



A Study to Identify Brain Processing of Motor Kinematics through Deep Neural Network and Explainable AI

설명가능한 인공지능과 인공 신경망을 통한 대뇌의 운동 처리 규명에 관한 연구

August 2023

Graduate School of Natural Science Seoul National University Brain and Cognitive Science Major

HongJune Kim

A Study to Identify Brain Processing of Motor Kinematics through Deep Neural Network and Explainable AI

지도 교수 정 천 기

이 논문을 박사 학위논문으로 제출함 2023년 8월

> 서울대학교 대학원 뇌인지과학과 김 홍 준

김홍준의 박사 학위 논문을 인준함 2023년 6월

위육	원 장	곽 지 한	<u>현 (</u> 인)
부위	원장	정 천 기	<u>기</u> (인)
위	원	이 상 품	<u> 후</u> (인)
위	원	이교 -	구 <u>(</u> 인)
위	원	염 홍 기	기 (인)

Abstract

Various motor components, including muscle and joint movements, force and momentum, and motor kinematic parameters such as speed, acceleration, and position of limbs, manifest motor behavior. Understanding how the human brain processes these motor components is one of the significant interests in neuroscience. For this interest, current findings suggested that the cortical processing of motor kinematics appears differently depending on its behavioral characteristics. These findings had a practical implication for improving the decoding performance of brain-computer interfaces (BCI) predicting limb movement.

Numerous single-neuron studies have found a relationship between neuronal activity and motor kinematics, such as acceleration, velocity, and position. Despite differences between behavioral characteristics of each motor kinematic, it is hard to distinguish neural representations of their characteristics with macroscopic functional images such as electroencephalography (EEG) and magnetoencephalography (MEG). The reason might be that cortical signals are not sensitive enough to segregate kinematic characteristics due to their limited spatial and temporal resolution. Considering the different roles of each cortical area in producing movement, there might be a specific cortical representation depending on acceleration, velocity, and position characteristics.

Recently, neural network modeling has been actively pursued in decoding. I hypothesized that neural features of each kinematic parameter could be identified with a high-performing model for decoding with an explainable AI method. Here, I aimed to introduce a novel method to identify brain processing of motor kinematics using a

i

Deep Neural Network (DNN) model and explainable artificial intelligence (XAI) technique.

First, I developed recurrent neural network (RNN) based DNN models to decode times series of motor kinematic parameters of directional reaching movements from the cortical source signals. The result shows that the DNN models could significantly decode the time series of motor kinematic parameters such as acceleration, velocity, and position of the hand-reaching movements from the cortical source activity (r > .811, p < .001).

Second, I developed a novel approach to extracting the contribution of cortical areas for predicting each motor kinematic parameter by decomposing the high-performing DNN models using the XAI technique. This approach could segregate the functional areas for processing each motor kinematic parameter. The result showed that the contribution of cortical areas within visuomotor areas appeared differently depending on the motor kinematic parameters. I identified that different and shared cortical areas existed for decoding each kinematic attribute. Shared areas included bilateral supramarginal gyri and angular gyri known to be related to the goal of movement and sensory integration. On the other hand, dominant areas for each kinematic parameter appeared differently (the contralateral motor cortex for acceleration, the contralateral parieto-frontal network for velocity, and bilateral visuomotor areas for position). Regarding the visuomotor reaching movement, the motor cortex was found to control the muscle force, the parieto-frontal network encoded reaching movement from sensory information, and visuomotor areas computed limb and gaze coordination in the action space.

Through those studies, I revealed that the motor behavior of humans comprises a combination of motor components processed in

ii

each cortical area. Such results imply that various motor variables should be acquired from cortical areas to develop motor BCI. However, in many cases, signals or activities of required brain areas cannot be acquired. Primarily, invasive recordings can provide high-resolution signals, but such methods can only measure the focal area's signal. I speculated that the generative neural network model would help address the limitation. Here, I provide a proof of concept for the inferring the cortical signals of the motor area essential for producing motor kinematics, such as acceleration, from the other cortical signals that participate in motor processing but do not in producing.

Keywords: Brain-computer interface (BCI), Deep neural network, movement prediction, Explainable AI, Neuroimaging **Student Number:** 2019–28861

Table of Contents

Abstract	i
Table of Contents	iv
List of Tables	viii
List of Figures	ix

PART I. INTRODUCTION

Chapter 1. Human Motor Behavior and Brain	L
1.1. Nervous Systems for Human Motor Behavior	L
1.2. Cortical Processings for Motor Control	5
1.3. Brain-Computer Interface System	3

Chapter 2. Application of Artificial Intelligence to Develop	
Neuroscience and Neuro-engineering Research	11
2.1. Artificial Intelligence and Machine Learning	11
2.2. Deep Learning (Deep Neural Network)	13
2.3. Explainable Artificial Intelligence (XAI)	18

Chapter 3. Purpose of the Study......20

PART II. Experimental Studies

Chapter 4. Decoding Motor Kinematics from Cortical Source		
Signals Using Deep Neural Network Model	.25	
4.1. Chapter Introduction	.25	
4.2. Materials and Methods	.27	
4.2.1. Data Information	27	
4.2.2. Signal Preprocessing	28	
4.2.3. Deep Neural Network Models	29	
4.2.4. Model Training and Evaluation	33	
4.3. Results	.35	
4.3.1. Decoding Accuracy	35	
4.3.2. Hand-reaching Trajectories in 3D spaces	36	
4.4. Discussion	.37	
4.4.1. Decoding time-series motor kinematics via DNN		
Model	37	
4.4.2. RNN Architecture for Decoding Time-series Neura	1	
Signals	38	
4.4.3. Limitation	40	

5.2.1. Data Information 4	12
5.2.2. Explainable Artificial Intelligence (XAI) 4	12
5.2.3. Statistical processing 4	ł5
5.3. Results	15
5.3.1. Shared areas 4	ŀ6
5.3.2. Acceleration-dominant areas 4	16
5.3.3. Velocity-dominant areas 4	ŀ6
5.3.4. Position-dominant areas 4	ł7
5.4. Discussion5	50
5.4.1. Mapping cortical areas for movement kinematics with	1
DNN and XAI technique5	50
5.4.2. Shared areas representing sensory processing and	
action goal selection for reaching movement5	54
5.4.3. Dominant areas showing unique characteristics for	
decoding each kinematic parameter in 3D space5	55
5.4.4. Limitation	52

PART III. General Conclusion and Perspectives

Chapter 6. General Conclusion and Perspectives	.64
6.1. General Conclusion	.64
6.2. Perspectives	.64

liography70

Abstract in Korean	81
--------------------	----

List of Tables

|--|

List of Figures

Part I. Introduction

Chapter 1. Human Motor Behavior and Brain

Figure 1-1	
Figure 1-2	
Figure 1-3	
Figure 1-4	5
Figure 1-5	7
Figure 1-6	

Chapter 2. Application of Artificial Intelligence to Develop

Neurosciecne and Neuro-engineering Research

Figure 2-1	
Figure 2-2	
Figure 2-3	
Figure 2-4	

Part II. Experimental Studies

Figure 4-2	
Figure 4-3	
Figure 4-4	
Figure 4-5	

Chapter 5. Identification of Cerebral Cortices for Processing Motor Kinematics Using DNN and Explainable Artificial Intelligence

Figure 5-1	
Figure 5-2	
Figure 5-3	
Figure 5-4	50

Part II. General Conclusion and Perspectives

Chapter 6. General Conclusion and Perspectives Signals Using Deep Neural Network Model

Figure 6-1	67
Figure 6-2	68

PART I. Introduction

Chapter 1. Human motor behavior and brain

Motor behavior includes every movement that controls the muscle, from involuntary twitches or reflexes to voluntary goal-directed actions, in every part of our body, from toe to head. Regardless of their type, the nervous system controls motor behavior. Therefore, understanding the control mechanism is one of the most significant interests in neuroscience. In addition, this is mandatory to develop a brain-computer interface system that controls robotic prosthetic limbs for disabled people. This chapter briefly introduces how our nervous system controls human motor behavior, specifically the cerebral cortex. Furthermore, I also introduce the basic concept of BCI for motor behavior.

1.1. Nervous Systems for Human Motor Behavior

The Human nervous system consists of two main parts: the central nervous system (CNS) and the peripheral nervous system (PNS) (Figure 1–1). CNS includes the brain and spinal cord. On the other hand, PNS comprises the afferent (sensory input channel) and the efferent (motor output channel) nerves. Afferent nerves in PNS translate extrinsic stimuli, such as light (vision), sound, or touch, and deliver them toward the CNS. The CNS processes the



Figure 1–1. Organization of the human nervous system.

extrinsic stimuli and then sends control commands toward muscles or organs via the efferent nerve of the PNS (Bear et al., 2020). This process is a fundamental mechanism of sensation and motor behavior.

Anatomically, the brain consists of three subdivisions: forebrain (prosencephalon), midbrain (mesencephalon), and hindbrain (rhombencephalon). The hindbrain contains the cerebellum and medulla, and the midbrain contains the brainstem. Lastly, the forebrain comprises the cerebrum, thalamus, and hypothalamus. The cerebrum can be divided into four lobes: occipital lobe, parietal lobe, temporal lobe, and frontal lobe



Figure 1–2. Lobes of human cerebrum (Bear, Connors, & Paradiso, 2020).

(Figure 1–2). Each can be segregated into anatomical or functional structures (Glasser et al., 2016) (Figure 1–3). For example, the motor cortex is the cerebrum's central structure for motor control. It comprises Brodmann area 4 (BA4), within the precentral gyrus, and BA6. The BA4 is referred to as the primary motor cortex (M1).

The CNS comprises a hierarchy system for motor control, with the forebrain at the top and the spinal cord at the bottom. In the higher level of the hierarchy, association areas of the neocortex and basal ganglia in involved in 'strategy', such as the goal of motor behavior or the optimal way to achieve the goal. In the middle level, the motor cortex and cerebellum control' tactics',



Figure 1–3. Dividing cerebral cortex into anatomical or functional organization. (upper) Brodmann area. The regions defined by its cytoarchitecture or histological structure and organization of cells. (lower) a multi-modal parcellations of HCPMMP1 atlas (Glasser et al., 2016), presented on the FreeSurfer Average inflated surface.

such as the spatiotemporal sequence of muscle control. Lastly, in the lower level, the brainstem and the spinal cord are involved with 'execution', which activates motor neurons (efferent). The brain communicates with the spinal cord via two major pathways in the hierarchical system (Figure 1–4). One is the lateral pathways. Most of the neurons in the pathways originate in the motor cortex (Kolb, Whishaw, Teskey, Whishaw, & Teskey, 2001). Through these pathways, the motor cortex can control motor behavior directly. Thus,



Figure 1–4. Hierarchical organization of descending spinal tracts and their origins with the motor loop of brain-basal ganglia-thalamus (Bear et al., 2020).

the pathways are involved in the voluntary movement of distal limbs. On the other hand, the other, named ventromedial pathways, originates in the brainstem and controls locomotion and posture (Purves et al., 2012)

1.2. Cortical Processing for Motor Control

The cerebral cortex controls overall motor behavior at the top of the central motor system. Penfield and Boldrey (1937) found that the electrical

stimulation on the precentral gyrus would elicit muscle cramps in a specific body part. Systemic investigation of the BA4 reveals that the M1 has a somatotopic organization (Figure 1–5). The cortical structure of M1 shows how this area receives inputs from other cortical areas and sends outputs toward the lower motor neurons. The neocortex comprises six layers. Most corticospinal cells within the layers originate in the cortical layer V (Kolb & Whishaw, 2009). Populations of pyramidal cells in cortical layer V (Betz cells) project their axons through the cortico–spinal tract toward the lower motor neurons. The axons monosynaptically excite the lower motor neurons.

Furthermore, the cortico-spinal axons can make branches toward the motor neurons. The branches of axons connect to motor neurons and inhibitory interneurons and enable the excitation of flexor motor neurons and inhibition of extensor muscle simultaneously (Evarts, 1968). The cortical layer V receives neural input from the other brain areas, such as BA6, the primary somatosensory cortex (S1), or the thalamus.

Studies on the electrophysiological recordings suggest that the cerebral cortex encodes various components for motor behavior: motor kinetics, such as force, and kinematics, such as direction, velocity, acceleration, and limb position. Early studies focused on the role of the motor cortex. For example, Georgopoulos, Schwartz, and Kettner (1986) found that some neurons in M1 fired for the reaching movement in one specific direction. Further studies found that the premotor area (BA6) processes the preparation



Figure 1–5. A somatotopic organization in the precentral gyrus. Such representation also is called as cortical homunculus.

of motor behavior (Churchland, Santhanam, & Shenoy, 2006; Churchland, Yu, Ryu, Santhanam, & Shenoy, 2006). Along with the studies on the motor cortex, other cortical areas also appear to process the component of motor behavior. For example, neuron activities of non-human primates in M1 and BA5, part of the posterior parietal cortex (PPC), correlate with motor kinematics during center-out hand-reaching movement (Ashe & Georgopoulos, 1994). Moreover, studies using non-invasive recordings, such as magnetoencephalogram (MEG), reveal that the human brain recruits a large-scale network for processing motor kinematics (Bourguignon, Jousmäki, Dalal, Jerbi, & De Tiège, 2019; Bourguignon et al., 2012; Jerbi et al., 2007).

