

音楽聴取時の脳波の周波数解析に基づくジャンル分類

佐藤 優稀* 齊藤 大智* Chen-Chieh Liao* 宮藤 詩緒* 吉村 奈津江* 小池 英樹*

概要. 本稿は、脳波に基づいた音楽のジャンル分類をブレイン・コンピュータ・インタフェース (BCI) へ応用することを目標とした研究である。音楽想起時の脳波の分類問題を解く前段階として、音楽聴取時の脳波の分類問題に取り組んでいる。被験者実験として 10 人に異なる 3 ジャンルの複数の音楽を聞かせている間の脳波の測定を行い、その脳波データに周波数解析的な処理を施した上で深層学習に入力した結果、チャンスレート (33.33 %) を平均 5.56 % 上回る分類精度を達成した。また、脳波の振幅値を入力した場合と比較して、平均 4.82 % 優れた分類精度を達成し、パワースペクトル密度値入力の方が脳波振幅値入力よりも優れている可能性を示した。今後は音楽想起時の脳波の分類に取り組み、将来的には VR 空間におけるアバター制御や、現実世界におけるロボットアームや車椅子等の操作などの HCI 分野への応用を目指していく。

1 はじめに

未来のユーザインタフェースとしてブレインコンピュータインタフェース (BCI) が注目されている。BCI として人間の脳活動を利用する上では、反応の顕著性という利点を持つ運動想起が主流となっている。例として Nishimoto らは、手指の開閉運動想起時に脳の運動野が活性化することを応用した BCI を開発し、リハビリテーションなどに活用している [8]。また、視聴覚情報の刺激受容時や想起時の脳活動も BCI に応用可能かもしれない。視覚に関する研究例として錯視画像を見たときの脳活動から画像を再構成する研究が挙げられる [3]。聴覚に関する研究例としては、単純音声の聴取時や想起時の脳波の方向定位に基づいた分類 [5] や、音声想起時の脳波からの母音「あ」「い」の再構成 [1]、fMRI 画像からの自然音声の再構成 [9] が挙げられる。これらの関連研究は、BCI をユーザが日常生活で娯楽として利用する視点から見た場合、想起対象が退屈でユーザが苦痛を感じてしまう可能性がある。

そこで我々は音楽を用いたアプローチを提案する。鼻歌のように音楽を想起する行為は日常生活で行われる一般的なものであり、運動想起や画像想起、単純音・母音・自然音の想起と比較して退屈なものではない。そのため、ユーザへの負担を軽減できるのではないかと考える。さらに我々は、音楽分類問題の中でも特にジャンル分類問題に着目している。ジャンルに着目することで無数に存在する音楽を、音楽間に共通する特徴に基づいてより少ない個数にまとめ上げることができる。そのため、分類クラスが少なくなり、分類問題として簡単にすることが可能だと考えている。

本研究では、BCI 制御に応用するための音楽想起時の脳波分類の前段階として、音楽聴取時の脳波のジャンル分類問題に取り組んでいる。本稿ではその途中経過として音楽聴取時の脳波を深層学習に入力する際の最適な形式を調べた結果を報告する。脳波データを深層学習に入力する際の主な形式としては、脳波の振幅値を入力する方法と、高速フーリエ変換 (FFT) でパワースペクトル密度値に変換してから入力する方法の 2 通りが考えられる。脳波振幅値をパワースペクトル密度値に変換することでデータの次元数が削減され、情報が凝縮されることで脳活動の反応がより顕著に表現され、分類精度向上に寄与するのではないかと考えた。

2 本研究のアプローチ

2.1 システム概要

BCI として人間の脳活動を利用するにあたり我々は脳波に着目している。脳波は非侵襲的な計測機器によって非専門家でも手軽に計測できる生理信号である。脳波には高い時間分解能と手軽さという利点と、低い空間分解能とノイズへの脆弱性という欠点がある。分類精度の向上だけが目標であれば、fMRI や侵襲的な脳活動計測手法を検討したほうが良いかもしれない。しかし我々は HCI への BCI の応用を最終目標としている以上、専門的な技術や大掛かりな機材が必要となるような計測手法は望ましくない。よって脳波に着目することにした。

脳波計は Smarting Pro (mBrainTrain, セルビア) を使用している。電極数は 32ch、サンプリングレートは 500Hz である。他の脳波計に比べ電極数が少ないため得られる情報は限られてしまうが、実験やシステム使用のための装着時間が短縮され、被験者の負担軽減に寄与している。また、被験者への音楽刺激提示は PC に接続した有線イヤホンによって行われる。

Copyright is held by the author(s). This paper is non-refereed and non-archival. Hence it may later appear in any journals, conferences, symposia, etc.

