Vol.2020-MUS-127 No.15
Vol.2020-SLP-132 No.15
2020/6/6

位相を含めた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雑音抑圧

竹中 幸輝[1]　小澤 賢司[1]

概要：本研究では，非差分型マイクロホンアレイとニューラルネットワークを用いて，所望の方向以外から到来した雑音を抑圧することを目的としている．マイクロホンアレイからの出力を時空間音圧分布画像に変換し，それを 2 次元高速フーリエ変換することによって得られる振幅と位相の 2 次元スペクトルを利用した雑音抑圧手法の提案と性能評価を行った．

Noise suppression by spectral subtraction based on instantaneous estimation of noise spectrum including phase information

Abstract: The purpose of this study is to suppress the noise that comes from the direction other than the target direction by using a non-differential microphone array and neural networks. We propose a noise suppression method using two-dimensional amplitude and phase spectra. These spectra are obtained by converting the output from the microphone array into a spatiotemporal sound pressure distribution image and performing a two-dimensional fast Fourier transform. We also evaluated the performance of the proposed method.

1. はじめに

複数の音源信号が混在している状況において，特定の音源信号を抽出する際には，それ以外の音源信号を抑圧する必要がある．特定の音源信号を抽出することは，雑音環境下での選択的収録や，多チャネル音響システム用の音源信号収録など，様々な場面で有効である．雑音抑圧の研究の一分野として，マイクロホンアレイとニューラルネットワーク（以下 NN と略記）を使用して，特定の方向以外から到来した信号を抑圧する研究が行われてきた [1], [2], [3]．以降では，直線状に配置したマイクロホンと直交する方向（0°）から到来した信号を目的音，その他の方向から到来した信号を雑音とする．

小澤ら [2] は，マイクロホンアレイの出力を時空間音圧分布画像 [4] に変換し，その 2 次元スペクトルと NN を用いて雑音を抑圧する手法を提案している．しかし，提案手法では振幅スペクトルのみに注目し，本来必要とされる位相については推定していなかった．

塩澤ら [3] は，差分型マイクロホンアレイ [1] と NN を用いて，雑音を抑圧し目的音を抽出するシステムを提案した．このシステムでは，差分画像から生成される 2 次元スペクトルの振幅と位相を NN によって推定している．

本研究では，小澤ら [2] の提案したシステムを発展させ，雑音の位相を含めたスペクトラルサブトラクションを行うシステムを提案する．具体的には，塩澤ら [3] が用いた位相推定の手法を，非差分型マイクロホンアレイに適用することで位相の推定を行う．位相の推定まで含めることにより，雑音抑圧性能の向上を図る．

2. 提案システムの概要

2.1 2 次元振幅スペクトル

時空間音圧分布画像 [4] とは，マイクロホンアレイで観測した信号の瞬時音圧の高低を，輝度と対応付けした 2 次元画像である．

8 個のマイクロホンから成るアレイに対して 0° から目的音が，90° から雑音が到来した場合，観測信号の時空間音圧分布画像を図 1 に示す．また，それを 2 次元フーリエ変換（2D FFT）することで，図 2 の 2 次元振幅スペクトルが得られる．本稿では，2 次元スペクトルにおける垂直周波数を空間周波数，水平周波数を時間周波数とし，時間周

