UNIVERSITY OF CALIFORNIA SAN DIEGO

# EEG-Based Assessment of Human Cognitive and Affective States in Real-World Scenarios

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Computer Science and Engineering

by

Kuan-Jung Chiang

Committee in charge:

Professor Tzyy-Ping Jung, Chair
Professor Chung-Kuan Cheng, Co-Chair
Professor Gert Cauwenberghs
Professor Garrison W. Cottrell
Professor Virginia de Sa

2023

The Dissertation of Kuan-Jung Chiang is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Professor Tzyy-Ping Jung for his invaluable guidance and support as my PhD advisor. His guidance and support were instrumental in my success throughout this journey.

I would also like to express my sincere appreciation for the unwavering support of Professor Chung-Kuan Cheng, without whom I would not have been able to complete this program.

Chapter 2, in full, has been submitted for publication of the material as it may appear in the *Journal of Neural Engineering*. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is currently being prepared for submission for publication of the material. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in part, is currently being prepared for submission for publication of the material. The dissertation author was the primary investigator and author of this paper.

This work was supported in part by award #1734883 from the National Science Foundation.

# VITA

2017 Bachelor of Science in Engineering, Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan

2019 Master of Science, Department of Bioengineering
University of California - San Diego, La Jolla, California, USA

2023 Doctor of Philosophy, Department of Computer Science and Engineering, University of California - San Diego, La Jolla, California, USA

ABSTRACT OF THE DISSERTATION

EEG-Based Assessment of Human Cognitive and Affective States in Real-World Scenarios

by

Kuan-Jung Chiang

Doctor of Philosophy in Computer Science and Engineering

University of California San Diego, 2023

Professor Tzyy-Ping Jung, Chair
Professor Chung-Kuan Cheng, Co-Chair

The availability of affordable and portable electroencephalogram (EEG) devices has sparked interest in using passive EEG-based brain-computer interfaces (BCIs) in real-world applications such as neuroergonomics and neuromarketing. These fields require objective measurement of human cognitive and affective states. Although studies have explored EEG features for different mental states and affective responses in these areas, there is still a gap between laboratory research and real-world implementation.

Two critical questions need to be addressed to bridge the gaps between laboratory research and real-world implementation. Firstly, can the EEG features identified in controlled

laboratory conditions be reliably detected in real-world settings? Secondly, how can transfer learning streamline the calibration process for new users or sessions of passive BCI features? Can laboratory-oriented tasks be employed to calibrate the model for real-world applications?

This dissertation aims to address the questions raised earlier by developing EEG signal-processing and feature-extraction methods, and exploring transfer learning techniques for assessing human cognitive and affective states in naturalistic environments. Chapter 2 describes a study demonstrating how EEG can be used in neuroergonomic research to monitor changes in an individual's memory workload during a regular office task. Chapter 3 presents a study on affective states, examining how EEG and eye-tracking can detect human interest levels in images of electronic products. These two chapters prove that robust EEG features found in laboratory settings can also be observed in real-world settings.

Chapter 4 investigates the transferability of EEG features in monitoring human cognitive loads. The study's outcomes can inform the development of transfer learning techniques for more effective BCI applications in real-world settings. Chapter 5 demonstrates the feasibility of cross-task transfer learning for passive BCIs and illustrates how EEG signatures from lab-controlled tasks can be applied to real-world scenarios. Finally, Chapter 6 concludes all the studies.

Overall, this dissertation offers valuable contributions to the EEG-based assessment of human cognitive and affective states in real-world settings and has significant implications for developing more practical and effective passive BCI applications.

# Chapter 1

# Bridging the Laboratory and Real World

A brain-computer interface (BCI) is a technology that offers an immediate and intuitive pathway for users, particularly those with disabilities, to convert their intentions into commands that can control external devices [1]. While BCI applications have traditionally focused on providing alternative pathways for patients with motor disabilities to communicate with machines or computers, or serving as monitoring systems for certain diseases in clinical environments such as seizure detection or the diagnosis of Alzheimer's disease [2, 3], researchers have also explored the development of BCI applications for the general public. Passive BCIs, which interpret users' mental and emotional states based on the measurement of spontaneous brain activities [4], have gained particular attention in this regard. Passive BCIs allow users to interact with the system passively, without requiring specific actions or stimulation, making them more suitable for general public applications.

Passive brain-computer interfaces (BCIs) have garnered considerable interest in monitoring human brain states across various applications. Two prominent fields that have particularly focused on this aspect are neuroergonomics and neuromarketing. Neuroergonomics is a sub-track of human factor study and aims to find optimal hardware or software settings where the user's cognitive workload is reduced. This includes monitoring changes in mental workload in different work environments or monitoring stress levels under multi-tasking conditions [5, 6]. On the other hand, neuromarketing aims to assess human affective states when receiving marketing

stimuli like advertisements or product displays. This field aims to gain insight into how humans make decisions about purchasing products [7]. Since these two fields require an assessment of human brain states, researchers in these fields have turned to passive BCIs to help obtain direct and objective measurements of human cognitive and affective states.

Although the signatures of brain-imaging signals for different cognitive states have been widely studied, researchers have faced technical challenges in translating laboratory-oriented BCI demos into real-world applications. These challenges include hardware issues, such as difficulty measuring brain activities in the real world. While several brain-imaging technologies, such as fMRI, fNIRS, and EEG, are available to explore brain functions and neural activity, they have limitations in applying to the real world.

FMRI and fNIRS assess changes in blood oxygen in brain vessels and have lower temporal resolution [8, 9]. While FMRI provides excellent spatial resolution and localization of brain areas' activities in a 3D space, its use is almost exclusively limited to laboratory settings due to the need for a strong magnet [10]. EEG monitors voltage fluctuations caused by the ionic current in the brain's neurons and has a low spatial resolution but a high temporal resolution. Traditional EEG and fNIRS devices were bulky and required a bunch of wires connected to the patient's head, so they are not practical to use in the real world either. However, in the last two decades, there has been a vast advancement in the development of cheaper and more portable EEG devices. As a result, EEG has been employed in many studies to investigate human mental states for various tasks or settings [11, 12, 13, 14, 15, 16, 17].

Although the measurement of brain activities has been greatly improved. There is still a gap between the laboratory and real-world setting as fundamental differences in human behavior or brain activities between the laboratory and naturalistic environment may exist. Therefore, it remains unclear how well the current knowledge of human brain function translates into the real world.

The first two chapters of this dissertation attempt to bridge the gap between laboratory research and real-world applications. This is achieved by designing experiments that simulate

real-world tasks and examining whether the EEG signatures of cognitive or affective states observed in a lab-controlled setting can also be identified in the real world.

In Chapter 2, we presented a study demonstrating how EEG can be used in neuroergonomic research to monitor changes in an individual's memory workload during a regular office task. The EEG features identified in this study provide valuable insights into tracking human cognitive states in real-life scenarios.

Chapter 3 discussed another study related to affective states, which examined how EEG combined with eye-tracking can be used to detect human interest levels for images of electronic products. The results of these two studies showed the feasibility of tracking human cognitive and affective states in real life. However, they also revealed huge EEG variability across subjects.

The variability in EEG features among individuals highlights the importance of using calibration data to fine-tune parameters in predictive models for monitoring cognitive or affective states [18, 19]. However, the need for recalibration before each use of BCI systems can severely hinder their practicality. To mitigate this issue, researchers have proposed various transfer learning techniques [18, 19].

Transfer learning is often required when training and testing data are extracted from different domains with inconsistent distributions, and collecting new data in the testing domain is expensive. In such cases, statistical models trained from training data may not effectively classify testing data. Transfer learning methods can reduce the disparity between domains and, thus, reduce the demand for recalibration [20]. For BCI applications, transfer learning methods can be applied so that existing data from a user can calibrate the model for a new session or data from other users can be leveraged for a new user. However, although transfer learning for BCIs has been widely studied, most previous works have been limited to traditional BCI paradigms [19].

To address the need to build practical models for passive BCIs, the last two chapters of this dissertation investigated how to apply transfer learning to improve BCI performance. Chapter 4 presented a study that explored the transferability of EEG features for monitoring

human cognitive loads under cross-day and cross-individual transfer learning scenarios. This study aimed to improve the robustness and generalizability of predictive models, leading to more effective BCI applications for real-world scenarios.

Finally, Chapter 5 explored the relationship between a classic memory task, the Sternberg task, and a natural speech listening task. The results of this study demonstrated the feasibility of cross-task transfer learning for passive BCIs. They highlighted how EEG signatures from laboratory-oriented tasks could be applied to real-world scenarios, facilitating the development of more effective BCIs in real-world settings.

# Chapter 2

# Memory Workload

## 2.1 Introduction

The goal of human factors research is to build the best possible human-machine interface [21]. The principles of human factors are applied to the design of systems to improve human comfort, productivity, and error reduction. Physical and cognitive ergonomics are two sub-fields of human factors research. Physical ergonomics is concerned with monitoring human peripherals such as muscle activity and tension. On the other hand, cognitive ergonomics, also known as neuroergonomics, assesses human cognitive states or brain activities, including memory and mental workload [5].

Neuroergonomics has received increasing attention and has been combined with neuroscience research as brain-imaging methods have improved in recent years. To obtain human cognitive states, neuroergonomics has always relied on performance evaluation, and subjective measurements [22]. Performance metrics are indirect methods that infer a subject's cognitive workload based on how well they complete specific tasks. [23]. Subjective measurements are commonly carried out by administering questionnaires to subjects. As a result, these procedures are prone to bias. Furthermore, subjective measurements may only be obtained after specific activities have been finished, making tracking the dynamic of human cognitive states during tasks difficult. [24]. Because of these constraints, human factor researchers have shifted their attention to physiological measures that provide continuous, direct, and objective measurements.

Even though much research has been undertaken to investigate EEG correlates of memory or mental workload as the physiological measurement for changes in cognitive states, only a few have focused on real-world scenarios. Participants in these studies completed tasks such as the n-back test, Sternberg task [25], or Simultaneous Capacity test [26], which were meant to induce various levels of memory or mental effort. On the other hand, humans are unlikely to undertake such jobs in real life. As a result, whether these neuroergonomic approaches can accurately represent the differences in cognitive states in a real-life scenario remains to be seen. Some studies looked into workload in the real world, although the tasks they used were mostly airplane control or human-robot interaction, which are still not typical daily tasks [27, 28]. Furthermore, only a small percentage of them [29, 30] are genuinely relevant to human factors research.

This study applied EEG analysis to a real-world scenario. More specifically, we designed an experiment to examine the memory workload when a subject performs office work under two different conditions: 1) a single-monitor setup and 2) a dual-monitor setup. Because the dual-monitor setup can lessen the frequency of switching between windows, we think the memory usage can be decreased when performing specific office tasks under the setup [31, 32]. We aimed to examine if the known EEG signatures of memory workload from literature [33, 34, 17, 28, 35] are indeed less phenomenal under the dual-monitor setup, which can be physiological evidence of the ergonomic design. Another goal of this study is to find more EEG features besides the known signatures from the literature by comparing EEG spectral power, mutual information, and coherence under two situations and using these features to train machine learning models to classify high and low memory workload states.

This study also analyzed EEG data from a Sternberg task recorded in [34] to validate whether the new workload-related EEG signatures found in the study were consistent with those found in the previous study.

## 2.2 Method

### 2.2.1 Number-copying Experiment Design

To study the workload while a user is conducting some office work, we designed an experiment in which participants performed a number-copying task under two conditions: single-monitor or dual-monitor configurations. This task seeks to mimic routine office tasks such as accounting and meeting scheduling, which need workers to ”copy” and paste numbers between multiple windows or files. During the experiment, participants were asked to sit in an office chair at a desk and operate a PC with two monitors. One monitor (the main/front monitor) was positioned directly in front of the subject, and the other one (the side monitor) was positioned to the right of the front monitor and pivoted 30 degrees to face the subject. On the PC, there were numerous PDF files containing a collection of random numbers. Fig. 2.1 shows the first page of an example PDF file. The participants had to use the keyboard to enter the same figures into an empty Excel page. Copying and pasting were not allowed. There were two sorts of PDF files in the single-monitor setup: 3 numbers per group (as in the example in Fig. 2.1) and 4 numbers per group. In the single-monitor condition, the side monitor was turned off, and subjects had to memorize several numbers at a time from the PDF file, switch between a PDF file and an excel sheet using Alter+Tab, and enter the numbers into the Excel sheet. In the dual-monitor arrangement, both monitors were turned on, the Excel sheet window was on the front, and the PDF file was on the side monitor. Subjects did not need to memorize the numbers because they could enter them while looking at a side monitor. The third type of PDF file was also given to participants in the dual-monitor setup, with half of the groups consisting of three numbers and the other half consisting of four. As a result, there were three different types of PDF files: (1) three numbers per group for the single-monitor setup, (2) four numbers per group for the single-monitor setup, and (3) a combination of three and four numbers per group for the dual-monitor setup. In the experiment, the subjects had to complete each type of file once as one cycle, and finish two cycles in total. In other words, each subject completed six files (two

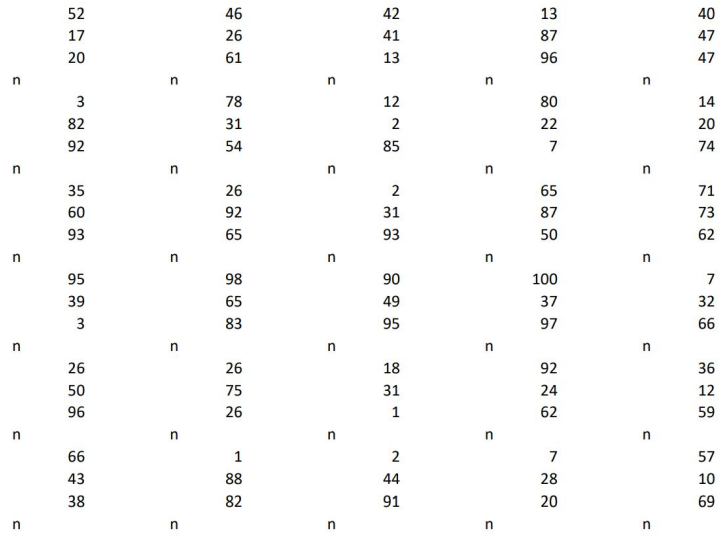



**Figure 2.1.** The first page of an example PDF file. The letter n indicated the end of a group of numbers. There were 48 groups of numbers in one file.

files per type) in total. Within each cycle, the order was either (1) $\rightarrow$ (3) $\rightarrow$ (2) or (2) $\rightarrow$ (3) $\rightarrow$ (1). That is, a file for the dual-monitor setup was always completed between two files of the single-monitor setup within a cycle. Each file took around 10 minutes to complete.

A total of ten people (4 females and 6 males, aging from 25-30 years old) took part in this investigation. The SMARTING mobile EEG amplifier (mBrainTrain, Belgrade, Serbia) was used to record EEG data using a Saline-based 24-channel EEG cap. The amplifier's output was transmitted over Bluetooth to the SMARTING acquisition program, and the signals were then broadcasted and recorded using Labstreaminglayer (LSL). Java software collected all of the keyboard inputs and synchronized them with the EEG signals using LSL.

### 2.2.2 Data Processing

All EEG channels were re-referenced to Cz after the EEG signals were filtered with a 0.1-50 Hz band-pass filter. To eliminate EEG artifacts, the function `clean_asr` in EEGLAB [36] was used to perform artifact subspace reconstruction (ASR) [37]. The calibration data for ASR were taken from a 30-second eye-close resting interval recorded at the beginning of the experiment. The signals were then downsampled to a sampling rate of 250 Hz.

This study focused on the EEG window preceding the participants' entry of the numbers. During this time, we believe the subjects went through a memory retrieval process to recall the numbers they had just memorized. An EEG trial began one second before the first digit of a series of numbers was entered. That is, if a subject's timestamp for entering the first digit of a series of numbers was $t$ second, the EEG trial corresponding to that group was epoched in the window $[t-1, t]$ sec. As a result, each subject should ideally have 288 trials (6 pdf files $\times$ 48 groups per file). However, several trials were missing from some subjects' experiments due to connection issues during recording. Table 2.1 lists the final number of trials for each subject. After the EEG trials were epoched, they were concatenated along the time dimension as a long trial. The EEG data were subsequently decomposed into statistically independent components using Independent Component Analysis [38]. We then used the ICLabel package [39] to identify the artifact and brain components in this long trial. The artifact components were removed, and the brain components were back-projected to the original scalp channels and reshaped back to the same sets of epochs.

**Table 2.1.** Number of trials in each subject's recording of the number-copying data set.

| Subject ID | Number of trials | Subject ID | Number of trials |
|---|---|---|---|
| 1 | 240 | 6 | 288 |
| 2 | 288 | 7 | 96 |
| 3 | 96 | 8 | 192 |
| 4 | 288 | 9 | 282 |
| 5 | 288 | 10 | 282 |

### 2.2.3 Sternberg Task Data Set

We also used a data set collected during a Sternberg task from a prior study [34] to validate the signatures we discovered in the number-copying data set. In the Sternberg task, subjects were asked to memorize a sequence of letters. At the beginning of each trial, there was a 5-second eye-open resting period. Then the first letter appeared on a computer screen for 1.4 seconds and disappeared. Right after the first letter disappeared, the second letter appeared for

another 1.4 seconds, and so on and so forth until eight letters were presented to the subjects for each trial. After the eighth letter disappeared, the screen stayed blank for a short random time between two to four seconds during which the subjects had to maintain the memory of the sequence of letters they just saw. Finally, after the random blank time, a probe letter was presented, and the subjects were asked to answer whether the probe letter was on the list of letters to memorize in the first part by clicking yes/no buttons.

Although eight letters were presented to the subjects, they didn't need to memorize all of them. Among the eight letters, some of the letters were labeled with green colors, indicating that they should be ignored. The others labeled with black were the ones they actually needed to memorize. The number of the green letters could be six, four, and two, making the number of black letters two, four, and six. We defined the trials with two/four/six black letters as easy/medium/hard trials because with more letters to be memorized, the trials should be more difficult. Each subject completed 50-100 trials. Each difficulty accounted for one-third of the trials and was randomly ordered. Due to missing channel labels in the recordings from some subjects, only eleven subjects' data were included in this study.

As mentioned in section 2.2.2, we focus on the memory retrieval period in this study. To validate the EEG signatures in the comparable two states: performing memory-retrieval (i.e., under single-monitor setup), and not performing memory-retrieval (i.e., under dual-monitor setup), we made two comparisons of different periods in the Sternberg task. The first comparison is comparing the period of resting states (not performing memory-retrieval) and the period between the subjects saw a probe and responded (performing memory-retrieval). For each trial, we extracted the eye-open resting period (five-second long) as epochs marked 'Rest'. As for the recalling period, one-second epochs after the onset of showing the probe letters were extracted and marked 'Recall'. For the second comparison, we further divided the 'Recall' trials into two labels by the difficulty of the trials. As mentioned above, trials had three levels of difficulty (easy/medium/hard) based on the number of letters to be memorized. We compared the EEG responses of the recalling periods of the easy trials (low memory-retrieval workload) versus the

hard trials (high memory-retrieval workload).

### 2.2.4 EEG Spectral Analysis

Several studies have reported that the EEG power in the delta (1-4 Hz), theta (4-8 Hz), and alpha (8-13 Hz) bands in the frontal area highly correlated with changes in mental and memory workload [33, 34, 17, 28]. Because some subjects had inconsistent missing channels in both datasets, and we want to preserve as many subjects as possible, only a subset of channels F3, F4, Fz, Pz, AFz, CPz, and POz were analyzed below. The power of delta, theta, and alpha bands of this subset of channels was extracted by summing the square of the Fourier transform magnitude within corresponding frequency ranges and taking the logarithm of the summation. Totally, 21 features (3 bands $\times$ 7 channels) were extracted from the EEG spectral analysis and referred to as band power (BP) features in the following sections.

### 2.2.5 Mutual Information

Mutual Information (MI) is the measure of how two random variables depend mutually. The mutual information of random variables $X$ and $Y$ is defined as:

$$\begin{aligned} & MI(X,Y) \\ = \quad & \sum_{y\in\mathscr{Y}}\sum_{x\in\mathscr{X}} p(X=x,Y=y)\log\frac{p(X=x,Y=y)}{p(X=x)p(Y=y)}, \end{aligned}$$

where $\mathscr{X},\mathscr{Y}$ are the spaces of $X,Y$, $p(X=x,Y=y)$ is the joint probability mass function of $X$ and $Y$, and $p(X=x), p(Y=y)$ are marginal probabilities. In practice, the joint probability and marginal probability are estimated by creating the histograms of observed samples. In EEG analysis, measurements of two channels over a certain period can be used to estimate the probabilities and the mutual information between them.

A variety of EEG paradigms have used mutual information-based features [22, 40, 41, 42]. This study also compared the mutual information between pairs of the same set of selected

channels in the spectral analysis under the two monitor setups. However, instead of measuring the mutual information of the raw signals for each channel pair, we used the square of the signals to represent their power. The number of bins for constructing the histograms was set to 64. An open-source implementation in MATLAB [43] was used.

The calculation of the mutual information value is slightly different between the number-copying data set and the Sternberg data set. In the number-copying data set, the entire 1-sec epochs were used as inputs to the MI calculation. In the Sternberg data set, as the lengths of epochs from the resting and recalling periods were not the same, to avoid the effect of the length differences, we calculated several values of mutual information across an epoch using a one-second non-overlapping sliding window. Then, the final mutual information was obtained by averaging across all five windows.

Because the symmetrical terms of the mutual information had the same values (e.g., MI of Fz and Pz is equal to MI of Pz and Fz), we only need to calculate 28 features (7 channels $\times$ $(7+1)/2$), which were referred to as mutual information (MI) features in the following sections.

### 2.2.6 Magnitude-squared Coherence

Magnitude-squared coherence (coherence) is a metric used to assess the relationship between two signals in the frequency domain. The coherence estimates the quotients between the cross-spectral densities of two signals and the product of their spectral densities. If two time-series signals are defined as $x(t)$ and $y(t)$, the coherence between these two signals $C_{x,y}$ is defined as:

$$C_{xy}(f) = \frac{|P_{xy}(f)|^2}{P_{xx}(f)P_{yy}(f)},$$

where $P_{xy}(f)$ is the cross power spectral densities between $x$ and $y$, and $P_{xx}(f)$, $P_{yy}(f)$ are their spectral densities [44].

This study first calculated the coherence between each pair of the selected channels (the same pairs as in the spectral and MI analysis) as a function of frequency. Then, we extracted

the delta (1-4 Hz), theta (4-8 Hz), and alpha (8-13 Hz) coherence by summing the coherence within the corresponding frequency. Again, to circumvent the effect of length differences, the calculation of the coherence in the Sternberg data set also used the sliding window approach, while the entire epochs were used in the number-copying data set. It's also worth noting that a channel's coherence with itself is one at all frequencies, so they weren't included in the feature space. In total, 63 features were extracted (3 bands $\times$ 7 channels $\times$ $(7-1)/2$) and referred to as coherence (Coh) features in the following sections.

### 2.2.7 Evaluation of Classification

This study aimed to evaluate how well the band power, mutual information, and coherence can track the changes in memory workload levels. We used each of the aforementioned features and a combined feature space to train machine learning models and examined their ability to estimate the participants' memory workload states. As mentioned in section 2.2.2 and 2.2.3, we tried to distinguish low/no memory-retrieval workload versus high memory-retrieval workload, so the machine learning models were used to solve the two-class classification problem. Because the design of the classifiers was not the focus of this study, we chose a common type of machine learning model, support vector machines (SVM) as our classifier. We trained the SVM classifiers with four different feature spaces: 1) Band powers only, 2) Mutual information only, 3) Coherence only, and 4) The concatenation of the first three.

For the number-copying data set, avoid temporal auto-correlations between consecutive EEG epochs, a cross-block evaluation within each subject was exploited. That is, at each iteration, epochs of one block (one PDF file) were used as testing data and all the epochs of other blocks were used as training data. No training and testing epochs were extracted within the same block. When assessing a subject's data that had no missing values, each iteration involved 240 training epochs and 48 testing epochs

Due to the short duration (one second) of each epoch, the single-epoch classification performance is limited. We used a sample-grouping method to boost the classification perfor-

mance. Within each iteration of the cross-block validation, after the data were split into training sets and testing sets, every $N_g \in \mathbb{N}$ samples of the same class were averaged and became a new sample. For example, if there are $n$ training samples of one class $j$, denoting as $\mathbf{x}_i^{(j)} \in C_j$, where $i = 1...n, C_j \in \{0,1\}$. The first new training sample after the grouping method will be $\hat{\mathbf{x}}_1^{(j)} = \sum_{k=1}^{N_g} \mathbf{x}_i^{(j)}/N_g$, and the second one will be $\hat{\mathbf{x}}_2^{(j)} = \sum_{k=N_g+1}^{2N_g} \mathbf{x}_i^{(j)}/N_g$, and so on. Note that the order of $i$ followed the same temporal order as they were recorded within each block. Also, if remaining samples are less than $N_g$, they are averaged and become the last new sample. This sample-grouping method mimics the scenario that the cognitive monitoring system can only predict by measuring a few epochs. This study showed the results of using different values of $N_g = [1,2,4,8]$.

