Vol.2019-MUS-122 No.20
Vol.2019-EC-51 No.20
2019/2/23

畳み込みニューラルネットワークを用いた脳波による音響信号再構成

及川大樹[†1]　饗庭絵里子[†1]　亀岡弘和[†2]

概要：本研究の最終目的は人間が音を聴取もしくは想起している時の脳波を計測し、その情報をもとに聴取もしくは想起していた音を再構成することである。聴覚分野においても脳情報デコーディングによる刺激の再構成の研究は行われているが、刺激の識別率は低く有効な手段は未だ確立されていない。本研究においては、音刺激を提示中の脳波を計測し、その脳波に含まれる特徴量について様々な側面から解析を行った。また、音想起をより頻繁に行っていると考えられる演奏家と頻度が低いと考えられる非演奏家についても比較を行った。これらの結果をもとに、脳波を用いた音響信号再構成手段について検証を行った。

キーワード：聴覚, EEG, 深層学習,

Acoustic Signal Reconstruction from EEG Using Convolution Neural Network

DAIKI OIKAWA[†1]　ERIKO AIBA[†1, 2]　HIROKAZU KAMEOKA[†3]

1. はじめに

人間の感覚情報は脳情報デコーディングによって再構成することが可能になりつつある. 視覚における研究では, マルチスケール局所画像デコーダーの組み合わせを用いた人間の脳活動からの視覚画像再構成[1], および睡眠中における視覚イメージのニューラルデコーディング[2]にて, 実際に見た視覚画像の再構成や想起された視覚画像の再構成が行われている. 聴覚においても聴取した音刺激の再構成は行われており, 非侵襲的な方法として Skoe ら（2010）による事象関連電位の一種である complex Auditory Brain Response（cABR）を用いた再構成がある[3]. また, 侵襲的な方法としては, 開頭した脳表面に電極を設置し, Electrocorticography (ECoG)と呼ばれる脳活動を記録する方法による再構成も行われている[4, 5]. 一方, 聴覚刺激を想起中の脳活動を用いた刺激の再構成は, 最近になって取り組みが始まったところであり[6, 7], 試行錯誤が行われている最中である.

脳活動から刺激の再構成をするために用いられる信号処理手法として, Skoe et al.（2010）は事象関連電位の取得に一般的な加算平均法を用いているが[3], いくつかの研究では機械学習も用いられている[1, 2, 5]. 前述の通り, Skoe et al.（2010）は, 非侵襲的な方法を用いて脳活動を取得しているが, その信号処理手法は加算平均法を主としている. しかしながら,「1.1 脳波」で述べる通り, 加算平均法は反応の取得に非常に時間がかかるため, 長い刺激は計測に不向きである. また, 将来的に刺激音想起時における脳波から想起刺激の再構成を試みる際, その起源やどのような反応としてあらわれるかが不明であることを考えると, 機械学習によるアプローチが有用である可能性が考えられる.

そこで本研究においては, 音刺激聴取時の人間の脳波から刺激音を再構成することを目指し, 機械学習の手法を用いて検証を行った.

1.1 脳波

本研究では脳情報デコーディングに用いる脳活動の計測方法として, 脳波を採用した. 脳波とは人間の頭部に電極を張り付けた際に生じる電位差を時間波形で記録したものである. 脳は多数の神経細胞から構成されており, それらの細胞が受けた刺激を電気信号へと変換し次から次へと伝達している. したがって, 頭表面に電極を貼付することによって, 脳における神経活動が電位差として記録され, 脳活動を時間波形として観察できる. 音刺激は画像刺激とは異なり, 必ず時間的変化を伴うものであり, 時間情報が必要不可欠である. 従って, 本研究のように音刺激の再構成を試みる場合, 高い時間分解能で反応を取得できる脳波が有用であると考えられる.

また, 脳波からは生命維持に伴う自発的に活動も含まれ

†1 国立大学法人 電気通信大学
The University of Electro-communications.
†2 電気通信大学技能情報学研究ステーション
Center for Art and Performance Science, University of Electro-Communications
†3 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
Nippon Telegraph and Telephone Corporation

