Discrimination by Deep Learning of 1Hz Difference in Auditory Cortex Using fMRI Activation Patterns

Yoshitaka Ooyashiki
Kochi University of Technology
Tosayamada, Kami, Kochi, 782-8502, Japan
e-mail: ooyashikiyoshitaka@gmail.com

Kyoko Shibata
Kochi University of Technology
Tosayamada, Kami, Kochi, 782-8502, Japan
e-mail: shibata.kyoko@kochi-tech.ac.jp

***Abstract*- Brain decoding is a technology that interprets physical and psychological states from brain activity, and it is expected to serve as a means of medical support and communication for people with disabilities. Recently, brain decoding has gained considerable attention, especially with the advent of deep learning techniques. This study builds on the concept of tonotopy in the auditory cortex and aims to develop a method to discriminate between two sounds with a 1 Hz difference, which is difficult for humans to distinguish, using brain activation images. In a previous work, the focus was on brain activation imaging acquisition methods, and research was conducted using the two main imaging designs in functional Magnetic Resonance Imaging (fMRI) experiments: event-related design and block design. The findings indicated that both designs were effective, and further improvements in accuracy are anticipated. Therefore, this report aims to further improve discrimination accuracy. To improve accuracy, this report focused on Region of Interest (ROI) expansion, hypothesizing that an increase in activation information contributes to improved accuracy of deep learning models. In this report involved the execution of experiments in which Brodmann Areas (BA) 22 was introduced as an additional ROIs, in conjunction with the existing ROIs, BA41 and BA42. The results demonstrated that expanding the ROI improved accuracy across both designs. Notably, the block design yielded an over 30% improvement, reaching 100% discrimination accuracy. The results demonstrated that ROI expansion is an effective method for enhancing accuracy.**

Keywords- fMRI; CNN; Brain decoding; Tonotopy; Region of interest.

I. INTRODUCTION

Brain decoding is a technology that decodes physical and psychological states from brain activity, and it is a subject of extensive research in the field of neuroscience. Recently, brain decoding has gained considerable attention, especially with the advent of deep learning techniques. A substantial body of research has been dedicated to investigating the visual cortex using fMRI. For instance, there are studies such as decoding emotional expressions from visual cortex images using Support Vector Machines (SVM) and Convolutional Neural Networks (CNN) [1], and decoding objects seen in dreams from visual cortex images using Deep Neural Networks (DNN) [2]. Conversely, research in fMRI-based auditory cortex decoding has lagged that of the visual cortex. This is primarily because approximately 80% of human perceptual information is visual, prioritizing research on the visual cortex. Moreover, operational noise during fMRI scanning interferes with detecting neural activity in the auditory cortex. However, advances in fMRI technology, including the implementation of noise-canceling mechanisms, have facilitated research on the auditory cortex.

Recent auditory decoding studies have reported decoding everyday sounds from brain activity [3][4], as these are directly related to people's daily lives. Studies on decoding brain activity for music have also progressed, with reports identifying genres and moods (e.g., cheerful, somber, uplifting) from neural responses [5][6]. However, most of these studies have focused on qualitative musical characteristics such as mood, whereas studies targeting quantitative musical characteristics (e.g., frequency or sound pressure) remain scarce.

Considering early disease detection and the identification of cognitive decline, quantitative analysis is required rather than qualitative examination. Therefore, our study group is focusing on the decoding of quantitative musical characteristics. In a previous work [7], an accuracy of 75% was achieved in discriminating two sounds with a 124.5 Hz difference using deep learning applied to brain activation images. However, investigating even finer frequency differences is necessary to advance the goals of early disease detection and the identification of cognitive decline. Therefore, this study addresses the recognition of finer frequency differences to enhance frequency resolution. To investigate brain responses to auditory stimuli, the auditory cortex was defined as the ROI, following the concept of tonotopy. Tonotopy refers to the spatial organization of frequency-specific responses within the auditory cortex. This phenomenon has been confirmed in many previous studies [8][9][10][11], particularly in BA 41 and 42 (primary auditory cortex). These studies typically employed a frequency range of 1–40 kHz and focused primarily on continuous frequency modulation. However, the investigation of the potential efficacy of tonotopy functions in instances of minor frequency discrepancies remains an uncharted territory. In this study, it was motivated by the hypothesis that "even frequency differences that cannot be perceived by humans could still lead to distinct patterns of activation in the auditory cortex". In general, healthy individuals can discriminate frequency differences of about 5 Hz, while 2–3 Hz differences are influenced by musical training and individual variability. Accordingly, this study defines "imperceptible frequency differences" as a 1 Hz gap between two pure sounds. The primary aim of this study is to develop a method capable of discriminating between two pure sounds differing by 1 Hz using fMRI data.

