様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　元 年　 ６ 月　 ３ 日現在

機関番号：１３５０１
研究種目：基盤研究(C)（一般）
研究期間：2016～2018
課題番号：１６Ｋ０６３８４

研究課題名（和文）音響信号処理と人工知能技術との融合による音響空間全周収録法の確立

研究課題名（英文）Formulation of a recording method of the entire sound field by the fusion of acoustical signal processing and artificial intelligence technology

研究代表者

小澤　賢司（OZAWA, Kenji）

山梨大学・大学院総合研究部・教授

研究者番号：３０２０４１９２

交付決定額（研究期間全体）：（直接経費）　　3,700,000円

研究成果の概要（和文）：本研究は，マイクロホンアレイの出力である時空間音圧分布画像の2次元フーリエ変換に基づいて，音源分離を行うことを目的とした。アレイ正面から到来する目的音については，時空間音圧分布画像は縦縞を形成するので，2次元スペクトルは空間軸方向の直流成分に局在するという特徴がある。
本研究では，雑音（目的音以外の音）の空間直流成分を，人工知能技術の1つである深層ニューラルネットワークを利用して瞬時に推定し，スペクトル減算法を実施することで雑音抑圧を行った。その結果，到来方向が異なる2音の分離に関し，従来法（遅延和法，MV法）より高い性能を示した。
本提案手法は，同一方向にある2音源の分離にも適用可能である。

研究成果の学術的意義や社会的意義
超高品位テレビ方式（スーパーハイビジョン）は，室内に22個のスピーカを配置することで音の空間情報を近似的に再生するものであり，真の意味で音の空間情報を記録するには不十分である。将来的に高次遠隔コミュニケーションを実現する，また歴史的瞬間の現場を完全に記録・再現するためには，全周にわたり精密に音の空間情報を記録する手法の開発が望まれている。これを小規模なシステムにより実現することは，スマートスピーカやスマートフォンの音声インタフェースが一般に普及している今日において大きな社会的意義がある。
本研究では，これを実現するために，急激に発展している人工知能技術を利用することに学術的な意義がある。

研究成果の概要（英文）：This project aims to achieve sound source separation based on the two-dimensional fast Fourier transform (2D FFT) of a spatio-temporal sound pressure distribution image consisting of the outputs of a microphone array. The target sound, which arrives from the front of the array, forms vertical stripes in the image. Therefore, its spectral components are perfectly localized as direct current (DC) components along the spatial frequency axis in the 2D-FFT spectrum.
In this study, noise suppression was performed by spectral subtraction after the DC components of noise were instantaneously estimated from the spectrum using an artificial intelligence technique, deep neural networks. As a result, the proposed method showed better performance than the conventional methods (delay and sum beamformer, MV beamformer) for separating the target sound from sound with a different direction of arrival.
The proposed method is also applicable for separating two sound sources in the same direction.

研究分野：音響信号処理

キーワード：マイクロホンアレイ　時空間音圧分布画像　音源分離　2次元フーリエ変換　ニューラルネットワーク　深層学習　L1正則化　スペクトル減算法

様　式　C－１９、F－１９－１、Z－１９、CK－１９（共通）

１．研究開始当初の背景

(1) 研究開始時点ではまだ本放送が開始されていなかったが，超高品位テレビ方式（8K 方式スーパーハイビジョン）では，部屋の床を除く 5 面に計 22 個のスピーカを配置することで，音の空間情報を近似的に再生する。その録音は 22 本の単一指向性マイクロホンを利用して行われるが，真の意味で音の空間情報を記録するには不十分である。将来的に高次遠隔コミュニケーションを実現する，また歴史的瞬間の現場を完全に記録・再現するためには，全周にわたり精密に音の空間情報を記録する手法の開発が望まれていた。しかし，空間情報の高精度な分離収録を行うためには，多数のマイクロホンから構成されるアレイを用いて，計算コストの高いアルゴリズムによる分析を行う必要があった。

(2) 人工知能技術が急激に発展しており，2045 年にはシンギュラリティ（特異点：人工知能が人類の知能を超える点）を迎えると予想されていた。そこで，人工知能技術との融合によって，計測技術を発展させる適時であると考えた。