As for the Sternberg data set, because the epochs of low and high memory workload appeared alternately in time, there should be no effect of temporal auto-correlation. Therefore, simple five-fold cross-validation was employed and repeated twenty times with different random seeds for shuffling.

To ensemble the comparison of sensitivity or specificity, we used the metric balanced accuracy, which is the mean of sensitivity and specificity. A balanced accuracy averaged across all iterations of cross-validation was obtained for each subject. Finally, the Wilcoxon signed-rank test [45] was applied to the samples of balanced accuracy of each subject (10 accuracy samples in the number-copying experiment and 11 samples in the Sternberg task study) to compare the models pair-wisely. Note that the distribution of accuracy is usually non-Gaussian. Therefore, the Wilcoxon signed-rank test as a non-parametric paired comparison method was employed. The Python package Scikit-learn [46] was used to implement the machine learning classifiers. Some parameter settings of the classifiers are noteworthy. The SVM models used the Radial Basis Function (RBF) kernel, and the class weights were set to balanced weights (which means the weight of a class is equal to one divided by the frequency of the class). The other parameters of the classifiers simply used the default values.

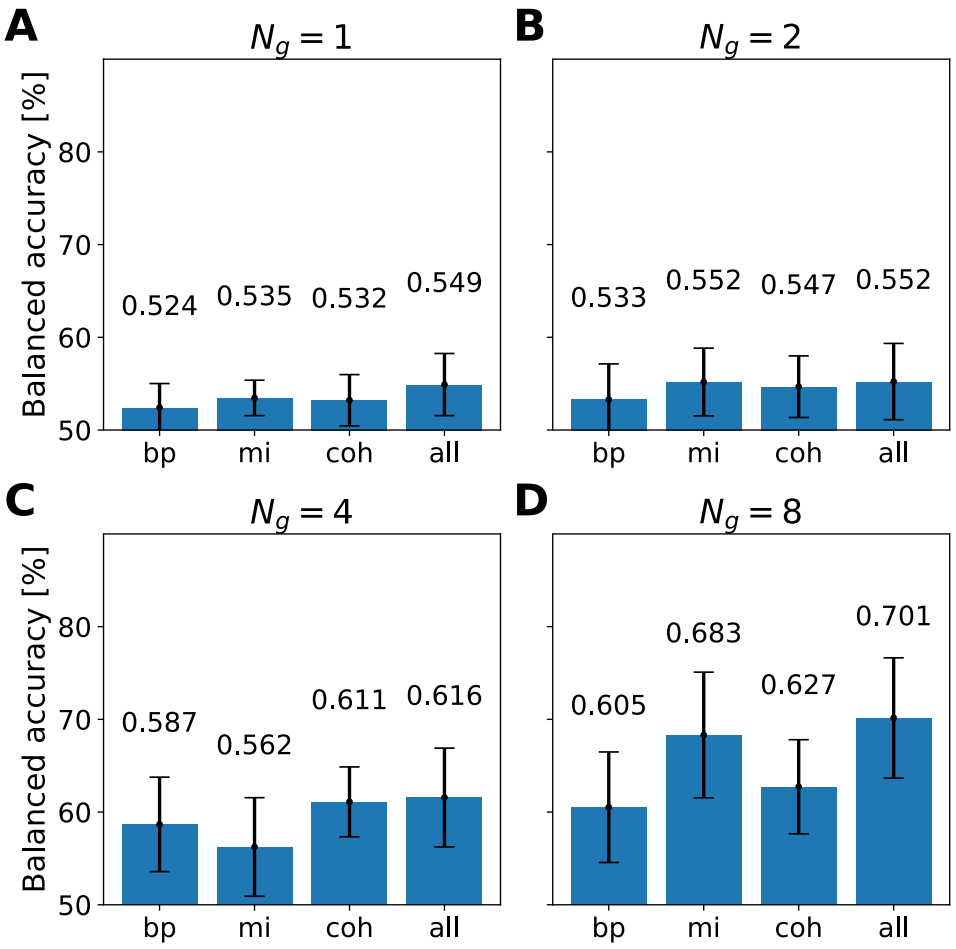



**Figure 2.2.** Classification results of the number-copying experiment. Bar plots of balanced accuracy when using different values of $N_g$. In each panel, the labels of x-axis show different feature spaces: 1) Band power (bp), 2) Mutual information (mi), 3) Coherence (coh), 4) All combined (all).

## 2.3 Result

Fig. 2.2 shows the results of the classification of the number-copying experiment using different values of $N_g$. Within each panel, the performance of the models trained with four different feature spaces: 1) Band power, 2) Mutual information, 3) Coherence, and 4) All combined is plotted. The heights of the bars represent the average balanced accuracy of 10 subjects. The Wilcoxon signed rank test didn't show a significant difference when comparing models trained with each of the first three feature spaces to those trained with the fourth (all

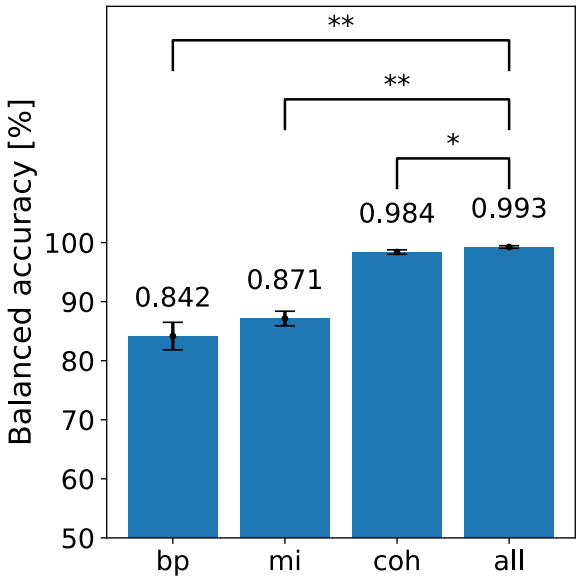



**Figure 2.3.** Classification results of the Sternberg task experiment. Bar plots of balanced accuracy when using different values of $N_g$. In each panel, the labels of x-axis show different feature spaces: 1) Band power (bp), 2) Mutual information (mi), 3) Coherence (coh), 4) All combined (all). The p-values of the Wilcoxon signed rank test are shown and the symbol */** indicates the case of a p-value smaller than 0.05/0.01.

combined). Yet, in all panels, combining all features yields the best results, and in most cases, using only band power features brings the worst. In panels A and B, when $N_g = 1$ and $N_g = 2$, the classification performance is near the chance level (50%). While in panels C and D, when $N_g = 4$ and $N_g = 8$, the accuracy of using all features combined can go up to 70%.

Fig. 2.3 consists of a similar plot as in Fig. 2.2 using the results from the Sternberg task experiment and classifying the Recall epochs versus Rest. Because the single-epoch classification accuracy is high enough, the sample-grouping method was not employed for this evaluation. Similar to the results of Fig. 2.2, in Fig. 2.3, using all features combined reaches the best performance, and using only band power features yields the worst. Unlike the results of the number-copying task, the Wilcoxon signed rank test showed a significant difference between using all features combined and the other three feature spaces.

The classification results of classifying Recall epochs with easy difficulty versus hard difficulty are not shown because the performance can only achieve around the chance level. The

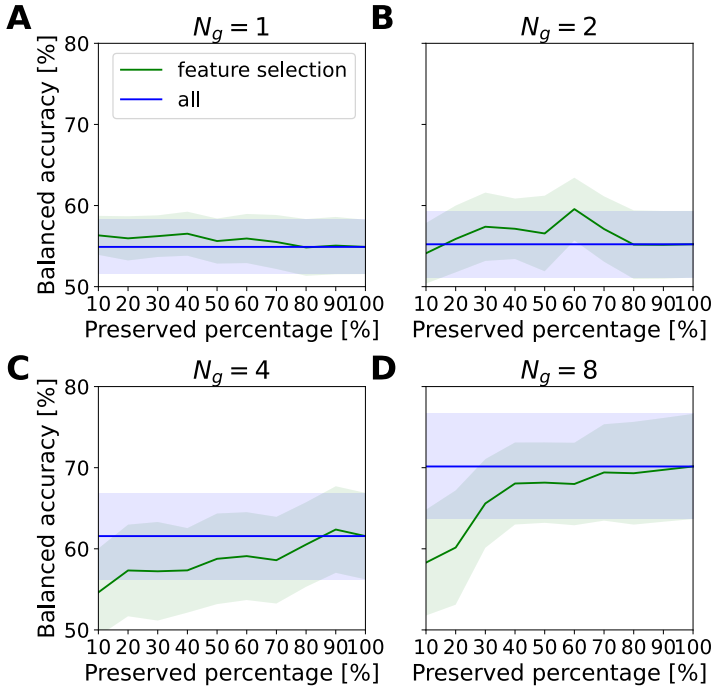



**Figure 2.4.** The balanced accuracy versus the percentage of preserved features (from the number-copying data set).

main reasons are the number of training trials in this scheme is much smaller (only one-third of the case when classifying Recall versus Rest) and the samples of two labels have higher similarity because they are all in recall period despite different difficulty levels.

Although combining all types of features brings the best performance, there may be redundant features among each type of feature space. To further investigate this, we exploited a second evaluation scheme in which a subset of features were selected. In this evaluation scheme, the cross-block validation for the number-copying data set and 5-fold validation for the Sternberg task data set was also employed, but within each fold, all features were ranked by ANOVA F-value scores, and then only the top $X\%$ features were selected from all features combined. We examined the performance of all four types of classifiers with $X$ ranging from 10 to 100, and the results are plotted in Fig. 2.4 and 2.5.

In Fig. 2.4, the performance of the classifiers is also limited when $N_g = 1$ and $N_g = 2$, but in panels C and D, with the larger values of $N_g$, we can see a clear trend that the accuracy

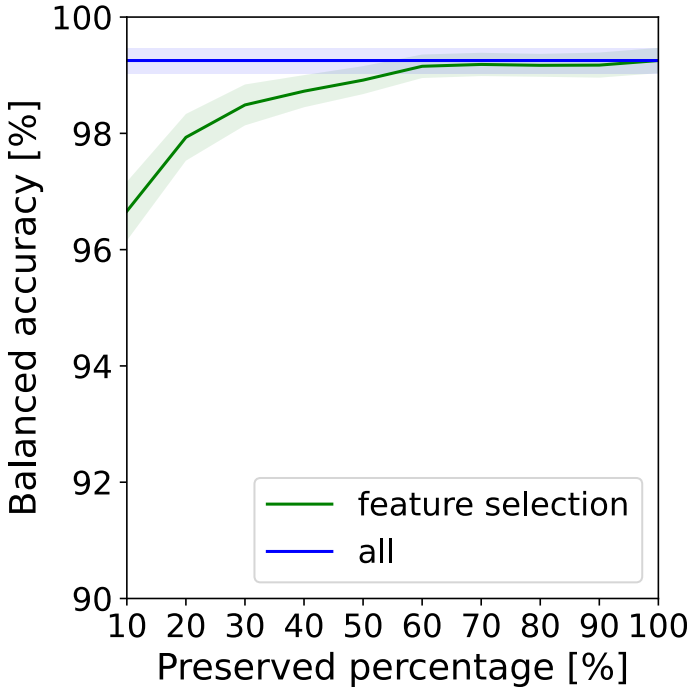



**Figure 2.5.** The balanced accuracy versus the percentage of preserved features (from the Sternberg data set).

increased as more features being preserved. Also, in Fig. 2.5, preserving more features brings higher classification accuracy. The green line reaches its plateau after preserving around 70% of the features, whereas the green line in Fig. 2.4D also reached its plateau around.

We further examined whether the top features of each feature space selected from the number-copying data set are robust across subjects and across data sets. For each feature space (i.e, BP, MI, and Coherence), we selected the top five features with the highest averaged ANOVA F-value scores (as we used to rank the preserved features in Fig. 2.4 and 2.5) across subjects of the number-copying data set. Within each subject, we normalized each of the top five features to zero mean and unit variance and summed them up for each feature space. Note that when calculating the sum, each normalized feature value could have either positive or negative mean differences between the two classes. Therefore, either $+1$ or $-1$ is multiplied by each normalized feature value to align the polarity. We averaged all summed feature values of the epochs of each label of each subject, and the distribution of the results is shown in Fig. 2.6.

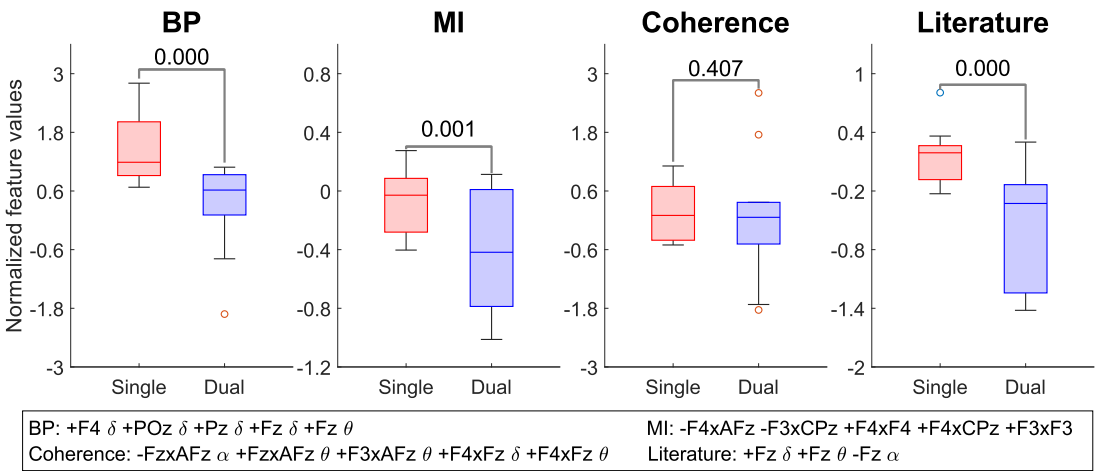



**Figure 2.6.** Box plots of the sum of the normalized top feature values across subjects from the number-copying data set. The numbers shown at the top are the p-values of the non-parametric paired bootstrap F-test.

The box plots in Fig. 2.6 represent the minimum, first quartile, median, third quartile, maximum values, and outlier values. The text in the lower area shows the selected top features. Besides the data-driven signatures, we also added a signature based on what previous literature defined as Fz $\delta$ + Fz $\theta$ − Fz $\alpha$ (referred to as the literature signatures) at the right-most column. [33, 34, 17, 28, 35]. The differences in the distribution of the two classes were tested with non-parametric paired bootstrap F-test, and the p-values are also shown in Fig. 2.6. The corrected p-values with the Benjamini-Hochberg procedure [47] to control the false discovery rate for multiple comparisons are $0.000, 0.002, 0.407$, and $0.000$ for the columns from the left to the right.

Last, similar comparisons for the Sternberg task data set were conducted and the results are plotted in Fig. 2.7 and 2.8. Fig. 2.7 was plotted to compare the Recall epochs versus the Rest epochs, while Fig. 2.8 further compared the Recall epochs with the hard difficulty versus the easy difficulty. Note that these results were obtained using the same sets of normalized selected features with the same polarity as in Fig. 2.6 in order to validate the data-driven signatures of the number-copying data set.

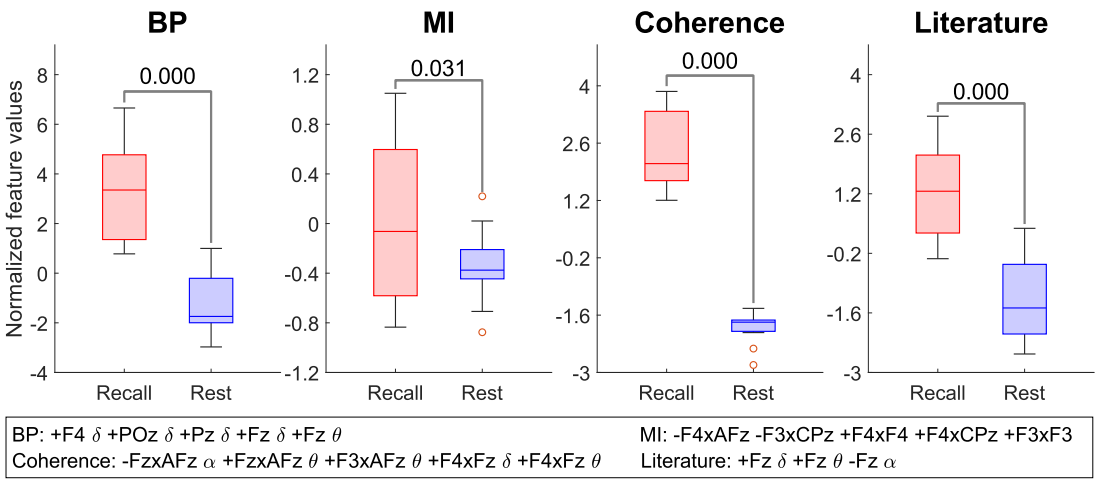



**Figure 2.7.** Box plots of the sum of the normalized top feature values across subjects comparing the Recall epochs versus the Rest epochs from the Sternberg task data set.

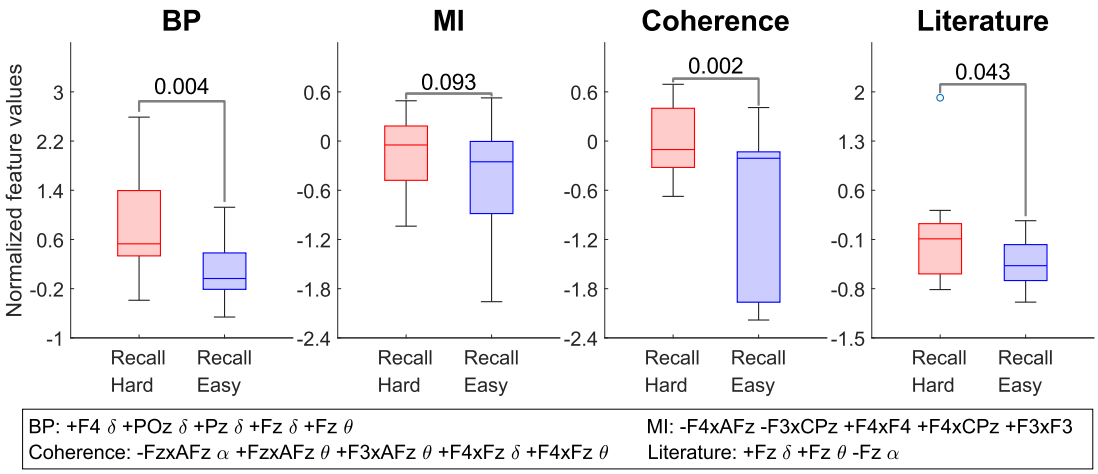



**Figure 2.8.** Box plots of the sum of the normalized top feature values across subjects comparing the Recall epochs with the hard difficulty versus the easy difficulty from the Sternberg task data set.

## 2.4 Discussion

The results in Fig. 2.2 show that when using $N_g = 8$ the performance for classifying high memory workload state (single-monitor setup) and low memory workload state (dual-monitor setup) can reach around 70% balanced accuracy. It indicates that in real-world scenarios, the band power, mutual information, and coherence features can be used to detect the changes in memory workload. However, it also shows the limitation that short-epoch prediction (i.e., $N_g = 1$

or 2) is not plausible using the feature spaces proposed in this study. Only when $N_g = 4$ or 8, the classification has reasonable performance. Considering each epoch before grouping is one second long, the case $N_g = 8$ is similar to making a prediction with eight-second-long epochs.

We achieved much higher classification accuracy in the Sternberg task data set (Fig. 2.3) than in our number-coping study; as for the Sternberg task we compared the high memory workload state to a more distinct state, the eye-open resting period. It is important to acknowledge that the Sternberg dataset was collected using a research-grade EEG device in a well-controlled laboratory environment, while our number-copying dataset was collected using a wireless and saline-based EEG headset on freely moving subjects in a real-world setting (a regular office). This difference in signal quality between the datasets may have contributed to the discrepancy in classification accuracy. Regardless of the exact values of the accuracy, the important finding is that the EEG features used in the present study, including band power, mutual information, and coherence, effectively distinguish between high and low memory workload states in both datasets.

Results in Fig. 2.4, 2.5 also imply that most of the features are useful for classifying the states high and low levels of memory workload. In Fig. 2.4C and D, and 2.5, these models have the best performance when 100% of features are preserved and preserving less features decreases the accuracy. Also, in Fig. 2.4D and 2.5, the accuracy does not increase after preserving more than around 70%, suggesting that part of the features might be redundant. Yet, including all of the features has no negative impact on performance.

Fig. 2.6 shows that the summed normalized feature values of all feature space in the single-monitor condition have higher first quartile, median, and third quartile than the dual-monitor condition. Although the statistic test doesn't show a significant difference for the Coherence signatures, which can be caused by the effect of the outliers, the differences in the signatures of all other feature spaces are significant. The results suggest that these signatures are robust across subjects.

The literature signatures derived from the frontal activities (the most right panels), the

most frequently reported features from the literature [33, 34, 17, 28, 35] significantly increase as the level of memory workload increases. It is proof of the efficacy of our number-copying experiment design. The selected BP features are also mostly about frontal Delta and Theta power, which are similar to the literature signatures with extra terms of parietal Delta. For the MI features, few studies have reported their correlations with memory workload. Mutual information within frontal channels, however, has been reported as a marker of Alzheimer's and schizophrenic disease [41, 48]. The MI signatures we discovered in the current study are also mainly based on the frontal channels. The changes in frontal-parietal mutual information could also result from changes in the frontal and parietal activities reported in [34, 22]. As for the coherence signatures, they also seem to capture the activities change in the frontal area.

It is not surprising to see significant differences between the feature values of two conditions in Fig. 2.6 because these are data-driven signatures extracted from this data set. To validate that the discovered signatures are not over-fitting features only for the number-copying data set, we tested if the significance still exists in another data set, the Sternberg data set. As shown in Fig. 2.7 and 2.8, these signatures are also robust across subjects in the Sternberg task. Except for the MI column in Fig. 2.8, the statistic tests show significance in all other columns in both figures. For the MI column in Fig. 2.8, although the p-value is not below 0.05, it's still quite low (lower than 0.1) considering there are only eleven samples (eleven subjects in the Sternberg data set). The results imply that these signature values robustly increase as the memory workload for each subject increases. Also, we can see that the distributions are more separable in Fig. 2.7 than in Fig. 2.8, this matches the fact that separating performing memory-retrieval (mixed of high and low memory workload) versus resting (zero memory workload) should be easier than separating hard Sternberg tasks (high memory workload) versus easy Sternberg tasks (low memory workload). Although this study has another limitation, that the model for single-epoch hard-versus-easy classification couldn't be successfully trained, we still find the significant difference by looking into the average of multiple epochs.

It is noteworthy that even though the non-parametric paired bootstrap F-test indicates

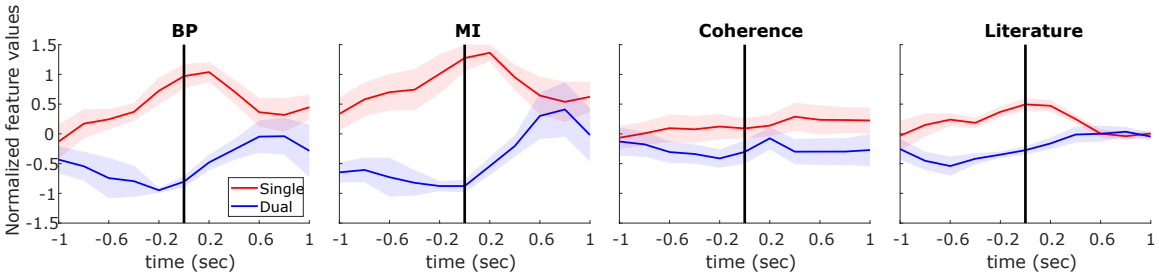



**Figure 2.9.** The time-course of the signature values in the number-copying data set. The solid lines are mean, and the shaded areas indicate the standard error across subjects. Trials were aligned to the moment when the first digit of a series of numbers was entered (the black vertical line).

a significant difference, some box plots show strongly overlapping distributions between the two conditions. For example, the MI signatures shown in Fig. 2.6 strongly overlapped in two setups, while the corresponding p-value is less than 0.01. This implies that the distributions of the feature values vary across participants even though the trend is consistent with changes in workload. Therefore, a transfer learning method is required to reduce the variations in the baseline across subjects if one wants to build a plug-and-play workload-monitoring system.