In this study, deep learning is employed to address frequency differences that are imperceptible to the human auditory system. Specifically, brain activation images are acquired while presenting two pure sounds differing by 1 Hz, and a deep learning model is used to discriminate which sound was presented based on the brain activity images. The primary challenges of this method pertain to the selection of a deep learning model and the method of acquiring brain activation images. To address the first challenge, a 3D Convolutional Neural Network (3DCNN) based model was adopted, considering the inherently three-dimensional structure of the brain. To address the second challenge, the focus will be on the two primary imaging designs employed in fMRI experiments: event-related design and block design. The event-related design is characterized by its limited capacity for image clarity due to the constrained temporal parameters allocated for imaging procedures. However, this design facilitates the acquisition of a substantial volume of data. Conversely, block design demonstrates superiority in terms of image clarity, a consequence of its prolonged imaging duration. However, this approach permits a more limited acquisition of data.

In this study, we have verified the imaging designs (event-related design and block design) that are effective for discrimination experiments using both designs. In previous works [12][13], an attempt was made to discriminate two sounds with a 1 Hz difference using deep learning, with the ROI set to BA41 and BA42 based on previous studies [8][9][10][11]. The findings indicated a discrimination accuracy of 55.90% for the event-related design and 63.41% for the block design, suggesting that both experimental designs are effective and hold promise for further enhancement of accuracy.

However, since there are no comparable previous studies for this study, we cannot rule out the possibility that accuracy is low when viewed as an absolute value of 63%. Therefore, this report aims to further improve accuracy compared to previous works [12][13]. Data augmentation has been recognized as an effective approach to enhancing accuracy. While a variety of data augmentation methods exist, the report focuses on ROI augmentation. The rationale for this focus is that the increased activation information obtained through ROI augmentation may enhance the performance of deep learning models. The proposed region for augmentation is BA22 (higher-order auditory cortex). As previously mentioned, tonotopy has been primarily confirmed in BA41 and 42; however, its presence has also been suggested in BA22 [14]. However, given the paucity of studies on tonotopy in BA22, the efficacy of BA22 in studies targeting frequency differences, as evidenced in this report, remains uncertain. Consequently, in this report, BA22 is additionally designated as a region of interest, assuming that tonotopy is also active in this area alongside BA41 and 42. In addition, given the utilization of two designs in previous works [12][13], this report employs two designs as well. This report is an individual analysis.

The structure of this report is as follows. Section II delineates the methodologies employed in brain activation imaging and frequency discrimination techniques. Section III presents the discrimination results obtained using the constructed deep learning model. Section IV investigates the factors contributing to improved discrimination accuracy and describes the discrimination techniques found to be effective based on the study's findings. Section V provides a summary.

II. METHOD

The procedure is outlined as follows. Brain activation images are obtained using an fMRI scanner while presenting two auditory stimuli differing by 1 Hz. These images are annotated using Statistical Parametric Mapping (SPM) 12 for input into deep learning. The annotated 3D data is then employed to train the model and perform discrimination using training data with a 3DCNN. A detailed explanation is provided in the subsequent section.