* 東京工業大学

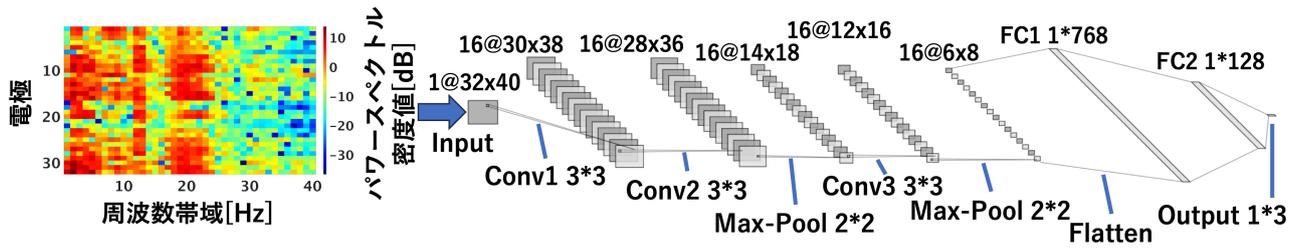


図 1. 実装した CNN

2.2 脳波データ処理

計測した脳波は、リアルタイムで瞬き等の筋電ノイズを削除する ASR[2] が適用されている。この脳波に 1~40Hz のバンドパスフィルタを適用して 50Hz の電源ノイズを除去した。その後音楽ごとに分割し、さらに 1 秒ずつに分割した脳波振幅値データとして 32×500 の csv を出力した。パワースペクトル密度値については、1 秒ごとに分割した脳波データに対して EEGLAB¹ の spectopo 関数を適用して算出し、 32×40 の csv を出力した。

2.3 深層学習モデル

脳波の分類については深層学習を使用した。データセットの分割割合は train(80%), validation(10%), test(10%) とした。脳波振幅値入力については EEGNet[7] をモデルとして使用した。パワースペクトル密度値入力については、図 1 に示す CNN を使用した。この CNN は Sonawane らの論文 [10] を参考に実装した。

3 実験

3.1 実験説明

被験者 10 人 (20 代から 60 代の男性 8 人, 女性 2 人) を対象に、音楽刺激提示時の脳波を計測する実験を行った。脳波信号へのノイズ混入防止のために、セッション中は常に目を閉じ、姿勢変更などは試行間の休憩時間に行ってもらった。

本実験の目的は、音楽聴取時の脳波から 3 ジャンルの分類が可能かを検証することである。分類対象の 3 ジャンルとしてロック、クラシック、ジャズを選択し、1 ジャンルにつき 10 曲ずつ選定し、30 秒ずつに切り取って使用した。1 セッション内で全曲がランダム再生され、合計で 4 セッション実施した。

3.2 実験結果

実験終了後、計測した脳波に対して 2.2 節のデータ処理を施し、深層学習に入力した。表 1 は、分類精度をまとめたものである。パワースペクトル密度値入力の方はチャンスレート (33.33%) を平均 5.56%

上回り、脳波振幅値入力時の分類精度と比較して平均 4.82% 高かった。脳波振幅値からパワースペクトル密度値への変換によって、音楽ジャンルごとの脳活動の違いがコンパクトに表現されていたことが分類精度の差に表れたのではないかと考えられる。

表 1. 入力形式ごとの分類精度 (%)。

被験者	脳波振幅値	パワースペクトル密度値
Sub1	36.33	36.00
Sub2	34.83	43.83
Sub3	33.75	42.33
Sub4	31.83	43.83
Sub5	34.25	36.75
Sub6	30.75	37.17
Sub7	33.67	43.33
Sub8	33.25	38.33
Sub9	38.67	31.67
Sub10	34.42	36.67

4 議論とまとめ

本稿では、深層学習に対する脳波の入力形式の比較を行い、脳波振幅値入力よりもパワースペクトル密度値入力の方が高い分類精度が得られることを示した。しかしながらこの結果から脳波振幅値入力が劣っているとは断言できない。例えば、MAEEG[4] や BENDR[6] といった先行研究では、脳波振幅値入力によって精度向上を試みている。そういった手法による分類精度向上効果やリアルタイム性なども含めて今後さらなる議論をしていく必要がある。

脳波計測時に使用する音楽のジャンルや選曲についても議論の余地がある。今回実験に使用した音楽はほとんどの被験者にとっては未知の曲であったため、各被験者に特化したジャンル選定と選曲で同様の実験を行った場合にどのような結果が出るのかは興味深い。