２．研究の目的

小規模なシステム構成により，全周にわたる音の繊細な空間情報を記録する技術を確立することを目的とした。

(1) これを実現するための基本原理として以下を考えた。まず，アレイを構成する各マイクロホンにおいて観測された時系列としての音圧の標本値を，輝度値に変換することで 1 ピクセル幅の濃淡画像とする。この画像を全てのマイクロホンについて並べることで構成した 2 次元画像を「時空間音圧分布画像」と定義した。この画像に 2 次元高速フーリエ変換(FFT: Fast Fourier Transform) を適用することで， 2 次元振幅スペクトルを得る。このスペクトルにおいては，到来方向が異なる平面波は，異なる直線上の成分として出現する。そこで，各直線上の成分のみを抽出し，逆 FFT することで目的音の波形だけを抽出することができるので，音源分離が可能である。このように，本研究では，画像処理のなかでも基本的な 2 次元 FFT を基本とすることで実用性の高いシステムの構築を目指した。

(2) 従来は音の物理現象を数式として表現し，それに基づいて音の分離アルゴリズムが考案されてきた。本研究では，深層学習に代表される人工知能技術を利用することで，必ずしも数式に頼ることなく，事前の学習結果に基づいて音源信号を分離するシステムの構築を目指すこととした。ただし，数式に基づく解法についての吟味も怠らないこととした。

３．研究の方法

研究は，純粋に音源分離アルゴリズムの性能を検討するため，環境雑音や室内反射音の影響などの外乱要因を排除できる計算機シミュレーションにより実施した。そして，以下に述べる手順で検討を行った。

(1) 少数マイクロホンで効率よく音情報を収集するための不等間隔マイクロホン配置の検証
(2) 不等間隔マイクロホンで得た時空間音圧分布画像から高精細画像を復元する技術の開発
(3) 少数マイクロホンアレイから得た時空間音圧分布画像による，到来方向が異なる音の分離
(4) 少数マイクロホンアレイから得た時空間音圧分布画像による，距離が異なる音の分離

４．研究成果

(1) 少数マイクロホンで効率よく音情報を収集するための不等間隔マイクロホン配置の検証

間隔が 4 cm，4 cm，15 cm と不等間隔な 4 個のマイクロホンからなるアレイを対象とした。それらの間隔の最大公約数は 1 cm であるため，1 cm 間隔でマイクロホンを配置したのと同様に，空間の標本化定理に基づけば 17 kHz までを収録可能であり，効率的な集音が可能であると考えた。

これを実証するために，図 1 にブロック図を示す差分型のマイクロホンアレイを構築した。4 マイクロホンのうち 1 つを参照マイクロホン RM (Reference microphone) として，他の 3 マイクロホンとの差分信号を処理の対象とした。ここでは，不等間隔マイクロホン配置の有効性の検証を目的としているので，処理には人工知能技術は用いず，雑音の方向・振幅・位相を連立方程式により解くことで推定する方法を提案した。 図 2 に雑音抑圧性能を示すとおり，従来法である

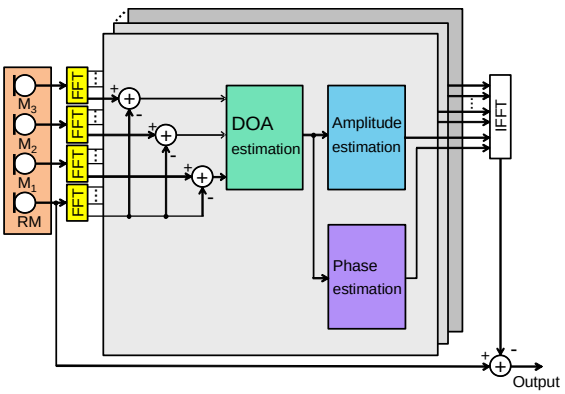

図 1　4 マイクロホンによる雑音抑圧システム

遅延和アレイ（Delay & sum array)に比べて良好な性能を実現し，不等間隔配置の有効性を明らかにした（成果の詳細は，学会発表⑯，⑰として公表した)。

さらに，人工知能技術のうち機械学習の 1 つの手法である L1 正則化(LASSO）を用いることで複数の雑音源が存在する場合でも目的音を分離抽出できることを示した（成果の詳細は，学会発表⑧，⑫として公表した)。