1.3. Brain-Computer Interface System

Brain-Computer interface (BCI) is a direct communication pathway between brain activity and external devices. Thus, the BCI system utilizes brain signals to control external devices, such as cursors or robotic limbs. Such technology will help disabled people interact with the world again (Wolpaw, 2013). Furthermore, BCI technology also can be employed to implement human experiences (Si-Mohammed et al., 2018).

BCI system comprises three modules. The first part is the acquisition module that measures the brain activities of BCI users. Measuring brain activity can be divided into invasive and non-invasive methods (Hagen et al., 2018) (Figure 1–6). The invasive method places sensors (electrodes) on the surface or inside the brain tissue. Thus, since this method measures brain activity directly in the brain, it can provide a high signal-to-noise ratio. However, this method can cause damage to the brain and be unsuitable for long-term implantation. Moreover, spatial coverage is limited to focal brain areas. In contrast, the non-invasive methods are highly accessible and safe since it does not require surgical procedures. Furthermore, although external noises can affect these methods, they can measure whole-brain activity. Therefore, the non-invasive method is currently more suitable than the

8



Figure 1–6. The methods to measure the brain signals for BCI (Hagen, Næss, Ness, & Einevoll, 2018). Non-invasive methods (EEG and MEG) measured the electrophysiological signals outside the scalp. On the contrary, the invasive methods measured it inside the skull, right over the brain surfaces or penetrates the cortices to measure the cell activity.

invasive method for developing a universal BCI system for ordinary people. The second is the signal processing module. This module process brain signals, measured by electrodes, into meaningful signals. The electrodes measure brain activity and transform them into a digital signal. However, the electrodes can also record external noises and brain signals simultaneously. The processing module filters these artifacts and leaves the brain signals. The general signal processing procedure includes detrending, artifact correction, signal filtering, and re–referencing.

Then, the prediction module, the third part, decodes the behavior

from the brain signals using an ML model. Traditional ML models included a manual feature engineering process. In this process, spatiotemporal features of brain signals, which seem related to behavior, are extracted before the prediction process. However, deep learning models learn the relationship between behavior and brain signals and automatically infer the signal features (Goodfellow et al., 2016) for the behavior. Although learning deep learning models requires large datasets compared to the traditional ML models, they have become critical methods for BCI with high accuracy and the advantage above.

Chapter 2. Application of Artificial Intelligence to Develop Neuroscience and Neuro-engineering Research.

In this chapter, I will briefly introduce the basics of AI and its subfields to promote understanding of this thesis. In addition, I also introduce some examples of the application of AI, especially deep neural networks (DNN), in neuroscience and neuro–engineering research.

2.1. Artificial Intelligence and Machine Learning

Artificial Intelligence (AI) is one of the fields of computer science that attempts to implement human learning, reasoning, and perceptual abilities artificially. Most of the works humans try to solve through AI are things humans can do intuitively and without difficulty. However, the critical challenge of artificial intelligence is to provide (teach) this intuitive and informal knowledge to computers (Goodfellow et al., 2016). Thus, artificial intelligence should be able to extract characteristics from original data and acquire knowledge independently.

Machine learning (ML), one of the subfields of AI (Figure 2–1), studies a computer algorithm that automatically improves through experience



Figure 2–1. A Venn diagram showing relationship between AI and its subfields and representative examples of each field (Goodfellow, Bengio, & Courville, 2016).

ML comprises two fundamental concepts. One is representation, the data evaluation, and the other is a generalization, which is the processing of unknown data. The primary processing workflow of ML consists of three parts. The first part is input. The data flows into the algorithm. The feature extraction algorithm extracts the hidden features from the input data in the second part. In the early days, this process was done manually through feature engineering. However, with the development of deep learning, the algorithm also learns and infers the feature. The third part is prediction. The ML algorithm predicts the output data based on the features.

There are two large categories of learning algorithms in ML:

supervised learning and unsupervised learning. Supervised learning learns the hidden relationship between the input and the output data (Russell, 2010). It learns the error (or loss) between predicted and real output (label) during the learning. Finally, after complete learning, supervised learning can predict the output of the new input data. Supervised learning allows us to regress or classify the input data.

On the other hand, the label does not be given in unsupervised learning. Unsupervised learning allows algorithms to infer patterns or clusters of data (Hinton & Sejnowski, 1999). Furthermore, this algorithm can learn the probability densities of given data. Thus, an unsupervised learning algorithm can generate novel data that resemble the given data.

2.2. Deep Learning (Deep Neural Network)

Among the various subfields of ML, one has recently attracted researchers and developers. Deep learning is one of the subfields of ML and is defined as a class of ML algorithms that uses multiple layers to progressively extract higher–level abstracts from the raw input data (Yoshua Bengio, Courville, & Vincent, 2013; Deng & Yu, 2014). The fundamental concept of deep learning, artificial neural network (ANN), was proposed more than 50 years ago. However, since 2012, deep learning has become a fundamental modern artificial intelligence technology.

The mechanism of the human neural network inspired the

1 3



Figure 2–2. Structures and computational mechanism of biological and artificial neurons. (upper) A basic structure of biological neuron. The neurons receive external input from the other neurons and fire action potential. (lower) computation mechanism of artificial neuron within deep neural network (DNN).

fundamental concept of the DNN model (Figure 2–2). The dendrites of biological neurons receive external input. When the external input exceeds the threshold, the neurons fire an action potential and transmit the signals through the axon (Kandel et al., 2000). Similarly, the artificial neurons in the DNN model receive input data (or features) (X), and then they multiply synaptic weights (W) on the data and add bias (b) (equation 2–1). The synaptic weights could correspond to the number of synapses made with the

other neurons. For the weighted inputs (Z), the artificial neurons transform them via activation function, g (equation 2–2).

$$(2-1) Z = WX + b$$

$$(2-2) Y = g(Z)$$

DNN models comprise several layers of artificial neurons. The most representative form of the DNN model is a multi-layer perceptron (MLP) (Figure 2–3). MLP consists of an input layer, a hidden layer, and an output layer. Like the traditional ML models, the input layer receives the input data (*X*), and the output layer makes predictions (\hat{Y}). However, the unique characteristic of DNN is in the hidden layers. The number of hidden layers between the input and output layers determines the depth of the DNN model. In the DNN model, the input data is transmitted to the output layer through the hidden layer (equation 2–3). The output of the hidden layer (*H*) may be input to another hidden layer again and, at the end of this chain, reaches the output layer (equation 2–4). Through this architecture, the DNN models can learn more complex representations.



Figure 2–3. Multi-layer perceptron (MLP). MLP comprises an input layer, hidden layers, and an output layer. As the hidden layers became deeper, the level of representation increased from low (contrast or edges) to high (eye, lips, or nose). Image source came from Google Tech Talk by Jeff Dean at Campus Seoul on March 7, 2016.

(2-3)
$$Z_1 = W_1 X + b_1$$
, $H_1 = g(Z_1)$
(2-4) $Z_2 = W_2 H_1 + b_2$, $\hat{Y} = g(Z_2)$

The learning (or fitting) of the neural network model means approximating the model's prediction (\hat{Y}) to label (Y). The output layer learns the error (loss) between the labels and predictions and then approximates the predictions to the labels. However, there are no explicit labels for the hidden layers, and the data did not suggest the desirable values that the layers should produce. Thus, the learning rules of deep learning propagate the error between prediction and label toward the parameters of each hidden layer.

Since the errors are propagated backward from the output layer to the first hidden layer, the learning algorithm is called backpropagation. This algorithm calculates the gradient of the error function into parameters of a target hidden layer (equation 2–5). Since the output is made through the chain of hidden layers, the values can be calculated following the chain rule of partial differentiation (equation 2–6). To utilize the backpropagation, the activation functions must be differentiable. The learning algorithm controls the amount of learning with the learning rate (η). Thus, the parameters are converged toward an optimized point by repeating the subtraction of the gradient.

(2-5)
$$W_n := W_n - \eta \frac{\partial}{\partial W_n} L(Y, \hat{Y})$$

(2-6) $W_n := W_n - \frac{\partial Z}{\partial W_1} \frac{\partial H_1}{\partial Z_1} \frac{\partial Z_2}{\partial H_1} L(Y, \hat{Y})$

Currently, neuroscience and neuro-engineering research adopt the DNN models. For example, for BCI development, DNN can apply to decode human behavior, such as hand-reaching movement, from complex brain signals (Kim, Kim, & Chung, 2023; Lawhern et al., 2018; Ma et al., 2021; Yeom, Kim, & Chung, 2020). In addition, since the DNN model is more biologically plausible than the traditional ML model, several studies have tried to understand human brain function and its mechanism by investigating the layers within the model (Kar, Kubilius, Schmidt, Issa, & DiCarlo, 2019; Richards et al., 2019; Yamins & DiCarlo, 2016).

2.3. Explainable Artificial Intelligence (XAI)

Deep learning allows us to enhance prediction performance compared to traditional ML algorithms. Furthermore, since the hidden layers learn the feature of the data, the deep learning model does not need a feature engineering process. However, understanding the inference behind the model's decision is almost impossible since millions of parameters are interwound between multiple layers. Thus, AI systems are considered 'black boxes' to users and developers, and how the features of input data influence the prediction output is hard to understand.

The explainable artificial intelligence (XAI) technique helps describe the relationship between model predictions and input data features. One of the aims of this technique is to turn the black boxes into glass boxes. Figure 2–4 illustrates the intuitive example of the XAI. Inception V1 (GoogLeNet) is one of the most potent image classifier models, consisting of 22 layers and about 11 million parameters. This model predicts the image as 'Giant Panda' with an 89.8% chance. The XAI can explain where the complex model seeks the attribution of Giant Panda. The attributions of the Giant Panda appear around the face, nose, and texture of the panda's fur.



Inception V1 top 3 class: giant panda 89.8%, teddy 0.3%, gibbon 0.3%

Figure 2–4. An example of explainable artificial intelligence (XAI) in the image classification model. The attribution mask (lower left) is where the areas contributed to classify the image as 'Giant Panda'. A XAI method for this example was integrated gradients (IG). Image source: https://www.tensorflow.org/tutorials/interpretability/ integrated_gradients?hl=ko

XAI technique can be utilized in functional neuroimaging research. Through the XAI, the neural attribution for behavior can be identified from brain activity. For example, Kim et al. (2023) decoded various motor kinematics of hand-reaching movement via the high-performing recurrent neural network (RNN) models. They identified the cerebral cortices in those models with XAI.

Chapter 3. Purpose of the study

Human beings make ceaseless and countless movements in their lives. Motor behavior is manifested by various motor characteristics, including muscle and joint movements, force and momentum, and motor kinematic parameters such as speed, acceleration, position, and direction of limbs (Ashe & Georgopoulos, 1994; Bourguignon et al., 2019; Humphrey, Schmidt, & Thompson, 1970; Jerbi et al., 2007; Kakei, Hoffman, & Strick, 1999). Therefore, the relationship between neural activity and various motor characteristics is essential for understanding the movement behavior of a person. It also has a practical implication for improving the decoding performance of brain–computer interfaces (BCI).

Previous studies using non-human primates have found that neuronal activity is closely related to various motor kinematic parameters. An early study on the relationship between neuronal activity and motor kinematic parameters has shown that neurons in the primary motor cortex (M1) can selectively respond to a specific direction of reaching movement (Georgopoulos et al., 1986). A further study has reported that neuronal activities in M1 and Brodmann area 5 (BA5) significantly correlate to timevarying motor kinematic parameters such as velocity, position, acceleration, and direction (Ashe & Georgopoulos, 1994). Two notable results were drawn from that study. First, neurons in M1 and BA5 preferred a specific kinematic parameter. Among neurons recorded in the M1 and BA5, the proportion of neurons that preferred velocity was the highest, followed by positionpreferred and acceleration-preferred neurons. This neuronal preference toward a specific motor kinematic characteristic has been validated in subsequent studies (Churchland, Santhanam, et al., 2006; Moran & Schwartz, 1999; Wang, Chan, Heldman, & Moran, 2007). Second, this preference varied across cortical areas. For example, acceleration-preferred neurons were found in BA5 but not in M1. These results suggest that the neural system of animals has the processor for motor kinematics even at the small brain level. Moreover, the neural process for kinematics may vary across brain areas. Since the spatial coverage to measure a single neuron activity was limited to a focal area (Sejnowski, Churchland, & Movshon, 2014), it was hard to identify neural representations that appeared throughout the brain. Thus, finding a cortical representation of each kinematic parameter with macroscopic neural data has been an important issue.

Non-invasive studies of humans have suggested that large-scale cortical activity also seems to be closely related to kinematic parameters. For example, Jerbi et al. (2007) found that reaching speed is significantly coherent with cortical source activity from M1, premotor cortex (PM), posterior parietal cortex (PPC), and dorsolateral prefrontal cortex (DLPFC). A similar coherence map was found at speed and acceleration (Bourguignon et al.,

 $2 \ 1$

2012). Taken together, motor kinematic parameters appeared to be handled in a large-scale network that centered around the motor-related cortex. However, since these two coherence maps were presented similarly, cortical areas related only to acceleration or velocity could not be identified from the abovementioned studies. I speculated that coherence-based mapping could not distinguish cortical areas for each kinematic parameter because the spatial resolution of a non-invasive method is too limited to segregate them. Instead of mapping a simple relation of cortical activity with kinematic parameters, an alternative approach is needed to map cortical areas contributing to the production of each kinematic parameter.

