Analysis of EEG Signals & Cognitive Activity in 3D Modeling for a Multi Modal Interface System

by

Muhammad Zeeshan Baig

A Dissertation Submitted In Fulfilment Of The Requirements For The Degree Of

DOCTOR OF PHILOSOPHY

Department of Computing
Faculty of Science and Engineering
Macquarie University
Sydney, Australia

August 15, 2019

Statement of candidate

I certify that the work in this thesis entitled "Analysis of EEG Signals & Cognitive Activity in 3D Modeling" has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree to any other university or institution other than Macquarie University.

I also certify that the thesis is an original piece of research and it has been written by me. Any help and assistance that I have received in my research work and the preparation of the thesis itself have been appropriately acknowledged.

In addition, I certify that all information sources and literature used are indicated in the thesis.

....

Muhammad Zeeshan Baig

January 19, 2020

Acknowledgements

First and foremost, I would like to express my sincere gratitude and appreciation to my supervisor, Associate Professor Manolya Kavakli, for her continuous support, supervision, encouragement, and suggestions throughout my Ph.D. candidature. Her wide knowledge and her passion for research and life have been of great value for me. I am extremely thankful to her for giving me the opportunity to work and to use facilities of the VR lab at Simulation Hub. Literally, without her continuous support and endless guidance, this dissertation would not have been possible. It has been a great honour to be her Ph.D. student.

In addition, my colleagues have helped me to develop this work. I wish to express my thanks to John Porte, Muhammad Touseef Afzal Rana, Charles Liu and Ahsan Ali for providing a friendly and enjoyable environment during these years. I sincerely thank all the people in the Department of Computing, Faculty of Science and Engineering, Macquarie University, for their warm support and help.

I am, and always will be, grateful to my parents and siblings for their love, support, blessing, and encouragement throughout my life. Their understanding, optimism, and persistence were invaluable, and to them, I offer my thanks and love. I also owe my deepest appreciation to my teachers from whom I have learned so many great lessons.

I am also indebted to the Macquarie University for their generous funding through the International Macquarie University Research Excellence Scholarship (iMQRES), and the Macquarie University Higher Degree Research Fund, which removed the financial concerns of my studies. Finally, above all, I owe it all to Almighty Allah for granting me the wisdom, persistence, health, and strength to undertake this research task and for enabling me to see it through to completion.

RELATED PUBLICATIONS

10 publications have been produced and listed here in reverse chronological order. 6 of these are conference papers and 4 of them are journal papers. 8 of these papers have been published. Mostly ERA A or CORE A conferences and high impact factor journals have been targeted.

Refereed research publications from this thesis:

1 **Baig, Muhammad Zeeshan**, and Manolya Kavakli. "Connectivity Analysis of Functional Brain Networks in Using Multi-modal Human-Computer Interaction. " *28th International Conference on Information Systems Development (ISD2019)*. Toulon, France. **Accepted for publication**. **CORE A**

2 **Baig, Muhammad Zeeshan**, and Manolya Kavakli. "Connectivity analysis using functional brain networks to evaluate cognitive activity during 3d modelling." *Brain Sciences* 9, no. 2 (2019): 24. **Impact Factor: 2.786**

3 **Baig, Muhammad Zeeshan**, and Manolya Kavakli. "A Survey on Psycho-Physiological Analysis Measurement Methods in Multimodal Systems." *Multimodal Technologies and Interaction* 3, no. 2 (2019): 37.

4 **Baig, Muhammad Zeeshan**, and Manolya Kavakli. "Expertise Classification using Functional Brain Networks and Normalized Transfer Entropy of EEG in Design Applications." *In Proceedings of the 2019 11th International Conference on Computer and Automation Engineering*, pp. 41-46. ACM, 2019. Brisbane, Australia.