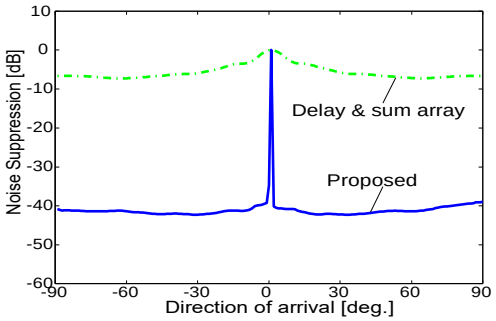

図 2 マイクロホンの不等間隔配置を利用したアレイによる雑音抑圧

(2) 不等間隔マイクロホンで得た時空間音圧分布画像から高精細画像を復元する技術の開発

多数マイクロホンからなるアレイを用いれば詳細な時空間音圧分布画像を得ることができる。しかし，システム規模を抑えるためには少数マイクロホンを用いて実現できることが望ましい。そこで，不等間隔 4 個のマイクロホンから高精細画像の復元を試みた。

音源分離システムとしては，入力音を周波数成分に分解して処理した後に，重ね合わせて出力を得る方式を取り上げた。この方式では，まず各マイクロホン出力である時間波形を，FFT を利用して正弦波成分に分解する。その後に逆 FFT を行うことで各周波数成分の波形を得て，音圧瞬時値を輝度に変換して時空間音圧分布画像を生成する。この疎なマイクロホン配置で得られる画像から，密なマイクロホン配置で得られる画像を復元する。

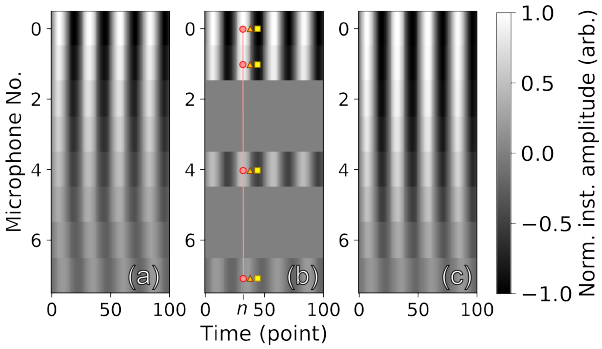

図 3　625 Hz に関する時空間音圧分布画像の復元。(a) μ^2ホンアレイの出力。(b) 4 個のマイクロホンからの出力。(c) 復元された時空間音圧分布画像

図 3(a) は，スマートフォンを想定して 2 cm 間隔で 8 個のマイクロホンを並べたアレイに，基本周波数 150 Hz と 200 Hz の高調波複合音（8 kHz 未満の高調波まで等振幅）が，それぞれ 0 ° と 90 ° から同時に到来した場合について，周波数 625 Hz ビンにおける時空間音圧分布画像を示している。標本化周波数は 16 kHz で，512 点 FFT を行った。

図 3(b) は，4 個のマイクロホンを不等間隔に配置したアレイ（マイクロホン番号：0，1，4，7）に関して，上記と同じ 2 音が到来した場合の時空間音圧分布画像を示している。この図 3(b)から図 3(a)を復元することを目的とした。

時刻 n における 8 点の観測値を，図 3(b) に○印で示した 4 点に加え，時系列として約 1/4 周期（△印)，1/2 周期（□印）だけ離れた位置にある点も含めた計 12 点を用いて復元した。復元には，L1 正則化（LASSO）を用いた。まず，予め −90 ° から 90 ° の範囲を 5 ° 刻みで平面波が到来する場合を想定して音圧の空間パターンを求めておき，辞書行列を構成した。観測された部分波形を，音源が疎に配置されているという制約の下で分解し，それらの線形和として全体を復元した。

図 3(c) に結果を示すとおり，十分な精度で復元が可能であった（復元誤差パワーは，入力音のパワーに比べて −22.3 dB)。周波数が高くなると復元誤差が大きくなるが，電話音声帯域に限定すれば十分に実用に耐えうる精度であることを確認した（成果の詳細は，学会発表⑦，⑨，⑩，⑬〜⑮，⑱として公表した)。

(3) 少数マイクロホンアレイから得た時空間音圧分布画像による，到来方向が異なる音の分離

当初は，時間軸上のデータを 2 点のみ用いる最小サイズの時空間音圧分布画像に対して，画像そのままをニューラルネットワークで処理することで，目的音のみを抽出するシステムを考案し，その性能を評価した。この小規模システムでも正弦波のような狭帯域な音に対しては十分な性能が得られるが，音声のような広帯域な音に対して効果を得るためにはシステムを大規模化する必要があることが明らかとなった（詳細は雑誌論文②，学会発表⑪として公表した)。

そこで，時空間音圧分布画像に対して 2 次元 FFT を施すことで得た，2 次元スペクトルを対象として検討した。音声帯域に対応する小規模システムを対象とし，標本化周波数 16 kHz で検討を行った。この帯域であれば，2 cm 間隔でマイクロホンを配置することで折り返し歪のないアレイが実現される。スマートフォンの長手方向を考えると，マイクロホン