Thus, this study aimed to identify cortical areas for each kinematic parameter. To achieve this aim, a novel approach was employed to complement the shortcomings of limited coverage of single neurons and limited resolution of large–scale approaches. Here, I developed state–of–art methods called deep neural network (DNN) model and explainable artificial intelligence (AI) to investigate the complex relationship of processing motor kinematics.

In Chapter 4, I developed deep neural network models to elucidate the relationship between various motor kinematics, such as acceleration, velocity and position, and brain signals. I implemented recurrent neural network (RNN) architecture to process sequential brain signal features for each motor kinematics. I hypothesized that the DNN models could predict

 $2 \ 2$

the time series of those motor kinematics.

DNN models were a powerful method to investigate the relationship between complex brain signals and time-series of motor kinematics since the decoding accuracy was high and hand-reaching movement toward the targets was accurately recovered. However, such models had a limitation in explainability due to their complex architecture. In Chapter 5, I tried to identify the attribution of motor kinematics in the cerebral cortices hidden inside the model. Using the XAI technique, I decomposed the highperforming DNN models that can decode motor kinematics. This approach enabled me to identify the cerebral cortices' contribution to decoding acceleration, velocity, and position. The findings in Chapters 4 and 5 have been published in NeuroImage (Kim et al., 2023).

It should be necessary to acquire signals of brain areas corresponding to motor components throughout such studies to develop a reliable motor BCI system. However, acquiring brain signals of desired areas was sometimes challenging. For example, invasive methods, such as iEEG, can provide highresolution signals, but such methods only could measure a focal brain area. Thus, In Chapter 6, I tried to provide a proof of concept for the inference of the brain activity of the motor cortex from the motor-related brain area via a generative adversarial neural network (GAN). I employed a signal generator network, MelGAN, to translate signals of the motor-related area (IPS) into the primary motor area (M1) and generate artificial signals of the area. I
hypothesized that the GAN model could translate the signals of the motorrelated area and generate artificial brain signals. Furthermore, the DNN model, which decodes hand-reaching acceleration from the M1 area, could also decode the acceleration from translated and generated signals. Chapter 4. Decoding Motor Kinematics from Cortical Source Signals Using Deep Neural Network Model

4.1. Chapter Introduction

Brain-computer interfaces (BCIs) represent a promising field of research and development in neuroscience that has the potential to change the way we interact with technology significantly. These devices establish a direct communication pathway between the human brain and external devices, allowing individuals to control them through their thoughts. As a result, BCIs offer a wide range of potential applications in various fields, such as neuroprosthetics (R. A. Andersen, Aflalo, & Kellis, 2019; Muller–Putz & Pfurtscheller, 2007), gaming and virtual reality (Kerous, Skola, & Liarokapis, 2018). Besides restoring mobility and independence for people with physical disabilities, BCIs could improve human cognitive abilities, enhance learning, and treat neurological disorders.

As I introduced in Chapter 1, a machine learning (ML) algorithm

exists in the center of the BCI system and mediates brain activity and behavior. First, ML receives the brain signals directly or refined neural features and behavior, then learns their hidden relationship. Then, the algorithm can decode the brain signals and predict the behavior that the users intended. Recently, deep learning algorithms have been drawing attention in AI. The deep neural network (DNN) model, the essential architecture of the algorithm, can significantly enhance the performance of BCIs by allowing for more accurate and efficient analysis of brain signals (Altaheri et al., 2021; Lawhern et al., 2018; Yeom et al., 2020).

Despite recent progress in ML algorithms for BCI, significant challenges remain in developing accurate and reliable systems. One of those challenges is decoding continuous behavior from a time-series of brain signals. Most human behavior, especially motor behavior, is continuous and expressed as continuous variables, such as time series of kinematics or kinetics. Thus, the algorithm should learn the relationship between the time-series of motor kinematics and brain signals. However, the current DNN-based BCI still focused on classifying several behavior classes (Lawhern et al., 2018; Ma et al., 2021), with few attempts to predict the time-series behaviors using traditional ML algorithm (Bradberry, Gentili, & Contreras-Vidal, 2009; Bundy, Pahwa, Szrama, & Leuthardt, 2016). Furthermore, although those attempts reported that the movement was significantly decoded, the prediction accuracy was too low for the practical application of the BCI system.

The present study tested that the DNN model could decode timeseries motor kinematics of hand-reaching movements, such as acceleration, velocity, and position, from the cortical source signals. I speculated that the recurrent neural network (RNN) architecture could process the sequential features of brain signals during hand-reaching movement. Therefore, I developed the RNN-based regression model to predict the time-series motor kinematics from the cortical source signals measured by magnetoencephalography (MEG).

4.2. Materials and Methods

4.2.1. Dataset information

I used the dataset of H. G. Yeom, Kim, and Chung (2013). Nine right-handed subjects participated in the study. Their mean age was 26.7 \pm 6.8 (mean \pm standard deviation) years (range, 19 to 37 years). The subject's handedness was evaluated through the Edinburgh Handedness Inventory. All participants were confirmed as right-handed with a performance of more than 80 points (87.2 \pm 5.7).

The experimental paradigm follows the center-out reaching task. Subjects were instructed to move their right hands toward the target represented in a 3D space with a visual movement cue during the experiment. Visual stimuli were presented on screen with a STIM2 system (Neuroscan, El Paso, USA). At the beginning of each trial, subjects were instructed to hold the index finger of the right hand on the sphere in the center of the screen for 4s. After that, a target sphere connected with the line was presented randomly in one of four corners for 1s. Participants were instructed to reach their right hand toward the target and return to the center as quickly and accurately as possible within the given time. This center–out task was repeated in 120 trials per session. All subjects repeated the session twice. Experiments were approved by the Institutional Review Board (IRB) of Seoul National University Hospital (approval number: 1105–095–363). They were performed under the Declaration of Helsinki.

During the experiment, 306 channels of the whole-head MEG system (VectorView TM, Elekta Neuromag, Finland) measured the subject's brain activities. The sensors were distributed in 102 locations. On each location, two planar gradiometers and one magnetometer were positioned. Simultaneously, a three-axis accelerometer on the subject's index finger measured the kinematics of hand-reaching movement. For both signals, the sampling frequency was 600.615Hz.

4.2.2. Signal Preprocessing

Although the whole-head MEG system measured the brain activity, I only used the 68 gradiometers in the 34 locations to decode the kinematics of hand-reaching movement. The gradiometers had a better signal-to-noise ratio (SNR) than magnetometers in measuring cortical signals (Hämäläinen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993). Furthermore, 34 locations around parietal regions were known to represent center-out reaching movement (Waldert et al., 2008; H. G. Yeom et al., 2013).

In the signal preprocessing, the signals of the MEG sensors were 0.8Hz high pass filtered. Then, the signal artifacts, such as electrooculography (EOG) and electrocardiogram (ECG), were corrected by independent component analysis (ICA). After artifact correction, the signals were filtered with a bandpass of 0.5Hz – 8Hz. Then, the filtered MEG signals were down-sampled to 50Hz (20ms of intervals). On the other hand, the accelerometer signals were filtered with a bandpass of 0.2Hz – 5Hz and down-sampled to 50Hz. Then, the linear trends were corrected. Thus, the individual data shapes were [240 trials X 141 time-points X 68–gradiometer channels] and [240 trials X 151 time-points X 3–axis of accelerometer].

Next, the cortical source activities of each trial were estimated from the preprocessed MEG sensor data. Since the individual MRI data were not acquired, I projected the MEG sensors on the template brain data. The template brain was the FreeSurfer average brain (Dale, Fischl, & Sereno, 1999; Fischl, Sereno, & Dale, 1999; Fischl, Sereno, Tootell, & Dale, 1999). I computed a forward solution of MEG sensor space and FreeSurfer average (FsAverage) source space. Then, the inverse solution was computed by a minimum–norm inverse operator. The estimation method was dynamic



Figure 4–1. Deep Neural Network based decoding model. DNN models for decoding kinematic trajectories from neural signal are described. The basic architecture is Long Short–Term Memory (LSTM), a specific type of Recurrent Neural Network (RNN) for processing time–series dynamics of neural signals. (A) The specific architecture of LSTM, (B) Decoding model structure, an 8–layerd deep neural network model.

statistical parametric mapping (dSPM) (Dale et al., 2000). The whole preprocessing procedures were done with a Python software package for M/EEG processing, MNE (Gramfort et al., 2013).

4.2.3. Deep Neural Network Model

A model with a high decoding ability was needed to accurately infer the relationship between cortical source activity and each kinematic parameter. To this end, I used a state-of-the-art model called DNN. Since DNN is more

biologically plausible than conventional linear regression, motor kinematic parameters can be more accurately decoded by the DNN. A recurrent neural network (RNN) architecture was selected among various neural network architectures. RNN is appropriate for decoding time-series or sequential data (Petneházi, 2018). The specific architecture type of RNN was Long-short term memory (LSTM; Figure 4-1A) (Hochreiter & Schmidhuber, 1997). LSTM can process time-series data by adjusting states cumulated from the past (memory) and updating them with current information. This process can effectively address the vanishing gradient problem when the model is fitted through error backpropagation through time. An overview of the LSTM operation can be summarized in the following equations. 'Gates' in the LSTM can update time-series (or sequencing) information and apply it to the current feature. First, LSTM concatenates current data (x_t) and past information (h_{t-1}) and then expresses them to gates. The forget gate notated in equation 4-1 adjusts memories (called 'state (C)') conveyed from the past state (C_{t-1}) by the forgetting ratio. Here, the function g has a sigmoid function. Thus, the forget gate adjusts the amount of data from the past. In equation 4-2, the input gate decides how much information should be saved in memories. Adding forgotten memories and new information from the input gate makes a new state of memories (C_t) (equation 4-3). Finally, the current decoded output is made by multiplying sigmoidal information and the vector of memories, as noted in equations 4-4.

$$(4-1) f_t = g(W_f[x_t, h_{t-1}]), \quad forget gate = f_t \odot c_{t-1}$$

$$(4-2) i_t = g(W_i[x_t, h_{t-1}]), \tilde{c}_t = \tanh(W_c[x_t, h_{t-1}]),$$

$$input gate = i_t \odot \tilde{c}_t$$

$$(4-3) c_t = input gate \oplus input gate$$

$$(4-4) o_t = g(W_o[x_t, h_{t-1}]) \qquad h_t = o_t \odot \tanh(c_t)$$

A bidirectional LSTM (bLSTM) was used to increase decoding performances so that time series computation by LSTM was done bidirectionally (Schuster & Paliwal, 1997). The model had eight layers, including an input layer with a sequencing function, two layers of bLSTM with two batch-normalization layers to cope with gradient vanishing and overfitting problems, and two dense layers. The last dense layer is the regression layer which made the 3-axis coordination of kinematic parameters in a 3D space. Figure 4–1B illustrates the structure of the bLSTM model. The bLSTM model comprises 3,083,267 parameters (weights) and includes two bLSTM with a layer normalization layer and two fully connected dense layers. The last dense layer predicted the three-axis coordination of each motor kinematics. The summary of the model is described in Table 4–1. The model's initial state was sampled from the Glorot normal distribution (Glorot & Bengio, 2010).

Layer name	Number of	Output shape of	Note
(activation)	parameters	the layer	Note
_		F- 4	[Batch size, time-
Input	—	[Batch, 151, 445]	points (tps), number
			of sources]
Lambda	-	[Batch, 141, 11, 445]	Sequencing time-
(Sequencing)			series signals into
(bequeneing)			20ms bin
bLSTM	1,437,696	[Batch, 141, 512]	Many-to-one
(tanh)			LSTM
Batch	_	[Batch, 141, 512]	
Normalization			
bLSTM	1,574,912	[Batch, 141, 512]	Many-to-many
(tanh)			LSTM
Batch	_	[Batch, 141, 512]	
Normalization			
Dense	65,664	[Batch, 141, 128]	
(ELU)			
Dense	387	[Batch, 141, 3]	Output layer, 3-axis
(linear)			coordinations
Total number	3,083,267		
of parameters			

Table 4–1. A detailed description of the DNN model's architecture

4.2.4. Model Training and Evaluation

For the model training, I split the individual dataset into five-fold. Four (196 trials) were used for training, and the other (48 trials) for testing. In addition, I applied five-fold cross-validation methods to evaluate the whole dataset. Thus, the accuracy of individual subjects was the average of five-fold cross-validation. Then, the input data were normalized through the Min-Max scaler. By the scaler, the range of input data was between 0.0 and 1.0. It helps

make the model's gradient follow the global minimum point. In addition to the data normalization, the weights of the DNN models were initialized.

The mean absolute error function (MAE(Y, \hat{Y}), (equation (5)) computed the loss between predicted (\hat{Y}) and real (Y) trajectories of each kinematics. The learning rule for optimization was Adaptive moment estimation (Adam) (Kingma and Ba, 2014). For the model training, I set the learning rate for optimization to 1E–3. In addition, beta1 and beta2, the exponential decay rates of the Adam optimizer, were set to 0.9 and 0.99.

