

科学研究費助成事業 研究成果報告書

平成 29 年 5 月 24 日現在

機関番号：94305

研究種目：若手研究(B)

研究期間：2014～2016

課題番号：26730100

研究課題名(和文) 時空間音響信号モデリングと機械学習に基づく音響情景分析の研究

研究課題名(英文) Acoustic scene analysis based on time-space acoustic signal modeling and machine learning

研究代表者

亀岡 弘和 (Kameoka, Hirokazu)

日本電信電話株式会社NTTコミュニケーション科学基礎研究所・メディア情報研究部・主任研究員

研究者番号：20466402

交付決定額(研究期間全体)：(直接経費) 2,200,000円

研究成果の概要(和文)：人間は両耳から入る音から、周囲の状況や気配を察知したり、どういう音がいつどこで鳴っているかを聴き分けることができる。本研究では、このような音環境理解(音響情景分析)能力を計算に備えさせることを目的とし、音の物理的性質をベースにした音響信号の時空間モデリングと、人間の聴覚機能をヒントにした音の認識プロセスの確率モデリングに基づく新たな音響情景分析手法を検討した。

研究成果の概要(英文)：Humans are able to recognize what kinds of sounds are present and which direction they are emanating from by using their ears. The aim of this work has been to develop a method that let machines imitate this kind of human auditory function through physical modeling of the generative process of acoustic waveforms and probabilistic modeling of human hearing perception.

研究分野：音響信号処理

キーワード：音響情景分析 深層学習 多重音解析 音響イベント検出 音源分離 到来方向推定 残響除去 高速学習アルゴリズム

1. 研究開始当初の背景

(1) 音響情景分析とは、音響信号から、どういう状況や場面か、どういう種類の音がどの方向からいつ鳴っているかを自動推定するための技術であり、この技術が実現されれば、ライフログ音響ダイアライゼーション、動画・音声コンテンツのインデキシング・検索、聴覚障害者のためのリアルタイム音響環境提示、高齢者の生活安全モニタリングなど、様々なシステムへの応用が可能である。本研究の要素となりうるトピックは、ブラインド音源分離、多重音解析、音源定位、残響・雑音除去、音声区間検出、音響シーン識別、音響イベント検出、音楽音響信号を対象とした場合は、自動採譜やジャンル識別など、多岐にわたる。関連する国際コンテストが最近10年間で数多く開催されているとおり、いずれの研究も音響情報処理に関連する研究領域の中でも重点トピックである。

(2) 音響情景分析の問題は、前景音と背景音（残響や雑音など）を分離する問題、多重に重畳した個々の音源の信号を前景音から分離・定位する問題、各音源の発音区間を検出する問題、各音源が何の音であるか（音響イベント名）を認識する問題、音響シーン（状況・環境）を識別する問題などからなる。例えばブラインド音源分離や音源定位は、異なる位置に配置された複数マイクロホンで取得した多チャンネル音響信号から、混在する複数の音源信号の波形と到来方向を推定することが目的であり、音の空間伝播の性質（球面波や平面波の方程式）や統計的性質（独立性やスパース性など）を手がかりとしたアプローチが主流である。音声区間検出、音響シーン識別、音響イベント検出は、時刻ごとに音の状態、種類、カテゴリを識別することが目的であり、音響的特徴（メル周波数ケプストラム特徴やデルタ特徴など）を手がかりとした識別的アプローチが主流である。なお、近年は深層ニューラルネットワーク（Deep Neural Network）に基づく識別的アプローチが画像認識をはじめとした様々な識別タスクにおいて無類の強さを誇っているが、音声区間検出、音響シーン識別、音響イベント検出への適用も徐々に試みられ始めている。多重音解析は、モノラル音響信号から多重に重畳する音源信号を分離・同定することが目的である。ブラインド音源分離と大きく異なるのはモノラル信号が対象である点で、このため空間的な手がかりが一切ない状況が想定される。多重音解析では空間的な手がかりの代わりに、人間の聴覚機能の性質（聴覚フィルタバンクや音脈分凝など）や音源の生成過程の物理的性質（調波性、ソースフィルタ理論など）を手がかりとしたアプローチが強力である。以上のように、何を入力として何を得たいかという問題設定や、音のどういう性質を手がかりとして解決を目指すのかに依って有効な方法論やアプローチは大きく異なり、これに応じて研究ト