5 **Baig, Muhammad Zeeshan**, and Manolya Kavakli. "Analyzing Novice and Expert User's Cognitive Load in using a Multi-Modal Interface

System." *In 2018 26th International Conference on Systems Engineering (ICSEng),* pp. 1-7. IEEE, 2018. Sydney, Australia. **CORE C**

6 **Baig, Muhammad Zeeshan**, and Manolya Kavakli. "EEG Signal Analysis in 3D Modelling to Identify Correlations Between Task Completion in Design User's Cognitive Activities." *In 25th International Conference on Neural Information Processing,* pp. 340-352. Springer, Cham, 2018. Siem Reap, Cambodia. **CORE A**

7 **Baig, Muhammad Zeeshan**, and Manolya Kavakli. "Qualitative analysis of a multimodal interface system using speech/gesture." *In 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pp. 2811-2816. IEEE, 2018. Wuhan, China. **ERA A**

8 Alibay, Farzana, Manolya Kavakli, Jean-Rémy Chardonnet, and **Muhammad Zeeshan Baig**. "The usability of speech and/or gestures in multimodal interface systems." *In Proceedings of the 9th International Conference on Computer and Automation Engineering,* pp. 73-77. ACM, 2017. Sydney, Australia.

Submitted Publication:

1 **Baig, Muhammad Zeeshan**, and Manolya Kavakli. Multimodal Systems: Taxonomy, Methods, and Challenges. Submitted in ***ACM Computing Surveys*. Impact Factor: 6.131**

2 **Baig, Muhammad Zeeshan**, and Manolya Kavakli. Classification of User Expertise Levels using EEG and Convolutional Neural Network in Modelling Application. Revision submitted in ***Expert Systems with Applications.* Impact Factor: 4.292**

ABSTRACT

The human brain uses a complex network of billions of neurons functioning together. Through learning and experience, the human brain establishes millions of connections between neurons. Although the individual functions of neurons are known, how these neurons work in a network to perform cognitive processes still requires research and investigation. Every human being has their own learning rate to understand things and develop a skill-set. There is a need for adaptive systems to change the pace of learning according to the user's competency level to have an impact on performance. This thesis explores the application of EEG signals to estimate the cognitive activity of competent and novice users in a design task. The main goal of this thesis is to identify the user's competency using cognitive activities acquired through EEG signals in an MMIS. We developed a multimodal interface system (MMIS) (xDe-SIGN v2) that allows the users to model a 3D object using speech and gesture modalities. We used Microsoft speech recognition API to detect and decode speech input and a Leap Motion sensor and API for gesture recognition.

Research questions are classified into 6 groups: input modality, psycho-physiological analysis, cognitive activity, information processing, and competency classification. The research questions are investigated in four major parts: a) the design and development of an MMIS (Chapter 4) b) qualitative evaluation of MMIS using speech and gestures for 3D modelling (Chapter 4) c) quantitative evaluation of MMIS using EEG signals (Chapter 6-9) d) classification of user's competency level for adaptive systems design (Chapter 10). We tested the usability of the system in 2 sets of experiments with 12 participants. We used EEG signals to record users' mental states and cognitive activity. First, we analyzed users' cognitive activity in a unimodal system (using keyboard and mouse inputs), and then, in a multimodal system (using

speech and gesture inputs). We used a combination of qualitative methods such as questionnaires and quantitative methods such as EEG bands, Power Spectral Density (PSD) and Functional Brain Networks (FBN) to investigate the cognitive activity of novice and competent users.

Our qualitative evaluation results supported by questionnaires indicate that speech and gestures were well-coordinated in human to human communication but not in human-computer interaction (HCI). However, speech and gestures could be used in HCI with proper pre-processing and optimization techniques, as 90% of the participants completed the given task with reasonable precision in xDe-SIGN v2. Our quantitative evaluation results supported by EEG power analysis showed that there are significant differences in the alpha, beta, and theta band activity of novice and competent users. The results also suggest that physical actions such as drawing, manipulation and moving 3D models have a direct impact on users' performance defined by task completion time, as competent users performed 1.5 times more physical actions than novices who had twice as many conceptual actions as competent users. These findings suggest that the structure of cognitive actions is the key to high performance.