(1)
$$\operatorname{MAE}(Y, \widehat{Y}) = \frac{1}{N} \sum |Y_i - \widehat{Y}_i|$$

For efficient training, mini-batch training methods were used (Li, Zhang, Chen, & Smola, 2014). I set the mini-batch size to eight. Thus, the optimizer updated the models' parameters per eight trials. The model was trained by iterating the whole train dataset 100 times. During the training, I only saved the parameters which produced the lowest validation error.

Furthermore, model ensemble methods were applied to increase the decoding performance (Zhou, Wu, & Tang, 2002). Here, I trained five DNN models for prediction. Each model had the same architecture but was initialized in different random states. The average prediction of five models determined the decoding result.



Figure 4– 2. Decoding accuracy. (A) Mean of normalized cross-correlation (NCC) between real and predicted motor kinematics. (B) Mean of Pearson's correlation. (C) Pearson's correlation of individual data.

4.3. Results

4.3.1. Decoding Accuracy

I calculated the decoding accuracy through correlation between real- and decoded time-series trajectories of motor kinematics (Figure 4–2). The first approach was calculating cross-correlation (Figure 4–2A). The result shows that the correlation peaks at the 0–lag point in every subject. Then, Pearson's

correlation was calculated between real kinematic parameters and decoded ones (Figure 4–2B and C) to evaluate the performances of DNN models. The mean correlations indicated that every motor kinematic parameter was decoded with a significantly high correlation (r > .811, p < .001). Velocity (r = .858, SEM = .026) showed the highest mean value of correlation, followed by position (r = .854, SEM = .024) and acceleration (r = .811, SEM = .025). Nevertheless, their differences were insignificant (F(2, 24) = 1.385, p = 0.270).

4.3.2. Trajectories of Motor kinematics Predicted by DNN models.

Figure 4–3 shows the predicted trajectories of acceleration and velocity, and the predicted trajectories of hand positions are presented in Figure 4–4. In both results, the predicted trajectories of motor kinematics seem to follow the real trajectories. Furthermore, in addition to the position directly predicted by DNN models, both the positions obtained by integrating the predicted velocity and acceleration were recovered. Figure 4–5 illustrates the predicted trajectories of the hand–reaching movement of individual subjects in 3D space. Overall results presented that the DNN models could decode the motor kinematics of hand–reaching movement from the cortical source activities.



Figure 4–3. The 2D plots of real and predicted hand-reaching acceleration (A) and velocity (B). The black solid lines represent the real trajectories of motor kinematics and the dotted color lines are predicted ones.

4.4. Discussion

4.4.1. Decoding time-series motor kinematics via DNN Model

This study demonstrates that DNN models can significantly decode the timeseries motor kinematics of hand-reaching movement from the cortical source signals. Furthermore, I could decode the various motor kinematics of handreaching movements, such as acceleration, velocity, and position, from the same cortical source signal. This result means that the DNN models significantly learned the time-series features of neural signals to process acceleration, velocity, and position. Furthermore, this result implies that the spatiotemporal features of brain signals might exist according to motor kinematics. The functional brain areas for processing motor components will be identified if these features can be visualized.



Figure 4–4. Normalized trajectories of hand-reaching position recovered from decoded kinematic parameters per axis and directions. I calculated position from the decoded acceleration and velocity by integrating them.

4.4.2. RNN Architecture for Decoding Time-series Neural Signals

Since Krizhevsky, Sutskever, and Hinton (2017) employed a deep learning architecture, the convolutional neural network (CNN) is the most frequently used. CNN architecture shows strong performance for image processing. Implementation of DNN models for BCI has also been developed based on CNN. For example, EEGNet (Lawhern et al., 2018) has shown that CNN– based architecture could decode various behavior and cognitive states from EEG features such as P300, the error–related negativity (ERN), and sensory– motor rhythm (SMR). In addition, this model showed better accuracy for the BCI tasks than the conventional ML models.



Figure 4–5. Recovered trajectories from motor kinematic parameters. The first on the left side shows real trajectories. Others are decoded. Decoded trajectories were calculated from decoded kinematic parameters by integrating them. For example, acceleration kinematics was decoded by Deep Neural Network (DNN) and then integrated into position coordination. White dots in the center of plots were initiation–point of movement.

I also considered CNN a candidate model but determined that the RNN architecture was more suitable for processing brain signals since characteristics of brain signals include spatial (brain areas) and temporal features simultaneously. However, the proposed CNN models for BCI are implemented based on the image processing model. Therefore, such models can process the spatial features of brain data but are limited in processing temporal features.

On the other hand, RNN has an advantage in processing long-term

temporal features (Petneházi, 2018; Schuster & Paliwal, 1997). However, processing long-term temporal features can cause a problem of long-term dependencies (Y. Bengio, Simard, & Frasconi, 1994), in which the early information in the hidden layer vanishes during propagation. To address the problem, I employ LSTM and bidirectional RNN architecture. This architecture helps the neural information of previous time points propagate toward the latest time points. Furthermore, such a design allows the DNN to decode time-series of motor kinematics from the complex brain signals.

4.4.3. Limitation

Although the DNN models learn the neural features for processing the motor kinematics, they are hard to identify since the model comprises multiple hidden layers and millions of parameters. Such limitation is one of the challenges in employing the DNN model for investigating the relationship between input and output data. In the next chapter, I will try to identify the neural features by decomposing the complex DNN models with explainable artificial intelligence (XAI) technique.

Chapter 5. Identification of Cerebral Cortices for Processing Motor Kinematics Using DNN and Explainable Artificial Intelligence

5.1. Chapter Introduction

The results of Chapter 4 suggest that the DNN model learns the relationship between cortical source activity and motor kinematics, such as acceleration, velocity, and position since the decoding performance was significant and high. Therefore, I speculated that analyzing the decoding features hidden inside the DNN models allows me to identify the cortical areas processing each motor kinematics. Since the decoding performance on each motor kinematics does not differ, I hypothesized that the cortical representation for the acceleration, velocity and position might differ.

However, one of the most noticeable shortcomings of the DNN model is that the model works as a 'black box'. The inner structure of the model is composed of complex connections between multiple layers consisting of millions of parameters. Although decoding performance is excellent, it is hard to identify neural features that the DNN model uses. To open the 'black box', explainable AI (XAI), I employed another state-of-art method in this study to evaluate the cortical area's contribution inside the DNN model for decoding each kinematic parameter. This method can evaluate the contribution of input neural data when DNN models decode an output (Hoffman, Mueller, Klein, & Litman, 2018). These approaches allowed us to construct a high-performing decoding model to decode each kinematic parameter and to identify large-scale cortical representation based on the cortical area's contributions inside the high-performing model for each specific kinematic parameter.

5.2. Materials and Methods

5.2.1. Data information

Here, I decomposed pre-trained DNN models introduced in Chapter 4. Those models decoded time-series motor kinematic parameters from the cortical source signals. In the decoding, I employed twenty-five models (five-fold cross-validation X five-ensemble models).

5.2.3. Explainable Artificial Intelligence (XAI)

Although a DNN model showed a powerful decoding performance, it has a shortcoming in explaining the model itself. The DNN model consisted of millions of trainable parameters (known as 'neurons'). Furthermore, the complex interactions between stacks of layers within the DNN models made it hard to explain their mechanism. For this shortcoming, the XAI technique is currently drawing attention. (Ancona, Ceolini, Öztireli, & Gross, 2017; Hoffman et al., 2018). XAI evaluates the contribution of each input data characteristic in computing the model's prediction. In the present study, I decomposed the DNN models for decoding the motor kinematics through the XAI technique. Thus, I could map the contribution of each cortical area for decoding each kinematic parameter.

I employed an axiomatic attribution method named 'Integrated Gradients' (IG) (Sundararajan, Taly, & Yan, 2017). Figure 1–5 illustrates an intuitive example of which the IG calculated the attribution of the 'giant panda' image from the neural network image classifier model. IG follows two axioms called 'Sensitivity' and 'Implementation invariance'. If the input data (x) and their baseline (\bar{x}) had a single different feature, but different outputs came out, the feature should have a non-zero attribution. This axiom is sensitivity. The attribution might be distorted because the gradient-based method (Ancona et al., 2017) could violate this axiom. Methods such as DeepLift and LRP can fulfill this axiom. They try to fulfill the axiom by employing a baseline.

Meanwhile, when two or more models predict the same outputs from the same input data concerning the model's structures, the attribution of the data should be the same. This axiom is the implementation invariance. Unfortunately, although the gradient-based methods could fulfill it, DeepLift and LRP violate this axiom.

IG was designed to follow both axioms. The IG could be calculated through equation (6). A fundamental concept of IG is to calculate the difference between the input and the baseline by integrating the parameter. Here, the F(x) is the DNN model.

(6) IG(
$$x_i$$
) = ($x_i - \bar{x}_i$) $\int_0^1 \frac{\partial}{\partial x_i} F(ax_i + (1-a)\bar{x}_i) da$

The equation adjusted the proportion of input data and the baseline by constant 'a'. This constant is an interpolation constant that confounds the feature of input data. The equation is transformed into equation (7) after applying Reimann integration for implementation. If a partition value is large enough, the calculation output is approximated to integration.

(7) IG^{approx}(x_i) = (x_i -
$$\bar{x}_i$$
) * $\sum_{k=1}^{a} \frac{\partial}{\partial x_i} F\left(\bar{x}_i + \frac{k}{a}(x_i - \bar{x}_i)\right) \frac{1}{a}$

Here, I set the number of steps in the Riemann approximation (*a*) to 100, which should be enough to approximate equation (5). In addition, the baseline (\bar{x}_i) was set to be zero-matrix, which could represent a non-movement state. Through this analysis, I could obtain twenty-five contribution matrices per subject.

5.2.3. Statistical analysis

The contribution matrices were expressed as [240 trials X 445 areas X 141 times] for every subject. These matrices were generalized with the following procedures. First, the trial dimension was averaged. Time-course contributions during reaching movement (0 ms onset from the visual cue to the offset of movement) were then averaged to reconstruct a representative contribution of each area. Representative contributions of 445 cortical areas were obtained per subject. Then, I tested the significance of the contribution of 445 cortical areas to decode each kinematic parameter. A one-sample t-test was conducted for each kinematic parameter. The comparison baseline was set to be the mean value of the contribution. Areas with significantly higher contributions than the mean (p < .05 level) were left. As a result, large-scale cortical maps were made according to each kinematic parameter.

Cortical areas within each map were categorized into two groups: shared areas and dominant areas. The notation 'shared' meant that the intersecting set of areas significantly contributed to all kinematic parameters. On the other hand, the notation 'dominant' indicated that the set of areas significantly contributed to specific kinematic parameters. The dominant areas were different sets of areas for the other two. These areas were marked on the FreeSurfer inflated surface for visualization (Figures 5–1 to 5–4).

5.3. Results

5.3.1. Shared areas

I could identify those 35 labels that significantly contributed to predicting all motor kinematics (p < .05) (Figure 5–1). These shared areas mainly consisted of the inferior parietal lobule (IPL), such as bilateral supramarginal gyrus (SMG, BA40), Brodmann area 5 (BA5), parts of left angular gyrus (AG, BA39). In addition, adjacent areas in the temporal lobe and occipital cortex, such as left occipital association areas (BA19), including V3A and right V2 area (BA18), were also included in the shared areas.

5.3.2. Acceleration-dominant areas

I could identify 40 labels in the contralateral motor cortex and ipsilateral visual areas (Figure 5–2), significantly contributing to acceleration decoding. Such areas included the contralateral arm and hand regions of M1 (BA4) and S1(BA2, BA3), a supplementary motor area (SMA, within BA6), and an inferior part of DLPFC (BA46). In addition, the ipsilateral visual association area (BA19) contributed to acceleration decoding.

5.3.3. Velocity-dominant areas

Twenty-five labels contributed only to velocity decoding. These areas were in the contralateral parieto-frontal reaching network, including the dorsal part of the premotor cortex (PMd within BA6), parts of BA5, the medial part of IPS (mIPS), an anterior part of DLPFC (BA10), and a part of bilateral AG.



Figure 5–1. Shared areas. Color-marked (green) areas are shared areas that significantly contribute to every kinematic parameter. Because the significance is different across the parameters, the level of significance is not presented. Lines on the cortical source surface is borders that separates Brodmann's area (Van Essen, 2005).

Like acceleration areas, some parts of the right BA18 within the visual areas were also significant (Figure 5–3). Ventral and posterior parts of the bilateral superior temporal sulcus (STS) also significantly contributed to velocity decoding.

5.3.4. Position dominant areas



Figure 5–2. Acceleration dominant areas that contribute to decoding for reaching acceleration only. Color-marked areas significantly contributed to decoding for reaching acceleration from source activity. Computed significance was plotted on the Freesurfer average source space. At the bottom of the figure, brightness means the significance level. Source space is displayed in split view. Columns correspond to each hemisphere. Frontal, lateral, parietal, and dorsal views of the brain are displayed in each row. Lines on the cortical source space is borders that separates Brodmann's area (Van Essen, 2005).