ピックも研究コミュニティも細分化されている。

2. 研究の目的

(1) 音源分離、音源定位、多重音解析、音響シーン識別、音響イベント検出などの音響情景分析の要素問題はいずれも「音を聴き、理解する」という意味では共通しており、決して独立なものではないはずである。しかし、先述のとおり、これらの要素問題は、問題設定や仮定に応じて細分化されており、異なる研究コミュニティで異なるアプローチで取り込まれているのが現状である。申請者は、これまで複数の研究分野を横断的に俯瞰することを常に意識し、さまざまな研究テーマを進めてきた。その過程で、人間と同等以上の能力をもつ究極的な音響情景分析手法を目指す上では、上記の要素問題を全体的（複眼的）に捉え、音響情景分析の問題を、各要素問題で有用な手がかりや視点を数理的に組み込んだ全体最適化問題として定式化することの必要性を強く感じてきた。こうした意識や経験から、音響情景分析問題を本質的に解決し、人間の音響環境理解能力に迫るレベルの性能を発揮するには上述のような複眼的アプローチが不可欠であろうという考えに至った。以上が本研究課題の動機である。

3. 研究の方法

本研究は、複眼的アプローチに基づく音響情景分析手法の実現に向けて、(1) 多重音解析のための新しい音源スペクトルモデルの構築、(2) 深層ニューラルネットワーク（DNN）の新しい学習アルゴリズムの検討、(3) (1)のモデルを用いた多重音解析・ブラインド音源分離・音源定位の融合アプローチの検討、(4) DNNに基づく音響イベント認識手法・音響イベント区間検出手法の検討、(5) (2)と(4)を統合化した音響情景分析手法の検討、の5つのテーマを段階的に取り組む。本研究は、申請者が東京大学で連携客員講座の客員准教授として指導した学生とともに進めた。

4. 研究成果

2014～2016年度の3年間で主に下記の検討を行った。

(1) 音源分離・残響除去・音響イベント検出・到来方向推定の統合的アプローチ：音源分離、残響除去、残響除去、音響イベント検出、到来方向推定の問題がそれぞれ相互依存関係にあるという点に着目し、これらを同時最適化問題として統合的に解決するアプローチを世界で初めて提唱した。

(2) 制約付きボルツマンマシンの高速学習アルゴリズム：人間は音環境を理解する上で、周囲の音響信号が何の音であるかを識別するプロセスが存在するはずであるが、このような識別能力を深層ニューラルネットワーク（DNN）により実現することを目的とし、DNNの一種であるDeep Belief Network (DBN)

の新しい高速学習アルゴリズムを提案した。DBN の学習アルゴリズムとして有名な Contrastive Divergence 法より早く停留点に収束することを確認した。

(3) 計算論的聴覚情景分析とスパースコーディングの統合的アプローチによる多重音解析：人間は音を聴き分ける際、調波性や近接性といったある特定の条件を満たす時間周波数成分をひとまとまりの音と捉える機能とともに、何度も生起する同一パターンの時間周波数成分をひとまとまりの音として捉える機能を有している。これらの二つの機能を同時実現する多重音解析アプローチを世界で初めて提唱した。

(4) 音声のスペクトル系列とケプストラム系列の同時強調：従来のスペクトル領域音声強調法では未知雑音環境下においても雑音や残響の重畳過程モデルを用いることで雑音・残響成分を大きく抑圧することができる利点がある一方で、雑音・残響成分とともに音声信号まで損傷する傾向があり、必ずしも聴感上品質の高い音声を得られないことが課題であった。本研究では、聴感上の品質の高さを保証しながら音声を強調することを目的とした新しいスペクトル領域音声強調手法を提案した。

(5) 音響信号の空間勾配スペクトルの確率モデル化に基づく瞬時音源定位：点音源からの球面波伝播を仮定すると、ある観測点における音圧の空間微分、時間微分と音源位置の間には物理的な拘束関係が成立する。本研究では、この拘束式に基づいて複数のマイクロホンで観測される音響信号の確率モデルを導出し、最尤法およびベイズ推論による高速高精度な瞬時音源定位手法を提案した。従来の音源定位手法は、単一音源のみを対象とした手法か、長時間観測を要する手法かのいずれかであった。提案法は短時間観測のもとで複数音源を対象とした初めての音源定位手法であり、聴覚障がい者のためのウェアラブル音響環境提示システムへの応用が期待される。