Directional FBN analysis also indicates significant differences in cognitive activity in both novice and competent users in various states. The cognitive activity is more intense while the participants use speech and gestures for 3D modelling. The frontal region of the brain is mostly active, which indicates the use of short-term memory. The thesis provides experimental evidence that EEG based measures can be used as a quantitative metric to analyze cognitive activity in HCI. Finally, we have proposed a method to classify user's competency levels using convolutional neural networks and EEG signals. We obtained a classification accuracy of more than 88%, which shows the effectiveness of the proposed method. Thus, we conclude that the proposed method has a clear potential for developing state-of-the-art adaptive systems that can adapt to users' competency levels.

Contents

List of Tables

List of Figures

Acronyms

2D	2 Dimensional
3D	3 Dimensional
4D	4 Dimensional
AAA	Adaptive Agent Architecture
ADHD	Attention Deficit Hyperactivity Disorder
ANS	Autonomic Nervous System
API	Application Program Interface
BCI	Brain Computer Interface
BVP	Blood Volumn Pulse
CAD	Computer Aided Design
CD	Connectivity Density
CE	Central Executive
CNN	Convolutional Neural Network
CSP	Common Spatial Patterns
CT	Completion Time
DTI	Diffusion Tensor Imaging
DTW	Dynamic Time Warping
ECG	Electrocardiogram
EEG	Electroencephalography
EOG	Electrooculogram
EMG	Electromyography
FBN	Functional Brain Network
FFT	Fast Fourier Transform
fMRI	functional Magnetic Resonance Imaging
FT	Fourier Transform
GC	Granger Causality
GKP	Glossokinetic Potential
GOMS	Goals,Operator,Methods,Selection rules
GSR	Galvanic Skin Response
GUI	Graphical User Interface
HCI	Human Computer Interaction
HHC	Human Human Communication
HMM	Hidden Markov models
HRI	Human Robot Interaction
HRV	Heart Rate Variability
ICA	Independent Component Analysis

IR	Infrared
kNN	K-nearest neighbors
LDA	Linear Discriminant Analysis
LTM	Long-Term Memory
MEG	Magnetoencephalography
MI	Mutual Information
MMIS	Multi-Modal Interface System
MRI	Magnetic Resonance Imagining
MRT	Multiple Resource Theory
MS	Microsoft
NTE	Normalized Transfer Entropy
NN	Neural Networks
OAA	Open Agent Architecture
PCA	Principle Component Analysis
PDC	Partial Directed Coherence
PL	Phonological Loop
PNS	Parasympathetic Nervous System
PPG	Photoplethysmography
PSD	Power Spectral Density
RR	Respiration Rate
SDK	Software Development Kit
SFS	Sequential Forward Search
sMRI	Structural Magnetic Resonance Imaging
SNS	Sympathetic Nervous System
ST	Skin Temperature
STM	Short-term Memory
SVM	Support Vector Machines
SWNT	Small World Network Theory
TE	Transfer Entropy
TTS	Text to Speech
VSS	Visuospatial Sketchpad
VR	Virtual Reality
WM	Working Memory
WWW	World Wide Web
XML	EXtensible Markup Language

Chapter 1

Introduction

"Teaching is the only major occupation of man for which we have not yet developed tools that make an average person capable of competence and performance."

-Peter Drucker-

The founder of modern management

1.1 Background

1.1.1 Competence

The emergence of the information society with its necessity for lifelong learning has brought the concept of "knowledge worker's productivity" to the frontier of management, which was coined by Drucker in 1959 [18]. Drucker's core concept, "management of objectives," has never been proven to work effectively. There is a fine balance between overemphasizing control and fostering creativity to meet the goals. Iazzolino and Laise [19] proposed a deep theoretical analysis based on Pulic's theory of human capital efficiency, which may primarily be used in measuring corporate performance.

How can we measure knowledge worker's productivity and competency?

Competence is the set of demonstrable characteristics and skills that enable, and improve the efficiency or performance of a task [20, 21]. Competence can be seen as a combination of practical and theoretical knowledge, cognitive skills, behavior and values used to improve performance. Competency is measurable and able to be broken into smaller criteria. It could be developed through training. Competency models are widely used in business for assessing competencies within the organization for performance management. Regardless of training, competency grows through experience and the extent of an individual's capacity to learn and adapt. In this thesis, our goal is to shed light into the relationship between competency and individual characteristics of knowledge workers.