Forty-seven areas for position decoding were identified in the bilateral visual area for spatial processing and the ipsilateral motor cortex for saccadic eyemovement control. Figure 5–4 illustrates the position-dominant areas. These areas consisted of parietal to occipital areas such as the lateral IPS (LIP), a part of the superior parietal lobule (BA7), and contralateral superior parieto-



Figure 5–3. Velocity dominant areas that contribute to decoding for reaching velocity only. Color–marked areas significantly contributed to decoding for reaching velocity from source activity. Computed significance was plotted on the Freesurfer average source space. Columns correspond to each hemisphere. Frontal, lateral, parietal, and dorsal views of the brain are displayed in each row. Lines on the cortical source space is borders that separates Brodmann's area (Van Essen, 2005).

occipital cortex (SPOC) known as the junction of SPL and lateral parietooccipital sulcus. In addition, I found the bilateral extra-striate cortex around the V3–V3A complex in BA19 only for position decoding. Furthermore, the ipsilateral motor cortex, such as the face region of M1 and S1, frontal eye field (FEF, BA8), superior frontal regions with supplementary eye field (SEF), and the posterior part of DLPFC (BA9) also significantly contributed to



Figure 5–4. Position dominant areas that contribute to decoding for reaching position only. Color-marked areas are significant parcellations that contribute to decoding for reaching position from source activity. Computed significance was plotted on the Freesurfer average source space. Columns correspond to each hemisphere. Frontal, lateral, parietal, and dorsal views of the brain are displayed in each row. Lines on the cortical source space is borders that separates Brodmann's area (Van Essen, 2005).

position decoding.

5.4. Discussion

5.4.1. Mapping cortical areas for movement kinematics with DNN and XAI technique

Through a DNN model and XAI, I could uncover the specific cortical

representation for each kinematic parameter, such as acceleration, velocity, and position. For the DNN models, I then used XAI to extract cortical representation embedded in the DNN model. Therefore, I could distinguish cortical maps for acceleration, velocity, and position.

This approach allows us to overcome the limitations of previous studies on single neurons and large-scale non-invasive cortical activity. Studies on single-neuron activity and motor kinematic parameters have found a relationship between motor neuron activity and kinematic parameters using a computational model (Ashe & Georgopoulos, 1994; Moran & Schwartz, 1999; Wang et al., 2007). Although such a method has a fine spatial resolution, the coverage was too focal to map the cortical activity across the whole brain corresponding to each motor kinematic parameter. To overcome this limitation, whole-brain activity was measured during reaching movement with a non-invasive method. DNN models learned relationships between large-scale cortical sources and kinematic parameters. XAI then identifies DNN models' attributions for kinematic parameters for large-scale cortical maps. Although the spatial resolution of this approach is lower than the single-neuron activity, it is enough to segregate large-scale cortical representation for each kinematic parameter.

Another method, the coherence between non-invasive MEG source signals and motor kinematic signals (Bourguignon et al., 2012; Jerbi et al., 2007), has the advantage of measuring the whole brain activity for motor

kinematic parameters. However, it is limited because coherence maps of kinematic parameters are superimposed and hard to distinguish. This might be because the spatial resolution is too low to segregate cortical representations according to each kinematic parameter.

Although a neural network model has a high decoding ability, it is difficult to know what happens inside them since it consists of multiple hidden layers of hundreds of thousands of parameters (here, about 3,000,000 parameters of 8 layers). Currently, there are approaches to explain the inside of a neural network model. One of them is the XAI method. It can map the attribution of models by calculating the contribution of each part of input data for decoding output results (Hoffman et al., 2018). Using such a method, features used by neural network models could be extracted and mapped from the input data. Based on this point, I employed an XAI method for neuroimaging data. One of various XAI methods, IG (Sundararajan et al., 2017), was used in this study. This method extracted cortical areas that contributed to decoding acceleration, velocity, and position from the DNN model. I mapped contributing areas and segregated them into dominant areas to identify cortical representations according to kinematic parameters. In cortical maps, there were also shared areas that appeared for all kinematic parameters. Therefore, it can overcome the limitation of previous large-scale coherence studies, which could not distinguish each kinematic parameter.

A machine learning (ML) model for decoding behavior from non-

invasive neural signals and a method to analyze neural features used by the model can provide more reliable neural features than conventional methodologies such as PCA (Hatamimajoumerd, Talebpour, & Mohsenzadeh, 2020). For example, hand direction (Wang et al., 2010), different hand movements (Belkacem, Nishio, Suzuki, Ishiguro, & Hirata, 2018), or invariant visual object information (Isik, Meyers, Leibo, & Poggio, 2014) could be classified from the MEG signals through LDA and support vector machine (SVM). The neural features used by the ML model reflected robust findings, such as the activity of motor cortices for movement or the role of posterior occipital regions for visual information. However, the conventional ML method also seems insufficient to sensitively identify the cortical representation of various components constituting a behavior. Beyond the conventional ML method above, this study increases the reliability and sensitivity by using the state-of-art and the biologically plausible method called the DNN model. The DNN model can increase the decoding accuracy higher than in a previous study using a conventional time-series ML model (H. G. Yeom et al., 2013). Furthermore, I can extract neural features used by the high-performing models through another state-of-art method, XAI. As a result, I can identify cortical areas for acceleration, velocity, and position from the directional reaching movement.

The results showed that cortical areas for acceleration, velocity, and position within the visuomotor network could be identified. Furthermore,

among these areas, both shared, and dominant characteristics were observed across areas of acceleration, velocity, and position. Thus, I segmented the cortical areas into shared and dominant and discussed them below.

5.4.2. Shared areas representing sensory processing and action goal selection for reaching movement.

Cortical areas of acceleration, velocity, and position could be obtained from DNN models and XAI. Some areas overlapped. We defined these overlapped cortices as shared areas. We speculated that there would be common functions in the neural system for all kinematic parameters. Sensory integration and goal selection functions would be essential to produce common kinematic attributes. First, we considered sensory integration as a common kinematics parameter. For successful reaching movement, our brain has to integrate various kinds of external sensory information from limbs and organs to minimize motor errors (Sober & Sabes, 2005). Goal–selecting behavior may represent a common attribute of all kinematic parameters. Although kinematic parameters have unique behavioral characteristics, their common purpose is to set and reach a goal of the movement.

My DNN models could identify cortical areas according to kinematic parameters and shared functions for them. The present study identified shared areas as SMG, superior parietal lobule, and AG within PPC by DNN models and XAI. PPC converts external sensory information of limbs to an internal

kinematic model and then locates them in the desired location (Buneo & Andersen, 2006). Since external sensory information input for each kinematic parameter is integrated into the internal kinematics model in the neural system, shared areas seem to contribute to all kinematic parameters. A distinct feature of shared areas is that these areas are centered around the bilateral SMG. This region is related not only to somatosensory integration but also to the function of goal selection. Fogassi et al. (2005) have found that neurons in the primate left IPL, which includes SMG and AG, represent action goals. Inducing a virtual lesion on this site by rTMS can cause a delay in goal–oriented action (Tunik, Lo, & Adamovich, 2008). Based on the functions of these areas, we speculate that shared areas might be identified for all kinematic parameters based on these two behaviors.

5.4.3. Dominant areas showing unique characteristics for decoding each kinematic parameter in 3D space.

I identified that each kinematic parameter was identified in the cortices of the motor-related network. In our study, acceleration areas consisted of the contralateral motor cortex, such as the hand and arm regions of M1 and S1, SMA, and a subregion of DLPFC (BA46). This result indicates that these motor cortices can best explain behavioral attributes of acceleration. Contrary to previous studies (Ashe & Georgopoulos, 1994; Bourguignon et al., 2012; Jerbi et al., 2007; Kadmon Harpaz, Ungarish, Hatsopoulos, & Flash, 2018;

Moran & Schwartz, 1999; Wang et al., 2007) that the motor area was related to all kinds of movement kinematic parameters, I identified that these motor– execution areas only represented acceleration. Although findings of the neural relationship between the M1 (or motor region) and movement acceleration have also been presented (Bourguignon et al., 2012; Kadmon Harpaz et al., 2018), they could not address the issue that arose in the overlapped relationship between motor cortices and all kinds of kinematic parameters since they did not include velocity or position. Unlike previous studies, our study showed dominant cortical areas to produce acceleration using DNN models and XAI methods.

In the reaching movement, behavioral attribution of acceleration is closely related to the control of muscle force. The force increases with the initiation of reaching movement from a stationary state. To reach the target, the force decreases for fine adjustment of limb action. The muscle force during a motor execution can be described as acceleration. M1 appears only in acceleration because the region is known to encode muscle force (Evarts, 1968; Kakei et al., 1999). In large–scale studies (Bourguignon et al., 2012; Jerbi et al., 2007), like M1, S1 is also found to be related to several kinematic parameters. I also found that S1 was significantly involved only in acceleration. It might be due to somatosensory processing for muscle control, such as proprioception (Tuthill & Azim, 2018). The interesting point was that different roles depended on the region within the premotor cortex (Brodmann's area 6, BA6), which involved movement preparation. SMA was shown to encode acceleration.

On the other hand, the premotor area (here, PMd) was shown to encode velocity. I noted that SMA was involved in the initiation and execution of movement (Eckert, Peschel, Heinze, & Rotte, 2006; Krainik et al., 2001). Compared to the premotor area, which prepares a visual–guided movement, SMA encodes the initiation and execution of upcoming motor sequences (Roland, Larsen, Lassen, & Skinhoj, 1980; Roland, Skinhoj, Lassen, & Larsen, 1980). Functions of contralateral sensorimotor areas and SMA can explain the behavior of humans to produce acceleration. These areas are presented in the acceleration dominant areas in our DNN models.

Considering a behavioral characteristic of the movement, velocity is presumed to be the most crucial variable for preparing a reaching movement and visuomotor transformation. Participants prepared a reaching movement based on the velocity (or speed) because they intended to reach their hand and drawback to an initial point as fast as possible. Based on the behavior, cortices that accounted for the velocity appeared in regions for visuomotor transformation and motor planning. Areas that account for these functions are mIPS and PMd within the parieto–frontal network, which is one of the fundamental systems for visual–guided reaching movement (Burnod et al., 1999). A parietal part of the network, mIPS, encodes reaching movement from visuospatial information of action space (Vesia & Crawford, 2012). The

encoded information is known to include action intentions (Michael Vesia & Davare, 2011). Considering the importance of mIPS in velocity decoding, it is presumed that the encoded information represents the velocity of movement rather than the position or acceleration. The encoded reaching movement was then transmitted to PMd through the parieto–frontal network. A frontal part, PMd, is a reaching region in the motor preparation area (Churchland, Yu, et al., 2006). For the velocity of movement, the velocity of upcoming reaching movement can be decoded from neural signals of PMd (Churchland, Santhanam, et al., 2006). After motor preparation based on the velocity was transmitted to areas for motor execution, the reaching movement could be executed using muscle control by acceleration.

Cortical areas for the position were also identified from large-scale visuomotor cortices such as SPOC, LIP, FEF, and SEF. These regions are known to play a role in visuospatial processing and eye (or saccadic) movement control (R. Andersen, Bracewell, Barash, Gnadt, & Fogassi, 1990; Purcell, Weigand, & Schall, 2012; Vesia & Crawford, 2012; Vesia, Prime, Yan, Sergio, & Crawford, 2010). Positional information on limb movement is mainly encoded by visual processing. The first behavior participants took was capturing the target based on spatial information. They then established a movement space from the hand to the target. After starting the movement, participants tracked their hands in the space and continuously updated the displacement between the moving hand and the target to reach it accurately. Since saccadic eye movement and a spatial process between a hand and the target in the action space are necessary for an accurate movement, visualmotor areas significantly contribute to position decoding. I found the contribution of a visuospatial area called V3a of visual cortices.

Along with visual space area, SPOC in the occipito-parietal junction (Vesia & Crawford, 2012) corresponds to a parietal reach region (PRR) of primates, which encodes reaching movement (Galletti, Kutz, Gamberini, Breveglieri, & Fattori, 2003; Vesia & Crawford, 2012). Disrupting the function of SPOC by rTMS stimulation can increase positional errors of reaching movement (Vesia et al., 2010). According to the gain fields theory, the receptive fields of this region can be modulated by gaze position, target, and hand position (Chang, Papadimitriou, & Snyder, 2009). PRR has encoded the displacement between the hand and gaze in a primate study (Blohm & Crawford, 2009). For hand position, saccadic eye movement information seems more important than motor information for reaching control. The saccadic eye movement along with the target and the reaching motion is also essential in addition to visuospatial information. Areas such as LIP, SEF, and FEF also appeared to contribute to position decoding. These areas are known to be related to the control of a saccadic eye movement. In the parietal part, LIP is known to encode a saccadic movement during reaching (R. Andersen et al., 1990). In the frontal part, SEF and ipsilateral FEF are known to initiate saccadic movement and oculomotor control (Purcell et al., 2012; Schall,
2004). Since ipsilateral motor-related areas are presented along with FEF, the contribution of these regions may also represent eye-related information. Identifying these areas for the position can also be due to peripersonal space processing between the hand and the target. An ablation study using primates has shown that lesions in FEF can cause inattention to stimuli presented in the space (Berti, Smania, & Allport, 2001). In humans, ipsilateral brain damage can cause neglect in the peripersonal space (Halligan & Marshall, 1991). Visuospatial areas in our results possibly reflect spatial processing between the participant's hand and the target.