Vol.2019-MUS-122 No.20
Vol.2019-EC-51 No.20
2019/2/23

ているが，刺激を受けたことに応じて生じる事象関連電位と呼ばれる活動電位も記録できる．事象関連電位は，刺激が物理的属性であるか心理的属性であるかに関わらず取得可能であることから[8]，取得できる反応の様相が異なったとしても，実際に音刺激を聴取している際と，音刺激を想起した際の両方の活動を取得できると考えられる．本研究では，まず Skoe et al.（2010）の方法にならい，脳幹における事象関連電位である聴性脳幹反応と同様の方法で脳活動の計測を行い，加算平均法を用いた解析も行うことで，再構成に必要となる成分が脳波に含まれているかどうかの確認を行う．一方で，加算平均法を行うためには同じ刺激を数千回にわたって繰り返し聴取し，その数千回分について加算平均を行う必要がある．従って，前述の通り実験時間が非常に長くなり，実験参加者の負担も大きい．単純に聴取中の事象関連電位を取得する場合，刺激間間隔は刺激長の 30%以上を挿入するにとどまるが，聞いた音を直後に想起することを考えると，短時間に次々と想起を行うのは困難であると考えられる.機械学習を活用することによって，計測時間が大幅に短縮され，より効率的な解析が行える可能性が考えられる．

1.2 機械学習

機械学習には様々な手法が存在するが，本研究においては畳み込みニューラルネットワークの一手法である Kameoka et al.（2018）の Convolutional Sequence to Sequence Voice Convertion（CONVS2S-VC）をベースとする手法を用いることとした[9]．畳み込みニューラルネットワークは，画像の認識に特化したニューラルネットワークである．可変な長さをとる時系列データのモデル化には Recurrent Neural Network（RNN）もよく用いられるが，畳み込みニューラルネットワークに比べると学習が難しく過学習が生じやすいとされている[10]．そこで，本研究においては提示した音刺激や取得した脳波の振幅スペクトログラム画像を入力として学習を行い，出力として得られた振幅スペクトログラム画像を Griffin-Lim 法[11]で波形に再構成する手法を採用した．

2. 実験

2.1 脳活動計測

音刺激聴取中および音刺激想起中の脳波について計測を行った．ただし，本研究においては，音刺激想起中のデータは解析対象としない．

2.2 実験参加者

実験参加者は 9 名(年齢=25.9∓4.8，楽器経験年数=13.1∓12.0，女:男=5:4)であった．

2.3 実験設備

音刺激はイヤホン（ER3C, Etymotic Research）を介して提示した．提示レベルは聴覚シミュレータ（Type4128-C ブリュエル・ケアー）を介し，騒音計（Type2270-A-D-00 ブリュエル・ケアー）で計測し，約 75dB SPL であった．生体信号は銀塩化銀電極を用い，両耳朶を基準電極として国際 10-20 法における Cz から 2CH で記録した．Fpz を設置電極とした．信号は生体アンプ（BA1008m，ニホンサンテク）により増幅し，サンプリング周波数 50kHz で A/D 変換を行った．提示刺激は MATLAB（Mathworks）を用いて作成し，Adobe Audition（Adobe Systems Incorporated）を用いて提示レベルの調節，提示タイミングを操作し，オーディオインタフェース（UA-1010，Roland）を通じて記録，提示を行った．音刺激の提示は PC（ThinkPad x240，Lenovo）を用い，脳波の記録には PC（OptiPlex 7010，Dell）を用いた．

2.4 方法

本実験で用いた音刺激は 262Hz の正弦波であった．この刺激が提示されてから次の刺激が提示されるまでを 1 試行とし，流れを図 1 に示した．一人の実験参加者に対し 1 試行が 200 回繰り返される約 5 分の音刺激を 10 回提示した．うち半分はアーチファクト低減の観点から逆位相の音源とした．ただし，音刺激には想起対象となる純音以外に電子楽器から録音した 440Hz のクラリネット音を含めた．これは実験参加者の実験への集中を持続させるための刺激であり，実験参加者にはこの音が聞こえた際にスイッチを押す課題を与えた．