(6) 識別的非負値行列因子分解の基底学習アルゴリズム：非負値行列因子分解(NMF)を用いた教師あり音源分離アプローチでは、音源サンプルの基底スペクトルを事前学習することで、混合信号に含まれる当該音源の成分を分離抽出することが可能となる。従来のこの枠組では基底学習において分離信号そのものが最適となるような規準とはなっていないが、分離信号と学習サンプルとの誤差を直接的に最適化規準として基底学習を行う、識別的 NMF と呼ぶ枠組が提案されている。識別的 NMF の学習規準は従来の NMF の学習規準に比べて解析的に複雑な形になるため、汎用的な最適化手法を用いた基底学習方式が採られていたが、この方式は停留点への収束性が保証されておらず識別的 NMF のポテンシャルを十分発揮できていなかった。本研究では、補助関数法という原理に基づく収

束性が保証された識別的 NMF の基底学習アルゴリズムを提案した。

(7) 複素 NMF による高精度音源分離：これまで我々は位相を考慮した NMF の複素領域拡張版である複素 NMF と呼ぶ高精度音源分離手法を提案し、その有効性を示してきた。本研究では、複素 NMF に「双対形式」が存在することを発見し、この「双対形式」に基づき複素 NMF の最適化規準の拡張と時間領域への拡張が可能になり、さらに、多重解像度化など従来の NMF では不可能だった拡張が可能となった。

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)

Hirokazu Kameoka, Takuya Higuchi, Mikihiro Tanaka, Li Li, “Non-negative matrix factorization with basis clustering using cepstral distance regularization,” *IEEE/ACM Transactions on Audio, Speech and Language Processing*, accepted. 査読有

Hirokazu Kameoka, Kota Yoshizato, Tatsuma Ishihara, Kento Kadowaki, Yasunori Ohishi, and Kunio Kashino, “Generative modeling of voice fundamental frequency contours,” *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 23, no. 6, pp. 1042-1053, Jun. 2015. 査読有

[学会発表](計57件)

Takuhiro Kaneko, Shinji Takaki, Hirokazu Kameoka, Junichi Yamagishi, “Generative adversarial network-based postfilter for STFT spectrograms,” in *Proc. 18th Annual Conference of International Speech Communication Association (Interspeech 2017)*, accepted. 査読有

Takuhiro Kaneko, Hirokazu Kameoka, Kaoru Hiramatsu, Kunio Kashino, “Sequence-to-sequence voice conversion with similarity metric learned using generative adversarial networks,” in *Proc. 18th Annual Conference of International Speech Communication Association (Interspeech 2017)*. 査読有

Kou Tanaka, Hirokazu Kameoka, Tomoki Toda, Satoshi Nakamura, “Physically constrained statistical F0 prediction for electrolaryngeal speech enhancement,” in *Proc. 18th Annual Conference of International Speech Communication Association (Interspeech 2017)*, accepted.

査読有

Li Li, Hirokazu Kameoka, Tomoki Toda, Shoji Makino, "Speech enhancement using non-negative spectrogram models with mel-generalized cepstral regularization," in Proc. 18th Annual Conference of International Speech Communication Association (Interspeech 2017), accepted. 査読有

Li Li, Hirokazu Kameoka, Shoji Makino, "Discriminative non-negative matrix factorization with majorization-minimization," in Proc. The 5th Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA2017), pp. 141-145, Mar. 2017. 査読有

Hirokazu Kameoka, Hideaki Kagami, Masahiro Yukawa, "Complex NMF with the generalized Kullback-Leibler divergence," in Proc. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2017), pp. 56-60, Mar. 2017. 査読有

Hideaki Kagami, Hirokazu Kameoka, Masahiro Yukawa, "A majorization-minimization algorithm with projected gradient updates for time-domain spectrogram factorization," in Proc. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2017), pp. 561-565, 2017. 査読有

Li Li, Hirokazu Kameoka, Takuya Higuchi, Hiroshi Saruwatari, "Semi-supervised joint enhancement of spectral and cepstral sequences of noisy speech," in Proc. The 17th Annual Conference of the International Speech Communication Association (Interspeech 2016), pp. 3753-3757, Sep. 2016. 査読有