Interestingly, the results segregated a specific cortical area into subregions according to kinematic parameters. It is typically found in the intraparietal sulcus, DLPFC, and occipital lobe. As discussed, DNN models used neural information from subregions within the intraparietal sulcus according to the velocity and position. The identified IPS regions matched previous findings (Andersen et al., 1990; Vesia et al., 2010). Within the DLPFC, an anterior part, a posterior part, and a more inferior part than the velocity part contributed to velocity, position, and acceleration, respectively. The well–known functions of DLPFC are attention and decision–making. However, DLPFC seems somewhat involved in movement control (Bourguignon et al., 2012; Jerbi et al., 2007; Ryun et al., 2014). As in the case of the intraparietal region, I speculate that a specific subregion of DLPFC might control corresponding kinematic parameters. To this end, further research is needed to clarify the relationship between the subregion of DLPFC and each kinematic parameter.

In cortical areas for kinematic parameters, functional lateralization across hemispheres is found according to motor-related functions. It seems likely that motor-related functions are processed in the contralateral cortex. In contrast, other functions, such as processes about the body's motion and space and goal-selection, are in bilateral or ipsilateral cortices. Current notions about motor behaviors are that limb is controlled by contralateral motor areas such as M1, PMd, S1, and BA7 (Bourguignon et al., 2012; Churchland, Santhanam, et al., 2006; Jerbi et al., 2007). In the present study, such contralateral motor cortex was valuable in cortices for acceleration and velocity. This is because acceleration is related to muscle control, and velocity is processed through parieto-frontal network for visuomotor transformation. I speculate that motor control or sensory-motor transformation is required for kinematic parameter processing. It may involve the role of contralateral cortices. On the other hand, concerning spatial processing or action goalselection, such lateralization seemed to be weaker. V3 and SMG were found bilaterally. Some areas in the ipsilateral hemisphere were found to be involved in kinematics processing. Such ipsilateral cortices were found in the visual association area of acceleration areas and the primary sensorimotor area within shared areas. Unfortunately, the roles of the ipsilateral hemisphere for visuomotor behavior are unclear. The contribution of ipsilateral areas might

be due to the processing of accelerating motion of biological limbs (Limanowski, Sarasso, & Blankenburg, 2018; Schlack, Krekelberg, & Albright, 2007) or an inter-callosal inhibition of motor cortices to facilitate kinematics control of contralateral limbs (Kobayashi, Hutchinson, Theoret, Schlaug, & Pascual-Leone, 2004). Further research on functional lateralization of the kinematics map would be necessary.

Current non-invasive neuroimaging methods have been based on the simple relationship between neural- and behavioral signals, such as corticalkinematic coherence (Bourguignon et al., 2012; Jerbi et al., 2007) or general linear model (Yeom, Kim & Chung., 2013). However, sensitivity is limited to segregating cortical representation according to the various cognitive components hidden in a behavior. Here, a state-of-art methodological scheme is presented to complement the limitation. With the DNN model and XAI, it is possible to sensitively map the kinematic representation of cerebral cortices hidden inside the biologically plausible neural model. Beyond the representation of motor components, I hope to apply this methodological scheme in exploring neural representations of various human behavior and cognition, such as semantic representation according to words, the neural representation of visual objects, or kinetics of movement.

5.4.4. Limitations

The reliability of the maps, which XAI could extract, depends on the

 $6\ 2$

performance accuracy of neural network models (Lawhern et al., 2018). If there is a model that can perfectly decode kinematic parameters from neural data, cortical areas for these parameters will be identified more precisely. However, our model did not reach that level. Nevertheless, since our DNN models significantly decoded all kinematic parameters above r = .80 (p $\langle .001 \rangle$), I speculate that this performance is the current state–of–art. Thus, I consider those cortical areas reliable enough to understand the kinematic parameters. In the future, a more biologically plausible and more accurate neural network model may produce a more reliable and sophisticated cortical contribution map.

Currently, many kinds of XAI methods are being implemented. Thus, results may vary depending on the XAI method (Ancona et al., 2017). In the current study, I implemented a method called IG. The method is known to effectively control issues derived from two axioms that could not be fulfilled by other previous methods (Sundararajan et al., 2017). I expect that more sophisticated imaging will be possible by developing XAI methodologies in the future.

PART III. General Conclusion and Perspectives

Chapter 6. General Conclusion and Perspectives

6.1 General Conclusion

Throughout studies, cortical areas for motor kinematics, such as acceleration, velocity, and position, were identified using DNN models and explainable AI. The cortical representation of the kinematic process was identified in two folds: shared and dominant areas. The findings regarding shared areas imply that multisensory integration and motor goal selection are crucial for producing a reaching movement. On the other hand, dominant areas represent unique characteristics of acceleration, velocity, and position in our brains. Each characteristic was presented as a cortical map which appeared according to a cortical characteristic for visuomotor functions. These findings suggest that our brain can process movement by decomposing it into various kinematic components, like how we describe the physical world.

6.2. Perspectives

Throughout the studies, I showed that the deep neural network (DNN) could

decode various motor kinematic parameters. In addition, explainable AI (XAI) enables me to decompose complex neural network models and to identify the cortical areas for each motor kinematic parameter. Such findings suggested that the cortical processing of motor kinematics appears differently depending on its behavioral characteristics. Along with the scientific implication, the findings also had practical implications in developing a practical motor BCI system to control a robotic limb.

The results imply two points. The first is that the human brain may process motor behavior by dispersing it into various motor variables. Second is that various motor variables can be acquired from various cortical areas to develop efficient motor BCI (e.g., M1–acceleration, IPS– velocity, and visual cortex–position). However, the spatial coverage in measuring brain signals might sometimes be limited. For example, invasive methods, such as electrocorticography (ECoG), ensure a high signal–to–noise ratio (SNR) (Lawhern et al., 2018; Waldert et al., 2008), but the spatial coverage is limited to several brain areas. Due to this limitation, researchers might have to acquire neural data from 'related areas', such as PPC (Vesia & Crowford, 2012), not the primary area like M1.

Furthermore, considering those findings, neural features contributing to decoding specific information might not exist even in the related areas. For example, since the IPS contributes to processing hand-reaching velocity and not to acceleration, the neural signals of IPS alone have a limit to predicting



Figure 6–1. Signal translation using MelGAN model. (A) Averaged signals (evoked response) of each subject. The shaded areas represent the standard deviation. The third and fourth rows present the generated signals. The third row (M1(IPS)) is the IPS– translated M1 signals, and the fourth row (M1(Noise)) is the nosie–generated M1 signals. (B) Cross–correlation between signals. The maker on the line represents the maximal point of the correlation. (C) Pearson correlation between M1 and the other signals (IPS, IPS–translate M1, and noise–generated M1). The asterisk mark over and under the bar represents significancy. (D) Mutual information. (E) Frechet Inception Distance (FID) score. For the visibility, I reverse the score to FID⁻¹.

the acceleration of the hand. In this case, the neural signals of M1 should be mandatory. However, in many cases, covering both areas using invasive methods could be challenging.

Generating the artificial data could address the limitation by augmenting spatial coverage. This approach also allows investigators to infer and utilize unobserved data from observed data. The most promising model structure is the generative adversarial neural network (GAN) proposed by Goodfellow et al. (2020). GAN comprises two networks: the generator and the discriminator. The generator synthesis artificial data and the discriminator identifies the authenticity of the data (fake or real). Repeating the competition between the generator and the discriminator could refine the artificial data like real data. Furthermore, current research demonstrated that the data 'style' could be translated into another data style by GAN. For example, GAN translates horses in an image into a zebra (Zhu, Park, Isola, & Efros, 2017), or a man's voice into a female (Yamamoto, Song, & Kim, 2020). Although several works tried to generate artificial brain signals, mostly EEG (Roy, Dora, McCreadie, and Prasad, 2020; Hartmann, Schirrmeister, and Ball, 2018; Kwon & Im, 2022), those studies employ image generation GANs, not the signal. Unlike the image, neural signals comprised spectrotemporal features. Thus, a model that generates the signal based on those features could be a more suitable approach than employing an image generator.

Thus, further works should aim to generating artificial brain signals for motor behavior from motor-related areas. I can provide a proof of concept for the idea and aim. I generate brain signals of M1 using a generative



Figure 6–2. Decoding real and generated brain signals through deep neural network (DNN) model. The architecture of DNN model is bidirectional Long-short term memory (bLSTM). (A) Individual subject's average decoding accuracy for five-fold cross-validation. (B) Decoded trajectories of hand-reaching acceleration of Subject 3. Dashed black lines are real trajectory, and colored lines are decoded trajectories.

adversarial neural network (GAN) designed for signal generation called MelGAN (Kumar et al., 2019).

I find that the model can learned the spectrotemporal features of M1 signals. At first, the signal waveforms of IPS and M1 appeared differently. However, when the model translated the IPS signal waveforms into M1, the waveforms changed similar to the M1's signal waveforms (Figure 7–1A). Moreover, measuring the similarity quantitively shows the signal waveforms of M1 and translated signals were linearly similar (Figure 7–1B and C). In addition, the mutual dependencies appeared higher between the brain signals (M1–IPS and M1–IPS–translated M1) than the noise (Figure 7–1D and E). Furthermore, I also find that the acceleration trajectories of the hand–reaching

movement can be decoded from the IPS-translated M1 signals (Figure 7-2).

Such a results illustrates the signal generation model can learn the spectrotemporal features of motor cortex and infer the signals from the motor-related area. Moreover, although the translated signals were the artificially generated ones, those signals can work as brain signals for motor BCI.

I know that the current the proof of concept had many limitations. For example, since the MelGAN model only can generate a single channel, the complex feature presented across the cerebral cortex are hard to generate. In addition, the dataset seemed small (240 samples per subject), so there might be a possibility of overfitting issue. Although I reduced the model and trainable parameters (4.6M to 1.23M) and stopped the training iterations to prevent overfitting, the fundamental approach to address current limitation should be increasing data samples. I speculate that designing GAN for multi– channel signal generation can address the one of limitations. In addition, the framework for addressing small amount of data, such as transfer learning or knowledge distillation (Hinton, Vinyals & Dean, 2015), may work. The future work aims to address those issue and testing the proof of concept to invasive data.

Bibliography

- Altaheri, H., Muhammad, G., Alsulaiman, M., Amin, S. U., Altuwaijri, G. A., Abdul, W., . . . Faisal, M. (2021). Deep learning techniqu es for classification of electroencephalogram (EEG) motor imager y (MI) signals: A review. *Neural Computing and Applications*, 1 –42.
- Ancona, M., Ceolini, E., Öztireli, C., & Gross, M. (2017). Towards bett er understanding of gradient-based attribution methods for deep neural networks. *arXiv preprint arXiv:1711.06104*.
- Andersen, R., Bracewell, R., Barash, S., Gnadt, J., & Fogassi, L. (1990).
 Eye position effects on visual, memory, and saccade-related activity in areas LIP and 7a of macaque. *The Journal of Neuroscience*, 10(4), 1176–1196. doi:10.1523/jneurosci.10–04–01176.1990
- Andersen, R. A. (1995). Encoding of Intention and Spatial Location in t he Posterior Parietal Cortex. Cerebral Cortex, 5(5), 457–469. http s://doi.org/10.1093/cercor/5.5.457
- Andersen, R. A., Aflalo, T., & Kellis, S. (2019). From thought to action
 : The brain-machine interface in posterior parietal cortex. *Proc Natl Acad Sci U S A*. doi:10.1073/pnas.1902276116
- Ashe, J., & Georgopoulos, A. P. (1994). Movement parameters and neur al activity in motor cortex and area 5. *Cereb Cortex, 4*(6), 590– 600. doi:10.1093/cercor/4.6.590
- Bear, M., Connors, B., & Paradiso, M. A. (2020). *Neuroscience: explori ng the brain, enhanced edition: exploring the brain*: Jones & Bar tlett Learning.
- Belkacem, A. N., Nishio, S., Suzuki, T., Ishiguro, H., & Hirata, M. (20 18). Neuromagnetic Decoding of Simultaneous Bilateral Hand Mo vements for Multidimensional Brain Machine Interfaces. *IEEE Tr ansactions on Neural Systems and Rehabilitation Engineering, 26*(6), 1301–1310. doi:10.1109/TNSRE.2018.2837003
- Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern anal* ysis and machine intelligence, 35(8), 1798–1828.
- Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term depen dencies with gradient descent is difficult. *IEEE transactions on ne*

ural networks / a publication of the IEEE Neural Networks Cou ncil, 5, 157–166. doi:10.1109/72.279181