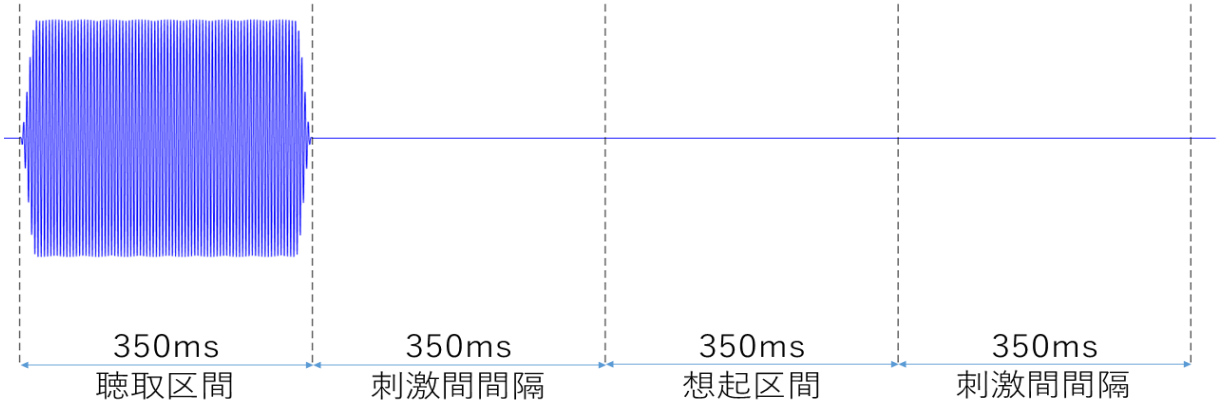

図 1　実験 1 試行分の流れ

Vol.2019-MUS-122 No.20
Vol.2019-EC-51 No.20
2019/2/23

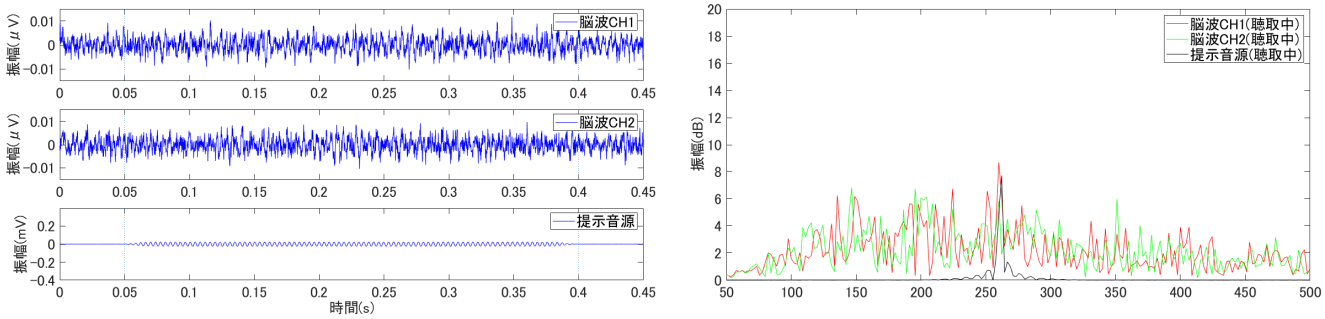

図 2　a. 脳波加算平均後の時間領域信号（左），b. 加算平均後のスペクトル（右）

3.　加算平均法による刺激の再構成

音刺激聴取中の脳波について，加算平均法による提示音刺激の再構成を試みた．

3.1　方法

脳波計測実験によって得られた脳波を提示刺激時間とその前後 50ms でそろえて加算平均を求めた．順位相のデータと逆位相のデータは加算平均の後，足し合わせた．また，それらの時間領域信号に FFT を行い，周波数領域信号を求めた．加算平均は実験参加者ごとに行った．

3.2　結果

結果の一例を図 2,3 に示す．図 2 は加算平均された 2 チャンネル分の脳波と提示した音刺激の波形である．各横軸と縦軸は，それぞれ時間と振幅を示している．刺激のオンセット，オフセットは破線で示した．先行研究においては[3]，呈示刺激と同様に目視でも周期的な反応が見られていたが，本データからは明確な周期性は観察できなかった．図 3 は図 2 のスペクトルである．波形からは明確な周期性を観察することができなかったが，スペクトルには提示した刺激音（黒実線）と同様の周波数成分が脳波（赤・緑実線）に含まれていることが明らかになった．一方で，個人差が大きく，明確な反応が見られない実験参加者もいた．

3.3　考察

先行研究においても，反応には音楽経験などによって個人差が生じるが報告されており[12, 13]，今後，実験参加者らの演奏経歴などを照会して，さらに検討の余地があると考えられる．また，本実験においては実験参加者の負担を考慮し，加算回数は 2000 回であった．一方で，先行研究[4]においては 4000～6000 回の加算平均を行っており，それに比べると半分以下の加算回数にとどまっている．従って，このことも明確な音信号波形の再構成に至らなかった原因の一つとして考えられる．しかし，音源と同じ周波数成分は含まれていたことから，畳み込みニューラルネットワークを用いた再構成を試みた．