数 8 個で 14 cm のアレイを搭載可能である。

音声 /a/ が到来方向 DOA (Direction of arrival) 0°（アレイ正面）から到来し，白色雑音が DOA 90° から同時に到来した場合の 2 次元振幅スペクトルを図 4 に示す（輝度が振幅に対応している）。このスペクトルでは，DOA 90° から到来する白色雑音が，右上に向かう帯として現れている。また，DOA 0° から到来した音声のスペクトルは，空間周波数 0 rad（空間直流成分）のビン上に局在して現れるが，この図では基本周波数成分のみが視認できる。これは，雑音のスペクトルが空間周波数 0 rad ビン（4 番ビン）に漏れ出しており，音声のスペクトルをマスクしている状態である。

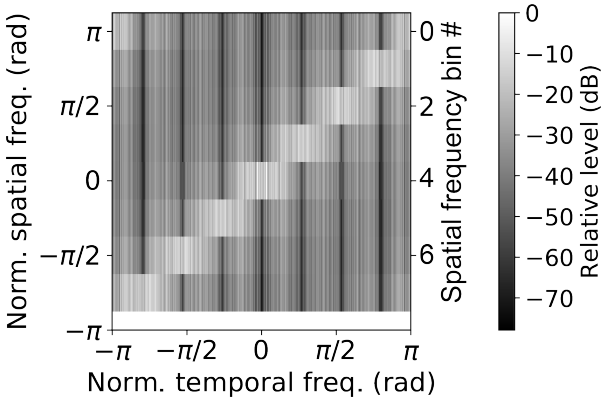

図 4 音声（DOA：0°）と白色雑音（DOA: 90°）が到来した場合の μ^2 ホンアレイスペクトル

そこで，空間周波数 4 番ビン（直流成分）に漏れ出した雑音の振幅値を，ビン 0〜3 および 5〜7 の計 7 個の値から推定し，それを減算することで正面から到来した音声のスペクトルを復元することとした。その推定は，深層ニューラルネットワーク（DNN: deep neural network）による回帰問題として行った。ここでは DNN の入力層・中間層（3 層）・出力層のユニット数はそれぞれ 7・15（3 層）・1 とし，活性化関数としては中間層：ReLU（Rectified liner unit），出力層：恒等関数を用いた。

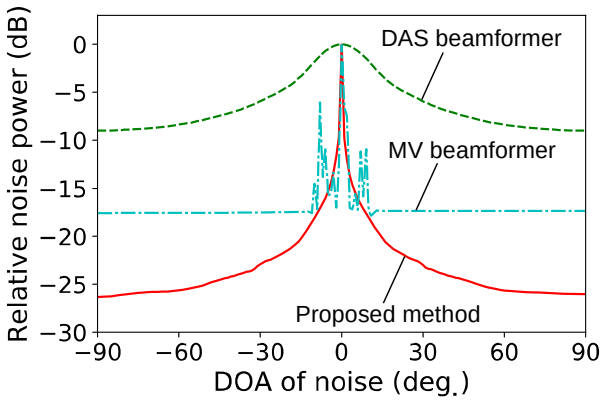

図 5 雑音の到来方向（DOA）の関数としての雑音抑圧の効果音。DOA: 0° には目的音が存在する。

このようにして分析の時間フレーム（512 点）ごとに推定した雑音の振幅値を，空間周波数ビン 4 の振幅値から引き去る瞬時推定スペクトル減算法を実施し，目的音への雑音の混入量の変化を調べた結果を図 5 に示す。図から，従来法である遅延和アレイ（Delay & sum array）や最小分散（MV）法に比べて，提案法は十分な雑音抑圧を実現していることが分かる。以上，ここでは正面から到来する平面波に焦点を合わせ，他方向からの雑音を抑圧する効果が十分であることから，提案原理により高性能な小型システム構築が可能であることを示した。（成果の詳細は学会発表②，⑤，⑦，⑬，⑭，⑱として公表した）。なお，妨害音源数を 1 個に限定すれば，数式に基づく推定も可能で，十分な雑音抑圧が可能であることも示した（詳細は学会発表⑥として公表した）。

以上は，直線上にマイクロホンを配置したアレイによる水平面内の音源分離であるが，マイクロホンを平面的に配置することで空間内の音源分離が可能であることも示した（成果の詳細は学会発表①として公表した）。