A. *fMRI experiment*

The fMRI experiment was conducted to obtain brain activation images for use in discrimination experiments. The fMRI apparatus utilized is the MAGNETOM Prisma 3T, manufactured by SIEMENS. The auditory stimuli consisted of pure sounds at 523 Hz and 524 Hz, with sound pressure levels ranging from 78 to 83 dB. Auditory stimuli were generated using Steinberg Nuendo 10.3 and delivered to participants via Opto ACTIVE thin headphones employing Active Noise Control to attenuate fMRI scanner noise. In this experiment, one 20-years-old healthy male subjects participated, who do not have abnormality in the simple hearing test. The imaging design will be a block design (Task 9 s, Rest 15 s) and an event-related design (Task 3 s, Rest 3 to 21 s in multiples of 3). This report uses brain activation images obtained in a previous work [12].

B. *Annotation*

The subsequent section will address the implementation of data analysis for deep learning. The conversion of the DICOM format to the NIfTI-1 format is necessary for the subsequent analysis using brain image analysis software SPM12. Subsequently, the preprocessing and individual analysis should be performed. Preprocessing included several steps: realignment to correct head motion, slice timing correction to adjust temporal differences across slices, coregistration with structural images, spatial normalization, and spatial smoothing. The objective of individual analysis is to extract brain activation characteristics through the random selection of multiple images, the implementation of statistical analysis, and the creation of contrast. In this report, regarding the training data, we created a single statistical image from two scans of brain activation images that had undergone preprocessing and obtained the following training data (per frequency) for each design by changing the combination of the two scans. There were 192 training data points for the event-related design and 80 training data points for the block design. The test data were obtained as a single statistical image from one scan, resulting in 24 test data points (per

 ISBN: 978-1-68558-312-5

TABLE I. NUMBER OF TRAINING DATA AND TEST DATA. LINE 1 IS EVENT-RELATED DESIGN. LINE 2 IS BLOCK DESIGN.

Designs	Training data	Test data
Event-related	192	24
Block	80	24

TABLE II. HYPER PARAMETER IN LEARNING.

Kernel size (Ks)	3, 4, 5, 6, 7
Filters (F)	16, 32
Batch size (Bs)	8, 16, 32

frequency) for each design. These are shown in Table 1. The ROIs are defined as BA41, 42, and 22, and for each contrast, the corresponding t-values and spatial coordinates are extracted. For each contrast, normalized values ranging from 0.0 to 1.0 are exported in CSV format. Using the spatial coordinates, the data are transformed into 3D arrays with dimensions H41×W50×D15 for input into the deep learning model. Voxels outside the ROI are assigned to a value of 0.0.

C. *Frequency discrimination method*

In this report, we utilize 3DCNN, a variant of deep learning, to discriminate auditory stimuli. 3DCNN represents a model that extends the capabilities of CNN, which are designed for image recognition, to three dimensions. 3DCNN utilizes convolution and pooling operations in three dimensions to extract features, thereby expanding the scope of image recognition in the third dimension. The architecture of the 3DCNN consists of sequential convolution and pooling layers, followed by a fully connected layer positioned directly before the output layer. To perform binary discrimination, the model employs two output neurons, with the softmax activation function applied to convert the outputs into probabilistic scores. The hyperparameters used for training are listed in Table 2. A grid search was performed to evaluate all possible combinations of parameter values specified in the table. Training was considered complete when the error rate dropped below the threshold of 0.1. The trained model was subsequently applied to discriminate the two auditory stimuli using test data.

III. RESULTS

The discrimination accuracy is defined as the number of correct answers obtained by inputting the test data into the trained model that has been successfully completed, divided by the total number of test data points, which is 24. A grid search was performed, and the discrimination accuracy and hyperparameters that achieved the highest accuracy after ROI expansion are shown in Table 3. Furthermore, as illustrated in Table 4, the discrimination accuracy and hyperparameters that were found to be most effective prior to ROI expansion are documented, as outlined in [13]. Table 4 presents the results for 192 training data and 24 test data.