Finally, we checked whether the workload-related signatures we found correlated with the subjects' behavior over time. Fig. 2.9 shows the dynamics of the BP, MI, Coherence, and the literature-suggested signatures in the number-copying data set. We first aligned all the trials to the moment when the first digit of each group was entered and then calculated the signature values within a 1-sec sliding window stepped at 0.2 seconds (centering at the onset time). Trials were averaged across subjects, and ten subjects' mean and standard errors were plotted.

Fig. 2.9 clearly shows that the BP, MI, and literature signatures increase before the onset of entering the first digit under the single-monitor setup and reach their peaks roughly around the onset time. The findings support our hypothesis that the subjects began retrieving numbers before entering them. However, we see a much smaller amplitude of the curve of the coherence signature, indicating that it doesn't have a clear trend. Although the coherence signature didn't correlate with the behavior well, it might be a slowly-reacting response that maintained at higher values during the whole blocks of the single-monitor setup since we also

see the strong significance in the Sternberg data set. The signatures in the dual-monitor setup, on the other hand, remain relatively flat before the onset of typing. We also can see an increase in BP, MI, and literature signature values of the dual-monitor setup after the time of tying. Because we don't expect any behavioral changes to coincide with the onset time in this situation, more research is required.

In summary, this work proposed a new neuroergonomics study that investigates the memory workload in a real-world setting. Our analysis demonstrated that band power, mutual information, and coherence features are effective in distinguishing between high and low memory workload states. Furthermore, the data-driven signatures are consistent across both the ecologically valid number-copying task and the standard Sternberg task from a previous study. The findings suggest that EEG recording might be used as an objective measure for human factor investigations in real-world scenarios.

Chapter 2, in full, has been submitted for publication of the material as it may appear in the *Journal of Neural Engineering*. The dissertation author was the primary investigator and author of this paper.

# Chapter 3

# Human Liking Responses

## 3.1 Introduction

Neuromarketing is a field that uses behavioral and cognitive science to study consumers' decision-making processes when purchasing goods. Like the advantage we discussed in the field of neuroergonomic, neuromarketing provides objective metrics of consumer responses to marketing stimuli or products without being influenced by a subjective bias that might be caused by the social situation or context [7, 49, 50, 51]. Various tools are used for neuromarketing, including non-brain activity recording such as galvanic skin response, facial expression, etc., and brain activity recording such as EEG, EMG, fNIRS, fMRI, etc. [52]. While measuring non-brain activities is usually more accessible, the measurement of brain activities provides the direct assessment of consumer natural or unconscious responses to the marketing stimuli without delay [51] Among the tools for brain activity recording, EEG is one of the most suitable modalities for studying the neuromarketing outside the laboratory and in real-world environments due to its relative affordability and high portability.

Previous neuromarketing studies that use EEG have mainly focused on analyzing brain responses to TV commercials or advertisements, which are video-based stimuli that make it difficult to capture instant brain responses to products themselves [53, 50, 54, 55, 56, 57]. Nonetheless, several studies have analyzed the EEG responses in response to product images presented to subjects [58, 59, 60]. For example, [59] studies subjects' EEG responses after

showing them images of shoes. [58] analyzed the EEG signals when participants selected clothing items from an online store based on their likes and dislikes. [60] built a predictive model based on EEG for the subject's choice of preferred consumer products like mice, USB sticks, or desk lights. However, in these studies, the product images were presented one at a time, using a fixed-viewing design . Although this is a more controlled method for studying the fundamental EEG reactions to viewing products and the onset time of each stimulus can be well-synchronized, it does not replicate real-world shopping experiences. In the real world, customers typically view multiple goods simultaneously in the free-viewing condition so they can compare them side-by-side. As a result, if it's not obvious which product has the consumer's attention, it can be difficult to link EEG responses to particular image stimuli. Combining EEG with eye-tracking is a promising approach to address this issue, as eye-tracking can help tag the event of consumer attention shifting between products for EEG analysis. However, the analysis combining these two modalities for neuromarketing studies is still in the exploratory phase. [61] is one of the earliest studies that combined EEG and eye-tracking to study neuromarketing. However, the stimuli used in the study were cartoon images of crackers rather than real product images. While [62] and [63] also combined EEG and eye-tracking to investigate brain activities during a real-world shopping experience, these two studies were primitive and lacked constructive results. Overall, combining EEG and eye-tracking is a promising approach, but more research is needed to explore its potential.

In our study, we designed an experiment where participants had to choose their preferred product after comparing several options in a free-viewing condition. This experimental design resembles merchandise comparisons by an online shopper. Our study is one of the few that uses eye-tracking and EEG to investigate human liking reactions in a practical environment.

## 3.2 Methods

### 3.2.1 Experiments

To thoroughly investigate the EEG features of human liking response, we implemented a two-stage experiment as illustrated in Fig. 3.1. The first stage aimed to capture basic event-related liking responses to stimuli of product images in the fixed-viewing condition. In this stage, participants were presented with a series of laptop images one at a time and instructed to rate their liking on a scale from 1 to 5 based on the appearance of the product. They used the corresponding numbers on the keyboard to rate the images, with higher scores indicating greater preference. Following each rating, the image disappeared and a blank screen appeared for a random period between 0.5 to 0.7 seconds before the next image stimulus appeared to avoid steady-state evoked potentials. Participants could take a break once every 15 trials and resume the experiment at their own pace. Each participant completed a total of 450 trials, with each trial presenting a unique image stimulus of a laptop during this stage.

In the second stage of the experiment, participants were presented with groups of nine images at a time arranged in a $3\times3$ grid to select their favorite image in the group by clicking on it with a mouse. More specifically, participants were asked to select the image that best matched their highest rating score based on the same criteria used in the first stage. Once a selection was made, the group of images disappeared, and a blank screen appeared for a fixed duration of 0.5 seconds before the next group of images was presented. Participants were allowed to take a break every five groups and proceed at their own pace. A total of 50 groups, consisting of 450 images in shuffled order, were presented, and the participants completed the same set of images used in the first stage. The aim of the second stage was to capture the liking responses to image stimuli under the free-viewing condition.

Participants completed the experiment while seated at a standard office desk and used a keyboard and mouse to complete the tasks. To record their EEG activity, participants wore a Quick-30 EEG cap from CGX Quick Systems, along with a Pupil Core eye-tracker from Pupil

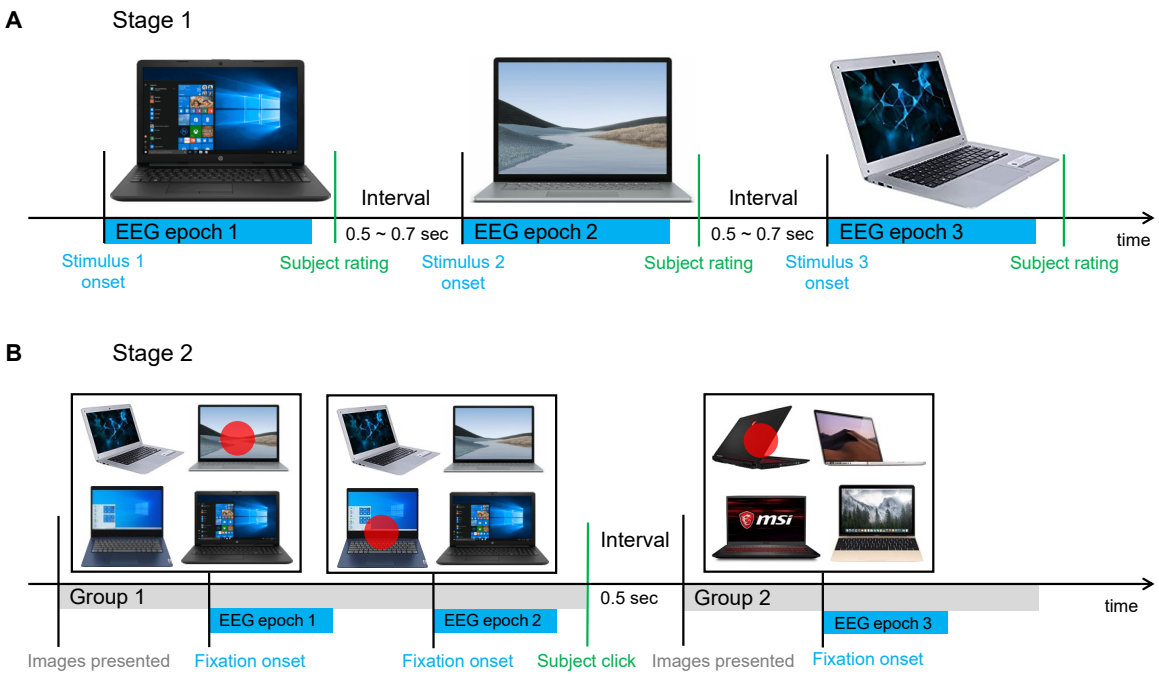



**Figure 3.1.** **A**. Flowchart depicting the procedure of the first stage of the experiment. **B**. Flowchart depicting the procedure of the second stage of the experiment. Note that a $2\times2$ grid of images was used for illustration purposes only, and the actual stimulus consisted of a $3\times3$ grid of images.

Labs. Although the eye-tracker was not used during the first stage of the experiment, it was set up beforehand to avoid any significant headset movement that might occur if set up during the experiment. Before the second stage, eye-tracking calibration was conducted. During the second stage, the Pupil Labs software [64] was used to detect fixations in real-time, and the fixation time and position on the screen were recorded along with the EEG signals.

A total of 14 individuals (9 females and 5 males) with ages ranging from 24 to 52 (first quartile: 27.0, median: 29.0, third quartile: 36.3) participated in the experiment.

### 3.2.2 EEG Preprocessing

The EEG signals recorded from participants were preprocessed using the offline version of the REST package [65, 66]. The preprocessing steps included applying a finite impulse response (FIR) bandpass filter between 1 and 15 Hz, performing online recursive independent component analysis [66] to obtain independent components, and using EyeCatch to remove

the artifacts of eye components [67]. For the first stage of the experiment, EEG epochs were extracted from 0.5 seconds before to 1 second after the onset time of each trial's image stimulus. Each EEG epoch was further baseline-corrected by subtracting the mean of the first 0.5 seconds (i.e., the time prior to the image onset) on a channel-wise basis. In the second stage, EEG epochs were extracted from 0.5 seconds before to 1 second after the onset of each fixation. Fixations with a movement of visual angle under the dispersion threshold for at least 300 milliseconds were detected, and the start of each epoch represented 200 ms before the actual start of the fixated gaze. Subsequent fixations without a gap from the first fixation were discarded. The EEG epochs in the second stage were also baseline-corrected by subtracting the mean of the first 0.2 seconds of each epoch.

### 3.2.3 Data Labeling

Each epoch in the first stage was associated with a rating score, which was subsequently converted into binary labels of low and high preference. However, as we did not calibrate the participants' ratings beforehand, a score of 3 may not always signify a neutral score (i.e., no liking or disliking) for every participant. To obtain the individual's neutral score, we examined the distribution of the scores given by each participant and defined the most frequently given score in the range of 2 to 4 as the neutral score. We selected the most frequently given score as the neutral score under the assumption that participants did not have strong feelings toward most of the image stimuli. Then, epochs with a rating score higher than the neutral score were labeled as a high preference, and epochs with a rating score lower than the neutral score were labeled as a low preference. To avoid missing any labels, the neutral score was restricted to the range of 2 to 4.

Epochs in the second stage were first associated with the ID of the image stimulus that was present at the position of the corresponding fixation. We then labeled each epoch based on the label given to the associated image in the first stage of the experiment.

**Algorithm 1.** Epoch shifting

Select the occipital channels of the unshifted epochs.
Define the $i$-th epoch (with the subset of channels) as $\mathbf{x}_i \in \mathrm{R}^{N_c \times N_t}, i = 1 \ldots N_e$, where $N_c$ is the number of selected channels, $N_t$ is the number of timestamps, and $N_e$ is the number of epoch.
Get the channel covariance matrix $\mathbf{C} = \frac{1}{N_e} \sum_{i=1}^{N_e} \mathbf{x}_i \mathbf{x}_i^T$, where $\mathbf{C} \in \mathrm{R}^{N_c \times N_c}$.
Get the eigenvector with the largest eigenvalue $\mathbf{e} = \mathrm{eigh}(\mathbf{C})$, where $\mathbf{e} \in \mathrm{R}^{N_c}$.
Project each epoch to the component as $\mathbf{p}_i = \mathbf{e}^T \mathbf{x}_i$, where $\mathbf{p}_i \in \mathrm{R}^{N_t}$.
Initialize the window start for each component $\mathbf{t} \in \mathrm{R}^{N_e} = 200$ ms for all elements.
**for** $iteration = 1, 2, \ldots, \mathrm{Max}_{\mathrm{iter}}$ **do**
    Get the mean across windows of each component $\mathbf{m} = \frac{1}{N_e} \sum_{i=1}^{N_e} \mathbf{p}_i[\mathbf{t}_i : \mathbf{t}_i + 300 \text{ ms}]$
    **for** $i_e = 1, 2, \ldots, N_e$ **do**
        Find $t'$ between [100, 250] ms such that $r(\tilde{\mathbf{m}}, \mathbf{p}_{i_e}[t' : t' + 300 \text{ ms}])$ is maximized, where $r(,)$ represents the Pearson correlation coefficient, and $\tilde{\mathbf{m}}$ is $\mathbf{m}$ excluding $\mathbf{p}_{i_e}$.
        Update the window start $\mathbf{t}_{i_e} = t'$.
    **end for**
**end for**
Return the updated window start $\mathbf{t}$.

### 3.2.4 Epoch Shifting

As mentioned above, a fixation was detected if it lasted longer than 300 ms. However, it should be noted that the minimum fixation duration can vary between 150 ms to 350 ms [68]. Consequently, epochs aligned with the onset of fixations may not be ideal for exploring time-locked responses. Furthermore, [69] also found that epochs based on fixation onset time are often affected by noise, making it difficult to identify specific time-locked patterns. To address this issue, they proposed the MLRd algorithm to shift and align epochs using P100 fixation-related potentials when slightly inaccurate fixation onsets are present.

Inspired by their approach, we also proposed an epoch-shifting algorithm that maximizes the peak signals of the averaged O1 and O2 channels $\tilde{1}00$ ms after the onset fixation (known as the P100 fixation-related potential, or FRP). The pseudo-code of our proposed epoch-shifting method is described in Algorithm 1. As the fixation-related potential mainly localizes in the occipital area, we extracted occipital channels (PO3, PO4, PO7, PO8, O1, O2 in this study) from each epoch. We then obtained the spatial filter that maximizes the variance of the occipital area

by solving for the eigenvector of the channel covariance matrix. We projected each epoch onto the component space using the spatial filter. Assuming that the FRP occurs around 100 ms after the onset of fixation (i.e., 200 ms of the epochs), the window of each component was initialized as [200, 500] ms of the epochs to include roughly 100 ms before and 100 ms after the FRP. We then shifted the start of the window of each component between [100, 250 ms] to keep the point at which the Pearson correlation coefficient between the new window and the mean across windows of other components was maximized. This procedure was repeated until the maximum iteration time (100 in this study) was reached or the window began to converge. We extracted each shifted epoch from the original signals by including the 500 ms before and 1 second after the corresponding updated window start. Finally, we shortened the shifted epochs so that the peak of the mean across all shifted epochs at Oz (average of O1 and O2) was centered at 100 ms. This epoch-shifting approach allowed us to get enhanced temporal signatures of liking responses.

### 3.2.5 Feature Extraction

We extracted features from the shifted epochs in the second stage and used them to train machine learning classifiers, exploring the potential use of these features in developing predictive models of like/dislike responses toward products.

To account for cases when participants viewed an image multiple times during selection, we combined epochs associated with the same image stimulus by averaging the preprocessed signals. Features were then extracted from each of the merged epochs. Prior studies have shown that frontal alpha asymmetry (defined as right frontal lobe minus left frontal alpha power) is associated with consumer responses to product attractiveness [70, 50], and that alpha band is associated with preference judgments [71, 50, 72, 60]. To capture these features, we extracted the alpha band power from selected channels AF3, F3, AF4, F4, Fz, and Cz for each epoch. In addition, the signals between 200 and 300 ms after image onset, a feature referred to as N200, was shown to correlate with subjective preferences in virtual shopping tasks [73, 60, 74], and was thus included as temporal features. Finally, we incorporated fixation-related features, such

as fixation revisit count and total fixation duration for each image stimulus. Fixation revisit count was defined as the total number of fixations reported for each image stimulus, while total fixation duration was the sum of the durations of all fixations for the image stimulus. These features were concatenated to form a 14-dimensional feature set (alpha power and N200 from 6 channels, plus 2 fixation-related features).

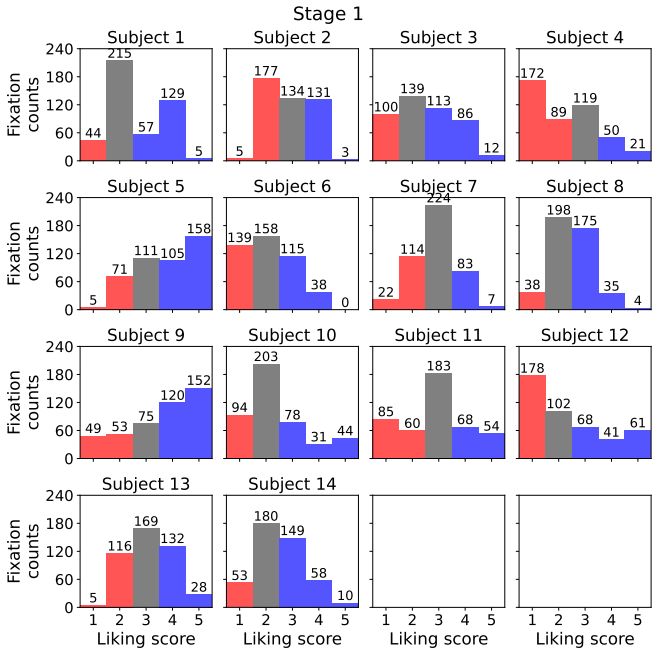
## 3.3 Results

In the analysis of the epochs from the first stage of the experiment, we focused on exploring EEG features associated with basic liking responses. To ensure that pressing keyboard keys did not influence these features, we examined the response times of each participant. Specifically, we calculated the means of the first quartile, median, and third quartile of response time for each participant, which were 1.2, 1.4, and 1.8 seconds, respectively. Based on these calculations, we determined that the 1-second epoch used in our analysis should not include the period when participants entered their rating scores using the keyboard.

To ensure the selection of an appropriate neutral score for each participant, we examined the frequency of each rating score received from each participant's image stimuli, as shown in Fig. 3.2A. The gray bars represent the number of epochs labeled as neutral, and the red and blue bars indicate the number of epochs labeled as low and high preference, respectively. For most participants, the method of determining the neutral score described in Subsection 3.2.3 was adequate, except for Subject 2. In this case, we manually selected the neutral score as 3 to avoid having too few low-preference epochs (only 5), which could have affected the validity of the results. Fig. 3.2B shows the frequency of epochs segmented based on the fixation onset in the second stage. As expected, the task was to select a participant's favorite, which resulted in fewer low-priority epochs overall, as participants tended to stick to higher-priority stimuli.

We then investigated the temporal features of liking responses by analyzing the mean of participant's averaged Fz and Cz signals across epochs of each label. Fig. 3.3A shows a

**A**




**B**

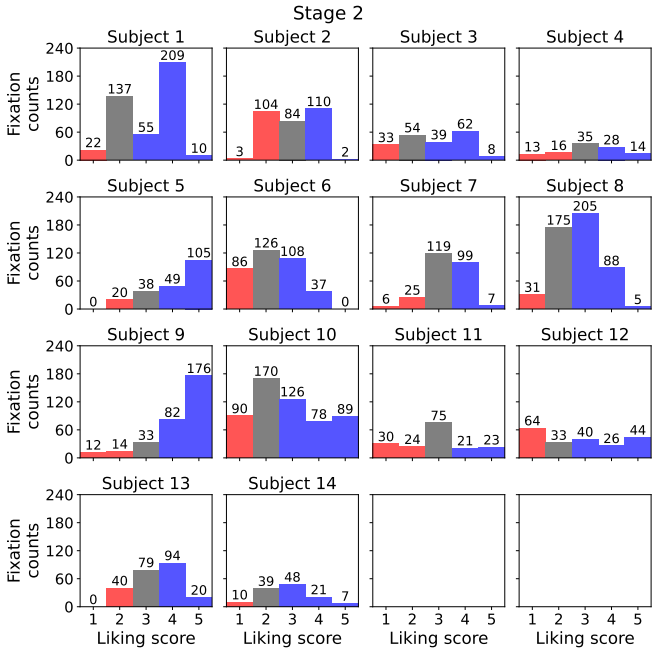
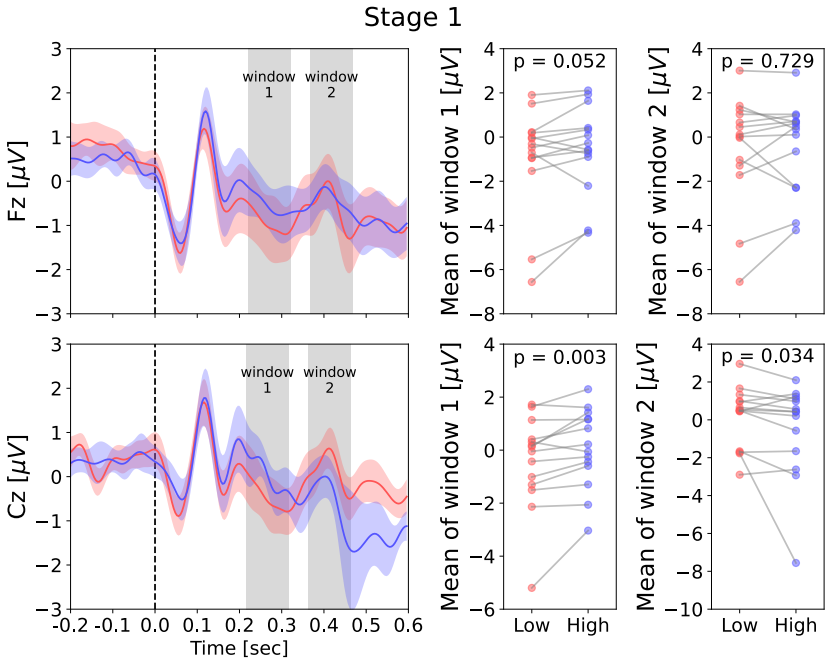



**Figure 3.2.** **A**. The frequency of rating scores given by each participant during the first stage of our experiment. **B**. The frequency of rating scores for epochs in the second stage of the experiment.

**A**




**B**

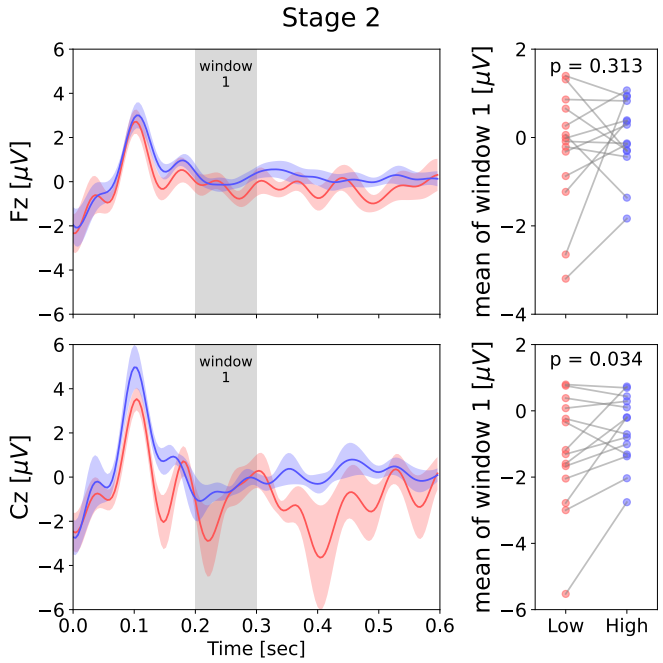



**Figure 3.3. A**. The temporal pattern of the epochs in the first stage, as well as the distribution of each participant's mean voltage during specific windows of interest. The numbers shown at the top of the scatter plot represent the p-value of the Wilcoxon signed-rank test. **B**. The temporal pattern of the epochs in the second stage.

