Kiyohiro Shikano, Kazunobu Kondo, “Musical-noise-

free speech enhancement based on optimized iterative spectral subtraction," 査読有, IEEE Transactions on Audio, Speech & Language Processing vol.20, no.7, pp.2080-2094, 2012.
DOI:10.1109/TASL.2012.2196513

〔学会発表〕(計 85 件)
① Hiroshi Saruwatari, "Statistical-model-based speech enhancement with musical-noise-free properties," 2015 IEEE International Conference on Digital Signal Processing (DSP2015), 2015 年 7 月 24 日, Singapore (Singapore). 招待講演

② Yuki Murota, Daichi Kitamura, Shoichi Koyama, Hiroshi Saruwatari, Satoshi Nakamura, "Statistical modeling of binaural signal and its application to binaural source separation," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2015), 2015 年 4 月 22 日, Brisbane (Australia).

③ Hiroshi Saruwatari, "Information-geometric optimization in nonlinear noise reduction systems," 2013 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS2013), 2013 年 11 月 12 日, 沖縄自治会館 (那覇). 招待講演

６．研究組織
(1)研究代表者
猿渡　洋 (SARUWATARI, Hiroshi)
東京大学・大学院情報理工学系研究科・教授
研究者番号：30324974

(2)研究分担者
鹿野　清宏 (SHIKANO, Kiyohiro)
奈良先端科学技術大学院大学・情報科学研究科・名誉教授
研究者番号： 00263426

(3)研究分担者
戸田　智基 (TODA, Tomoki)
奈良先端科学技術大学院大学・情報科学研究科・准教授
研究者番号： 90403328

(4)研究分担者
川波　弘道 (KAWANAMI, Hiromichi)
奈良先端科学技術大学院大学・情報科学研究科・助教
研究者番号： 80335489

(5)研究分担者
小野　順貴 (ONO, Nobutaka)
国立情報学研究所・情報学プリンシプル研究系・准教授
研究者番号： 80334259

(6)研究分担者
宮部　滋樹 (MIYABE, Shigeki)
筑波大学・大学院システム情報工学研究科・助教
研究者番号： 50598745

(7)研究分担者
牧野　昭二 (MAKINO, Shoji)
筑波大学・大学院システム情報工学研究科・教授
研究者番号： 60396190

(8)研究分担者
小山　翔一 (KOYAMA, Shoichi)
東京大学・大学院情報理工学系研究科・助教
研究者番号： 80734459