IV. DISCUSSION

Given that this report constitutes a two-classification discrimination, the probability of a correct guess by chance is 50%, and thus the chance level is also 50%. Previous studies indicate that an accuracy exceeding 50% can be interpreted as successful discrimination [15], while an accuracy above 60%

TABLE III. HYPERPARAMETERS AND DISCRIMINATION ACCURACY IN TWO DESIGNS. (ROIS: BA41, 42, 22)

Designs	Discrimination accuracy	Hyper parameter
Event-related	60.42%	Ks:6, F:16, Bs:16
Block	100%	Ks:4,7, F:32, Bs:8

TABLE IV. HYPERPARAMETERS AND DISCRIMINATION ACCURACY IN TWO DESIGNS. (ROIS: BA41, 42) [13]

Designs	Discrimination accuracy	Hyper parameter
Event-related	55.90%	Ks:6, F:6, Bs:16
Block	63.41%	Ks:3, F:14, Bs:16

is considered sufficiently reliable [16]. As demonstrated in Tables 3 and 4, an enhancement in discrimination accuracy was observed with ROI expansion in both designs, particularly in the block design. This enhancement is likely attributable to the incorporation of activation information from BA22, which contributed to the enhancement in accuracy. Although methodological differences from a previous study [14] preclude definitive conclusions, the results suggest a potential presence of tonotopic organization in BA22.

A substantial discrepancy in the degree of discrimination accuracy enhancement was observed between the two designs. While the event-related design demonstrated a 5% enhancement in accuracy, the block design exhibited a substantial 35% improvement. This discrepancy in accuracy enhancement is hypothesized to be attributable to the clarity of the images. The disparity in image clarity between the two designs can be attributed to the following. The fMRI detects changes in cerebral blood flow induced by variations in oxygen demand following external stimuli, a phenomenon known as the BOLD effect, to acquire brain activation images. In essence, longer stimulus duration (presented for 9 s in the block design, three times longer than in the event-related design) enhances the BOLD effect, enabling clearer brain activation imaging in the block design. Thus, the notable 35% improvement in discrimination accuracy observed in the block design is considered to result from ROI expansion in clearer images, which substantially enhances the inclusion of activation-related information. In a previous work [13], the ROI was set to BA41 and BA42, the same as in [12], and accuracy was examined by doubling the training data in the block design. While this resulted in a 15% increase from approximately 48% to 63%, in absolute terms, it only achieved a marginally more reliable level of accuracy. In this report, the training data remains the same 80 data points as in [12]. However, the accuracy improvement achieved through ROI expansion is twice as much as previous work. This result demonstrates that ROI expansion yields higher accuracy than simply increasing the training data. In contrast, the lower image clarity in event-related designs likely limit the effectiveness of ROI expansion in adding activation-related information, compared to the block design. To achieve further improvements in discrimination accuracy, potential strategies include increasing the amount of training data.

Based on the results of previous works [12][13] and this report, it was determined that the following approach is effective for discriminating two tones differing by 1 Hz in brain activation images of individuals using deep learning: implementing the imaging design as a block design, creating statistical images from two scans (as in a previous work [12]), and selecting ROIs BA41, 42, and 22. Therefore, the objective of this study was achieved. The results of this study demonstrate that discrimination within the 500 Hz band is possible and highly accurate for individuals. However, since sounds encountered in daily life exist beyond the 500 Hz band, investigation of other frequency bands is also important. Furthermore, as this study involved only one examinee and is a preliminary investigation, increasing the number of participants and verifying generalization to untrained individuals is necessary. Furthermore, the test data used in this report is limited to 24 samples. From the perspective of generalization performance in the recognition model, there is room for discussion, such as increasing the test data to verify performance.

V. CONCLUSIONS AND FUTURE WORK

Given the limited number of studies decoding quantitative musical characteristics, this study addressed the discrimination of two sounds differing by only 1 Hz a difference imperceptible to humans. The aim of this study was to improve discrimination accuracy beyond that reported in previous works [12][13]. ROI expansion was proposed as a method, and discrimination experiments were conducted using two fMRI experimental designs, event-related and block, as in a previous works [12][13]. The results showed an accuracy improvement of approximately 5% with the event-related design and over 30% with the block design. These findings demonstrate that ROI expansion is an effective approach for improving accuracy. Moreover, based on both the present results and those of previous works [12][13], it was determined that, for discriminating two sounds with a 1 Hz difference using brain activation images and deep learning, an effective strategy is to employ a block design for imaging, generate statistical images from two scans, and select ROIs in BA41, BA42, and BA22. This study provides new evidence that the brain responds even when humans are unable to perceive the difference.