Competencies are the skills that a novice user attains by going through various phases such as education, training, and experience. The difference between competent and novice users is the way of representing and processing knowledge. The training and learning time varies from person to person, but some human-dependant factors can reduce the training time and increase the learning rate of the user. Various researchers suggest that use of human-like ways for interaction can improve user learning rate [3]. Many efforts have been made to bring this hypothesis to life. However, the systems are still immature due to technological and psychological limitations. Investigating the user-dependent factors to overcome the technological and psychological barriers could make the system robust, flexible, and efficient. Development of adaptive systems is another solution to increase the learning rate of the user by changing the system according to the mental state of the user, but the parameters on which the system will adapt are still unknown.

Researchers have suggested that the **multimodal interaction systems can provide a high degree of freedom and reduce the mental effort of the user [1]**, but research is quite limited to back this argument. There are two possible ways to validate the argument. First, we can evaluate the system using qualitative methods such as questionnaires and interview, but the results could be biased. Second, we can use quantitative methods to evaluate the interaction, such as to measure cognitive load, emotional response, and brain activity.

1.1.2 Expertise

An expert is a person with extensive knowledge or ability based on their research, experience, or occupation in a particular area of study. A novice who undergoes extensive training and research becomes an expert, and many studies focus on understanding expertise in a specific field such as sports or music [22]. There is a general understanding that expertise develops over time when a person experiences and learns things, but after achieving the peak performance, an inevitable decline period begins. The peak performance age varies from person to person or profession to profession. For example, in science, the peak performance age may be in the thirties, while it may be in the forties in art [23], but one thing that is clear in studies of expertise is that it requires a minimum period of practice and continuous involvement before getting to the peak performance stage.

Expertise development usually has various phases. At the start, a novice must train and educate himself in a chosen field and later on becomes an expert by accumulating experience. Experts and novices have different ways to represent knowledge (extent, organization, abstraction, and consolidation) which determine how they retrieve information [24] and solve problems [25]. To some extent, novices utilize a depth-first search approach to solve a problem, whereas an expert predominantly uses breadth-first or top-down strategies. Most of the literature is filled with the study of game-playing environments or problem-solving to analyze differences between novices and experts. A problem-solving study between novices and experts [26] suggests that experts use explicit problem-decomposing schemes and sometimes apply a bottom-up approach to solve the problem, different from novices. Some examples such as creative writing and programming also show contradictions from the standard results of expertise, e.g., an expert will generally solve an ill-defined problem in the easiest possible way or with ease compared to a novice [22]. Protocol analysis of junior and senior industrial design students show that some students get stuck on information gathering [27]. They have found that junior students do not gather necessary information, and tend to solve the problem, totally unaware of the criteria and difficulties.

1.1.3 Research Problem

One of the major problems in Multimodal Interface System (MMIS) and adaptive system development is the detection of levels of proficiency. In this thesis, we study the differences in information processing between novices and competent users using a Multimodal Interface System (MMIS) for a 3D object modeling task. There are some questions that needs answering such as what are the differences between novices and competent or expert users' cognitive activities? Can certain methods make the transition from novice to expert more effective or efficient [28]? These questions motivates us to conduct this PhD project. There are 5 levels of proficiency:

1 Novice

2 Advanced beginners

3 Competent

4 Proficient

5 Expert

There is a fine line between expertise and competence. In our terminology, *what a person knows refers to expertise and what a person does refers to competence.* A novice has no experience in the situation in which they are expected to perform. It is a well-established fact that training contributes to the progression of a novice to an expert.

Expertise is developed by someone who has gone through intense and prolonged learning and practice in a specific field. To become an expert, a novice user has to get through a long period of learning and training to become an expert. The learning and training time might be different for different users and depends upon several factors [28]. These factors define how long a user takes to become an expert. Researchers have suggested that using a human-like way of interaction with the computers can decrease the training time and increase the learning rate of novice users [29]. Various efforts have been made to use human-like inputs such as speech, touch, and gesture to control a digital environment, but due to technological limitations, these systems have not come up to expectations.