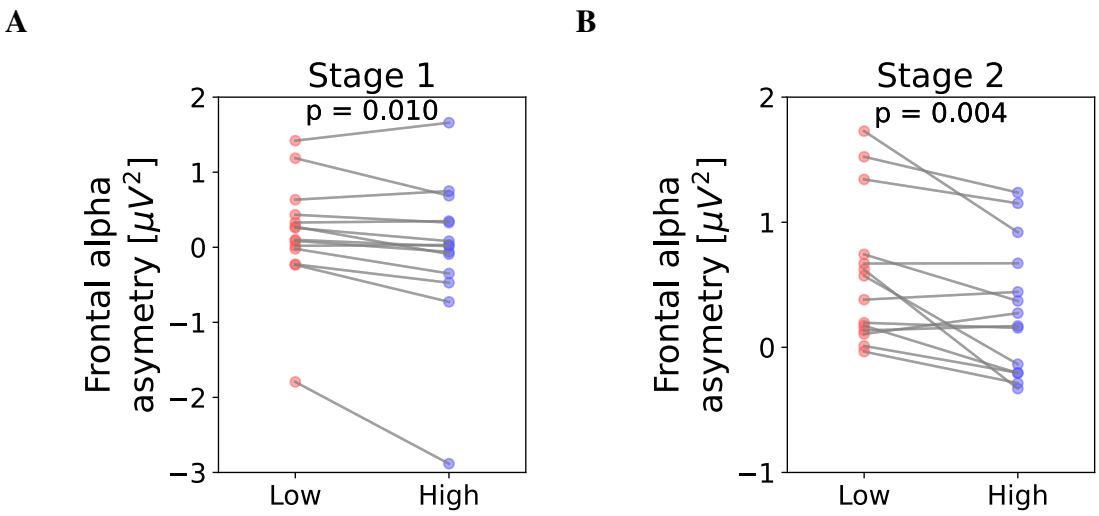



**Figure 3.4. A**. The distribution of each participant's mean frontal alpha asymmetry in the first stage. The numbers shown at the top of the scatter plot represent the p-value of the Wilcoxon signed-rank test. **B**.The distribution of each participant's mean frontal alpha asymmetry in the second stage.

noticeable peak at around 100 ms after the stimulus onset in both low and low preference conditions. Moreover, a stronger negative peak around 250 ms was present in low-preference epochs at both channels, and a stronger positive peak around 400 ms was apparent in low-preference epochs at Cz. We then calculated the mean voltage over the intervals [100, 200] ms and [250, 350] ms following the detected P100 peak time of Cz for each participant and compared the distribution pairwise in the right two panels of Fig. 3.3A. The Wilcoxon signed rank test revealed a significant difference in the mean voltage of the first window at both channels and a significant difference in the second window at Cz.

We further examined if comparable temporal patterns were present in the second stage, as shown in Fig. 3.3B. Notably, we shifted the epochs in the second stage to ensure that the peak was precisely at 100 ms. The results showed that both channels exhibited a similar P100 peak, and an N200 peak was present at Cz. Moreover, the distributions of the participant's mean voltage of the window [100, 200] ms after the peak of both labels were also significantly different. However, the temporal patterns after 300 ms differed from those in the first stage.

**Table 3.1.** The balanced accuracy of the 10-fold cross-validation within each participant.

| Subject Id | Balanced Accuracy | Subject Id | Balanced Accuracy |
|---|---|---|---|
| 1 | 73.22% | 8 | 59.03 % |
| 2 | 62.90% | 9 | 70.76 % |
| 3 | 56.42% | 10 | 53.79 % |
| 4 | 65.42% | 11 | 55.19 % |
| 5 | 69.14% | 12 | 60.85 % |
| 6 | 60.93% | 13 | 67.77 % |
| 7 | 61.69% | 14 | 47.05 % |
| | | Average | 61.73 % |

We calculated frontal alpha asymmetry (FAA) To investigate frequency-domain features. For each epoch, we subtracted the averaged alpha power across AF3 and F3 from that across AF4 and F4. We obtained the means of FAA across epochs for each participant for both low and high preference conditions. The alpha power was calculated over the window [30, 530] ms after the P100 peak time of the temporal signals. Fig. 3.4 shows the distributions of the participant's mean FAA across epochs. The results demonstrate a robust and significant shift in FAA from low preference epochs to high preference epochs in both stages.

Finally, we explored the possibility of using EEG features to train a predictive model for classifying like/dislike responses. We combined shifted epochs corresponding to the same image stimulus and extracted both frequency- and time-domain features (see Subsection Feature Extraction). For calculating the alpha asymmetry, we selected the same window range used in Fig.3.4B ([130, 630] ms). The time-domain feature capturing the N200 was calculated as the mean voltage over the window [200, 300] ms, which is aligned with the window used in Fig.3.3B. We trained a linear discriminant analysis (LDA) classifier with the EEG features, along with the fixation features, for binary classification of low or high preference. We employed 10-fold cross-validation within each participant to assess the model's performance. Fig. 3.2B provides information about the number of raw epochs for each label in each participant before epochs corresponding to the same image stimulus were combined. Due to the imbalanced data, we

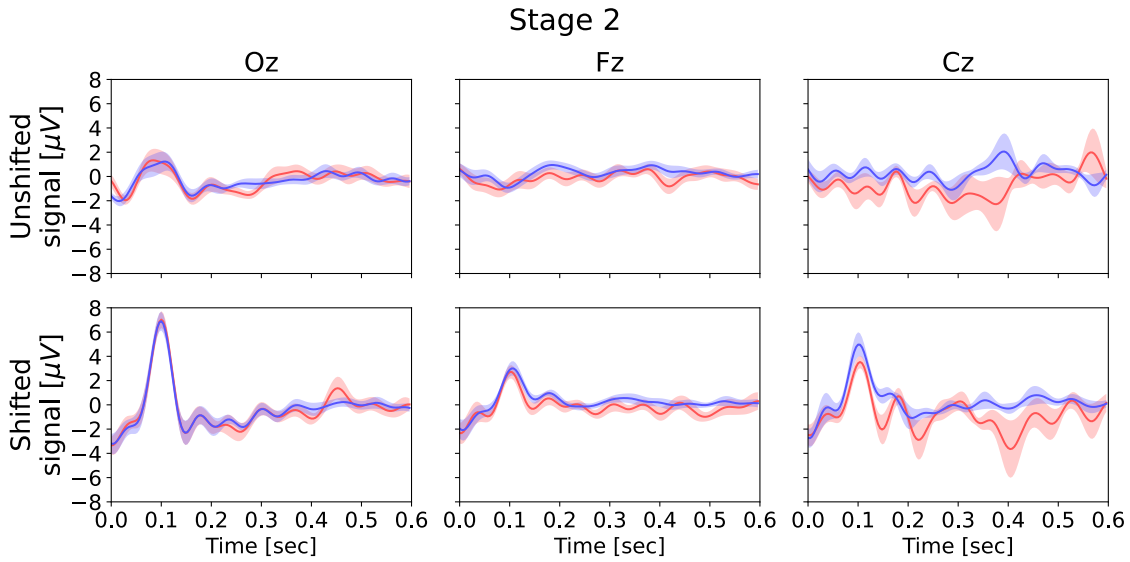



**Figure 3.5.** The temporal patterns of shifted epochs versus unshifted epochs.

used balanced accuracy (i.e., the average of sensitivity and specificity) as the metric for model performance. Table 3.1 shows the results of the analysis. The results indicate that the model achieves a balanced accuracy of 60% for most participants, with the best-performing participant reaching 70%.

## 3.4 Discussion

The temporal patterns shown in Fig. 3.3 are consistent with findings reported in [60], which also investigated the temporal dynamics of preference responses. Our results confirm the existence of a stronger N200 peak in low preference epochs. While previous studies have also reported the involvement of P300 in preference responses [75, 76], they reported stronger P300 in the high preference condition, which contradicts our results. In our study, the stronger P300 in the low preference epochs might be caused by the less frequent occurrence of low rating stimuli (see Fig. 3.2A) as stronger P300 is a typical signature of the odd-ball task.

The significant difference in the mean voltage of the first window in both stages further confirms the robustness of the signature. As for the mean voltage of the second window, inconsistent patterns were observed, and it might be due to the influence of subsequent fixations

on the signals after 300 ms in the second stage.

To validate whether the epoch-shifting method was necessary for improving the temporal signatures, we compared the unshifted epochs with the shifted epochs in addition to comparing the temporal patterns between the first and second stages. Fig. 3.5 compares the averaged shifted epochs and unshifted epochs. It should be noted that we put the FRP peak of unshifted epochs at 100 ms for easier comparison, as we forced the peak of FRP at Oz of the shifted epochs to be located at 100 ms. We observed that in the unshifted epochs, the P100 FRP was still visible at Oz but not at Fz and Cz, which is consistent with previous works [77, 69]. However, the P100 and N200 at Cz, which were the signatures found in the first stage (see Fig.3.3A) and reported in [60], were not observable. In contrast, With the epoch-shifting, the fixation-related potential at Oz was more prominent, and the P100 signature at Fz and Cz became apparent. These findings suggest that our epoch-shifting method for aligning P100 was effective in improving the temporal signature.

The consistent results in Fig. 3.4 confirm the robustness of frontal alpha asymmetry (FAA) as a signature of liking response. Nevertheless, it is important to acknowledge that while the contrast in FAA between the two labels was consistent within each participant, there was substantial individual variability in the FAA distribution. This suggests that calibration may be needed when developing a preference-predicting model that can be applied to new individuals. Alternatively, transfer learning techniques could be used to mitigate inter-individual variability and build a calibration-free model.

The classification results presented in Table 3.1 demonstrate that EEG features combined with fixation features can be useful for building a preference-predicting model for most participants, as the model performance is robustly higher than the chance level. However, there is still big room for improvement before achieving satisfactory performance in the real world. Nonetheless, our analysis demonstrated that the feature values, when averaged across multiple epochs, were significantly different between liked and disliked image stimuli, indicating that the effect is present. Nevertheless, better hardware or signal-processing techniques are necessary to

improve the signal-to-noise ratio.

To summarize, our study has validated several signatures previously reported in the literature related to liking responses, both in the time and frequency domains, and has demonstrated that we can also observe these signatures in a free-viewing condition. The consistent signatures observed in both stages have validated our methods of time-locking EEG signals with fixation onset and shifting the epochs. These findings indicate that the observed signatures can be utilized to develop a predictive model for like responses, but further improvement is still needed.

Chapter 3, in full, is currently being prepared for submission for publication of the material. The dissertation author was the primary investigator and author of this paper.

# Chapter 4

# Transfer Learning for Passive BCIs

## 4.1 Introduction

Studies have shown that electroencephalography is one of the most reliable modalities for predicting cognitive states. Many previous studies [78, 79, 80, 15] demonstrated that the spectral power of EEG, which pertains to power in different sub-bands of EEG, can be highly correlated to different mental workloads or memory states. However, the results from previous studies also showed that EEG band power could exhibit considerable variations, which may be caused by different contextual factors such as individual differences or different times within an individual. This human variability typically has a negative impact on the performance of the machine learning models used in the passive BCIs. In general, acquiring calibration data for a new session is necessary to develop a predictive model for human cognitive states that performs adequately, as it aids in learning the distribution of EEG features and optimizing the parameters in the model. However, this calibration process can be time-consuming and impede the practicality of passive BCIs..

To overcome the challenge of calibration in BCIs, transfer learning techniques have been proposed. These methods attempt to leverage data from other domains, such as other sessions or users, to learn the appropriate data distribution for a new session. A recent review by Zhang et al. [19] evaluated the use of transfer learning techniques in the BCI field. However, most of the transfer learning approaches for EEG data have focused on traditional BCI paradigms

like SSVEP, MI, and ErP, rather than passive BCIs. Additionally, transfer learning techniques for passive BCIs typically involve instance-transferring or model-transferring based methods [79, 81, 82, 83]. Methods that transfer instances assign weights to data from source domains based on their similarity to the target domain's data. Similarly, model transferring techniques give weights to models trained with data from multiple source domains and ensemble the models' output using weighted voting. This study is one of the few to explore feature-transferring techniques for transfer learning in passive BCIs. As an initial investigation in this field, we proposed simple and intuitive feature-transferring methods and evaluated the transferability of EEG band power features. Specifically, we examined the transferability of different EEG band powers in cross-session, cross-subject, and cross-task paradigms.

Moreover, our study considers a transfer learning scenario that reflects the practical constraints of acquiring calibration data. Unlike previous studies that assume access to calibration data of all task conditions (e.g., data of low and high mental workload), we assumed that only baseline (low mental workload state) data were available, which is more realistic and applicable in real-world scenarios. For example, a passive BCI that monitors office workers' cognitive load during their workday may find it challenging to provide tasks that genuinely induce cognitive load during calibration. Inducing additional cognitive load before work can also reduce the user's willingness to use the device. Therefore, it is more practical for the device to require only baseline data, such as the user relaxing or performing easy tasks, which can be readily obtained in real-world settings.

## 4.2 Methods

### 4.2.1 Experiments

This study used a block-design experiment to assess mental workload. Figure 4.1 depicts the flow of the experiment. During a recording session, participants were instructed to complete three tasks: mental math, Stroop, and image caption tasks, each consisting of an easy and a hard

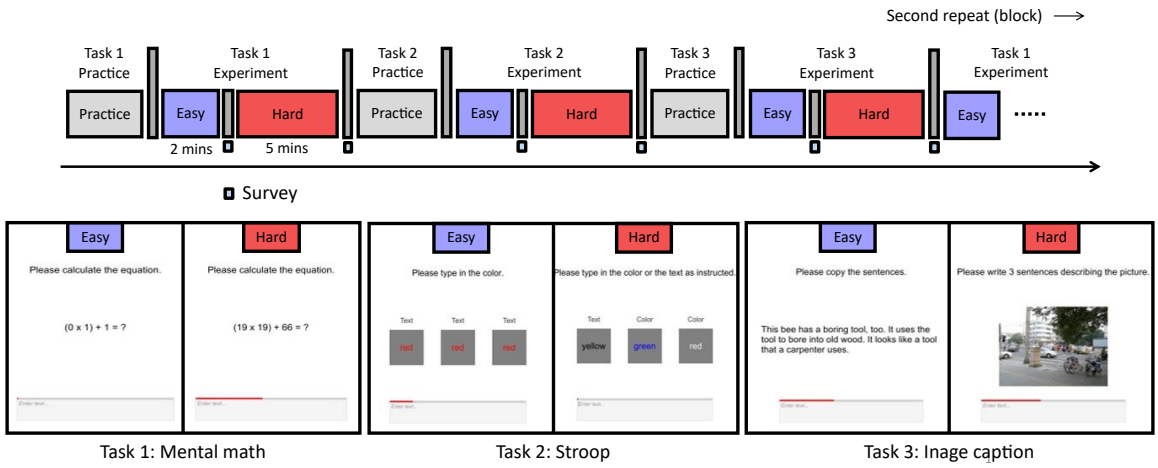



**Figure 4.1.** The procedure of the cognitive load experiment.

mode. A session always starts with the easy mental math task, followed by the hard mental math task. After the mental math task, participants performed the Stroop task in the easy mode, then in the hard mode. Finally, after completing the Stroop task, they finished the image caption task, which also had an easy and a hard mode. The completion of the three tasks was considered a single block. Following the first block, they were asked to repeat the same sequence of tasks in a second block. After each difficulty mode for each task, participants rated the mental demand of the task on a scale of 1 to 10, with 1 indicating no mental demand and 10 indicating maximum mental demand.

During the hard math task, participants were instructed to solve mental math equations involving the multiplication of a number between 1 and 20 with another number between 1 and 20, and then adding a number between 1 and 100. The easy math task required participants to solve similar equations, but the numbers were limited to 0 or 1. For the hard Stroop task, participants are presented with a series of color names written in different ink colors (e.g., the word ”RED” written in blue ink). The task requires the participant to identify the color of the ink in which the word is written while ignoring the actual word itself. The easy Stroop task had matching color names and text colors. During the hard image caption task, participants were shown an image stimulus and asked to come up with a short story describing the image. The easy version of the task required the participants to simply look at and copy a random paragraph

from a child's story.

Throughout the experiment, participants sat comfortably at a standard office desk and used a keyboard to complete the tasks on a computer monitor. The participants' EEG and ECG were recorded while they performed the tasks. EEG signals were recorded using a Saline-based 24-channel EEG cap and the SMARTING mobile EEG amplifier (mBrainTrain, Belgrade, Serbia). The ECG was measured and recorded using the ProComp Infiniti system.

After completing the first session, participants were asked to return for the second session on another day. A total of 14 individuals (6 females and 8 males, 11 aged 35-44 years old) participated in the experiment, and 10 of them returned for the second session.

### 4.2.2 Data Preprocessing

The EEG signals were subjected to a bandpass filter ranging from 0.5 to 50 Hz, and average re-referencing was performed. To remove artifacts, the ASR was calibrated with the eye-close data of each session, using a cutoff parameter of 10. Following the artifact reduction process using ASR, we utilized Independent Component Analysis (ICA) [84] to decompose the EEG data into independent components. We then classified these components using ICLabel [65] and removed any components related to eye artifacts. The resulting EEG signals for each task and condition were divided into non-overlapping epochs of ten seconds. This process resulted in roughly 24 "easy"-mode epochs and 60 "hard"-mode epochs for each task in each session.

### 4.2.3 Evaluation Schemes

The main objective of this research was to assess the effectiveness of transfer learning models in predicting whether the participants were in a state of low cognitive load (i.e., performing a task in easy mode) or high cognitive load (i.e., performing a task in hard mode). The testing data and labels used to evaluate the model's performance were the ten-second epochs and the corresponding binary labels from the target session. This study evaluated three different schemes for cross-domain transfer learning: cross-session, cross-subject (i.e., cross-participant),

and cross-task scenarios.

In the cross-session scheme, the second-day session for each participant was considered the target domain, while the first-day session of the same participant was considered the source domain. This scenario represented a one-to-one domain transfer.

In the cross-subject scheme, the first-day session of each participant was considered the target domain. For the case of each participant being the target domain, the first-day sessions of all the other participants were considered the source domain. This scheme represented a multi-to-one domain transfer.

The cross-task scheme aimed to explore the feasibility of transferring data from a well-established task to a more real-world task. In this scheme, the image caption task in the second-day session of each participant was considered the source domain, while the mental math task in the first-day session of the same participant was considered the target domain. It is worth noting that in both the cross-session and cross-subject schemes, we extracted and evaluated the mental math task and the image caption task. However, the target and source domains always involved the same task, and as a result, the results of evaluating each task will be presented separately.

As mentioned in Section 4.1, this study considered a real-world transfer learning setting where only data of low mental workload state (i.e., the easy-mode data in the present study) were accessible during calibration, which were used for fine-tuning the models. We only used the data from the first block of the target session for fine-tuning to prevent overfitting that could arise from the temporal auto-correlation of EEG signals within the same block. We reserved the data from the second block for testing. Specifically, the epochs of both the easy mode and hard mode conditions from the second block were used as the testing data. As for the first block, the easy-mode epochs were used as fine-tuning data, while the hard-mode epochs were used to train an unattainable model with the upper-bound performance, as explained in more detail below.

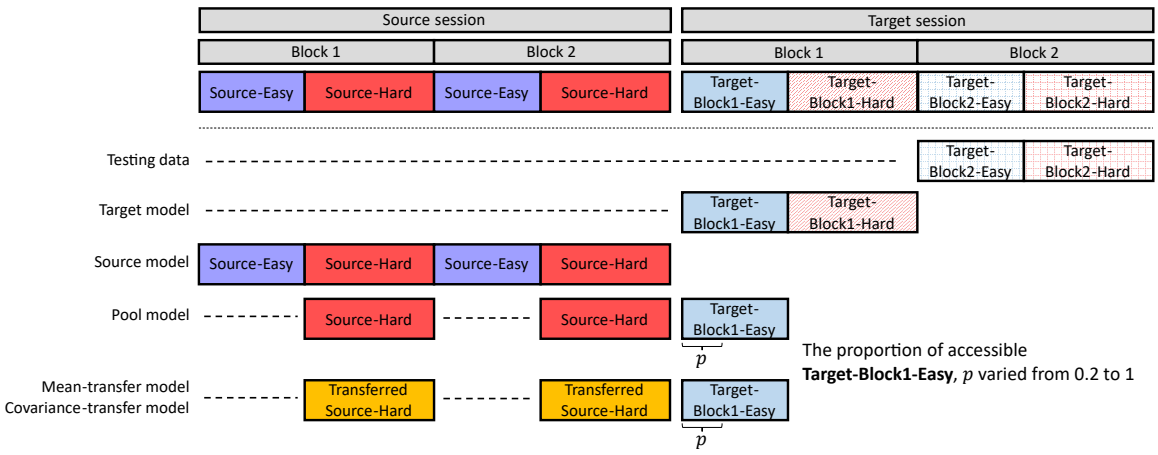



**Figure 4.2.** The illustration of the scheme used to split data from a source session and a target session for validation and for training different types of models.

### 4.2.4 Transferring Models

As mentioned in the preprocessing section, band power features were extracted from each epoch. To simplify the description, we introduce the following terms for different groups of band powers:

- **Source-Easy**: The band powers of easy-mode epochs from both blocks of the source session.
- **Source-Hard**: The band powers of hard-mode epochs from both blocks of the source session.
- **Target-Block1-Easy**: The band powers of easy-mode epochs from the first block of the target session.
- **Target-Block2-Easy**: The band powers of easy-mode epochs from the second block of the target session.
- **Target-Block1-Hard**: The band powers of hard-mode epochs from the first block of the target session.

- **Target-Block2-Hard**: The band powers of hard-mode epochs from the second block of the target session.

We did not separate the band powers from the first and second blocks of the source session, so they don't need separate terms.

We defined five types of models with different training data to compare our transfer learning methods. The scheme used to split data from a source session and a target session for validation and for training different types of models was illustrated in Fig. 4.2 and described as below:

**1. Target model**

The target model was trained with **Target-Block1-Easy** and **Target-Block1-Hard**. This model represents the ideal scenario where data from both labels are attainable during the calibration, and its performance serves as an upper bound for all other models. Note that, in our real-world setting, **Target-Block1-Hard** is not accessible.

**2. Source model**

The source model was trained solely with **Source-Easy** and **Source-Hard**, without any fine-tuning from the target session. This model represents the case where a pre-trained model from the source session is directly used, and no calibration process is involved. We expect the performance of this model to be lower than other models, as it does not consider any differences in the data distributions between the source and target sessions.

**3. Pool model**

The pool model was trained with **Source-Easy** and a portion of **Target-Block1-Easy**. This model uses the most naive transferring method, pooling the calibration data with the source data.

**4. Mean-transfer model**

This model applied a mean transfer method on **Source-Hard** and trained on the transferred samples together with a portion of **Target-Block1-Easy**. Fig. 4.3 illustrates the process of the mean transfer method. The mean transfer method assumes there is a transformation between the distribution of **Target-Block1-Hard** and **Source-Hard**, and this transformation is consistent with that between the distribution of **Target-Block1-Easy** and **Source-Easy**. Based on the assumption, the method obtains a mean transfer vector by subtracting the mean of **Source-Easy** from the mean of **Target-Block1-Easy**. The estimated **Target-Block1-Hard** is then obtained by transferring **Source-Hard** with this mean transfer vector. Finally, the model can be trained on the target domain with **Target-Block1-Easy** and the estimated **Target-Block1-Hard**.

**5. Covariance-transfer model**

This model builds upon the mean transfer model by adding a covariance transfer process before the mean transfer process. My previous work [85] has demonstrated that a linear channel-wise projection matrix can reduce individual variability in time-locked EEG signals. This work aimed to expand this concept to non-time-locked EEG signals. For time-locked signals, we assume that an epoch from the source domain $\mathbf{x}_s$ can be transformed by a transfer matrix $\mathbf{P}$ to be comparable to an epoch from the target domain $\mathbf{x}_t$, expressed as $\mathbf{x}_t = \mathbf{P}\mathbf{x}_s$, where $\mathbf{x}_s, \mathbf{x}_t \in \mathrm{R}^{N_c \times N_t}$ and $\mathbf{P} \in \mathrm{R}^{N_c \times N_c}$ ($N_c$ and $N_t$ represent the number of channels and timestamps). In extending to non-time-locked signals, linear channel-wise projection is not applicable. Therefore, we explored the possibility of finding the transfer matrix $\mathbf{P}$ in the covariance space using the equation $\mathbf{x}_t\mathbf{x}_t^T = (\mathbf{P}\mathbf{x}_s)(\mathbf{P}\mathbf{x}_s)^T = \mathbf{P}\mathbf{x}_s\mathbf{x}_s^T\mathbf{P}^T$, which can also be expressed as $\mathbf{C}_t = \mathbf{P}\mathbf{C}_s\mathbf{P}^T$. Then $\mathbf{P}$ can be solved as $\mathbf{C}_t^{\frac{1}{2}}\mathbf{C}_S^{\frac{-1}{2}}$.