今後は脳波の前処理、深層学習のモデル、実験構成の再検討などを通して分類精度のさらなる向上を図るとともに、音楽想起時の脳波の分類問題にも取り組んでいきたい。

¹ <https://scn.ucsd.edu/eeglab/index.php>

謝辞

本研究は JST ムーンショット型研究開発事業 JP-MJMS2012 の支援を受けている。

参考文献

- [1] W. Akashi, H. Kambara, Y. Ogata, Y. Koike, L. Minati, and N. Yoshimura. Vowel Sound Synthesis from Electroencephalography during Listening and Recalling. *Advanced Intelligent Systems*, 3(2):2000164, 2021.
- [2] C.-Y. Chang, S.-H. Hsu, L. Pion-Tonachini, and T.-P. Jung. Evaluation of Artifact Subspace Reconstruction for Automatic Artifact Components Removal in Multi-Channel EEG Recordings. *IEEE Transactions on Biomedical Engineering*, 67(4):1114–1121, 2020.
- [3] F. Cheng, T. Horikawa, K. Majima, M. Tanaka, M. Abdelhack, S. C. Aoki, J. Hirano, and Y. Kamitani. Reconstructing visual illusory experiences from human brain activity. *bioRxiv*, 2023.
- [4] H.-Y. S. Chien, H. Goh, C. M. Sandino, and J. Y. Cheng. MAEEG: Masked Auto-encoder for EEG Representation Learning. In *NeurIPS Workshop*, 2022.
- [5] Y. Koike, Y. Hiroi, Y. Itoh, and J. Rekimoto. Brain-Computer Interface Using Directional Auditory Perception. In *Proceedings of the Augmented Humans International Conference 2023*, AHs '23, p. 342–345, New York, NY, USA, 2023. Association for Computing Machinery.
- [6] D. Kostas, S. Aroca-Ouellette, and F. Rudzicz. BENDR: Using Transformers and a Contrastive Self-Supervised Learning Task to Learn From Massive Amounts of EEG Data. *Frontiers in Human Neuroscience*, 15, 2021.
- [7] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 15(5):056013, jul 2018.
- [8] A. Nishimoto, M. Kawakami, T. Fujiwara, M. Hiramoto, K. Honaga, K. Abe, K. Mizuno, J. Ushiba, and M. Liu. Feasibility of task-specific brain-machine interface training for upper-extremity paralysis in patients with chronic hemiparetic stroke. *Journal of Rehabilitation Medicine*, 50(1):52–58, Sep. 2017.
- [9] J.-Y. Park, M. Tsukamoto, M. Tanaka, and Y. Kamitani. Sound reconstruction from human brain activity via a generative model with brain-like auditory features, 2023.
- [10] D. Sonawane, K. P. Miyapuram, B. Rs, and D. J. Lomas. GuessTheMusic: Song Identification from Electroencephalography Response. In *Proceedings of the 3rd ACM India Joint International Conference on Data Science & Management of Data (8th ACM IKDD CODS & 26th COMAD)*, CODS-COMAD '21, p. 154–162, New York, NY, USA, 2021. Association for Computing Machinery.

未来ビジョン

我々は、BCIとVRを融合し、いずれはコントローラー等を使うことなく、仮想空間上の自分のアバターの制御などを、頭で考えただけでできるようになることを目指して研究に取り組んでいる。本稿で記述してきた内容は、音楽聴取時の脳波を計測して、そのジャンルを計測が終わったあとで分類するというものであったが、この方法では分類精度をいくら高めたところで、自分が能動的に発生させた思考ではない以上、何らかのアプリケーション操作に応用することは難しい。では音楽聴取時の脳波のジャンル分類をすることは無意味であるのかと言われれば、そうではないと考えている。アプリケーション操作をする上では能動的に発生させた思考による脳波を分類できるようになる必要があり、音楽においては楽曲のメロディや歌詞を頭の中で想起することが該当する。同じ音楽を聞くことと想起するこ

との間には何らかの関連性があると考えられるのは自然であり、それは脳活動にも現れているのではないかと仮説を立てた。そして、その関連性は、深層学習のモデルの学習という形で応用できないかと期待している。アプリケーション使用の度に音楽想起時の脳波データを大量に収集してモデルの学習を行わないといけないようなシステムでは、アプリケーション使用開始前にユーザが疲労してしまう。しかし、音楽を聞くという行為であれば、もしかしたら日常的に脳波を計測してデータを収集しておくことができるような手軽さが実現できるかもしれない。そしてその大量のデータでモデルを事前学習させておき、アプリケーション使用開始時に数分程度の音楽想起の脳波データでファインチューニングできるようになれば、ユーザの負担を軽減し、より多くの人が使いたいと思えるようなシステムにできるのではないかと考えている。