- Berti, A., Smania, N., & Allport, A. (2001). Coding of far and near spa ce in neglect patients. *Neuroimage, 14*(1), S98-S102.
- Blohm, G., & Crawford, J. D. (2009). Fields of Gain in the Brain. *Neur* on, 64(5), 598-600. doi:https://doi.org/10.1016/j.neuron.2009.11. 022
- Bourguignon, M., Jousmäki, V., Dalal, S. S., Jerbi, K., & De Tiège, X. (2019). Coupling between human brain activity and body moveme nts: Insights from non-invasive electromagnetic recordings. *Neuro image, 203*, 116177. doi:10.1016/j.neuroimage.2019.116177
- Bourguignon, M., Jousmaki, V., Op de Beeck, M., Van Bogaert, P., Gol dman, S., & De Tiege, X. (2012). Neuronal network coherent wi th hand kinematics during fast repetitive hand movements. *Neuro image*, *59*(2), 1684–1691. doi:10.1016/j.neuroimage.2011.09.022
- Bradberry, T. J., Gentili, R. J., & Contreras-Vidal, J. L. (2009, 3-6 Sep t. 2009). Decoding three-dimensional hand kinematics from electr oencephalographic signals. Paper presented at the 2009 Annual In ternational Conference of the IEEE Engineering in Medicine and Biology Society.
- Bundy, D. T., Pahwa, M., Szrama, N., & Leuthardt, E. C. (2016). Deco ding three-dimensional reaching movements using electrocorticogr aphic signals in humans. *Journal of Neural Engineering*, 13(2), 0 26021.
- Buneo, C. A., & Andersen, R. A. (2006). The posterior parietal cortex: sensorimotor interface for the planning and online control of visu ally guided movements. *Neuropsychologia*, *44*(13), 2594–2606.
- Burnod, Y., Baraduc, P., Battaglia-Mayer, A., Guigon, E., Koechlin, E.,
 Ferraina, S., . . . Caminiti, R. (1999). Parieto-frontal coding of r
 eaching: an integrated framework. *Exp Brain Res, 129*(3), 325-3
 46. doi:10.1007/s002210050902
- Chang, S. W. C., Papadimitriou, C., & Snyder, L. H. (2009). Using a C ompound Gain Field to Compute a Reach Plan. *Neuron, 64*(5), 744–755. doi:https://doi.org/10.1016/j.neuron.2009.11.005
- Churchland, M. M., Santhanam, G., & Shenoy, K. V. (2006). Preparato ry activity in premotor and motor cortex reflects the speed of th e upcoming reach. *J Neurophysiol*, *96*(6), 3130–3146. doi:10.115

2/jn.00307.2006

- Churchland, M. M., Yu, B. M., Ryu, S. I., Santhanam, G., & Shenoy, K. V. (2006). Neural variability in premotor cortex provides a si gnature of motor preparation. *J Neurosci, 26*(14), 3697–3712. do i:10.1523/jneurosci.3762–05.2006
- Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage*, 9(2), 179–194. doi:10.1006/nimg.1998.0395
- Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., & Halgren, E. (2000). Dynamic statistical paramet ric mapping: combining fMRI and MEG for high-resolution ima ging of cortical activity. *Neuron, 26*(1), 55–67. doi:10.1016/s089 6–6273(00)81138–1
- Deng, L., & Yu, D. (2014). Deep learning: methods and applications. *F* oundations and trends® in signal processing, 7(3-4), 197-387.
- Donahue, C., McAuley, J., & Puckette, M. (2018). Adversarial audio sy nthesis. *arXiv preprint arXiv:1802.04208*.
- Eckert, T., Peschel, T., Heinze, H. J., & Rotte, M. (2006). Increased pre -SMA activation in early PD patients during simple self-initiated hand movements. *J Neurol, 253*(2), 199-207. doi:10.1007/s00415 -005-0956-z
- Evarts, E. V. (1968). Relation of pyramidal tract activity to force exerte d during voluntary movement. *J Neurophysiol, 31*(1), 14–27. doi: 10.1152/jn.1968.31.1.14
- Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate s ystem. *Neuroimage*, 9(2), 195–207. doi:10.1006/nimg.1998.0396
- Fischl, B., Sereno, M. I., Tootell, R. B., & Dale, A. M. (1999). High-re solution intersubject averaging and a coordinate system for the c ortical surface. *Hum Brain Mapp*, 8(4), 272–284. doi:10.1002/(si ci)1097-0193(1999)8:4<272::aid-hbm10>3.0.co;2-4
- Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolat ti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science*, 308(5722), 662–667. doi:10.1126/science. 1106138
- Galletti, C., Kutz, D. F., Gamberini, M., Breveglieri, R., & Fattori, P. (2 003). Role of the medial parieto-occipital cortex in the control o

f reaching and grasping movements. *Exp Brain Res, 153*(2), 158-170. doi:10.1007/s00221-003-1589-z

- Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neurona l population coding of movement direction. *Science, 233*(4771), 1 416–1419. doi:10.1126/science.3749885
- Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell , J., Yacoub, E., . . . Jenkinson, M. (2016). A multi-modal parce llation of human cerebral cortex. *Nature*, 536(7615), 171–178.
- Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. Paper presented at the Procee dings of the Thirteenth International Conference on Artificial Inte lligence and Statistics, Proceedings of Machine Learning Research. https://proceedings.mlr.press/v9/glorot10a.html
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*: MIT press.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., . . . Bengio, Y. (2020). Generative adversarial network s. *Communications of the ACM, 63*(11), 139-144.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., . . . Hämäläinen, M. (2013). MEG and EEG data analysis with MNE–Python. *Front Neurosci*, *7*, 267. doi:10.3389/ fnins.2013.00267
- Griffin, D., & Lim, J. (1984). Signal estimation from modified short-tim e Fourier transform. *IEEE Transactions on acoustics, speech, and signal processing, 32*(2), 236-243.
- Hagen, E., Næss, S., Ness, T. V., & Einevoll, G. T. (2018). Multimodal Modeling of Neural Network Activity: Computing LFP, ECoG, EEG, and MEG Signals With LFPy 2.0. *Frontiers in Neuroinform atics*, 12. doi:10.3389/fninf.2018.00092
- Halligan, P. W., & Marshall, J. C. (1991). Left neglect for near but not far space in man. *Nature, 350*(6318), 498–500. doi:10.1038/3504 98a0
- Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa,
 O. V. (1993). Magnetoencephalography---theory, instrumentatio
 n, and applications to noninvasive studies of the working human
 brain. *Reviews of Modern Physics*, 65(2), 413-497. doi:10.1103/
 RevModPhys.65.413

- Hartmann, K. G., Schirrmeister, R. T., & Ball, T. (2018). EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals. *arXiv preprint arXiv:1806.01875*.
- Hatamimajoumerd, E., Talebpour, A., & Mohsenzadeh, Y. (2020). Enha ncing multivariate pattern analysis for magnetoencephalography th rough relevant sensor selection. *International Journal of Imaging Systems and Technology, 30*(2), 473–494. doi:https://doi.org/10.1 002/ima.22398
- Hinton, G., & Sejnowski, T. J. (1999). Unsupervised learning: foundation ns of neural computation: MIT press.
- Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Ne ural computation*, *9*(8), 1735–1780.
- Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). Metrics for explainable AI: Challenges and prospects. arXiv preprint arXi v:1812.04608.
- Humphrey, D. R., Schmidt, E., & Thompson, W. (1970). Predicting mea sures of motor performance from multiple cortical spike trains. S cience, 170(3959), 758–762.
- Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in neural information proc essing systems, 30.
- Isik, L., Meyers, E. M., Leibo, J. Z., & Poggio, T. (2014). The dynamic s of invariant object recognition in the human visual system. *Jour nal of Neurophysiology*, 111(1), 91–102. doi:10.1152/jn.00394.20 13
- Jang, S. J., Yang, Y. J., Ryun, S., Kim, J. S., Chung, C. K., & Jeong, J. (2022). Decoding trajectories of imagined hand movement using e lectrocorticograms for brain-machine interface. *Journal of Neura 1 Engineering*, 19(5), 056011.
- Jerbi, K., Lachaux, J. P., N'Diaye, K., Pantazis, D., Leahy, R. M., Garne ro, L., & Baillet, S. (2007). Coherent neural representation of ha nd speed in humans revealed by MEG imaging. *Proc Natl Acad Sci U S A, 104*(18), 7676–7681. doi:10.1073/pnas.0609632104

Kadmon Harpaz, N., Ungarish, D., Hatsopoulos, N. G., & Flash, T. (20

18). Movement Decomposition in the Primary Motor Cortex. *Cer ebral Cortex, 29*(4), 1619–1633. doi:10.1093/cercor/bhy060

- Kakei, S., Hoffman, D. S., & Strick, P. L. (1999). Muscle and movemen t representations in the primary motor cortex. *Science*, *285*(5436), 2136–2139. doi:10.1126/science.285.5436.2136
- Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S., Hudspeth,A. J., & Mack, S. (2000). *Principles of neural science* (Vol. 4): McGraw-hill New York.
- Kar, K., Kubilius, J., Schmidt, K., Issa, E. B., & DiCarlo, J. J. (2019). E vidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature Neuroscien ce, 22*(6), 974–983.
- Kerous, B., Skola, F., & Liarokapis, F. (2018). EEG-based BCI and vide o games: a progress report. *Virtual Reality, 22*, 119-135.
- Kim, H., Kim, J. S., & Chung, C. K. (2023). Identification of cerebral c ortices processing acceleration, velocity, and position during direct ional reaching movement with deep neural network and explaina ble AI. *Neuroimage, 266*, 119783.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimiz ation. *arXiv preprint arXiv:1412.6980*.
- Kobayashi, M., Hutchinson, S., Theoret, H., Schlaug, G., & Pascual-Le one, A. (2004). Repetitive TMS of the motor cortex improves ips ilateral sequential simple finger movements. *Neurology*, *62*(1), 91 -98.
- Kolb, B., & Whishaw, I. Q. (2009). *Fundamentals of human neuropsych ology*: Macmillan.
- Kolb, B., Whishaw, I. Q., Teskey, G. C., Whishaw, I. Q., & Teskey, G.C. (2001). An introduction to brain and behavior: Worth New Y ork.
- Krainik, A., Lehéricy, S., Duffau, H., Vlaicu, M., Poupon, F., Capelle, L ., . . Marsault, C. (2001). Role of the supplementary motor are a in motor deficit following medial frontal lobe surgery. *Neurolo* gy, 57(5), 871–878. doi:10.1212/wnl.57.5.871
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classifica tion with deep convolutional neural networks. *Communications o f the ACM*, 60(6), 84–90.
- Kumar, K., Kumar, R., De Boissiere, T., Gestin, L., Teoh, W. Z., Sotelo,

J., . . . Courville, A. C. (2019). Melgan: Generative adversarial networks for conditional waveform synthesis. *Advances in neural information processing systems, 32*.

- Kwon, J., & Im, C.-H. (2022). Novel Signal-to-Signal translation meth od based on StarGAN to generate artificial EEG for SSVEP-base d brain-computer interfaces. *Expert Systems with Applications, 2* 03, 117574. doi:https://doi.org/10.1016/j.eswa.2022.117574
- Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2018). EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. *J Neur al Eng*, 15(5), 056013. doi:10.1088/1741-2552/aace8c
- Li, M., Zhang, T., Chen, Y., & Smola, A. J. (2014). Efficient mini-batc h training for stochastic optimization. Paper presented at the Proc eedings of the 20th ACM SIGKDD international conference on K nowledge discovery and data mining, New York, New York, US A. https://doi.org/10.1145/2623330.2623612
- Limanowski, J., Sarasso, P., & Blankenburg, F. (2018). Different respons es of the right superior temporal sulcus to visual movement feedb ack during self-generated vs. externally generated hand movemen ts. *Eur J Neurosci, 47*(4), 314-320. doi:10.1111/ejn.13824
- Ma, T., Chen, W., Li, X., Xia, Y., Zhu, X., & He, S. (2021). fNIRS si gnal classification based on deep learning in rock-paper-scissors imagery task. *Applied Sciences, 11*(11), 4922.
- Moran, D. W., & Schwartz, A. B. (1999). Motor cortical representation of speed and direction during reaching. *J Neurophysiol, 82*(5), 26 76–2692. doi:10.1152/jn.1999.82.5.2676
- Muller-Putz, G. R., & Pfurtscheller, G. (2007). Control of an electrical prosthesis with an SSVEP-based BCI. *IEEE transactions on biome dical engineering*, *55*(1), 361–364.
- Penfield, W., & Boldrey, E. (1937). Somatic motor and sensory represent ation in the cerebral cortex of man as studied by electrical stimul ation. *Brain*, 60(4), 389-443.
- Petneházi, G. (2018). Recurrent Neural Networks for Time Series Foreca sting.
- Purcell, B. A., Weigand, P. K., & Schall, J. D. (2012). Supplementary Ey e Field during Visual Search: Salience, Cognitive Control, and Per formance Monitoring. *The Journal of Neuroscience*, 32(30), 1027

3-10285. doi:10.1523/jneurosci.6386-11.2012

- Purves, D., Augustine, G., Fitzpatrick, D., Hall, W., LaMantia, A., & W hite, L. (2012). Neuroscience, {Fifth} {Edition}. In: Sinauer Associat es, Inc., Sunderland, Massachusetts USA.
- Ray, S., Crone, N. E., Niebur, E., Franaszczuk, P. J., & Hsiao, S. S. (2 008). Neural correlates of high-gamma oscillations (60-200 Hz) in macaque local field potentials and their potential implications i n electrocorticography. *Journal of Neuroscience, 28*(45), 11526-1 1536
- Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., C hristensen, A., . . . Kording, K. P. (2019). A deep learning frame work for neuroscience. *Nat Neurosci, 22*(11), 1761–1770. doi:10. 1038/s41593-019-0520-2
- Roland, P. E., Larsen, B., Lassen, N. A., & Skinhoj, E. (1980). Supplem entary motor area and other cortical areas in organization of vol untary movements in man. *Journal of Neurophysiology*, 43(1), 11 8–136. doi:10.1152/jn.1980.43.1.118
- Roland, P. E., Skinhoj, E., Lassen, N. A., & Larsen, B. (1980). Different cortical areas in man in organization of voluntary movements in extrapersonal space. *Journal of Neurophysiology*, 43(1), 137–150 . doi:10.1152/jn.1980.43.1.137
- Roy, S., Dora, S., McCreadie, K., & Prasad, G. (2020). *MIEEG-GAN:* generating artificial motor imagery electroencephalography signals. Paper presented at the 2020 International Joint Conference on N eural Networks (IJCNN).
- Russell, S. J. (2010). *Artificial intelligence a modern approach*: Pearson Education, Inc.
- Ryun, S., Kim, J. S., Lee, S. H., Jeong, S., Kim, S.-P., & Chung, C. K. (2014). Movement type prediction before its onset using signals f rom prefrontal area: an electrocorticography study. *BioMed resea rch international, 2014*.
- Schall, J. D. (2004). On the role of frontal eye field in guiding attention and saccades. *Vision Res, 44*(12), 1453–1467. doi:10.1016/j.visre s.2003.10.025
- Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter,M., Eggensperger, K., Tangermann, M., . . . Ball, T. (2017). Deep learning with convolutional neural networks for EEG decoding

and visualization. Human brain mapping, 38(11), 5391-5420.