4.　畳み込みニューラルネットワークによる刺激の再構成

今回用いた音刺激は 262Hz であり，学習の効率化の観点から刺激と関連の低い高周波成分を取り除くため，記録した提示音刺激と脳波は 2kHz にリサンプリングした．続いて畳み込みニューラルネットワークで学習させるため，短時間 Fourier 変換により振幅スペクトログラムに変換することで画像化した．

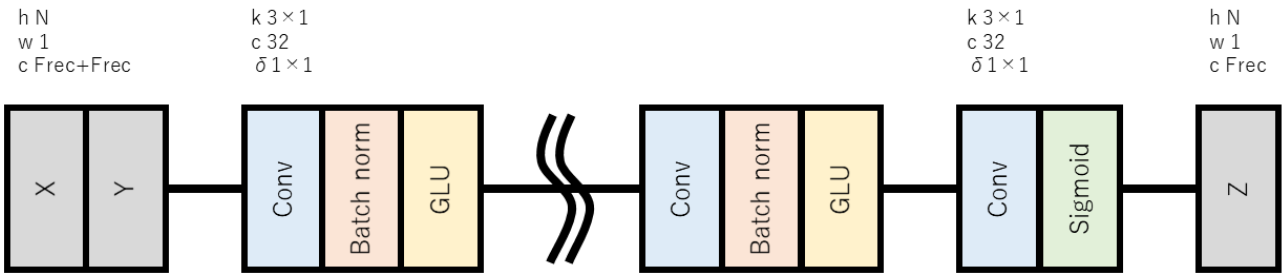

図 3　ネットワーク構造図

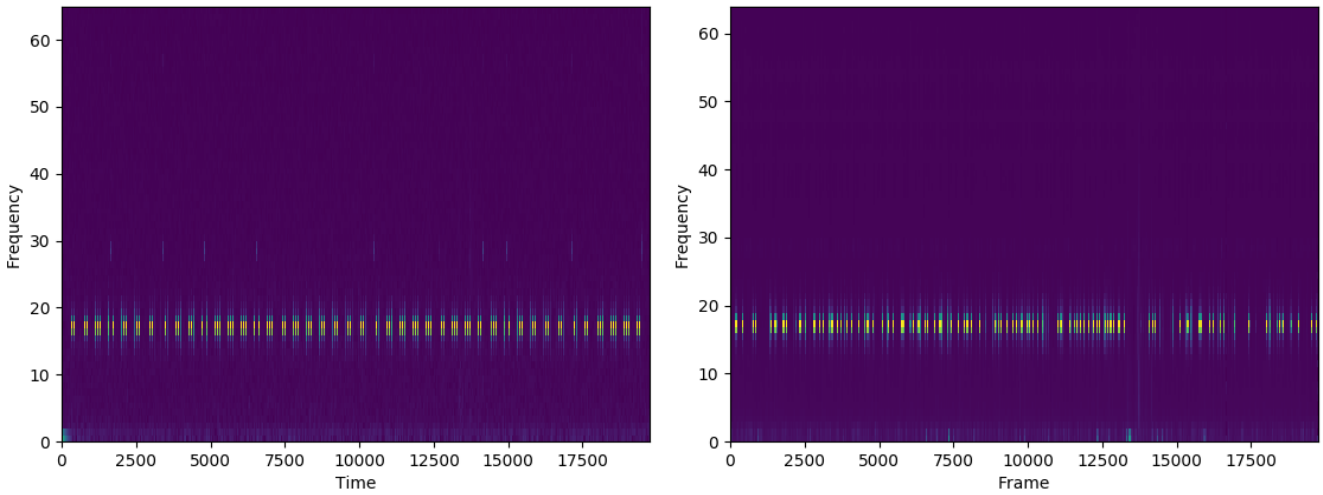

図 4　a. 提示音刺激の振幅スペクトログラム（左），b. 再構成音の振幅スペクトログラム（右）

4.1 方法

学習は実験参加者ごとにそれぞれ 10 回提示した実験のうち，9 回分を学習用データとして行った．残り 1 回分はテストデータとした．図 3 に今回用いた畳み込みニューラルネットワークの構造を示す．ネットワークの出力も入力と同形式で得られ，これを Griffin-Lim らによる方法[5]により音響信号に変換した．ここで，学習の入出力は画像として解釈され，「h」，「w」，「c」はそれぞれ高さ，幅，チャンネル数を表す．「N」，「Frec」はそれぞれ振幅スペクトログラムのサンプル点数，周波数である．「Conv」，「Bach norm」，「GLU」，「Sigmoid」はそれぞれ畳み込み，バッチ正規化，geted linear unit，シグモイド層である．「k」，「c」，「δ」はそれぞれ畳み込み層のカーネルサイズ，出力チャネル数，および膨張率を表す．