(4) 少数マイクロホンアレイから得た時空間音圧分布画像による，距離が異なる音の分離

研究開始当初は，到来方向が異なる音を分離することのみを目指していたが，上記 (1)〜(3) の検討を進める中で，音源までの距離が異なる音の分離が可能であることに気がついた。

4 個の実体マイクロホン（M_1〜M_4）を，目的音源の位置を中心とする円弧上に配置する（図 6）。また y 軸に関して対称な位置に，鏡像として仮想マイクロホン（M_{-1}〜M_{-4}）を考える。仮想マイクロホンは実際には配置せず，その出力は鏡像位置のマイクロホンの出力と同一とする。全てのマイクロホンは，x 軸方向に関しては等間隔 d_x = 2 (cm)で配置する（実体マイクロホンのアレイ長は 6 cm である）。仮に y 軸上で 2 m の位置を焦点と考えると，マイクロホン M_1〜M_4

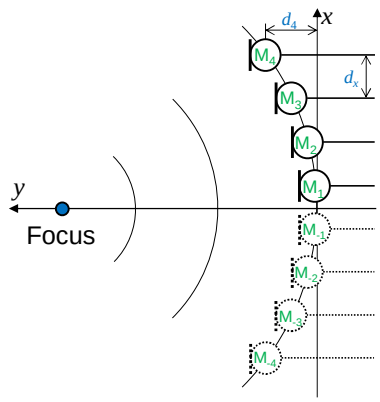

図 6 実体マイクロホン（M_1〜M_4）と仮想マイクロホン（M_{-1}〜M_{-4}）の配置

の y 座標 d_1～d_4 はそれぞれ 0.025, 0.225, 0.625, 1.23 mm となる。

焦点位置から到来する球面波は，全マイクロホンに同時に到来するので，その 2 次元スペクトルは空間軸方向の直流成分に局在する。一方，焦点以外の音源のスペクトルは広がるので，上記の平面波の場合と同様な手順によって，焦点位置にある音のスペクトルを抽出することが可能である。

提案法の性能評価のために計算機実験を行った．目的音源（女性音声“Thank you very much.”）は焦点（2 m）に固定し，白色雑音を放射する雑音源を 0～5 m の範囲で 2 cm ずつ移動させ，雑音減衰量を測定した。測定結果を図 7 に示すとおり，同じ 8 マイクロホンを用いた遅延和法では雑音低減はアレイ近傍にわずかに見られるだけである。それに対して，提案法では約 25 dB の低減が実現されているので，提案法は有効であると考える（成果の詳細は雑誌論文①，学会発表③，④として公表した）。

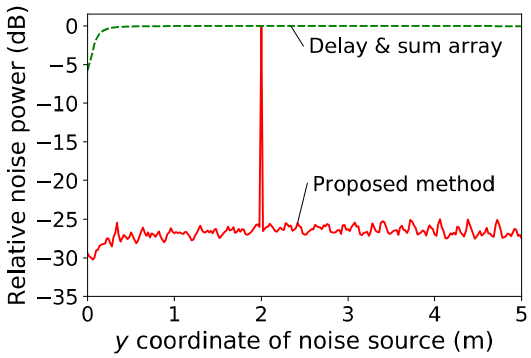

図 7　2 m 位置に焦点を合わせた場合の雑音の低減量

5．主な発表論文等

〔雑誌論文〕（計２件）

① 小澤賢司、焦点位置からの音のみを収録するマイクロホンアレイ、ケミカルエンジニアリング、査読無、Vol. 64、No. 3、2019、pp. 161－165

② A. Iseki、K. Ozawa、and Y. Kinoshita、Neural-network-based microphone-array system trained with temporal-spatial patterns of multiple sinusoidal signals、Acoustical Science and Technology、査読有、Vol. 38、No. 2、2017、pp. 63－70
https://doi.org/10.1250/ast.38.63

〔学会発表〕（計１８件）

① 小澤賢司、森勢将雅、坂本修一、平面マイクロホンアレイと時空間音圧分布画像を用いた音源分離に関する考察、日本音響学会春季研究発表会、2019

② K. Ozawa、M. Morise、S. Sakamoto、Sound source separation by instantaneous estimation-based spectral subtraction、The 5th International Conference on Systems and Informatics (ICSAI2018)、2018

③ K. Ozawa、Y. Koshimizu、M. Morise、S. Sakamoto、Separation of two sound sources in the same direction by image signal processing、The 7th IEEE Global Conference on Consumer Electronics (GCCE2018)、2018