- Schlack, A., Krekelberg, B., & Albright, T. D. (2007). Recent history of stimulus speeds affects the speed tuning of neurons in area MT. *Journal of Neuroscience*, 27(41), 11009–11018.
- Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural net works. *Trans. Sig. Proc.*, 45(11), 2673 – 2681. doi:10.1109/78.650 093
- Sejnowski, T. J., Churchland, P. S., & Movshon, J. A. (2014). Putting b ig data to good use in neuroscience. *Nature Neuroscience*, 17(11) , 1440-1441. doi:10.1038/nn.3839
- Si-Mohammed, H., Petit, J., Jeunet, C., Argelaguet, F., Spindler, F., Evai n, A., . . . Lécuyer, A. (2018). Towards BCI-based interfaces for augmented reality: feasibility, design and evaluation. *IEEE transa ctions on visualization and computer graphics, 26*(3), 1608-1621.
- Sober, S. J., & Sabes, P. N. (2005). Flexible strategies for sensory integr ation during motor planning. *Nature Neuroscience*, 8(4), 490–49 7. doi:10.1038/nn1427
- Sosnik, R., & Zheng, L. (2021). Reconstruction of hand, elbow and sho ulder actual and imagined trajectories in 3D space using EEG cur rent source dipoles. *Journal of Neural Engineering, 18*(5), 056011
- Sundararajan, M., Taly, A., & Yan, Q. (2017). *Axiomatic attribution for deep networks.* Paper presented at the International Conference on Machine Learning.
- Tunik, E., Lo, O. Y., & Adamovich, S. V. (2008). Transcranial magnetic stimulation to the frontal operculum and supramarginal gyrus di srupts planning of outcome-based hand-object interactions. *J Ne* urosci, 28(53), 14422-14427. doi:10.1523/jneurosci.4734-08.2008
- Tuthill, J. C., & Azim, E. (2018). Proprioception. *Curr Biol, 28*(5), R194 -r203. doi:10.1016/j.cub.2018.01.064
- Van Essen, D. C. (2005). A Population-Average, Landmark- and Surfac e-based (PALS) atlas of human cerebral cortex. *Neuroimage, 28*(3), 635-662. doi:10.1016/j.neuroimage.2005.06.058
- Vesia, M., & Crawford, J. D. (2012). Specialization of reach function in human posterior parietal cortex. *Exp Brain Res, 221*(1), 1–18. d oi:10.1007/s00221-012-3158-9

- Vesia, M., & Davare, M. (2011). Decoding action intentions in parietofr ontal circuits. *Journal of Neuroscience*, 31(46), 16491–16493.
- Vesia, M., Prime, S. L., Yan, X., Sergio, L. E., & Crawford, J. D. (2010).
 Specificity of human parietal saccade and reach regions during transcranial magnetic stimulation. *J Neurosci, 30*(39), 13053–130 65. doi:10.1523/jneurosci.1644–10.2010
- Waldert, S., Preissl, H., Demandt, E., Braun, C., Birbaumer, N., Aertsen, A., & Mehring, C. (2008). Hand Movement Direction Decoded from MEG and EEG. *The Journal of Neuroscience, 28*(4), 1000– 1008. doi:10.1523/jneurosci.5171–07.2008
- Wang, W., Chan, S. S., Heldman, D. A., & Moran, D. W. (2007). Mot or cortical representation of position and velocity during reaching . J Neurophysiol, 97(6), 4258–4270. doi:10.1152/jn.01180.2006
- Wang, W., Sudre, G. P., Xu, Y., Kass, R. E., Collinger, J. L., Degenhart , A. D., . . . Weber, D. J. (2010). Decoding and cortical source l ocalization for intended movement direction with MEG. *J Neuro physiol*, 104(5), 2451–2461. doi:10.1152/jn.00239.2010
- Wolpaw, J. R. (2013). Brain-computer interfaces. In *Handbook of clini* cal neurology (Vol. 110, pp. 67–74): Elsevier.
- Yamamoto, R., Song, E., & Kim, J.-M. (2020). Parallel WaveGAN: A f ast waveform generation model based on generative adversarial n etworks with multi-resolution spectrogram. Paper presented at th e ICASSP 2020–2020 IEEE International Conference on Acoustics , Speech and Signal Processing (ICASSP).
- Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, 19(3), 356-365.
- Yang, Y. J., Jeon, E. J., Kim, J. S., & Chung, C. K. (2021). Characteriz ation of kinesthetic motor imagery compared with visual motor i mageries. *Scientific Reports*, 11(1), 3751.
- Yeom, H. G., Kim, J. S., & Chung, C. K. (2013). Estimation of the vel ocity and trajectory of three-dimensional reaching movements fro m non-invasive magnetoencephalography signals. *J Neural Eng*, 1 *O*(2), 026006. doi:10.1088/1741-2560/10/2/026006
- Yeom, H. G., Kim, J. S., & Chung, C. K. (2020). LSTM Improves Accuracy of Reaching Trajectory Prediction From Magnetoencephalogr aphy Signals. *IEEE Access*, 8, 20146–20150.

- Zhou, Z.-H., Wu, J., & Tang, W. (2002). Ensembling neural networks: many could be better than all. *Artif. Intell.*, *137*(1-2), 239-263. doi:10.1016/s0004-3702(02)00190-x
- Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-t o-image translation using cycle-consistent adversarial networks. P aper presented at the Proceedings of the IEEE international confe rence on computer vision.

Abstract in Korean

설명가능한 인공지능과 인공 신경망을 통한 대뇌의 운동 처리 규명에 관한 연구

인간의 운동은 근육과 관절의 동작과 힘, 특히나 가속도, 속도, 위치로 표현되는 운동학적 변인과 같이 운동을 구성하는 다양한 요소들로 구성되어 있다. 인간의 뇌가 이러한 운동 요소들 처리하는 기전을 이해하는 것은 신경과학 연구에서도 중요한 관심사 중 하나이다. 이러한 기전을 밝히는 것은 사지의 운동을 예측하며 동작하는 뇌-기계 인터페이스(Brain-Computer interface, BCI)의 개발과 발전에 대해서도 실용적인 활용이 가능하기에 중요하다.

대뇌 신경세포의 활동을 관찰했던 단일-뉴런 기반의 고전적인 연구들은 개별 뉴런들과 그 집합들의 활동과 운동의 속도, 위치, 그리고 가속 간의 상관이 있음을 밝혔다. 하지만 각각의 운동학적 변인들이 요하는 운동적 특징이 다르고, 피질 단위에서 선호하는 뉴런이 비율이 다른 점으로 보아 대뇌 피질 전반에 걸쳐서 그 표상이 다르게 나타남을 예측할 수 있으나, 고전적인 접근 방법은 측정 범위의 한계가 있어 운동학적 변인에 따라 나타나는 대뇌 전반의 표상을 알 수 없었다. 뇌자도(MEG) 및 뇌전도(EEG)와 같은 비침습적 방법론으로 측정한 가속도, 그리고 속력과의 상관 관계는 이러한 표상을 어느정도 제시하나, 피질 영역에 따라 처리하는 변인을 알기에는 측정가능한 면적 대비 공간적 해상도가 낮아 상호간에 중첩되어 보이는 문제가 있다.

최근, 인공 신경망 기반 모델링을 통해 뇌 신호로부터 운동을 예측

가능하게 됨으로써, 본 연구에서는 운동학적 변인에 따른 신경 특징 또는 표상을 고성능 인공신경망 모델 및 설명가능한 인공지능 (Explainable AI, XAI) 방법론을 통해 식별할 수 있을 것으로 가정했다. 따라서 본 연구는 심충신경망(Deep neural network, DNN)모델과 XAI 기술을 활용하여 대뇌가 운동학적 변인들을 처리하는 기전을 식별하는 방법론을 제시하는 것을 목표로 한다. 이를 위해 하기의 연구들을 진행하였고 그 결과를 제시하였다.

먼저, 첫번째 실험 연구에서는 순환 신경망(Recurrent neural network, RNN)기반의 DNN 모델을 개발하여 네 개의 대상을 향한 팔 뻗기 운동에 대한 운동학적 변인들의 시계열 궤적을, 운동 중에 측정한 MEG 신호에서 추정된 피질 전파원 신호로부터 해독하였다. 본 연구의 결과로, DNN 모델은 전파원 신호로부터 팔 뻗기 운동의 가속도, 속도, 그리고 위치에 대한 시계열 궤적을 정확하게 예측하였으며, 그 결과는 통계적으로 유의하였다 (r > .811, p <.001).

두번째 실험 연구에서는 XAI 기술을 사용하여 상기 연구에서 쓰인 DNN 모델을 해석하여 각 피질 영역들이 각각의 운동학적 변인들을 예측하는데 기여한 정도를 수치로 표현하는 새로운 접근법을 제시하였다. 이러한 접근을 통해 하나의 피질 활동 지도로부터 가속도, 속도, 위치를 처리하기 위한 기능적 영역들을 분리할 수 있었다. 본 연구의 결과로 운동의 가속, 속도, 그리고 위치를 처리하기 위해 기여한 대뇌 피질 영역 별 기여도가 서로 다르게 나타났다. 각각의 운동학적 변인들의 처리에 관여하는 것으로 나타난 피질 영역들을 집합으로 묶었을 때, 모든 경우에서 유의했던 공유영역과 특정 변인에서만 유의했던 지배 영역으로 식별되었다. 공유영역은 주로 양반구의 연상회(supramarginal gyrus, SMG)와 대측 각회(angular gyrus, AG)

영역으로 이루어져 있으며, 운동 연구에서 운동을 위한 다중 감각 정보를 통합하고 운동의 목표를 설정하는 인지 기능을 담당하는 것으로 알려져 있다. 반면, 운동학적 변인에 따른 지배 영역들은 상호 배타적인 집합에 속하며, 서로 다른 기능적 영역을 포함했다 (가속 영역-힘 및 근육 제어를 위한 대측 운동 피질; 위치 영역-공간 처리 및 안구 동작 제어를 위한 시각-운동영역; 속도-운동 정보 처리 및 운동 제어를 위한 두정-전두 네트워크). 이러한 영역들은 운동학적 변인의 처리를 위한 운동의 성격을 명확히 반영하고 있었다.

전반적인 연구들의 함의는 인간의 운동은 다양한 운동 요소들의 계산 을 통해 만들어지며, 이를 처리하는 피질 영역이 있음을 밝혔다. 이러한 영역들 의 활동으로 인해 다양한 운동 변인들이 분산되어 처리되기에 인간의 운동이 매끄럽게 이루어질 수 있는 것이라 생각한다. 또한 BCI 개발에 있어서 본 연구 가 가지는 의미는, 다양한 운동 요소들을 예측하기 위해서는 대뇌 전반이 아니 라 특정 영역에서 특정한 정보를 얻어서 계산한다면 효율적인 BCI 개발로 이어 질 수 있다는 것이다. 이는 좁고 정해진 영역에서 고해상도의 신호를 측정할 수 있는 침습적 BCI 시스템을 만드는데 있어 중요하게 활용할 수 있겠으나, 필요 한 영역의 신호를 얻지 못하면 그 정보 역시 얻을 수 없음을 의미하기도 한다. 나는 이러한 한계를 극복하는 방법으로써, 인공지능 생성 모델 중 하나인 적대 적 생성 신경망 (Generative Adversarial Network, GAN)을 통해 운동 피질의 신호의 시간주파수성 특징을 운동 관련 영역으로부터 추론하는 방법을 통해 해 결할 수 있을 것이라 생각하였고, 이에 대한 개념적 증명을 뒷받침 할 수 있는 증거들을 제시하였다.

주요어: 뇌-컴퓨터 인터페이스, 인공신경망, 신경 영상, 설명가능한 인공지능 학번: 2019-28861