In the covariance-transfer model, we learn $\mathbf{P}$ using the epochs from **Source-Easy** and **Target-Block1-Easy**, and apply it to both the **Source-Easy** and **Source-Hard** epochs. Then, we calculate the band power from these transferred epochs and proceed with the remaining steps, as in the mean transfer model.

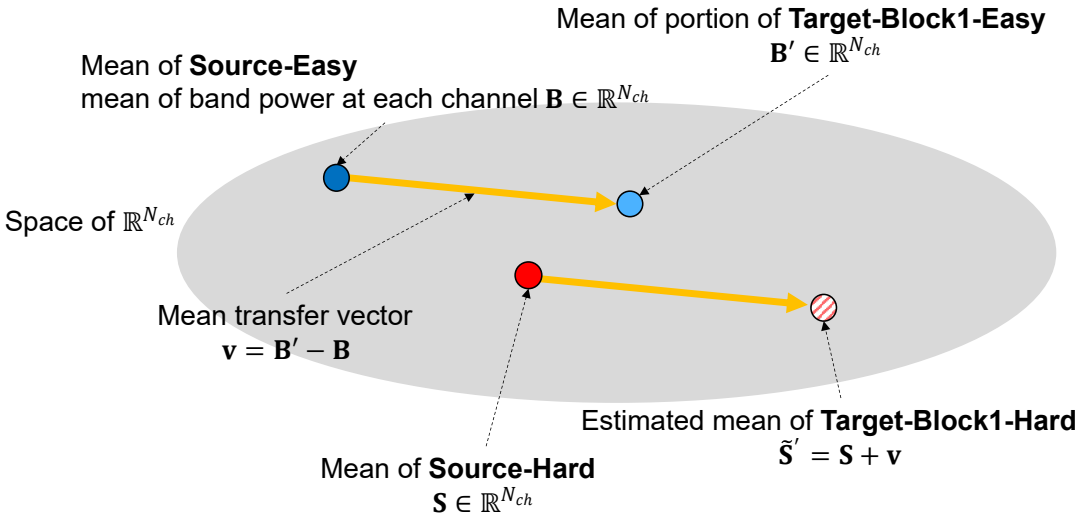



**Figure 4.3.** The illustration of the mean transfer method.

It is worth noting that in the pool model, mean-transfer model, and covariance-transfer model, the portion $p$ of **Target-Block1-Easy** used as fine-tuning data varied from 0.2 to 1.0 to examine the model's performance with different lengths of the calibration process. Additionally, in the cross-subject scheme, the multi-to-one domain transfer was employed, resulting in multiple source domains. In the pool model, raw **Source-Hard** data from each source domain was pooled together. In the mean-transfer and covariance-transfer models, transferred **Source-Hard** data from each source domain was pooled together.

To evaluate how changes in the training data distributions of the source domain impact classification performance, all five types of models used the Gaussian Naive Bayes classifier and were assessed by the testing data comprising both **Target-Block2-Easy** and **Target-Block2-Hard**.

## 4.3 Results

Fig. 4.4 shows each participant's mental demand rating scores, indicating the level of cognitive load required to complete the task. The scores were averaged across two blocks for each mode of the same task, and each session of a participant was plotted separately. The results

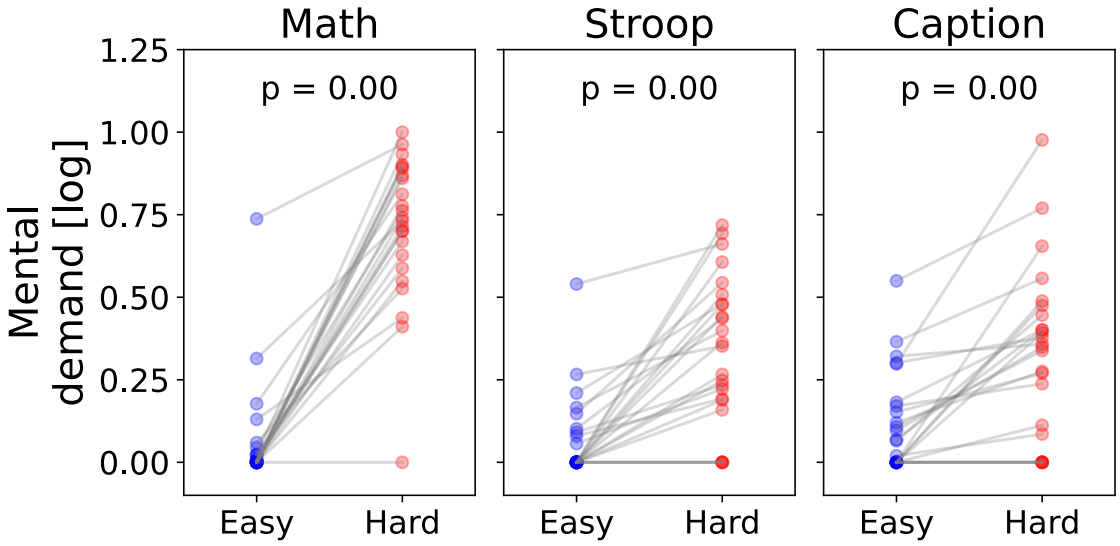



**Figure 4.4.** The rating scores of the mental demand for completing the task from each participant. The p-value of the Wilcoxon signed rank test is shown on top of each panel.

revealed that the hard mode of the three tasks imposed a greater cognitive burden on participants, as indicated by their self-reported mental demand scores. The differences in scores between the two modes were statistically significant, as determined by the Wilcoxon signed-rank test. While all three tasks exhibited significant differences in self-reported mental demand scores, the hard mode of the math and caption task had a higher average score than the Stroop task. Additionally, participants reported that the Stroop task only elicited a short burst of effort upon initially seeing the question and did not require sustained effort throughout the five-minute task. To focus our analysis on tasks that impose a more prolonged cognitive load, we selected the math and caption task, using one mode as a lab-controlled task and the other as a real-world task.

We first verified whether the ECG and EEG signals could reflect different cognitive load levels. Previous literature has commonly reported heart rate variability as the most relevant ECG feature. Fig. 4.5A compares each participant's average standard deviation of heart rate across epochs during the mental math task in two conditions. We also examined the efficacy of the heart rate variability feature when the signals were divided into short-window (10-second) and long-window (120-second) epochs. The most commonly reported mental workload feature

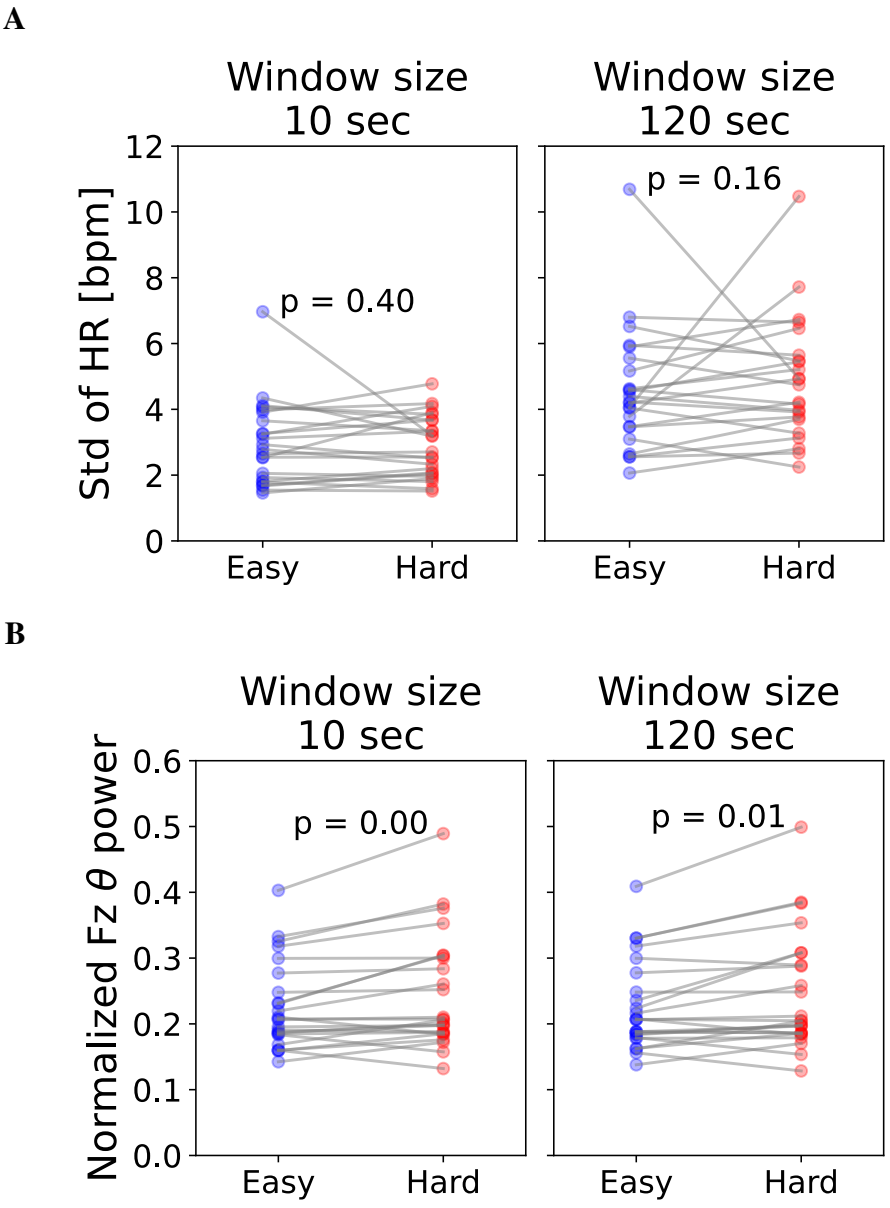



**Figure 4.5. A**. The standard deviation of each participant's heart rate × session during the mental math task in two conditions, where the ECG signals were divided into epochs with 10 seconds and 120 seconds window sizes. **B**. The normalized Fz theta power of each participant × session during the mental math task in two conditions, where the EEG signals were divided into epochs with 10 seconds and 120 seconds window sizes. The p-value of the Wilcoxon signed rank test is shown on top of each panel.

for the EEG analysis is Fz's theta power. Therefore, we also compared each participant's averaged normalized theta power at Fz, as shown in Fig. 4.5B. The findings revealed that heart

rate variability highly depended on the window size. The paired difference between the two conditions became more apparent (though still insignificant) only with a longer window size. In contrast, the EEG feature Fz theta power remained consistent, regardless of the window size. Consequently, our analysis focused on the transferability of the EEG features.

To validate the efficacy of our proposed transferring method in reducing the difference in the distribution of band power features between the source and target domains, we used the Euclidean distance between the means of the distributions as the metric of the transformation. Specifically, we calculated the Euclidean distance between the mean of **Target-Block1-Hard** and the mean of **Source-Hard** for each target session as the native distance between the target and source domains. Then this distance was further compared with the distance between **Target-Block1-Hard** and **Source-Hard** after applying either the mean or covariance transfer method. Moreover, to examine the sensitivity of the transferring method to the session effect, we also compared the distances between the native or transferred source domain and **Target-Block2-Hard**. As we assumed the mental math task was more controlled, and the theta band power is the most robust indicator of mental workload, this analysis focused on the theta band power features during the mental math task. The results under the cross-session and cross-subject scenarios are presented in Fig. 4.6. The results show that the mean and covariance transfer methods significantly reduced the difference in distributions between the source and target domains in all cases. Moreover, under the cross-session scenario, the covariance transfer method showed a significantly greater reduction in distance to **Target-Block1-Hard** than the mean transfer method. However, overall, the efficacy of the two methods was comparable.

Finally, we evaluated the effectiveness of the proposed transfer methods by examining their impact on classification performance. We trained five types of models with different sets of training data as described in subsection Transferring Models. Instead of using accuracy, we used the point biserial correlation coefficient (i.e., the Pearson correlation coefficient for binary variables) between the prediction outputs of the target model (model 1) and the outputs of each model to evaluate their performance. This metric indicates how closely a model's behavior

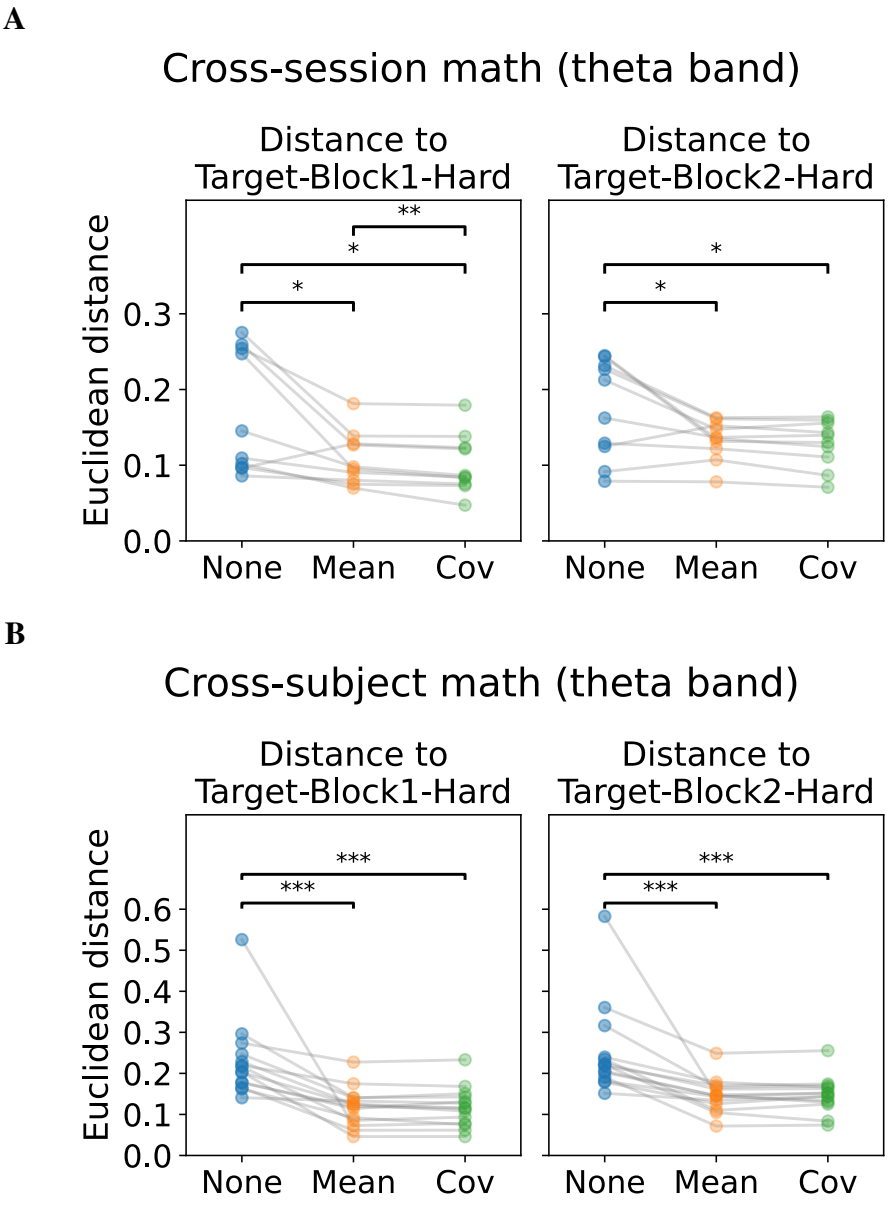



**Figure 4.6. A**. Left panel: The euclidean distance between **Target-Block1-Hard** and **Source-Hard** without data transfer (None), applied with mean transfer (Mean), and applied with covariance transfer (Cov) under the cross-session transferring scenario. Right panel: The distance to **Target-Block2-Hard**. **B**. The same results as in **A** but under the cross-subject transferring scenario.

matches the target model, which represents the ideal model that has been fully calibrated. Figure 4.7 and 4.8 show the model performance results evaluated with the math task data and caption task data under the cross-session and cross-subject transfer scenarios. For the mental

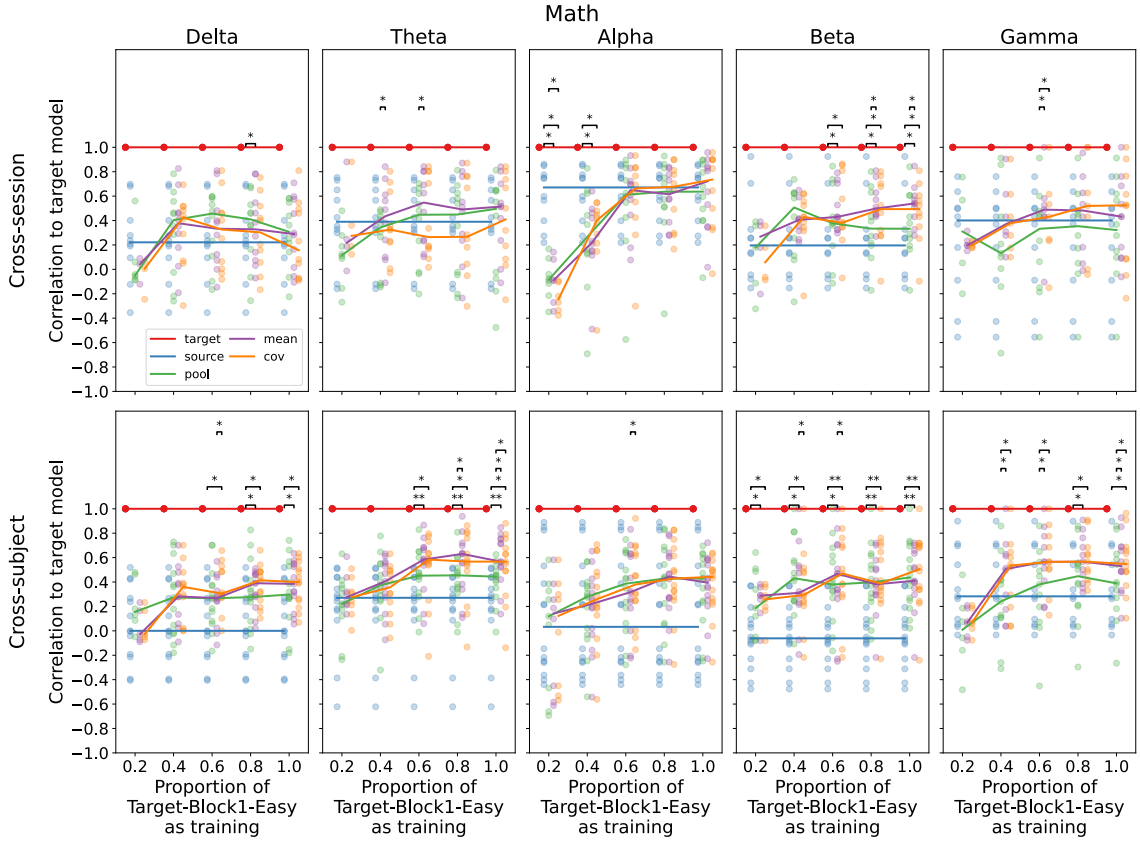



**Figure 4.7.** The results of the correlation between the outputs of each model and the target model in the math task.

math results, the mean transfer model performed better in most cases, especially in the cross-subject scenario when the proportion $p$ was larger than 0.6. Additionally, mean transferring on theta band power in the cross-subject scenario showed the strongest efficacy. Regarding the caption task results, the beta band had the strongest efficacy, while transferring in the theta band did not work well.

Figure 4.9 further shows the results of the cross-task scenario where the data were transferred from the domain of the math task to the domain of the caption task. Overall, we observed lower correlation scores for the transferring models in this scenario than in the cross-session and cross-subject scenarios.

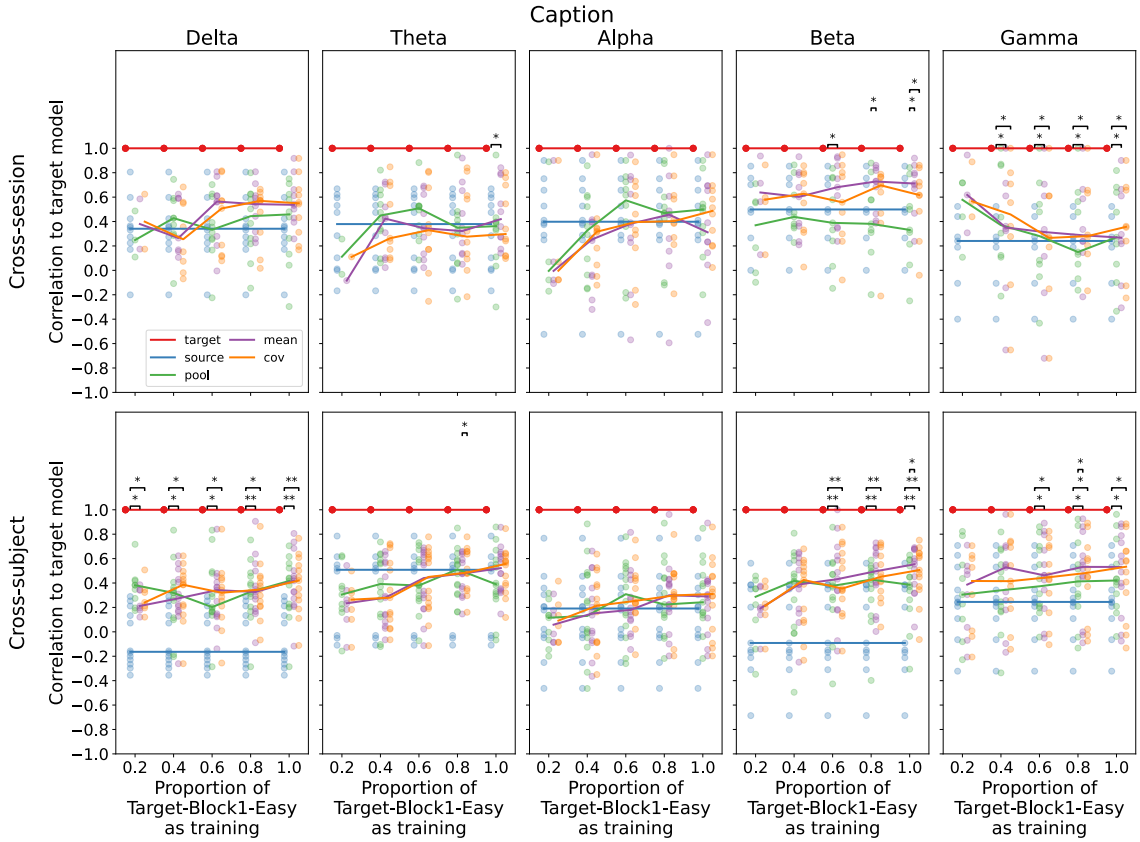



**Figure 4.8.** The correlation results between the outputs of each model and the target model in the caption task.

## 4.4 Discussion

Fig. 4.7 and 4.8 show that the source model performed poorly in most cases, indicating that the source and target domains have different distributions and that transfer learning is necessary. The performance of the transfer models increases as the proportion $p$ increases, as expected, because more easy-mode data from the target domain can be used to learn better transformations. The mean-transfer model ranks at the top or has the equal performance to the best model in most cases when $p$ is set to 0.8 or 1, suggesting that it can either improve performance or, at the very least, not negatively impact it. It is important to note that setting $p$ to 1 is equivalent to having 3 minutes of low mental workload state calibration.