4.2 結果

提示音刺激と再構成した音双方の振幅スペクトログラムの一例を図 5,6 に示す．画像の解像度は縦:横＝65:19751)であった．学習の結果，ターゲットとして用意した音源の周波数 262Hz が再構成された．また発音タイミングは提示音刺激では 1050 ms おきに 350 ms の長さで発音されるが，再構成された音では音の長さ，間隔ともに不規則なものとなっていた．したがって，加算平均よりも周波数的な再現性は高かったが，時間的な再現性は改善の余地があると考えられる．

4.3 考察

周波数に関する再現性が高かった理由としては，使用した音刺激の周波数が一種類であったことが挙げられる．追加実験として，別の周波数の刺激による実験も実施していることから，今後，両データを用いた学習を行う予定である．発音タイミングに関しては，学習に最適なパラメータを用いられているかが未検証であることから，これについても今後検証を行っていく．

5. 結論

本研究においては，音刺激聴取時の人間の脳波から刺激音を再構成することを目指し，これまでの研究で一般的に行われてきた加算平均法および機械学習の手法を用いて検証を行った．加算平均法を用いて再構成された波形は，データ量不足が要因となって，明確な周期性が目視できるほどの結果ではなかった．しかし，畳み込みニューラルネットワークによる再構成では周波数について再現度が高い音が再構成された．従って，畳み込みニューラルネットワークは，加算平均法に比べて少ないデータでも特定の音響特徴量抽出が行える可能性が示唆された．

今後，音刺激の時間的側面の再構成精度を上げるために最適なパラメータを検証し，また様々な周波数を用いた学習を行う予定である．

謝辞

本研究は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との共同研究である．

参考文献

[1] Youichi Miyawaki, Hajime Uchida, Okito Yamashita, Masa-aki Sato, Yusuke Morito, Hiroki C. Tanabe, Norihiro Sadato and Yukiyasu Kamitani. Visual Image Reconstruction from Human Brain Activity using a Combination of Multiscale Local Image Decoders, Neuron, 2008 Nov, 60, p. 915-929.

[2] T. Horikawa, M. Tamaki, Y. Miyawaki,Y. Kamitani. Neural Decoding of Visual Imagery During Sleep, Science, 2013 April, 340(6132), p. 639-642.

[3]Skoe, E. & Kraus, N. Auditory brain stem response to complex sounds: a tutorial. Ear Hear. 31, 302–324 (2010).

[4]Pasley, B. N. et al. Reconstructing speech from human auditory cortex. PLoS Biol. 10, e1001251 (2012).

[5]Wang, R., Wang, Y. & Flinker, A. Reconstructing Speech Stimuli From Human Auditory Cortex Activity Using a WaveNet Approach. *arXiv [cs.SD]* (2018).

[6]宇澤志保美, 滝口哲也, 有木康雄 & 中川誠司. 脳磁界データによる想起音声の識別 -次元数削減による精度向上の検討-. in 日本音響学会講演論文集 337–340
[7]宇澤志保美, 滝口哲也, 有木康雄, 添田喜治 & 中川誠司. 音想起に伴う脳磁界反応：等しいエンベロープをもつ音声と純音の比較. in 日本音響学会講演論文集 1291–1294
[8]入戸野宏. 心理学のための事象関連電位ガイドブック. (北大路書房, 2005).
[9]Hiroazu Kameoka, KouTanaka, Takuhiro Kaneko,Nobukatu Hojo. CONVS2S-VC:Fully Convolutional Sequence to Sequence Voice Convertion. arXiv1811.01609v2, 2018 Nov.
[10]阪上大地. 音楽情報処理のための深層学習. in 情報処理学会研究報告 **2018-MUS-119**, 1–6 (情報処理学会音楽情報科学研究会).
[11] D. W. Griffin and J. S. Lim. Signal estimation from modified short-time Fourier transform, IEEE, Trans. ASSP, vol. 32, no. 2, pp. 236{243, 1984.
[12]Musacchia, G., Strait, D. & Kraus, N. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. Hear. Res. 241, 34–42 (2008).
[13]Aiba, E. *et al.* Accuracy of Synchrony Judgment and its Relation to the Auditory Brainstem Response: the Difference Between Pianists and Non-Pianists. *JACIII* **15**, 962–971 (2011).