[1] 山梨大学
University of Yamanashi

Vol.2020-MUS-127 No.15
Vol.2020-SLP-132 No.15
2020/6/6

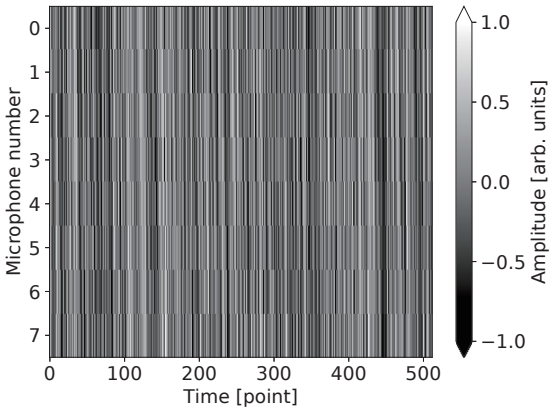


図 1　観測信号の時空間音圧分布画像（0° と 90° から白色雑音）

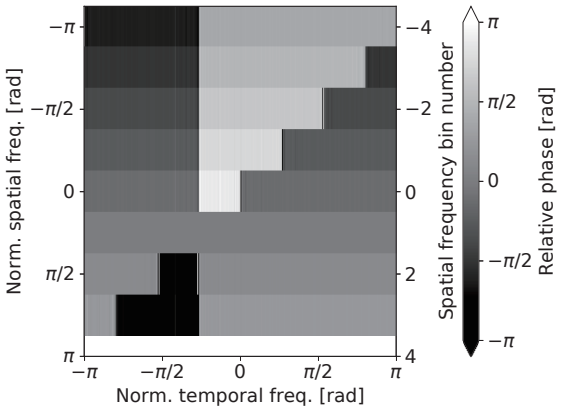


図 3　空間周波数 1 番ビンを基準とした相対位相スペクトル

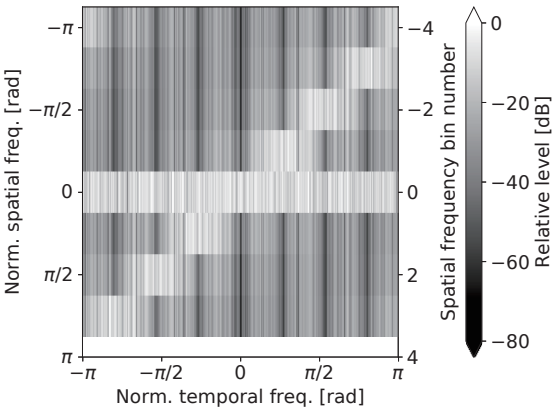


図 2　図 1 の 2 次元振幅スペクトル

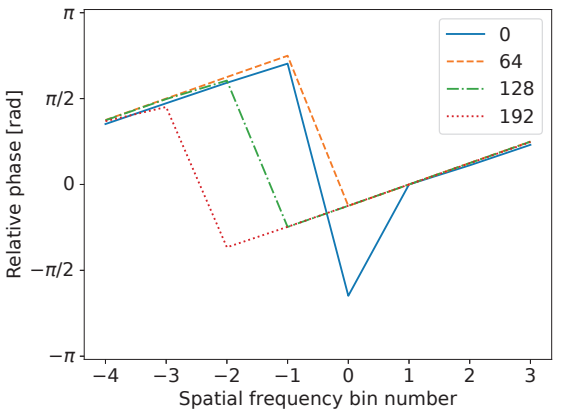


図 4　空間スペクトルの相対位相（パラメタは時間周波数ビンの番号）

波数ビンごとのスペクトルを空間スペクトルと呼ぶ．

目的音は全てのマイクロホンに同相で到来するため，時空間音圧分布画像は縦縞を形成し，そのスペクトルは空間周波数の 0 番ビンに局在する．一方，雑音はマイクロホン間で位相差が生じるため，時空間音圧分布画像は斜め縞を形成し，ある空間周波数ビンに頂点を持つスペクトルを描く．このとき，空間周波数 0 番ビンには，目的音のスペクトルと，雑音のスペクトルの 0 番ビン成分が混在していることになる．

小澤ら [2] は，このスペクトルの特徴と時間周波数に伴う変化の系統性に着目し，NN を利用した音源分離手法を提案した．雑音の空間周波数 0 番ビンの振幅値を NN によって推定し，引き去ることで目的音の抽出を行った．

2.2　2 次元位相スペクトル

小澤ら [2] は振幅スペクトルのみに注目して雑音抑圧を行ったが，本来は雑音の位相も推定する必要がある．雑音の信号のみからなる時空間音圧分布画像を 2 次元フーリエ変換して得られる 2 次元位相スペクトルを図 3 に示す．図 3 では，各空間周波数ビンの位相値に対して，1 番ビンとの位相差を求めることで相対位相を表示している．また，図 3 のいくつかの時間周波数ビンについて，空間スペクトルを取り出したものを図 4 に示す．相対位相を表示したことで，位相は空間周波数ビンに関して系統的な変化を示すので，位相についても NN による推定が可能となる．

2.3　提案システム

本研究では，標本化周波数を 16 kHz とした．このときのナイキスト周波数は 8 kHz であり，それに対応する波長は約 4.25 cm である．空間折返し歪が発生しないようにするため，マイクロホンの間隔は 2 cm とした．また，本システムはスマートフォンに搭載することを想定し，マイクロホン 8 個の 14 cm 長のアレイで構成するものとした．提案システムのブロック図を図 5 に示す．

提案手法では，空間周波数 0 番ビンに含まれる雑音成分の振幅と位相を，独立した NN によって推定し引き去ることで，目的音のスペクトルを抽出する．これを逆 FFT し時間波形に戻すことで，雑音抑圧された目的音を得る．

3.　提案システムによる雑音抑圧

3.1　推定の方針

提案システムに関して，計算機シミュレーションによる雑音抑圧実験を行った．学習データの選定や NN モデルの構築は，先行研究 [2] に基づいて行った．具体的には，以