Furthermore, the fact that the mean-transfer model outperforms the pooled model in

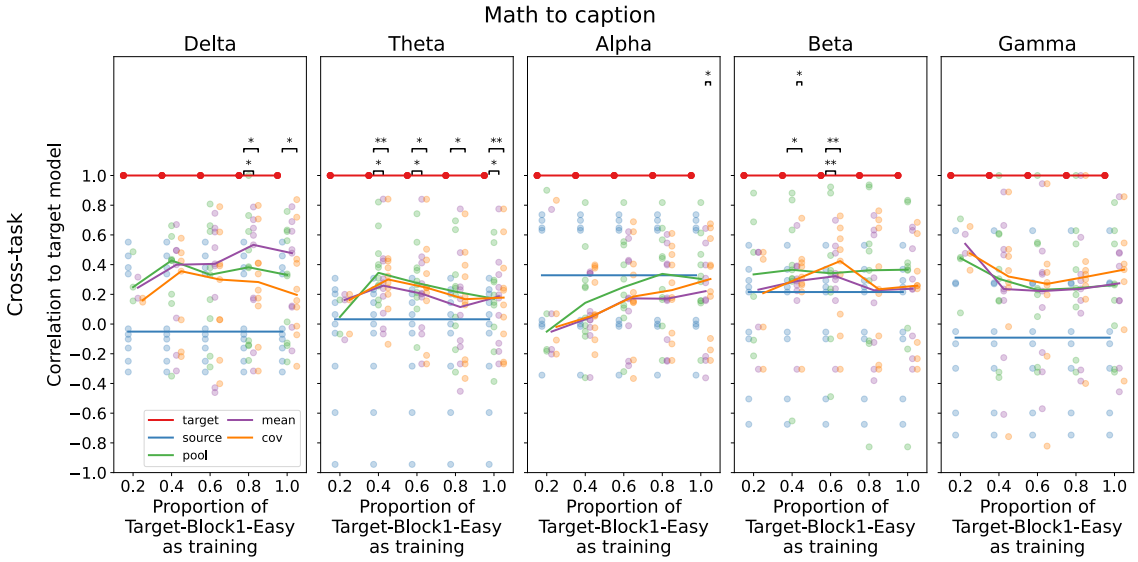



**Figure 4.9.** The results of the correlation between the outputs of each model and the target model in the cross-task (math to caption task) transferring scenario.

different bands suggests that different bands have different transferability. Specifically, the theta, beta, and gamma bands have better transferability for the mental math task, while the beta band has better transferability in the image caption task. Additionally, it is unclear why the source model performs better in the alpha band in cross-session transfer scenarios for math tasks. However, the performance is very poor in the cross-subject scenario, suggesting that the alpha band may be robust across days but not individuals. Further research is needed to understand this phenomenon.

We had anticipated incorporating covariance transferring with the mean-transfer approach would yield better results, but our findings were not as encouraging as expected. Despite noticing a substantial reduction in the Euclidean distance to the target domain depicted in Fig. 4.6A, the covariance-transfer model's performance was comparable to the mean-transfer model in most scenarios and even worse in some cases. Therefore, it is imperative to conduct further exploration and refinement of this method to unlock its full potential.

The difference in performance between the source and pool models in the theta band of the mental math task is less evident in the cross-session scenario. This may be attributed to

the original robustness of the same participant's band power features across days. Nonetheless, significant transfer method effects are still noticeable in the cross-subject scenario. The success in transferring the theta band power features is in accordance with our assumption and suggests that this signature is consistent across participants. Upon closer examination of the correlation scores in the theta band, we found that the transfer models outperformed in the cross-subject scenario compared to the cross-session scenario. This implies that the transferred theta band power features are more resilient than the features of the same individual across different days.

The effectiveness of transferability in the theta band can also be elucidated by the topoplot of the theta band power in the math task. As shown in Figure 4.10A, the topoplot reveals a strong frontal theta band in most cases, which is even more pronounced in the target domain. Interestingly, the topoplots for each participant are very similar in both the source and target domains, which explains why the performance of the pool and mean-transfer models in the cross-session scenario is comparable. However, even though all participants show strong frontal theta power, the theta power of the surrounding channels is not identical, creating room for improvement for the mean-transfer model in the cross-subject scenario.

The disparity in the effectiveness of transferring theta band power features between the math and caption tasks suggests that these two tasks might elicit distinct mental efforts or cognitive states. For example, mental math may involve intense working memory demands, while the image caption task may not. The induced brain activities in the two tasks appeared more complex than initially thought. Nonetheless, we did observe similar effectiveness in transferring beta band features in the cross-session scenario, indicating that there may be some shared changes in brain states between the two tasks that can be observed in the beta band. This argument is further supported by the topoplots of the theta and beta band power, as shown in Fig. 4.10 and 4.11. The topoplots in Fig. 4.11A show that the localized frontal theta power can also be observed in the image caption task. However, it is comparable between the easy and hard condition, unlike the stronger frontal theta power in the hard condition shown in Fig.4.10A In contrast, the elevation in beta band power around the temporal area from the easy to the hard

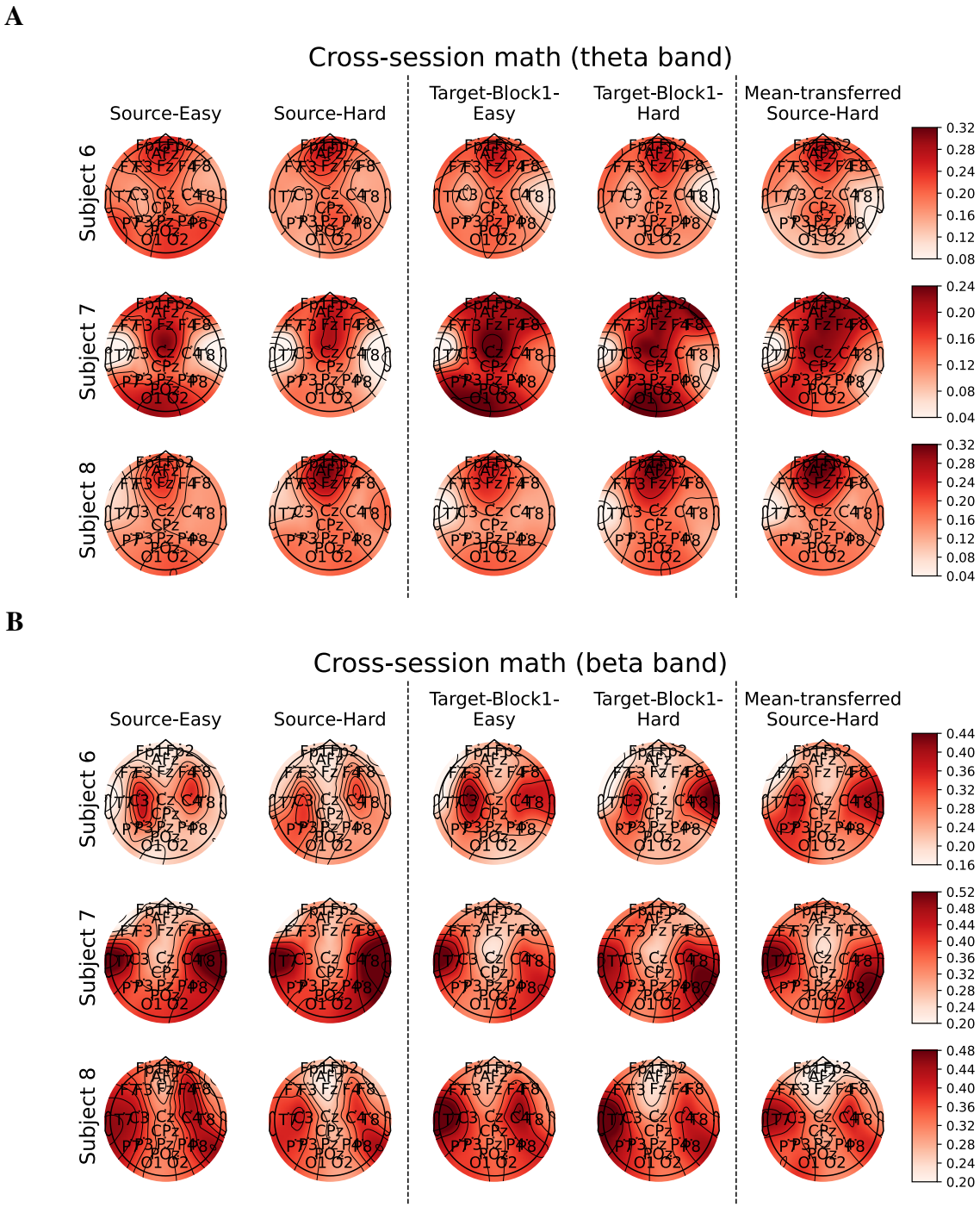



**Figure 4.10.** **A**. The example topoplot of theta band power in the math task. **B**. The example topoplot of beta band power in the math task.

condition is observable in both the math task and the caption task, as illustrated in Fig.4.10B and 4.11B We may link this shared increase in beta band power to attention to attention, as beta band

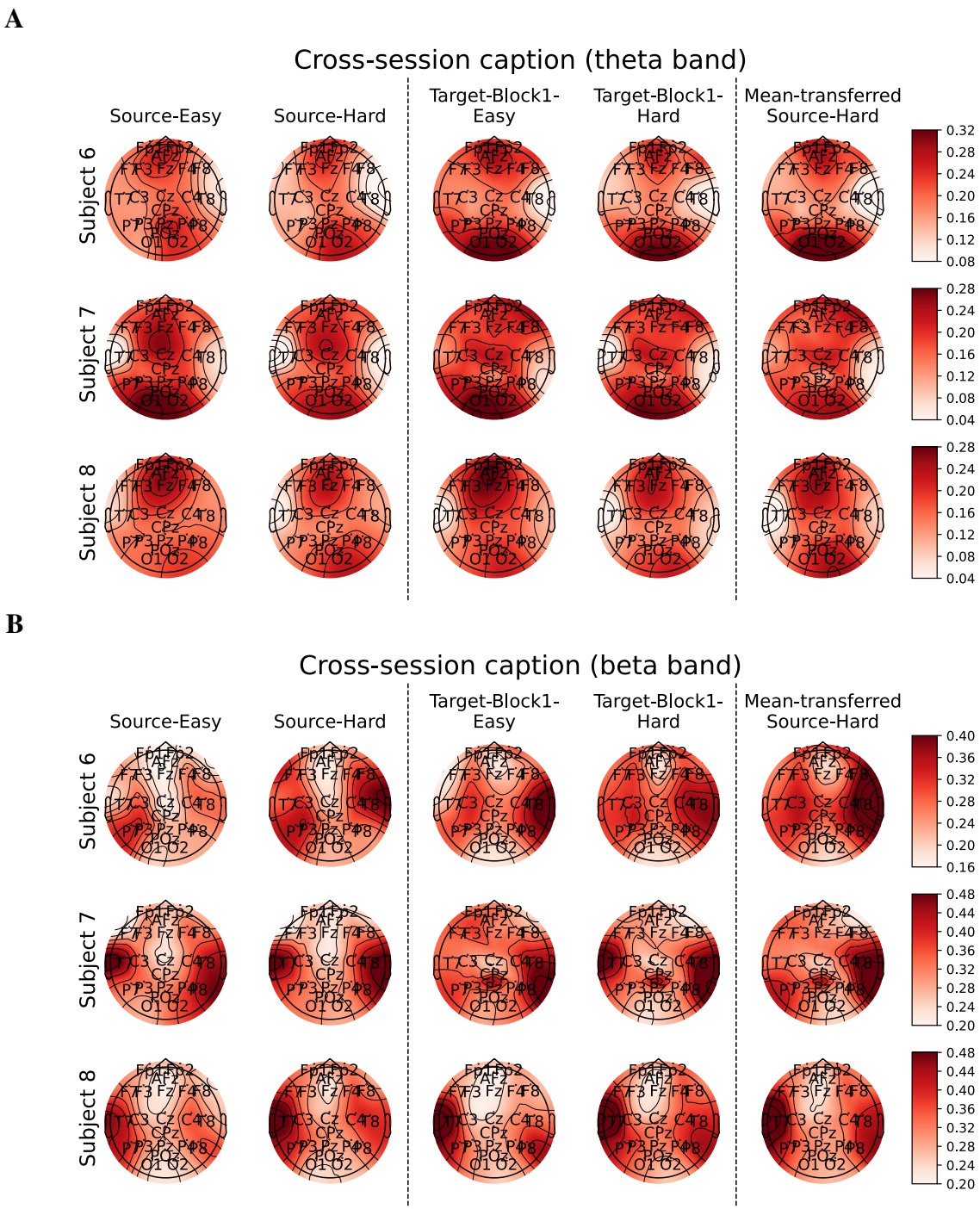



**Figure 4.11.** **A**. The example topoplot of theta band power in the caption task. **B**. The example topoplot of beta band power in the caption task.

power has been reported in the literature to be linked to attention [86, 87].

The inability to transfer features from the mental math task to the image caption task

indicates that the process of cross-task transferring is more complicated than originally assumed, highlighting the importance of more comprehensive investigations into the underlying cognitive processes and experimental designs. Nevertheless, this study has shown that the transferability of band power features across sessions or subjects varies among bands. The simple mean transfer method can effectively reduce the difference in the distribution of band power features between the source and target domains in most cases, given sufficient easy-mode data for fine-tuning. This suggests that the mean transfer method might improve the practicality of EEG-based passive BCIs.

Chapter 4, in part, is currently being prepared for submission for publication of the material. The dissertation author was the primary investigator and author of this paper.

# Chapter 5

# Memory Workload and Listening Effort

## 5.1 Introduction

This study investigated the relationship between a classic memory task, the Sternberg task, and a natural speech listening comprehension task, a common everyday challenge. Specifically, we explored the possibility of inferring human mental workload during real-world tasks using knowledge acquired from lab-controlled tasks. Our focus was assessing the mental effort or workload required for listeners to comprehend natural speech in the presence of background noise, a typical daily challenge such as conversing in a coffee shop or on public transit. Even healthy individuals and those with hearing loss can find it difficult to comprehend speech in noisy environments. To mitigate these challenges, various listening enhancement technologies have been proposed [88, 89, 90]. Still, there are significant individual differences in vulnerability to these difficulties, and sometimes the technologies can distort or alter the sound more than necessary. Therefore, assessing an individual's sensitivity to background noise interference is essential to provide personalized listening enhancement solutions.

The role of working memory in speech perception, particularly in noisy environments, is crucial. Working memory allows individuals to temporarily hold and manipulate information for complex tasks. It is essential for filtering out background noise and focusing on the target speech in noisy situations. Some studies have suggested that individuals who are more susceptible to background noise may have limited working memory capacity, which makes it more challenging

to disentangle the target sound from the acoustic mixture and attend to it [91, 92]. However, most of these studies have relied on behavioral evidence to establish the relationship between working memory and auditory tasks. One potential limitation of this approach is the ”smarter people perform better” phenomenon, which could influence performance metrics.

The objective of this study was to obtain EEG measurements during a natural speech listening task, and to validate whether EEG signatures for increased memory workload during the task can be observed. Furthermore, if such signatures were identified, the study aimed to investigate their correlation with the signatures of increased memory workload that can be observed during a classic memory task, the Sternberg task.

## 5.2 Methods

### 5.2.1 Experiments

To ensure an unbiased comparison of common EEG features between the memory and listening tasks, we conducted two separate sessions on different days. In the first session, participants completed the listening task, while in the second session, they performed the working memory task. A total of 15 participants (7 females) with an average age of 28.5 years (standard deviation 6.2 years) participated in the study. All participants reported normal hearing and normal or corrected-to-normal vision. During both sessions, participants were seated comfortably in a quiet room and fitted with an EEG cap to record scalp electrical activity. We used a 64-channel BrainWave cap from MFI BV and a SAGA 32+/64+ amplifier from TMSi to continuously record EEG signals throughout the sessions. However, we only applied gel to and recorded from 32 out of the 64 available channels.

#### Opposite-ear listening task

During the first session, participants engaged in an opposite-ear listening task, a method utilized to understand the origin of interference sources in situations where masking occurs beyond the cochlea. This task involves listening to two auditory stimuli simultaneously, one

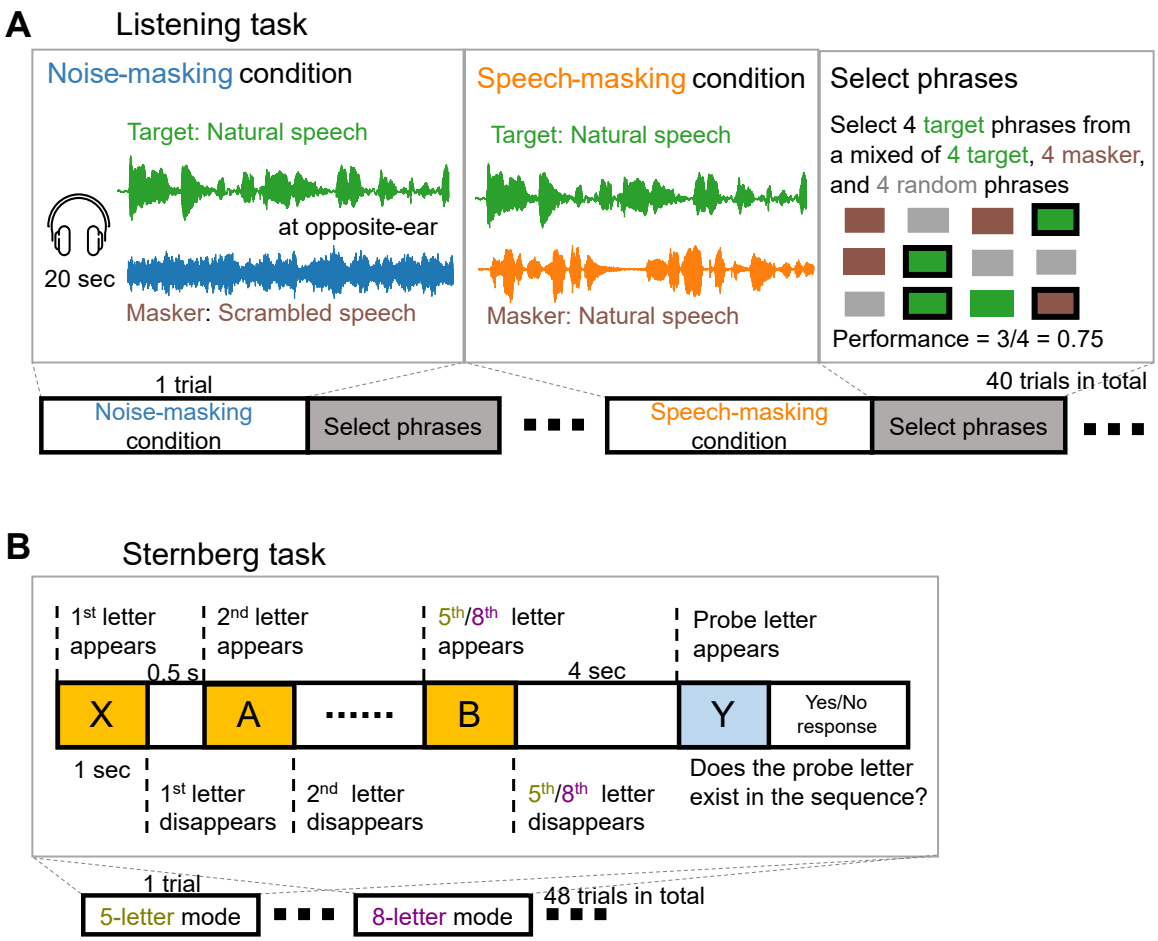



**Figure 5.1.** **A**. The flowchart of the opposite-ear listening task in the first session of the experiment. **B**. The flowchart of the modified Sternberg task in the second session of the experiment..

presented to each ear, and focusing attention on one of the stimuli.

The task was further divided into two conditions: the ”noise” masking condition and the ”speech” masking condition. In the noise masking condition, the two auditory stimuli delivered to each ear consisted of one natural speech and one scrambled speech that sounded like babble noise, and the participants were instructed to focus on the natural speech. This allowed us to establish the control level of the participant's memory workload when comprehending a single speech. In the speech masking condition, both auditory stimuli were natural speeches, and the participants were required to focus on one of them. This condition allowed us to measure the participant's memory workload when informational masking occurred. Informational masking

refers to the interference caused by an intelligible source, which induces additional effort in the central processing function.

Fig. 5.1A depicts a sequence of events for each trial in the opposite-ear listening task. At the beginning of each trial, a visual cue was presented, indicating the target's location as either "left" or "right." Then, after a 1.5-second delay, the experimental program played two audio tracks simultaneously: one in the left ear and one in the right ear, each lasting for 20 seconds. One track was the target speech, and the other was the masker consisting either of scrambled natural speech (noise masking condition) or natural speech (speech masking condition). The target speech was randomly assigned to either ear, with a masker in the other ear. Participants were instructed to focus on the target speech and close their eyes once the cue was displayed. After each 20-second listening task, a phrase-selection closed-set response matrix appeared on a computer monitor in front of the participant. The matrix contained four phrases extracted from the target, four from the masker, and four random phrases from a third audiobook. Participants were required to select four phrases from the target speech. Once the selection of four phrases was submitted, the program provided feedback on the correctness of the response.

The first session consisted of five blocks of eight trials each, alternating between four in the noise and four in the speech masking conditions. Each participant completed 20 trials in the noise and 20 trials in the speech masking conditions.

During the recording, participants were instructed to keep their eyes closed and minimize head and body movements to reduce the impact of eye movements and blinks on the EEG signals. A pair of headphones delivered the speech stimuli binaurally to both ears at a comfortable loudness level of 75 dB SPL (DT 990 PRO, Beyerdynamic GmbH & Co. KG).

To create the stimuli for the opposite-ear listening task, we utilized natural, continuous speech from the LibriSpeech project [93]. Specifically, we selected groups of consecutive sentences from audiobooks that lasted about 20 seconds and used them as stimuli. To automatically extract four key phrases from the transcription of each speech stimulus, we employed an unsupervised machine-learning approach for keyword extraction. Specifically, the Yake tool was used

with the deduplication threshold parameter set to 1, and 1- or 2-grams were extracted as key phrases, with all other parameters set to their default values [94]. Subsequently, the transcribed key phrases were manually screened to eliminate colloquialisms.

The target and masker in each trial were always spoken by talkers of the same gender, with the male or female voice selected randomly for each trial. The scrambled speech for the noise masker was generated by time-domain scrambling natural speech within multiple frequency bands, using a window length of 500 ms and a shuffling radius of 2000 ms [95].

**Sternberg task**

During the second session, we used the Sternberg task [25, 96] to assess visual memory in each participant. In this task, a list of items (e.g., letters or numbers) is presented briefly, followed by a probe item. Participants are then asked whether the probe item was part of the list. The list length is varied across trials to manipulate the working memory demands, and reaction time is measured as an index of performance. The Sternberg task is a commonly used tool to study the cognitive and neural mechanisms involved in working memory and to investigate individual differences in working memory capacity.

We used a modified version of the Sternberg task to evaluate working memory capacity (shown in Fig. 5.1B). Participants were presented with a sequence of letters in two conditions: 1) ”Easy” mode, which consisted of sequences of five letters, and 2) ”Difficult” mode, which had sequences of eight letters. Each trial displayed the letters one at a time for 1 second, with a 0.5-second interval between each presentation. After the final letter was shown, the screen remained blank for 4 seconds before displaying a probe letter. Participants were instructed to indicate whether the probe letter was part of the previously presented sequence by pressing a corresponding key on the keyboard. The task consisted of three blocks containing eight trials for each condition (Easy and Difficult). The Easy trials were always presented before the Difficult trials. Participants completed 24 trials for each condition, resulting in a total of 48 trials.

### 5.2.2 Data Processing

EEG signals were filtered using a 1-50 Hz finite impulse response filter, re-referenced to the average, and downsampled to a 100 Hz sampling rate. For the first session, we extracted EEG epochs lasting 20 seconds for each onset time of audio track playback. This resulted in 40 epochs (20 noise-masking and 20 speech-masking) for each participant.

For the second session, we segmented the EEG signals into epochs, starting from the onset of the first letter display and ending at the disappearance of the last letter. For the easy mode, we generated 24 epochs lasting 7 seconds, while for the hard mode, we created 24 epochs lasting 11.5 seconds. In total, this resulted in 48 epochs for the second session.

To analyze the EEG signals, we calculated the spectrogram for each epoch using consecutive Fourier transforms with 2-second wide and 1.8-second overlapping sliding windows. To normalize the power spectral density for each channel, we divided it by the total power at each time point. We then calculated the normalized power of the EEG bands ($\delta$, $\theta$, $\alpha$, $\beta$) by integrating the normalized power spectral density over the corresponding frequency ranges ($\delta$: 1-4 Hz, $\theta$: 4-8 Hz, $\alpha$: 8-12 Hz, $\beta$: 12-30 Hz) using the trapezoidal method.

### 5.2.3 Theta Power Change

Previous research has demonstrated a strong correlation between frontal midline theta power and memory workload [97, 98, 34, 80]. The current work, therefore, focuses on frontal theta power by analyzing the EEG time series from Fz. Next, we introduce an index of memory workload, called "frontal theta ($\theta$) power change," or $\Delta\mathrm{Fz}\theta$, where the subscripts "L" and "S" denote "L"istening" vs "S"ternberg task ($\Delta\mathrm{Fz}\theta_L$ vs $\Delta\mathrm{Fz}\theta_S$).

In the Sternberg task, normalized frontal $\theta$ power reaches a plateau after memorizing the fifth letter. We, therefore, estimated the maximum change, $\Delta\mathrm{Fz}\theta_S$, by dividing the normalized theta power during the memorization of the fifth letter by that during the memorization of the first letter.