In the case of conceptual design, such as modeling a 3D object, speech and gesture inputs can provide a more humane way of communicating with the computers but to develop such a system, we have to understand the human dependent factors that affect the performance of a user. These factors can be psychological or technological. Some systems use speech and gesture recognition techniques to draw naturally in a 3D environment, but due to technological limitations, these systems are not up to expectations when it comes to conceptual design.

In the existing systems, differences in information processing can be observed between males and females, novices and experts and left-handed and right-handed people while describing a simple 3D object using speech and gestures [30]. The study of these user-dependent factors makes the interaction robust in a multi-modal interaction system and enhances system flexibility, efficiency, naturalness. It also provides the user with a high degree of freedom to choose modality based on situation, task, context, and comfort. Multi-modal interaction systems are expected to reduce the overall cognitive load of the user compared to unimodal systems.

1.1.4 Multimodal Interface Systems

The goal of an MMIS is to narrow the gap between Human-Computer Interaction (HCI) and Human-Human Communication (HHC). A Multi-Modal Interface System (MMIS) provides input flexibility, adaptability, and accessibility to a wider range of users than unimodal interfaces [30]. The first MMIS dates back to Bolt's "Put That There" experiment in 1980, which uses simple speech and hand-pointing commands as input [5]. With advancements in technologies, the focus has shifted towards speech and hand-gesture based MMIS. Over the past decades, many researchers have worked on improving the recognition rate of speech and gestures. For example, a detection rate of 95% has been achieved for small vocabularies [31], a recognition rate of 90% for discrete hand gestures [32] and 95% for vision-based hand gesture-recognition [33]. However, for designing an efficient MMIS, each level of modality should be investigated and combined properly with other modalities to achieve a better experience. Figure 1.1 shows a typical structure of an MMIS. The first part of MMIS architecture is to recognize the various user's modalities, such as speech, gesture, and facial expressions. The second task is to integrate and interpret the recognition information. The last part is to generate MMIS output based on the interpretation of the input.

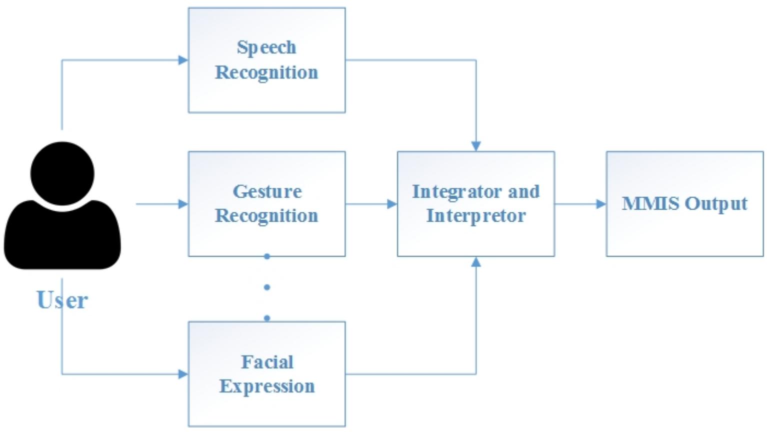


Figure 1.1: Multi-modal Input System structure

Each phase of the MMIS structure affects performance. The user's gender, culture, experience, and age may influence the interaction with an MMIS. Currently, researchers design MMISs for a range of applications, devising ways to minimize the user-dependent variables and develop an efficient MMIS system.

1.1.5 User-dependent Factors

How user-related factors affect the design of an MMIS requires investigation. There is a need to understand the reasons behind these user-dependent factors to address them in an MMIS. The most common way to study the user-dependent factors is to use survey, questionnaires, and interviews and get user feedback. These qualitative evaluation methods have many drawbacks, such as biases based on gender, competency, and culture [34]. A quantitative evaluation metric is needed to analyze user activity during an interaction. One possible solution is to use brain signals such as Electroencephalography (EEG) to analyze the user's cognitive activity during the interaction. In this thesis, we compare variations in user' cognitive activities while using unimodal and multimodal systems using EEG signals.