Naoki Murata, Hirokazu Kameoka, Keisuke Kinoshita, Shoko Araki, Tomohiro Nakatani, Shoichi Koyama, Hiroshi Saruwatari, "Reverberation-robust underdetermined source separation with non-negative tensor double deconvolution," in Proc. 2016 24th European Signal Processing Conference (EUSIPCO 2016), pp. 1648-1652, Aug. 2016. 査読有

Tomohiko Nakamura, Hirokazu Kameoka, "Shifted and convolutive source-filter non-negative matrix factorization for

monaural audio source separation," in Proc. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2016), pp. 489-493, Mar. 2016. 査読有

Hirokazu Kameoka, "Modeling speech parameter sequences with latent trajectory hidden Markov model," in Proc. The 25th IEEE International Workshop on Machine Learning for Signal Processing (MLSP2015), Sep. 2015. 査読有

Takuya Higuchi, Hirokazu Kameoka, "Unified approach for audio source separation with multichannel factorial HMM and DOA mixture model," in Proc. The 2015 European Signal Processing Conference (EUSIPCO 2015), Aug. 2015. 査読有

Hirokazu Kameoka, "Multi-resolution signal decomposition with time-domain spectrogram factorization," in Proc. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2015), pp. 86-90, Apr. 2015. 査読有

Tomohiko Nakamura, Hirokazu Kameoka, "Lp-norm non-negative matrix factorization and its application to singing voice enhancement," in Proc. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2015), pp. 2115-2119, Apr. 2015. 査読有

Takuya Higuchi, Hirokazu Kameoka, "Unified approach for underdetermined BSS, VAD, dereverberation and DOA estimation with multichannel factorial HMM," in Proc. of The 2nd IEEE Global Conference on Signal and Information Processing (GlobalSIP 2014), Dec. 2014. 査読有

Tomohiko Nakamura, Kotaro Shikata, Norihiro Takamune, Hirokazu Kameoka, "Harmonic-Temporal Factor Decomposition incorporating music prior information for informed monaural source separation," in Proc. The 15th International Society for Music Information Retrieval Conference (ISMIR 2014), pp. 623-628, Oct. 2014. 査読有

Hirokazu Kameoka, Norihiro Takamune, "Training restricted Boltzmann machines with auxiliary function approach," in

Proc. The 24th IEEE International Workshop on Machine Learning for Signal Processing (MLSP2014), Sep. 2014. 査読有

Norihiro Takamune, Hirokazu Kameoka, "Maximum reconstruction probability training of restricted Boltzmann machines with auxiliary function approach," in Proc. The 24th IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2014), Sep. 2014. 査読有

Takuya Higuchi, Hirokazu Kameoka, "Joint audio source separation and dereverberation based on multichannel factorial hidden Markov model," in Proc. The 24th IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2014), Sep. 2014. 査読有

Takuya Higuchi, Hirofumi Takeda, Tomohiko Nakamura, Hirokazu Kameoka, "A unified approach for underdetermined blind signal separation and source activity detection by multichannel factorial hidden Markov models," in Proc. The 15th Annual Conference of the International Speech Communication Association (Interspeech 2014), pp. 850-854, Sep. 2014. 査読有

21 Tomohiko Nakamura, Hirokazu Kameoka, "Fast signal reconstruction from magnitude spectrogram of continuous wavelet transform based on spectrogram consistency," in Proc. The 17th International Conference on Digital Audio Effects (DAFx-14), pp. 129-135, Sep. 2014. 査読有

22 Takuya Higuchi, Norihiro Takamune, Tomohiko Nakamura, Hirokazu Kameoka, "Underdetermined blind separation and tracking of moving sources based on DOA-HMM," in Proc. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2014), pp. 3215-3219, May 2014. 査読有

〔図書〕(計4件)

Hirokazu Kameoka, Hiroshi Sawada, Takuya Higuchi, "General formulation of multichannel extensions of NMF variants," in Audio Source Separation, S. Makino (Ed.), Springer, to appear in 2017.

Alexey Ozerov, Hirokazu Kameoka, "Gaussian model based multichannel separation," in Audio Source Separation and Speech Enhancement, E. Vincent (Ed.), Springer, to appear in 2017.

Hirokazu Kameoka, "Non-negative matrix factorization and its variants for audio signal processing," in Applied Matrix and Tensor Variate Data Analysis, T. Sakata (Ed.), Springer Japan, Feb. 2016.