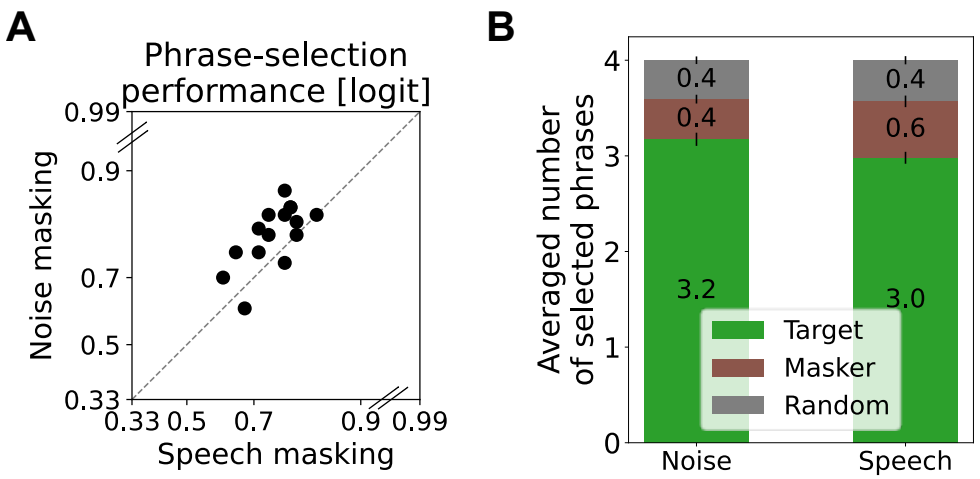
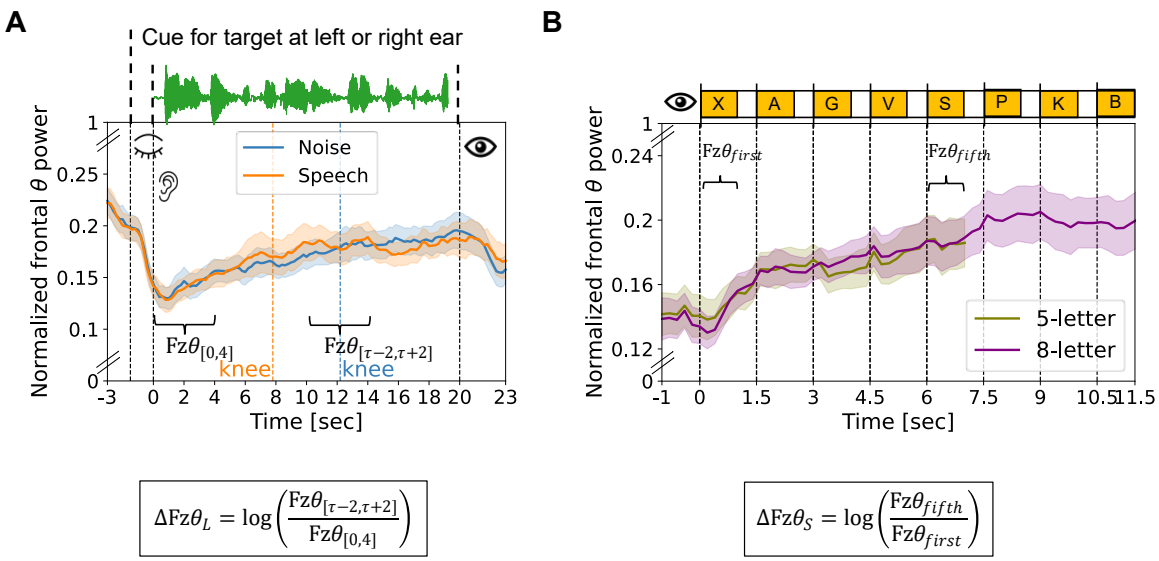



**Figure 5.2. A**. The distribution of each participant's averaged phrase-selection performance in the noise-masking versus speech-masking condition. **B**. The average frequency of each type of response was selected across all participants.

To make a meaningful comparison of theta changes between the two tasks, the analysis for the listening task also compared the change of normalized $\theta$ power averaged over a window of 4 seconds, corresponding to 20% of the total duration for each task. This window size was chosen to match the ratio of the one-letter period (for both the duration of memorizing the first letter and the fifth letter) versus the total of five letters in the Sternberg trials before reaching the plateau. Fz theta power averaged over the first four seconds after stimulus onset serves as the baseline. The memory workload index $\Delta \mathrm{Fz}\theta_L(\tau)$ is then calculated as the normalized Fz theta power averaged over a 4-second window $[\tau-2,\ \tau+2]$, divided by the baseline. The value of $\tau$ was selected to capture the change of the frontal theta power as it reached the plateau. More specifically, $\tau$ was set to the knee point of the curve representing the averaged normalized theta power over time in both the noise-masking and speech-masking conditions. The knee points were obtained numerically by fitting each curve to a 3-degree polynomial and then applying a knee point detection algorithm (Kneedle, [99]) to identify the precise knee point in the polynomial.




$$\Delta \mathrm{Fz}\theta_L = \log\left(\frac{\mathrm{Fz}\theta_{[\tau-2,\tau+2]}}{\mathrm{Fz}\theta_{[0,4]}}\right)$$

$$\Delta \mathrm{Fz}\theta_S = \log\left(\frac{\mathrm{Fz}\theta_{fifth}}{\mathrm{Fz}\theta_{first}}\right)$$

**Figure 5.3. A**. The dynamic of the normalized Fz theta power over time in the listening task. **B**. The dynamic of the normalized Fz theta power over time in the Sternberg task.

## 5.3 Results

### 5.3.1 Results of the Listening Task

Fig. 5.2A shows the behavioral performance of each participant in a scatter plot, with raw scores ranging from 0 to 100%. To mitigate the issue of ceiling and flooring effects, the scores were transformed into logit scores using the formula $\mathrm{logit}(p) = \ln(p) - \ln(1-p)$. To prevent undefined values, the input value of $p$ was capped between 33% (chance-level performance) and 99%s. The results indicated that 13 out of 15 participants had equal or better performance in the phrase-selection task than in the speech-masking trials, as evidenced by the dots located to the right of the equal-performance line in the scatter plot.

Two types of response errors can occur with opposite-ear listening. When the participant misses a target phrase, they must guess the response, leading to a random response. In contrast, if they attended to the wrong talker, they would report the masker, resulting in a masker response. To elucidate whether missed trials mainly limited the performance as opposed to source selection, Fig. 5.2B shows the mean number of selected phrases of the type of responses participant gave. The mean was first obtained by taking the average across noise-/speech-masking trials, and then

averaged across all participants. Note that in each trial, the participants had to select exactly four phrases, so the sum of the means of each type was always four. The p-values of the Wilcoxon signed rank test between the number of selected phrases of target/masker/random response in the noise-masking condition versus in the speech-masking condition are $p < 0.01$ for target responses, $p < 0.05$ for masker responses and $p > 0.05$ for random responses. This shows that the increased numbers of incorrectly selected phrases in the speech-masking condition were due to masker errors.

We proceeded to explore the correlation between memory capacity and speech comprehension with background noise by analyzing the frontal theta power dynamics (see subsection Data Processing). To achieve this, we computed the normalized Fz theta power for each 2-second time window in every EEG epoch and then averaged the results across both noise-masking and speech-masking conditions for each participant.Fig. 5.3A shows the mean and standard error of the averaged normalized Fz theta power in each condition.

The results show that, in both conditions, frontal theta power decreased before the onset of the stimulus, reached a minimum at the onset, and then increased as the tasks progressed. Although there were differences between the two curves, they were not statistically significant evaluated by the cluster-level statistical permutation test [100, 101, 102]. It is worth noting that the initial decrease in theta power before stimulus onset is likely due to participants closing their eyes at -1.5 sec and resetting their workload after completing the previous trial's phrase-selection task.

In the noise-masking condition, the curve reached its peak around the stimulus offset and then decreased, while in the speech-masking condition, the curve appeared to reach a maximum of around 10 sec, plateau, and then decreased after the offset. We numerically identified the knee point at 7.8 seconds after the stimulus onset for the speech-masking condition and 12.2 seconds for the noise-masking condition, as mentioned in subsection Theta Power Change. As shown in Fig. 5.3A, the normalized frontal theta power increased over time during the listening trials. To quantify the increase from baseline to plateau, we calculated the frontal theta power change

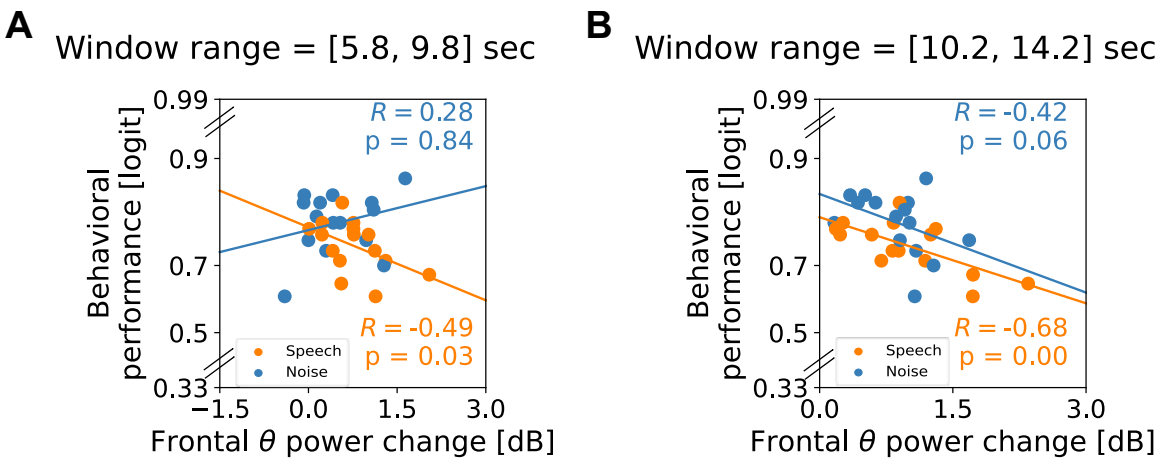



**Figure 5.4. A**. The correlation between $\Delta Fz\theta_L$ and the phrase-selection performance when the window is set to the knee point of the speech-masking condition. **B**. The correlation between $\Delta Fz\theta_L$ and the phrase-selection performance when the window is set to the knee point of the noise-masking condition.

$\Delta Fz\theta_L$ using the method described in subsection Theta Power Change, with $\tau$ set to 7.8 and 12.2 seconds for the speech-masking and noise-masking conditions, respectively.

We investigated further to determine whether $\Delta Fz\theta_L$ could reflect speech intelligibility. For each knee point, we calculated the mean $\Delta Fz\theta_L$ and mean phrase-selection performance by averaging across trials in each condition (noise-masking or speech-masking) for each participant. This resulted in 15 pairs of samples for each condition, representing each participant's mean $\Delta Fz\theta_L$ and mean phrase-selection performance. We then applied the least-squares linear regression to the 15 pairs of samples and obtained Pearson's correlation coefficients (R) and the $p$-value of the Wald Test. Fig. 5.4 shows the results. At each condition's respective knee point, we observed a negative correlation between the $\Delta Fz\theta_L$ and behavioral performance, with the speech-masking condition showing a stronger correlation.

### 5.3.2 Results of the Sternberg Task

We then examined how the memory workload impacted participants' behavior during the Sternberg task. We hypothesized that more effort would be required to remember a greater number of letters, resulting in longer response times for the probe letter recall in the eight-letter

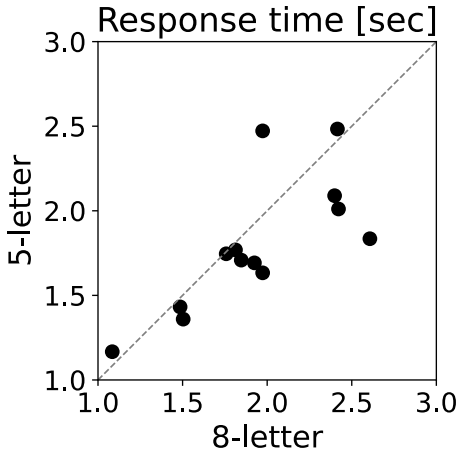



**Figure 5.5.** Each participant's averaged response time in the 5-letter trials versus in the 8-letter trials in the Sternberg task.

trials compared to the five-letter trials. We calculated the average response times within each participant, which is presented in Fig. 5.5. The left panel displays the scatter plot of each participant's average response time, while the right panel shows the mean and standard error of the response times across all participants. The results revealed a significant difference in the average response time between the two conditions, with 12 out of 15 participants exhibiting longer response times in the eight-letter trials. This was further supported by a significant $p$-value of 0.013 in the Wilcoxon signed-rank test.

We also used a topoplot to examine whether the increase in the theta power was localized in the frontal area rather than a result of signal propagation from other channels. Fig. 5.6 shows the topoplot of the mean normalized theta power averaged over different time windows during the tasks, and further averaged across all participants. The topoplots show that the theta power increase in both tasks was centered around the frontal area, consistent with previous research on memory workload [103, 34, 98].

Next, we investigated the dynamics of normalized Fz theta power during the Sternberg task. We calculated the power spectral density using 2-sec sliding windows with 90% overlap (see subsection Data Processing) in each trial and obtained the average for each participant.

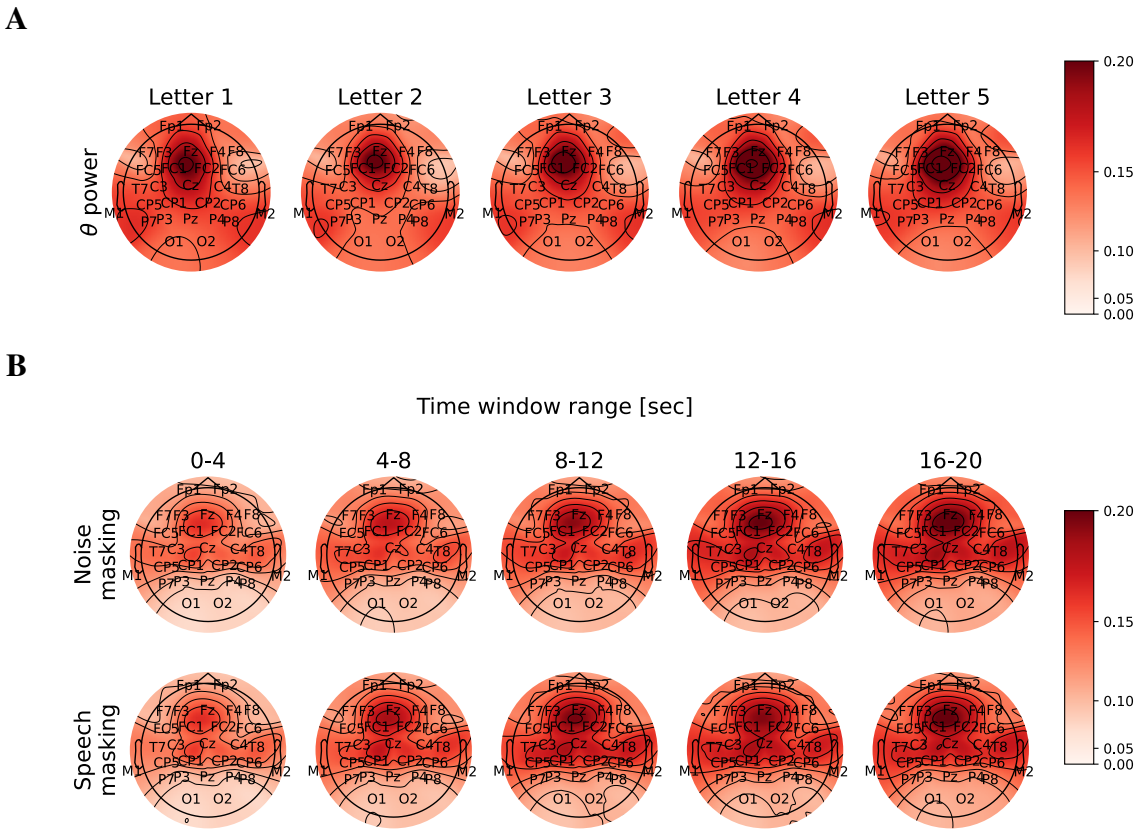



**Figure 5.6. A**. The topoplot of the averaged theta power of all participants when memorizing letters at different orders in the Sternberg task. **B**. The topoplot of the averaged theta power of all participants at different time window ranges.

Fig. 5.3B shows the average of participants' means. Upon visual inspection, we observed that the normalized frontal theta power initially decreased to its minimum at the onset of the first letter and then gradually increased as more letters appeared. The curves for the 5-letter and the first half of the 8-letter conditions showed a high degree of overlap. Therefore, we simplified the analysis by extracting the period of the first five epochs of the 8-letter condition and combining the two conditions. To measure the progressive increase in frontal theta power over time, we defined the change in frontal theta power during the Sternberg task as $\Delta \mathrm{Fz}\theta_S$ (refer to subsection Theta Power Change), and investigated the relationship between $\Delta \mathrm{Fz}\theta_S$ and $\Delta \mathrm{Fz}\theta_L$ in the listening task (see below).

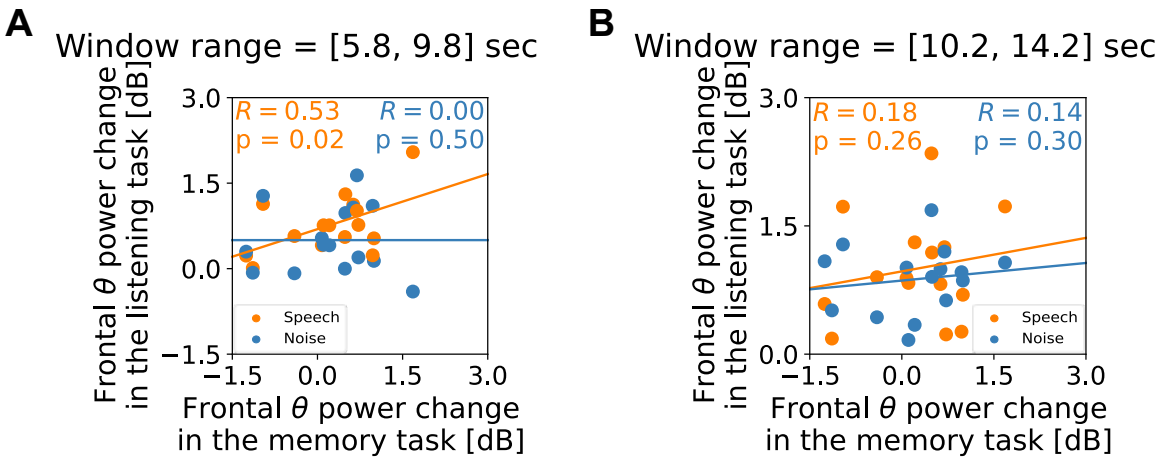



**Figure 5.7. A**. The correlation between $\Delta \mathrm{Fz}\theta_S$ and $\Delta \mathrm{Fz}\theta_L$ when the window is set to the knee point of the speech-masking condition. **B**. The correlation between $\Delta \mathrm{Fz}\theta_S$ and $\Delta \mathrm{Fz}\theta_L$ when the window is set to the knee point of the noise-masking condition.

### 5.3.3 Link between Informational Masking and Working Memory

The primary objective of this study was to investigate the dynamics of memory workload in both the listening and the Sternberg tasks by examining changes in frontal theta power. We calculated the average change in frontal theta power $\Delta \mathrm{Fz}\theta_S$ for each participant across all Sternberg trials. We then paired the resulting 15 samples with each participant's average $\Delta \mathrm{Fz}\theta_L$ in the listening task, separately for both speech- and noise-masking conditions. Finally, we applied least-squares linear regression to the 15 pairs of samples and presented the results in Fig. 5.7. The results reveal that $\Delta \mathrm{Fz}\theta_L$ in the speech-masking condition was positively correlated with $\Delta \mathrm{Fz}\theta_S$ at its knee point. In contrast, in the noise-masking condition, $\Delta \mathrm{Fz}\theta_L$ did not correlate significantly with $\Delta \mathrm{Fz}\theta_S$ in any case.

## 5.4 Discussion

We employed a novel phrase-selection task to assess speech intelligibility in the presence of masking speech, which was either intelligible or unintelligible, presented in the opposite ear. The results presented in Fig. 5.2 show that speech intelligibility was compromised when the target speech was presented with intelligible masking speech, as evidenced by reduced

performance in the phrase-selection task. This result is consistent with prior research suggesting that speech intelligibility is primarily affected by the information in the masking speech.

We found that changes in frontal theta power during the opposite-ear listening task, specifically in the speech-masking condition, could predict speech intelligibility. The results presented in Fig. 5.4 revealed a negative correlation between changes in frontal theta power and speech intelligibility, indicating that participants with greater changes in frontal theta power tended to have lower speech intelligibility. Furthermore, the negative correlation in the speech-masking condition emerged earlier and was more robust than that observed in the noise-masking condition, suggesting that informational masking imposed an additional cognitive workload on the participants.

The results in Fig. 5.3B revealed that frontal theta power increased as participants retained more letters during the Sternberg task, consistent with prior research [34] reporting that frontal theta power increases with memory workload. This investigation adds to previous research by demonstrating a comparable pattern of increasing frontal theta power during the listening task, as depicted in Fig. 5.3A. Notably, the time scales and overall dynamic range of normalized frontal theta power were comparable across the task periods of the Sternberg task and the speech- or noise-masking conditions, suggesting that the cognitive demands of the two tasks were similar.

Further examination of Fig. 5.7 provides more compelling evidence for the association between the memory task and the listening task in the speech-masking condition. Panel A shows a positive correlation between the Sternberg task and the speech-masking condition, indicating that individuals who exert more effort into the memory task also devote more effort to listening to speech with informational masking. Furthermore, the fact that the positive correlation is only significant in panel A but not in panel B indicates that the additional cognitive load caused by the informational masking in the early period of the trials is similar to the cognitive load induced by the Sternberg task. Later on, after reaching a plateau, the cognitive workload may be dominated by the effort to comprehend a single speech, as in the noise-masking condition, where the frontal theta power also attained a plateau at a later stage, and the change was not correlated with the

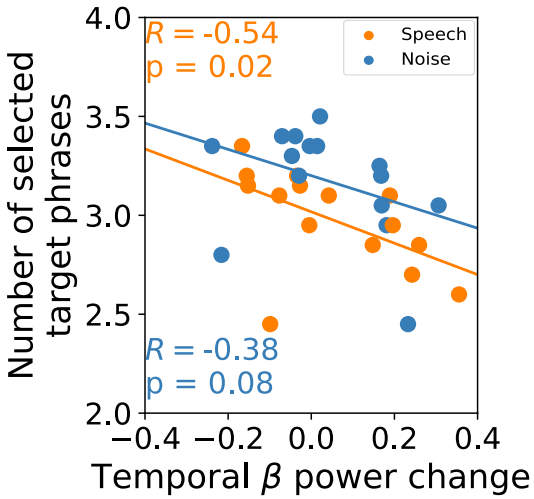



**Figure 5.8.** Each participant's averaged number of selected phrases versus mean temporal beta power.

change in the Sternberg task.

In addition to the observed correlation in frontal theta power, we also found an intriguing association between behavioral performance and temporal beta band power. To investigate this, we calculated the correlation between each participant's mean behavioral performance (measured by the exact number of selected phrases rather than the percentage) and their mean beta power at the mastoid channels, as displayed in Fig. 5.8. We selected the mastoid channels for this analysis because we were interested in non-hair-bearing areas, and they are typically labeled as TP9 and TP10 in the 10-20 system. Instead of calculating the knee point of the temporal mastoid, we normalized the beta power averaged across [10, 20] seconds after the stimulus onset by dividing it with the [0, 4] second baseline. The results imply that beta band power can function as an indicator of a particular type of mental workload, as it exhibits a substantial negative correlation with behavioral performance. This result may be related to previous research [104]. It is also worth noting the outcomes from Chapter 4, where an increase in temporal beta band power was identified in the hard mode condition during both the mental math and image caption tasks. These results are unexpectedly related and provide additional evidence for the potential utilization of beta power as a marker of mental workload.

The results depicted in Fig. 5.7A suggest potential implications for cross-task transfer learning. Specifically, the positive correlation between a participant's performance in a lab-controlled task and their performance in a task within a more real-world setting suggests the possibility of training predictive models using lab-controlled tasks that can later be applied to real-world scenarios. This finding sheds light on potential avenues for future research in the field of cross-task transfer learning.

# Chapter 6

# Contribution of This Dissertation

My research explored different facets of EEG-based studies in real-world scenarios. Specifically, I conducted a neuroergonomics investigation of memory workload in real-world environments in the first study. The results revealed that band power, mutual information, and coherence features were effective in distinguishing between high and low memory workload states. These findings can be applied to the development of working memory monitoring systems in real-world settings.