1.1.6 Methodologies

Although various functions of the brain have been explored in detail in the literature, the underlying neural processes that contribute towards human cognition still need investigation [35]. Generally, cognition is a combination of various low and high-level cognitive processes; the phenomena itself involves highly complex networking of billions of neurons making the whole process quite complicated [36]. Despite all these complexities, there are many non-invasive technologies available that can record human brain activity in terms of electrical signals such as electroencephalography (EEG), magnetoencephalography (MEG), magnetic resonance imagining (MRI) and functional magnetic resonance imaging (fMRI) [36]. Choosing the technique for recording brain activity is governed by the type of analysis. MRI/fMRI offers high resolution and very considerable at-source localization, but faster changes in cognition activity take some time to appear in fMRI/MRI but can be observed in EEG easily because of its high temporal resolution. EEG is more affordable than other technologies and is therefore considered the most suitable choice for contemporary research [37].

From the past decade, neuroscientists have shifted the research focus from the conventional ways of analyzing EEG signals, such as power analysis of various EEG bands, to connectivity analysis [38]. Functional connectivity gives a much better understanding of 'when and how' of the cognitive activity. Graph theory is commonly used to study these functional connectivities to find factors such as strength and flow of information [39]. To compute the connectivity, statistical information measures are commonly used, such as Pearson's correlation coefficient, coherence, Granger Causality, partial directed coherence, mutual information, and transfer entropy [40]. Although these measures have been used extensively to analyze complex networks such as social networks and the World Wide Web, they did not incorporate the direction of data flow but recently the directional network analysis has gained interest because it captures some subtleties and essential information about the network [41].

1.2 Motivation

This research is a union of neuroscience, information science, and computing to study the user behavior in multimodal HCI system. This trans-disciplinary research reveals some intriguing insights and subtleties that were not observed in the literature. It has direct implications in the field of education and games and indirect implications in psychology, and behavioral studies.

Many attempts have been made to use speech and gestures in designing multi-modal interfaces (MMIS). We [42, 43] have developed a 3D object modeling system (xDe-SIGN), but many limitations degrade the performance of this MMIS. The main reason is the complexity of vocabulary used to draw a 3D object. As stated in [42], to create a simple 3D object is difficult for even a competent user (e.g.a CAD expert) using speech and gesture modes of input. The system must be able to accommodate the communication mode desired by the user and adapt to the user.

The available metrics to evaluate the interactions are qualitative such as questionnaires and interviews [34]. Most of the times these qualitative measures are biased and to improve the system evaluation metrics and decrease the user's bias, there is an urgent need to develop a quantitative measure that can remove the user-related effects. If a validated evaluation metric could be developed that uses the measures to estimate the user's mental effort through cognitive activities, it will lay the basis for the development of adaptive intelligent systems.

Every human being is different in their ways of communication. Diversity plays a different role in a complex system, where it merely produces variation around the mean for performance measures. Through learning and experience, the human brain establishes millions of connections between neurons. Every human being has their own learning rate to understand things and develop a skill-set. There is a need for adaptive systems to change the pace of learning according to the user's competency level to have an impact on performance. Cognitive activities could be exploited to develop adaptive learning systems.

1.3 Significance

According to a report published in 2012 by the Australian Bureau of Statistics, the design and manufacturing industry is worth $100 billion with a continuous growth [44]. Professional designers and architects are hired to construct prototypes by using a 3D modeling software. If the end user wants to change some aspects of the design, this comes at an extra cost. Researchers try to find novel ways to design 3D models, either by using virtual reality (VR) tools, or using multimodal interface systems.

There are computer graphics software products that allow non-experienced users to create and manipulate 3D objects without any help or prior knowledge of the software but these systems are not up to the mark. To develop an ideal system, the system should have the ability to obtain information from the humans effectively, but there are technological, physiological, and psychological limitations to overcome [3]. Humans utilize all available modalities to convey information, such as speech, gestures and facial expressions. The human brain's perceptual and cognitive functions are synced to accommodate human-human interaction; this is not the case in human-computer interaction (HCI).