Vol.2020-MUS-127 No.15
Vol.2020-SLP-132 No.15
2020/6/6

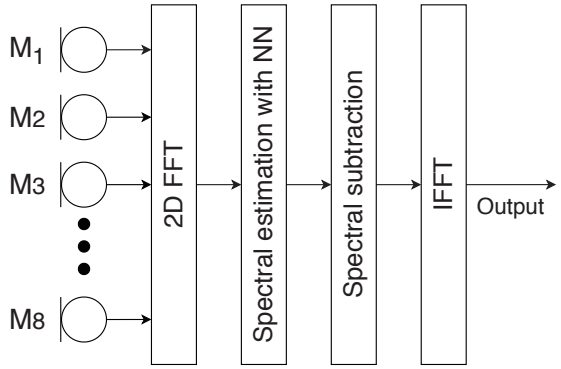


図 5　2 次元スペクトルを用いた雑音抑圧システムのブロック図

下のとおりである．

学習データは，90° と 15° からマイクロホンアレイに到来した白色雑音の時空間音圧分布画像に対して求めた 2 次元スペクトルを用いた．到来した白色雑音の信号をそれぞれ 512 点 (32 ms) のハニング窓で切り出し，2D FFT により 2 次元スペクトルに変換した．その後，空間スペクトルごとに学習データを作成した．

NN モデルの構築には，Python [5] のライブラリである Chainer V4 [6] を用いた．NN の入力層・中間層（振幅推定用，位相推定用共に 3 層）・出力層のユニット数はそれぞれ 7・15・1 とした．また，活性化関数には ReLU（Rectified Linear Unit），最適化アルゴリズムには Adam を使用し，時間周波数ビンの合計である 1024 個の学習データに対してバッチサイズ 2，100000 エポックの学習を行った．

3.2　振幅推定用 NN の学習

振幅スペクトルの最大値に対する相対レベルを計算し，最小値が 0，最大値が 1 となるように正規化を行った．0 番以外のビンの値を入力データ，0 番ビンの値を教師データとし学習を行った．

学習データに関するクローズドテストの結果として，観測値と推定値の散布図を図 6 に示す．観測値が 1 の場合，つまり低域の信号については推定に失敗する場合が多い．低域の空間スペクトルは 0 番ビンに頂点を持つが，それ以外の周波数では 0 番以外のビンに頂点を持つ．したがって，0 番ビン以外の小さな値から 0 番ビンの大きな値を推定しなければならないことが，低域の信号で推定が失敗する原因であると考えられる．全体としては $r = 0.999$ といった高い相関係数を示していることから，高い精度で推定がなされているものと考えている．

3.3　位相推定用 NN の学習

位相スペクトルの各ビンに対して，1 番ビンとの位相差を求めることで正規化を行う．1 番ビンの値は必ず 0 になるため，0，1 番以外のビンの値を入力データ，0 番ビンの値を教師データとし学習を行った．

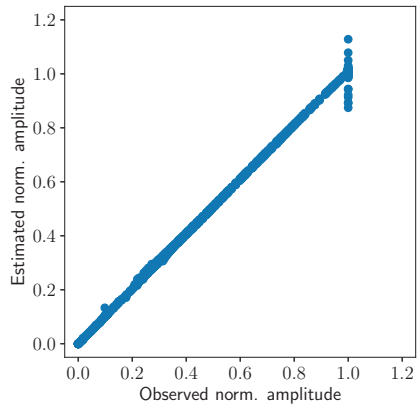


図 6　振幅推定を行う NN についてのクローズドテストの結果（$r = 0.999$）

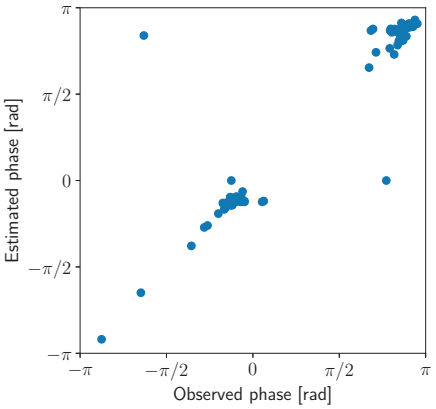


図 7　位相推定を行う NN についてのクローズドテストの結果（$r = 0.993$）

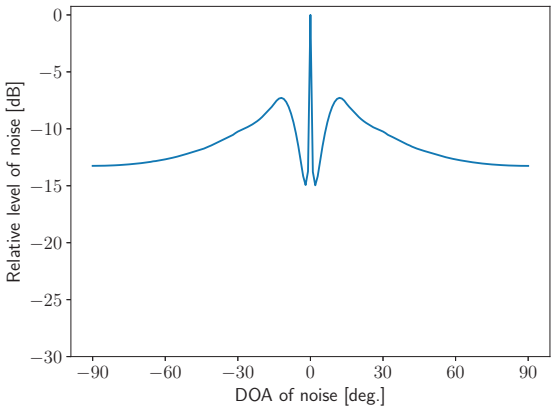
また，位相差と同時に振幅の情報を入力する．空間周波数の −1, 1 番ビンの振幅値を比較し，−1 番ビンの方が大きければ 1，そうでなければ 0 を入力データとして NN に与える．これは，DOA が正の角度であるか，負の角度であるかの情報を与えることと等価である．これら 6 つの位相差と振幅の情報を合わせた 7 つのデータを入力データ，0 番ビンの値を教師データとして学習を行った．