The second study examined the utilization of EEG to measure human liking responses in real-world environments. The results showed that time-domain and frequency-domain signatures were consistently observed in both fixed-viewing and free-viewing conditions, highlighting EEG's potential as a valuable tool for measuring human liking responses in real-world settings. Furthermore, these two studies demonstrate the feasibility of detecting the same EEG signatures in real-world situations as in laboratory-controlled conditions.

In subsequent research, I devised feature-transferring methods specifically for passive BCIs and investigated the transferability of different EEG bands commonly used in such systems. The results demonstrated that data from various domains could be used to learn the appropriate data distribution for a new session, thus reducing the human variability and requirement for extended calibration and enhancing the feasibility and efficiency of EEG-based passive BCI systems.

Lastly, I utilized a phrase-selection task to evaluate speech intelligibility in the presence of masking speech. The results indicated that frontal theta power rose with memory workload during the listening and Sternberg tasks. Notably, this study is one of the few with the same participants performing diverse tasks. It detected a robust correlation of EEG signatures between them, underscoring the potential of cross-task transfer learning.

Overall, this dissertation makes a significant contribution to the EEG-based evaluation of human cognitive and affective states in real-world settings. The results have meaningful implications for developing more practical and effective passive BCI applications.

# Bibliography


[1] Jonathan R Wolpaw, Niels Birbaumer, William J Heetderks, Dennis J McFarland, P Hunter Peckham, Gerwin Schalk, Emanuel Donchin, Louis A Quatrano, Charles J Robinson, Theresa M Vaughan, et al. Brain-computer interface technology: a review of the first international meeting. *IEEE transactions on rehabilitation engineering*, 8(2):164–173, 2000.

[2] Pietro Aricò, Gianluca Borghini, Gianluca Di Flumeri, Nicolina Sciaraffa, and Fabio Babiloni. Passive bci beyond the lab: current trends and future directions. *Physiological measurement*, 39(8):08TR02, 2018.

[3] Shanbao Tong and Nitish V Thankor. *Quantitative EEG analysis methods and clinical applications*. Artech House, 2009.

[4] Pietro Arico, Gianluca Borghini, Gianluca Di Flumeri, Nicolina Sciaraffa, Alfredo Colosimo, and Fabio Babiloni. Passive bci in operational environments: insights, recent advances, and future trends. *IEEE Transactions on Biomedical Engineering*, 64(7):1431–1436, 2017.

[5] Ranjana K Mehta and Raja Parasuraman. Neuroergonomics: a review of applications to physical and cognitive work. *Frontiers in human neuroscience*, 7:889, 2013.

[6] Lucas Paletta, Martin Pszeida, Bernhard Nauschnegg, Thomas Haspl, and Raphael Marton. Stress measurement in multi-tasking decision processes using executive functions analysis. In *Advances in Neuroergonomics and Cognitive Engineering: Proceedings of the AHFE 2019 International Conference on Neuroergonomics and Cognitive Engineering, and the AHFE International Conference on Industrial Cognitive Ergonomics and Engineering Psychology, July 24-28, 2019, Washington DC, USA 10*, pages 344–356. Springer, 2020.

[7] Nick Lee, Amanda J Broderick, and Laura Chamberlain. What is 'neuromarketing'? a discussion and agenda for future research. *International journal of psychophysiology*, 63(2):199–204, 2007.

[8] Ranganatha Sitaram, Andrea Caria, Ralf Veit, Tilman Gaber, Giuseppina Rota, Andrea Kuebler, and Niels Birbaumer. Fmri brain-computer interface: a tool for neuroscientific research and treatment. *Computational intelligence and neuroscience*, 2007, 2007.

[9] Noman Naseer and Keum-Shik Hong. fnirs-based brain-computer interfaces: a review. *Frontiers in human neuroscience*, 9:3, 2015.

[10] Nikos K Logothetis. What we can do and what we cannot do with fmri. *Nature*, 453(7197):869–878, 2008.

[11] Haleh Aghajani, Marc Garbey, and Ahmet Omurtag. Measuring mental workload with eeg+ fnirs. *Frontiers in human neuroscience*, 11:359, 2017.

[12] Scott C Bunce, Kurtulus Izzetoglu, Hasan Ayaz, Patricia Shewokis, Meltem Izzetoglu, Kambiz Pourrezaei, and Banu Onaral. Implementation of fnirs for monitoring levels of expertise and mental workload. In *International Conference on Foundations of Augmented Cognition*, pages 13–22. Springer, 2011.

[13] Mickaël Causse, Zarrin Chua, Vsevolod Peysakhovich, Natalia Del Campo, and Nadine Matton. Mental workload and neural efficiency quantified in the prefrontal cortex using fnirs. *Scientific reports*, 7(1):1–15, 2017.

[14] Christian Herff, Dominic Heger, Ole Fortmann, Johannes Hennrich, Felix Putze, and Tanja Schultz. Mental workload during n-back task—quantified in the prefrontal cortex using fnirs. *Frontiers in human neuroscience*, 7:935, 2014.

[15] Chris Berka, Daniel J Levendowski, Michelle N Lumicao, Alan Yau, Gene Davis, Vladimir T Zivkovic, Richard E Olmstead, Patrice D Tremoulet, and Patrick L Craven. Eeg correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine*, 78(5):B231–B244, 2007.

[16] Chris Berka, Daniel J Levendowski, Caitlin K Ramsey, Gene Davis, Michelle N Lumicao, Kay Stanney, Leah Reeves, Susan Harkness Regli, Patrice D Tremoulet, and Kathleen Stibler. Evaluation of an eeg workload model in an aegis simulation environment. In *Biomonitoring for physiological and cognitive performance during military operations*, volume 5797, pages 90–99. International Society for Optics and Photonics, 2005.

[17] Anne-Marie Brouwer, Maarten A Hogervorst, Jan BF Van Erp, Tobias Heffelaar, Patrick H Zimmerman, and Robert Oostenveld. Estimating workload using eeg spectral power and erps in the n-back task. *Journal of neural engineering*, 9(4):045008, 2012.

[18] Dongrui Wu, Yifan Xu, and Bao-Liang Lu. Transfer learning for eeg-based brain–computer interfaces: A review of progress made since 2016. *IEEE Transactions on Cognitive and Developmental Systems*, 14(1):4–19, 2020.

[19] Kai Zhang, Guanghua Xu, Xiaowei Zheng, Huanzhong Li, Sicong Zhang, Yunhui Yu, and Renghao Liang. Application of transfer learning in eeg decoding based on brain-computer interfaces: a review. *Sensors*, 20(21):6321, 2020.

[20] Ahmed M Azab, Jake Toth, Lyudmila S Mihaylova, and Mahnaz Arvaneh. A review on transfer learning approaches in brain–computer interface. *Signal Processing and Machine Learning for Brain-Machine Interfaces*, pages 81–98, 2018.

[21] EJ MS and Mccormick Ej. Human factors engineering design. *National Defense Industry Press*, 1992.

[22] Shouyi Wang, Jacek Gwizdka, and W Art Chaovalitwongse. Using wireless eeg signals to assess memory workload in the *n*-back task. *IEEE Transactions on Human-Machine Systems*, 46(3):424–435, 2015.

[23] Julien Cegarra and Aline Chevalier. The use of tholos software for combining measures of mental workload: Toward theoretical and methodological improvements. *Behavior Research Methods*, 40(4):988–1000, 2008.

[24] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

[25] Saul Sternberg. High-speed scanning in human memory. *Science*, 153(3736):652–654, 1966.

[26] O Bratfisch and E Hagman. Simkap–simultankapazität/multi-tasking. *Mödling: Schuhfried GmbH*, 2008.

[27] Evan M Peck, Daniel Afergan, Beste F Yuksel, Francine Lalooses, and Robert JK Jacob. Using fnirs to measure mental workload in the real world. In *Advances in physiological computing*, pages 117–139. Springer, 2014.

[28] Guofa Shou and Lei Ding. Frontal theta eeg dynamics in a real-world air traffic control task. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5594–5597. IEEE, 2013.

[29] Gianluca Di Flumeri, Gianluca Borghini, Pietro Aricò, Nicolina Sciaraffa, Paola Lanzi, Simone Pozzi, Valeria Vignali, Claudio Lantieri, Arianna Bichicchi, Andrea Simone, et al. Eeg-based mental workload assessment during real driving: A taxonomic tool for neuroergonomics in highly automated environments. In *Neuroergonomics*, pages 121–126. Elsevier, 2019.

[30] Andrea Giorgi, Vincenzo Ronca, Alessia Vozzi, Nicolina Sciaraffa, Antonello Di Florio, Luca Tamborra, Ilaria Simonetti, Pietro Aricò, Gianluca Di Flumeri, Dario Rossi, et al. Wearable technologies for mental workload, stress, and emotional state assessment during working-like tasks: A comparison with laboratory technologies. *Sensors*, 21(7):2332, 2021.

[31] Ayako Hashizume, Masaaki Kurosu, and Takao Kaneko. Multi-window system and the working memory. In *International Conference on Engineering Psychology and Cognitive Ergonomics*, pages 297–305. Springer, 2007.

[32] Tsung-Sheng Cheng, Yu-Chun Lu, and Chu-Sing Yang. Using the multi-display teaching system to lower cognitive load. *Journal of Educational Technology & Society*, 18(4):128–140, 2015.

[33] WL Lim, O Sourina, and LP Wang. Stew: simultaneous task eeg workload data set. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(11):2106–2114, 2018.

[34] Julie Onton, Arnaud Delorme, and Scott Makeig. Frontal midline eeg dynamics during working memory. *Neuroimage*, 27(2):341–356, 2005.

[35] Sébastien Puma, Nadine Matton, Pierre-V Paubel, Éric Raufaste, and Radouane El-Yagoubi. Using theta and alpha band power to assess cognitive workload in multitasking environments. *International Journal of Psychophysiology*, 123:111–120, 2018.

[36] Arnaud Delorme and Scott Makeig. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21, 2004.

[37] Chi-Yuan Chang, Sheng-Hsiou Hsu, Luca Pion-Tonachini, and Tzyy-Ping Jung. Evaluation of artifact subspace reconstruction for automatic eeg artifact removal. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1242–1245. IEEE, 2018.

[38] Scott Makeig, Tzyy-Ping Jung, Dara Ghahremani, and Terrence J Sejnowski. Independent component analysis of simulated erp data. *Institute for Neural Computation, University of California: technical report INC-9606*, 1996.

[39] Luca Pion-Tonachini, Ken Kreutz-Delgado, and Scott Makeig. Iclabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, 198:181–197, 2019.

[40] Tian Lan, Deniz Erdogmus, Andre Adami, Misha Pavel, and Santosh Mathan. Salient eeg channel selection in brain computer interfaces by mutual information maximization. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pages 7064–7067. IEEE, 2006.

[41] Jaeseung Jeong, John C Gore, and Bradley S Peterson. Mutual information analysis of the eeg in patients with alzheimer's disease. *Clinical neurophysiology*, 112(5):827–835, 2001.

[42] Justin A Blanco, Michael K Johnson, Kyle J Jaquess, Hyuk Oh, Li-Chuan Lo, Rodolphe J Gentili, and Bradley D Hatfield. Quantifying cognitive workload in simulated flight using passive, dry eeg measurements. *IEEE Transactions on Cognitive and Developmental Systems*, 10(2):373–383, 2016.

[43] Jose Delpiano. Fast mutual information of two images or signals. MATLAB Central File Exchange. Retrieved August 20 2021.

[44] L Mandel and E Wolf. Spectral coherence and the concept of cross-spectral purity. *JOSA*, 66(6):529–535, 1976.

[45] RF Woolson. Wilcoxon signed-rank test. *Wiley encyclopedia of clinical trials*, pages 1–3, 2007.

[46] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[47] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.

[48] Sun Hee Na, Seung-Hyun Jin, Soo Yong Kim, and Byung-Joo Ham. Eeg in schizophrenic patients: mutual information analysis. *Clinical Neurophysiology*, 113(12):1954–1960, 2002.

[49] Guillermo González-Mena, Carolina Del-Valle-Soto, Violeta Corona, and Jafet Rodríguez. Neuromarketing in the digital age: The direct relation between facial expressions and website design. *Applied Sciences*, 12(16):8186, 2022.

[50] Mashael Aldayel, Mourad Ykhlef, and Abeer Al-Nafjan. Deep learning for eeg-based preference classification in neuromarketing. *Applied Sciences*, 10(4):1525, 2020.

[51] Vaishali Khurana, Monika Gahalawat, Pradeep Kumar, Partha Pratim Roy, Debi Prosad Dogra, Erik Scheme, and Mohammad Soleymani. A survey on neuromarketing using eeg signals. *IEEE Transactions on Cognitive and Developmental Systems*, 13(4):732–749, 2021.

[52] AHMED H Alsharif, NZM Salleh, ROHAIZAT Baharun, and MOHD EFFANDI Yusoff. Consumer behaviour through neuromarketing approach. *Journal of Contemporary Issues in Business and Government*, 27(3):344–354, 2021.

[53] Himaanshu Gauba, Pradeep Kumar, Partha Pratim Roy, Priyanka Singh, Debi Prosad Dogra, and Balasubramanian Raman. Prediction of advertisement preference by fusing eeg response and sentiment analysis. *Neural Networks*, 92:77–88, 2017.

[54] Rafal Ohme, Dorota Reykowska, Dawid Wiener, and Anna Choromanska. Application of frontal eeg asymmetry to advertising research. *Journal of economic psychology*, 31(5):785–793, 2010.

[55] Liwei Hsu and Yen-Jung Chen. Neuromarketing, subliminal advertising, and hotel selection: An eeg study. *Australasian Marketing Journal*, 28(4):200–208, 2020.

[56] Parnaz Golnar-Nik, Sajjad Farashi, and Mir-Shahram Safari. The application of eeg power for the prediction and interpretation of consumer decision-making: A neuromarketing study. *Physiology & behavior*, 207:90–98, 2019.

[57] M Murugappan, Subbulakshmi Murugappan, Celestin Gerard, et al. Wireless eeg signals based neuromarketing system using fast fourier transform (fft). In *2014 IEEE 10th international colloquium on signal processing and its applications*, pages 25–30. IEEE, 2014.

[58] Mahendra Yadava, Pradeep Kumar, Rajkumar Saini, Partha Pratim Roy, and Debi Prosad Dogra. Analysis of eeg signals and its application to neuromarketing. *Multimedia Tools and Applications*, 76:19087–19111, 2017.

[59] Bülent Yılmaz, Sümeyye Korkmaz, Dilek Betül Arslan, Evrim Güngör, and Musa H Asyalı. Like/dislike analysis using eeg: determination of most discriminative channels and frequencies. *Computer methods and programs in biomedicine*, 113(2):705–713, 2014.

[60] Ariel Telpaz, Ryan Webb, and Dino J Levy. Using eeg to predict consumers' future choices. *Journal of Marketing Research*, 52(4):511–529, 2015.

[61] Rami N Khushaba, Chelsea Wise, Sarath Kodagoda, Jordan Louviere, Barbara E Kahn, and Claudia Townsend. Consumer neuroscience: Assessing the brain response to marketing stimuli using electroencephalogram (eeg) and eye tracking. *Expert systems with applications*, 40(9):3803–3812, 2013.

[62] B Li, Y Wang, and K Wang. Data fusion and analysis techniques of neuromarketing. *WIT Transactions on Engineering Sciences*, 113:396–404, 2016.

[63] Piotr Chynał, Janusz Sobecki, Michał Rymarz, and Barbara Kilijańska. Shopping behaviour analysis using eyetracking and eeg. In *2016 9th International Conference on Human System Interactions (HSI)*, pages 458–464. IEEE, 2016.

[64] Moritz Kassner, William Patera, and Andreas Bulling. Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '14 Adjunct, pages 1151–1160, New York, NY, USA, 2014. ACM.

[65] Luca Pion-Tonachini, Sheng-Hsiou Hsu, Scott Makeig, Tzyy-Ping Jung, and Gert Cauwenberghs. Real-time eeg source-mapping toolbox (rest): Online ica and source localization. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4114–4117. IEEE, 2015.

[66] Sheng-Hsiou Hsu, Tim R Mullen, Tzyy-Ping Jung, and Gert Cauwenberghs. Real-time adaptive eeg source separation using online recursive independent component analysis. *IEEE transactions on neural systems and rehabilitation engineering*, 24(3):309–319, 2015.

[67] Nima Bigdely-Shamlo, Ken Kreutz-Delgado, Christian Kothe, and Scott Makeig. Eyecatch: Data-mining over half a million eeg independent components to construct a fully-automated eye-component detector. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5845–5848. IEEE, 2013.

[68] David E Irwin and Gregory J Zelinsky. Eye movements and scene perception: Memory for things observed. *Perception & psychophysics*, 64(6):882–895, 2002.

[69] Dennis Wobrock, Andrea Finke, Thomas Schack, and Helge Ritter. Using fixation-related potentials for inspecting natural interactions. *Frontiers in human neuroscience*, page 447, 2020.

[70] Benjamin Touchette and Seung-Eun Lee. Measuring neural responses to apparel product attractiveness: an application of frontal asymmetry theory. *Clothing and Textiles Research Journal*, 35(1):3–15, 2017.

[71] Enrica Modica, Giulia Cartocci, Dario Rossi, Ana C Martinez Levy, Patrizia Cherubino, Anton Giulio Maglione, Gianluca Di Flumeri, Marco Mancini, Marco Montanari, Davide Perrotta, et al. Neurophysiological responses to different product experiences. *Computational Intelligence and Neuroscience*, 2018, 2018.

[72] Giulia Cartocci, Myriam Caratù, Enrica Modica, Anton Giulio Maglione, Dario Rossi, Patrizia Cherubino, and Fabio Babiloni. Electroencephalographic, heart rate, and galvanic skin response assessment for an advertising perception study: application to antismoking public service announcements. *JoVE (Journal of Visualized Experiments)*, (126):e55872, 2017.

[73] Qingguo Ma, Kai Wang, Xiaoyi Wang, Cuicui Wang, and Lei Wang. The influence of negative emotion on brand extension as reflected by the change of n2: a preliminary study. *Neuroscience letters*, 485(3):237–240, 2010.

[74] Nobuhiko Goto, Faisal Mushtaq, Dexter Shee, Xue Li Lim, Matin Mortazavi, Motoki Watabe, and Alexandre Schaefer. Neural signals of selective attention are modulated by subjective preferences and buying decisions in a virtual shopping task. *Biological psychology*, 128:11–20, 2017.

[75] Jing Wang and Weiwei Han. The impact of perceived quality on online buying decisions: an event-related potentials perspective. *Neuroreport*, 25(14):1091–1098, 2014.

[76] Qingguo Ma, Xiaoyi Wang, Liangchao Shu, and Shenyi Dai. P300 and categorization in brand extension. *Neuroscience letters*, 431(1):57–61, 2008.

[77] PIA Rämä and Thierry Baccino. Eye fixation–related potentials (efrps) during object identification. *Visual Neuroscience*, 27(5-6):187–192, 2010.

[78] Sridhar Raghavachari, Michael J Kahana, Daniel S Rizzuto, Jeremy B Caplan, Matthew P Kirschen, Blaise Bourgeois, Joseph R Madsen, and John E Lisman. Gating of human theta oscillations by a working memory task. *Journal of Neuroscience*, 21(9):3175–3183, 2001.

[79] Mahsa Bagheri and Sarah D Power. Simultaneous classification of both mental workload and stress level suitable for an online passive brain–computer interface. *Sensors*, 22(2):535, 2022.

[80] Marta Z Zakrzewska and Aneta Brzezicka. Working memory capacity as a moderator of load-related frontal midline theta variability in sternberg task. *Frontiers in human neuroscience*, 8:399, 2014.

[81] Miguel Arevalillo-Herráez, Maximo Cobos, Sandra Roger, and Miguel García-Pineda. Combining inter-subject modeling with a subject-based data transformation to improve affect recognition from eeg signals. *Sensors*, 19(13):2999, 2019.

[82] Chun-Shu Wei, Yu-Te Wang, Chin-Teng Lin, and Tzyy-Ping Jung. Toward drowsiness detection using non-hair-bearing eeg-based brain-computer interfaces. *IEEE transactions on neural systems and rehabilitation engineering*, 26(2):400–406, 2018.

[83] Yuan-Pin Lin. Constructing a personalized cross-day eeg-based emotion-classification model using transfer learning. *IEEE journal of biomedical and health informatics*, 24(5):1255–1264, 2019.

[84] Scott Makeig, Anthony Bell, Tzyy-Ping Jung, and Terrence J Sejnowski. Independent component analysis of electroencephalographic data. *Advances in neural information processing systems*, 8, 1995.

[85] Kuan-Jung Chiang, Chun-Shu Wei, Masaki Nakanishi, and Tzyy-Ping Jung. Boosting template-based ssvep decoding by cross-domain transfer learning. *Journal of Neural Engineering*, 18(1):016002, 2021.

[86] Ching-Tai Chiang, Chen-Sen Ouyang, Rei-Cheng Yang, Rong-Ching Wu, and Lung-Chang Lin. Increased temporal lobe beta activity in boys with attention-deficit hyperactivity disorder by loreta analysis. *Frontiers in Behavioral Neuroscience*, 14:85, 2020.

[87] Mateusz Gola, Mikołaj Magnuski, Izabela Szumska, and Andrzej Wróbel. Eeg beta band activity is related to attention and attentional deficits in the visual performance of elderly subjects. *International Journal of Psychophysiology*, 89(3):334–341, 2013.

[88] W Bastiaan Kleijn, Joao Bernardo Crespo, Richard Christian Hendriks, Petko Petkov, Bastian Sauert, and Peter Vary. Optimizing speech intelligibility in a noisy environment: A unified view. *IEEE Signal Processing Magazine*, 32(2):43–54, 2015.

[89] Bastian Sauert and Peter Vary. Near end listening enhancement: Speech intelligibility improvement in noisy environments. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, volume 1, pages I–I. IEEE, 2006.

[90] Harold H Kim and David M Barrs. Hearing aids: a review of what's new. *Otolaryngology—Head and Neck Surgery*, 134(6):1043–1050, 2006.

[91] Sven L Mattys, Matthew H Davis, Ann R Bradlow, and Sophie K Scott. Speech recognition in adverse conditions: A review. *Language and Cognitive processes*, 27(7-8):953–978, 2012.

[92] Rebecca E Millman and Sven L Mattys. Auditory verbal working memory as a predictor of speech perception in modulated maskers in listeners with normal hearing. *Journal of speech, language, and hearing research*, 60(5):1236–1245, 2017.

[93] Vassil Panayotov, Guoguo Chen, Daniel Povey, and Sanjeev Khudanpur. Librispeech: an asr corpus based on public domain audio books. In *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 5206–5210. IEEE, 2015.

[94] Ricardo Campos, Vítor Mangaravite, Arian Pasquali, Alípio Jorge, Célia Nunes, and Adam Jatowt. Yake! keyword extraction from single documents using multiple local features. *Information Sciences*, 509:257–289, 2020.

[95] DPW Ellis. Time-domain scrambling of audio signals in matlab. *Natick, Massachusetts: Mathworks Inc*, 116401, 2010.

[96] Ole Jensen and John E Lisman. An oscillatory short-term memory buffer model can account for data on the sternberg task. *Journal of Neuroscience*, 18(24):10688–10699, 1998.

[97] Liang-Tien Hsieh and Charan Ranganath. Frontal midline theta oscillations during working memory maintenance and episodic encoding and retrieval. *Neuroimage*, 85:721–729, 2014.

[98] Adrian P Burgess and John H Gruzelier. Short duration synchronization of human theta rhythm during recognition memory. *Neuroreport*, 8(4):1039–1042, 1997.

[99] Ville Satopaa, Jeannie Albrecht, David Irwin, and Barath Raghavan. Finding a" kneedle" in a haystack: Detecting knee points in system behavior. In *2011 31st international conference on distributed computing systems workshops*, pages 166–171. IEEE, 2011.

[100] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti S. Hämäläinen. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7(267):1–13, 2013.

[101] Eric Maris and Robert Oostenveld. Nonparametric statistical testing of eeg-and meg-data. *Journal of neuroscience methods*, 164(1):177–190, 2007.

[102] Jona Sassenhagen and Dejan Draschkow. Cluster-based permutation tests of meg/eeg data do not establish significance of effect latency or location. *Psychophysiology*, 56(6):e13335, 2019.

[103] Urs Maurer, Silvia Brem, Martina Liechti, Stefano Maurizio, Lars Michels, and Daniel Brandeis. Frontal midline theta reflects individual task performance in a working memory task. *Brain topography*, 28:127–134, 2015.

[104] Inga M Schepers, Till R Schneider, Joerg F Hipp, Andreas K Engel, and Daniel Senkowski. Noise alters beta-band activity in superior temporal cortex during audiovisual speech processing. *Neuroimage*, 70:101–112, 2013.