④ 小澤賢司、輿水雄太、森勢将雅、坂本修一、画像信号処理を用いた同一到来方向の 2 音源の分離に関する一考察、日本音響学会秋季研究発表会、2018

⑤ 小澤賢司、森勢将雅、坂本修一、雑音スペクトルの瞬時推定に基づくスペクトラルサブトラクションによる音源分離に関する一考察、日本音響学会秋季研究発表会、2018

⑥ 小澤賢司、森勢将雅、坂本修一、雑音スペクトルの瞬時推定に基づくスペクトラルサブトラクションによる音源分離に関する考察、電子情報通信学会応用音響研究会、2018

⑦ K. Ozawa、M. Ito、G. Shimizu、M. Morise、S. Sakamoto、Proposal of a sound source separation method using image signal processing of a spatio-temporal sound pressure distribution image、2018 AES International Conference on Spatial Reproduction、2018

⑧ 小澤賢司、横打詩音、森勢将雅、眼鏡フレーム上のマイクロホンアレイによる雑音抑制～LASSO アルゴリズムによる優勢な雑音源の方向推定、日本音響学会春季研究発表会、2018

⑨ 小澤賢司、森勢将雅、坂本修一、マイクロホンアレイにより得た時空間音圧分布画像の復元に関する考察、電子情報通信学会エンリッチメントメルチメディア研究会、2018

⑩ 小澤賢司、伊藤将亮、清水源也、森勢将雅、坂本修一、時空間音圧分布画像の復元に関する考察、日本音響学会秋季研究発表会、2017

⑪ K. Ozawa、Y. Akishika、M. Morise、A. Iseki、Y. Kinoshita、Broadbanding of a NN-based microphone-array system by decomposing into frequency components、The 6th IEEE Global Conference on Consumer Electronics (GCCE 2017)、2017

⑫ K. Ozawa、S. Yokouchi、M. Morise、Noise reduction using an eyeglass-frame microphone array based on DOA estimation by LASSO、The 1st International Conference on Challenges in Hearing Assistive Technology (CHAT-2017)、2017

⑬ 小澤賢司、マイクロ間隔マイクロホンアレイの基本特性に関する一考察、日本音響学会春季研究発表会、2017

⑭ 小澤賢司、マイクロ間隔マイクロホンアレイの基本特性に関する考察、電子情報通信学会

エンリッチメントメルチメディア研究会、2017
⑮ M. Ito、K. Ozawa、M. Morise、G. Shimizu、S. Sakamoto、Sound source separation using image signal processing based on sparsity of sound field、2016
⑯ K. Ozawa、Superdirective microphone array based on DOA and waveform estimations of noise、the 5th IEEE Global Conference on Consumer Electronics (GCCE 2016)、2016
⑰ 天野拳志、清水源也、小澤賢司、森勢将雅、大出訓史、雑音の到来方向及び波形推定に基づく超指向性マイクロホンアレイシステムの構築、日本音響学会秋季研究発表会、2016
⑱ 伊藤将亮、小澤賢司、森勢将雅、清水源也、坂本修一、時空間音圧分布画像に対する画像信号処理による音源分離法の提案、日本音響学会秋季研究発表会、2016

〔図書〕（計０件）

〔産業財産権〕
○出願状況（計２件）

① 名称：音源分離システム、音源位置推定システム、音源分離方法および音源分離プログラム
発明者：小澤賢司
権利者：国立大学法人山梨大学
種類：特許
番号：特許願 2018-147470 号
出願年：平成３０年
国内外の別：国内
② 名称：音源分離装置、及び音源分離方法
発明者：小澤賢司，森勢将雅，伊藤将亮，清水源也
権利者：国立大学法人山梨大学
種類：特許
番号：特許願 2016-167815 号
出願年：平成２８年
国内外の別：国内

○取得状況（計０件）

〔その他〕
ホームページ：山梨大学工学部コンピュータ理工学科メディア感性工学研究室
http://www.ccn.yamanashi.ac.jp/~ozawa/lab.htm

６．研究組織

(1)研究分担者
研究分担者氏名：坂本　修一
ローマ字氏名：(SAKAMOTO, Shiuichi)
所属研究機関名：東北大学
部局名：電気通信研究所
職名：准教授
研究者番号（8 桁）：60332524

研究分担者氏名：森勢　将雅
ローマ字氏名：(MORISE, Masanori)
所属研究機関名：山梨大学
部局名：大学院総合研究部
職名：准教授
研究者番号（8 桁）：60510013

(2)研究協力者
なし