学習データに関するクローズドテストの結果として，観測値と推定値の散布図を図 7 に示す．$r = 0.993$ といった高い相関係数を示していることから，位相推定に関しても良好な推定がなされていると考えている．

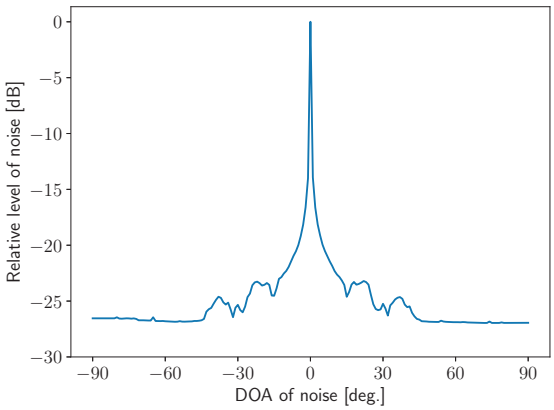
3.4　性能評価

提案システムの雑音抑圧量を調べるため，目的音（0°）と雑音（−90° ～ 90°）の 2 つの白色雑音を入力した．図 8 (a) は振幅推定のみを行った場合，図 8 (b) は振幅推定と位相推定を行った場合の雑音抑圧量を示している．

ここで，NN によって推定した振幅値を観測値から引き去った結果が負になった場合は，過大推定が発生したと見なせる．過大推定が発生した場合には，振幅値を引き去っ



(a) 振幅推定のみの雑音抑圧量



(b) 位相推定も行った場合の雑音抑圧量

図 8 提案手法による白色雑音に対する雑音抑圧量

た結果を 0 とした．

振幅推定のみの場合，雑音抑圧量は最大 −13 dB 程度である．NN による位相の推定は行わず，観測信号の位相スペクトルの値を用いてスペクトラルサブトラクションを行っている．ここで，図 8 (a) より，雑音の到来角が 0° に近い場合，雑音抑圧量が大きいことがわかる．雑音の到来角が 0° に近い場合，雑音と観測信号の位相スペクトルの値も近くなることが原因ではないかと考えられる．

NN によって雑音の位相スペクトルも推定した場合には，雑音抑圧量は最大 −26 dB 程度である．これにより，雑音の位相スペクトルを推定することで，雑音抑圧システムにおける全体の性能が向上することが示されたといえる．

なお，先行研究 [2] では振幅推定のみで最大 −26 dB 程度という抑圧量を示している．これは，先行研究では雑音のみを入力し抑圧量の計算を行っているため，雑音と観測信号の位相スペクトルが一致していることが原因であると推察される．

4. まとめ

本稿では，マイクロホンアレイと NN を使用した雑音抑圧システムについて実装と評価を行った．雑音の振幅と位相のスペクトルを推定しスペクトラルサブトラクションを行うことで，白色雑音に対して最大 −26 dB 程度の抑圧量を示した．

本研究では，雑音源が 1 つの場合を想定して性能評価を行った．今後の課題として，NN の学習条件などを再検討することによる複数雑音源への対応などが挙げられる．また，複素 NN を用いることにより，計算量の削減や性能の向上が期待される．

参考文献

[1] 森田亘, 小畑秀文：非線形多層構造を持つ超指向性マイクロホン・アレイ・システム，日本音響学会誌, Vol. 49, No. 1, pp. 28–33 (1993).
[2] Ozawa, K. Morise, M. and Sakamoto, S: Sound source separation by instantaneous-estimation-based spectral subtraction, Proc. of the 5th International Conference on Systems and Informatics (ICSAI 2018), pp. 870–875 (2018).
[3] 塩澤光一朗, 小澤賢司, 伊勢友彦：差分型マイクロホンアレイと 2 次元スペクトルの機械学習による雑音抑制に関する考察，信学技報, EA2019-31, pp. 19–24 (2019).
[4] Ito, M. Ozawa, K. Morise, M. Shimizu, G. and Sakamoto, S.: Sound source separation using image signal processing based on sparsity of sound field, *J. Acost. Soc. Amer.*, Vol. 140, No. 4, p. 3058 (2016).
[5] Python Software Foundation: Python: Welcome to Python.org(online), 入手先 ⟨https://www.python.org/⟩ (2020–2–26).
[6] Prefered Networks: Chainer: A powerful, Flexible, and Intuitive Framework for Neural Networks(online), 入手先 ⟨https://chainer.org/⟩ (2020–2–26).