Hirokazu Kameoka, "Probabilistic modeling of pitch contours towards prosody synthesis and conversion," in Speech Prosody in Speech Synthesis: Modeling and generation of prosody for high quality and flexible speech synthesis, K. Hirose, J. Tao (eds.), Springer-Verlag Berlin Heidelberg, 2015.

〔産業財産権〕

出願状況(計22件)

名称: 信号推定装置、方法、及びプログラム
発明者: 亀岡弘和, 関翔悟, 戸田智基
権利者: 同上
種類: 特許
番号: 特願 2017030173
出願年月日: 2017/02/21
国内外の別: 国内

名称: 信号解析装置、方法、及びプログラム
発明者: 亀岡弘和, 李莉
権利者: 同上
種類: 特許
番号: 特願 2017028843
出願年月日: 2017/02/20
国内外の別: 国内

名称: 音源定位装置、方法、及びプログラム
発明者: 亀岡弘和, 植野夏樹
権利者: 同上
種類: 特許
番号: 特願 2017037404
出願年月日: 2017/02/28
国内外の別: 国内

名称: 声道スペクトル推定装置、方法、及びプログラム
発明者: 亀岡弘和, ソウユンハン
権利者: 同上
種類: 特許
番号: 特願 2017037402
出願年月日: 2017/02/28
国内外の別: 国内

名称: 基本周波数モデルパラメータ推定装置、方法、及びプログラム
発明者: 亀岡弘和, 平松薫, 柏野邦夫, 佐藤遼太郎
権利者: 同上
種類: 特許
番号: 特願 2016240303

出願年月日：2016/12/12
国内外の別： 国内

名称：信号解析装置、方法、及びプログラム
発明者：亀岡弘和，岸田拓也
権利者：同上
種類：特許
番号：特願 2016168309
出願年月日：2016/08/30
国内外の別： 国内

名称：信号解析装置、方法、及びプログラム
発明者：亀岡弘和，鏡英章
権利者：同上
種類：特許
番号：特願 2016168322
出願年月日：2016/08/30
国内外の別： 国内

名称：信号解析装置、方法、及びプログラム
発明者：亀岡弘和，鏡英章
権利者：同上
種類：特許
番号：特願 2016168332
出願年月日：2016/08/30
国内外の別： 国内

名称：音声合成学習装置、方法、及びプログラム
発明者：金子卓弘，亀岡弘和，平松薫，柏野邦夫
権利者：同上
種類：特許
番号：特願 2016168356
出願年月日：2016/08/30
国内外の別： 国内

名称：信号解析装置、方法、及びプログラム
発明者：亀岡弘和，田尻祐介，戸田智基，中村哲
権利者：同上
種類：特許
番号：特願 2016032414
出願年月日：2016/02/23
国内外の別： 国内

名称：信号解析装置、方法、及びプログラム
発明者：亀岡弘和，李莉
権利者：同上
種類：特許
番号：特願 2016032396
出願年月日：2016/02/23
国内外の別： 国内

名称：基本周波数パターン予測装置、方法、及びプログラム
発明者：亀岡弘和，田中宏，戸田智基，中村哲
権利者：同上
種類：特許
番号：特願 2016032411
出願年月日：2016/02/23
国内外の別： 国内

名称：音響信号解析装置、音響信号解析方法、及びプログラム
発明者：亀岡弘和，村田直毅
権利者：同上
種類：特許
番号：特願 2016031801
出願年月日：2016/02/23
国内外の別： 国内

名称：音源定位装置、方法、及びプログラム
発明者：亀岡弘和，鈴木惇
権利者：同上
種類：特許
番号：特願 2016032364
出願年月日：2016/02/23
国内外の別： 国内

名称：音源分離装置、方法、及びプログラム
発明者：林亜紀，亀岡弘和，松林達史，澤田宏
権利者：同上
種類：特許
番号：特願 2016014692
出願年月日：2016/01/28
国内外の別： 国内

名称：信号解析装置、方法、及びプログラム
発明者：亀岡弘和，樋口卓哉，竹田裕史
権利者：同上
種類：特許
番号：特願 2014-166903
出願年月日：2014.8.19
国内外の別： 国内

〔その他〕
ホームページ等
<http://www.kecl.ntt.co.jp/people/kameoka.hirokazu/index.html>

6. 研究組織
(1) 研究代表者
亀岡 弘和 (KAMEOKA, Hirokazu)
日本電信電話株式会社・NTT コミュニケーション科学基礎研究所・主任研究員
研究者番号：20466402