Future work will include verifying generalization performance for untrained participants. With further progress, this line of this study is expected to contribute to the early detection of disease and to improvements in hearing aid performance.

REFERENCES

[1] Y. Liang, K. Bo, S. Meyyappan, and M. Ding, “Decoding fMRI data with support vector machines and deep neural networks”, Journal of Neuroscience Methods, vol. 401, pp. 110004, 2024.

[2] T. Horikawa and Y. Kamitani, “Hierarchical Neural Representation of Dreamed Objects Revealed by Brain Decoding with Deep Neural Network Features”, Frontiers in Computational Neuroscience, vol. 11, pp. 00004, 2017.

[3] M. Zhao and B. Lin, “An fMRI-based auditory decoding framework combined with convolutional neural network for predicting the semantics of real-life sounds from brain activity”, Applied Intelligence, vol. 55, pp. 118, 2025.

[4] I. R. Khan, S. L. Peng, R. Mahajan, and R. Dey, “Short-window EEG-based auditory attention decoding for neuroadaptive hearing support for smart healthcare”, Neuroscience Informatics, vol. 5, pp. 100222, 2025.

[5] M. Guilhem, M. D. L. Giovanni, and A. S. Shihab, “The Music of Silence: Part I: Responses to Musical Imagery Encode Melodic Expectations and Acoustics”, Journal of Neuroscience, vol. 41 (35), pp. 7435-7448, 2021.

[6] I. Daly, “Neural decoding of music from the EEG”, Nature, Scientific Reports. 13, pp. 624, 2023.

[7] N. Shigemoto, H. Satoh, K. Shibata, and Y. Inoue, “Study of Deep Learning for Sound Scale Decoding Technology from Human Brain Auditory Cortex”, IEEE, Global Conference on Life Sciences and Technologies, pp. 212-213, 2019.

[8] D. R. M. Langer, W. H. Backes, and P. V. Dijk, ” Representation of lateralization and tonotopy in primary versus secondary human auditory cortex”, NeuroImage, vol. 34, pp. 264-273, 2007.

[9] W. Guo et al., “Robustness of Cortical Topography across Fields, Laminae, Anesthetic States, and Neurophysiological Signal Types”, The Journal of Neuroscience, vol. 32(27), pp. 9159–9172, 2012.

[10] V. A. Kalatsky, D. B. Polley, M. M. Merzenich, C. E. Schreiner, and M. P. Stryker, “Fine functional organization of auditory cortex revealed by Fourier optical imaging”, PNAS vol. 102 No. 37, pp. 13325-13330, 2005.

[11] S. Romero et al., “Cellular and Widefield Imaging of Sound Frequency Organization in Primary and Higher Order Fields of the Mouse Auditory Cortex”, Cerebral Cortex, vol. 30, pp. 1603-1622, 2020.

[12] Y. Ooyashiki, K. Shibata, and H. Satoh, “Identification of Small Frequency Differences in Primary Auditory Cortex of Human Brain by Deep Learning Using fMRI”, JSME Chugoku-Shikoku Student Association 54th Student Member Graduation Research Presentation Lecture, pp. 01c2, 2023, (in Japanese).

[13] Y. Ooyashiki, K. Shibata, and H. Satoh, “Discrimination fMRI Learning of 1Hz Difference in Primary Auditory Cortex of Human Brain using fMRI”, JSME Mechanical Engineering Congress, pp. J162p-21, 2024, (in Japanese).

[14] J. B. Issa et al., “Multiscale optical Ca2+ imaging of tonal organization in mouse auditory cortex”, Neuron, vol. 83, pp. 944-959, 2014

[15] Thomas A. Carlson, Tijl Grootswagers, and Amanda K. Robinson;“An introduction to time-resolved decoding analysis for M/EEG”, arXiv, q-bio.NC, 1905.04820, (2019).

[16] A. K. Robinson, G. L. Quek, and T. A. Carlson, “Visual Representations: Insights from Neural Decoding”, Annual Review of Vision Science, vol. 9, pp. 313-335, 2023.