The research aims to investigate the cognitive activities of various user in a 3D modeling task. The better understanding of underlying cognitive activities can help us understand the mental load of a user and to develop an adaptive 3D modeling system that can adapt to the variations in the user cognitive activities. The information about mental load and cognitive activities can be used to help 2-4% of the people with learning disabilities or difficulties by developing adaptive learning systems [45]. In addition, EEG signals can be used to develop the adaptive games in which the difficulty level can be changed based on the user mental state.

1.4 Goals and Research Questions

The main goal of this thesis is to identify the user's competency using cognitive activities acquired through EEG signals in an MMIS. The goal is achieved by experimenting with various participants while they are performing a 3D modeling task. Both qualitative and quantitative (EEG) data were recorded for analysis. The research questions can be classified in 6 groups as follows:

RQ 1. Input Modalities:

RQ 1.1 What modalities are suited the most to the development of an MMIS for

3D modeling?

RQ 1.2 What kind of integration techniques should be used to fuse the multimodal inputs?

RQ 1.3 Is it possible to develop a multi-modal 3D object manipulation system xDe-SIGN v2 using speech and gestures?

RQ 1.4 What are the limitations of using speech and gestures in MMIS?

RQ 2. Psycho-physiological Analysis:

RQ 2.1 Can we use psycho-physiological analysis in an HCI system?

RQ 2.2 Which EEG parameters can be used for evaluating the cognitive activity?

RQ 3. Cognitive Activity:

RQ 3.1 Why do some novice users perform better than others?

RQ 3.2 What are the factors that affect novice users' performance?

RQ 4. Information Processing:

RQ 4.1 Are there any differences in information processing and cognitive activity between novice and competent users?

RQ 4.2 Can Functional Brain Networks (FBNs) be used to identify the information flow patterns?

RQ 5. Modality Comparison:

RQ 5.1 What are the differences in cognitive activity between multimodal and unimodal systems?

RQ 5.2 Does competency play a role when a new set of inputs are used for a predefined task?

RQ 6. Competency Classification:

RQ 6.1 Can EEG signals be used in classifying a user's competency level?

RQ 6.2 Which features contribute the most towards the classification of competence?

To address the research questions mentioned above, the main tasks are listed as follows:

Task. 1: Investigate the usability of speech and gesture modalities in a 3D multimodal interface system.

Task. 2: Identify the differences in cognitive activities among novice and competent users.

Task. 3: Explore the use of EEG signals in developing a cognitive measure for performance evaluation.

Task. 4: Quantify the changes in binary and weighted FBNs of various users.

Task. 5: Estimate the information flow patterns in various modeling stages.

Task. 6: Build a model for the classification of user's competency levels using EEG signals.

1.5 Major Contributions

The thesis makes three major contribution in the domain of MMIS. These three domains are multimodal systems, analysis of cognitive activity and classification of competency.

- **MMIS development:** A multimodal interface system is developed and evaluated to investigate the usability of speech and gesture in a 3D modeling task. We evaluated the system through questionnaires and video log and found that using a multimodal system is exhausting compared to a traditional unimodal system. Our evaluation results supported by questionnaires showed that speech and gestures are not coordinated effectively in human-computer interaction (HCI). However, using proper processing and optimization techniques, speech and gestures may be successfully used in an MMIS.

- **Analysis of cognitive activities:** The thesis has explored the use of EEG signals for the development of a cognitive measure for performance evaluation in HCI. Various techniques have been used in the thesis to study the cognitive activities of novice and competent users. The applicability of techniques such as power analysis and functional brain networks (FBNs) was evaluated to quantify the variation in user's cognitive activities when performing a modeling task. The findings demonstrate that competent users perform better because they are engaged in physical (drawing) actions more than conceptual actions. We have found significant differences between novice and competent users cognitive activities. The power analysis show that competent users in a 3D modeling software (AutoCAD) were experiencing difficulties in using a multimodal system compared to novices. The results show that EEG power analysis and FBNs may be used as a cognitive measure for user-based HCI evaluation.

- **Classification of competency:** The thesis has also proposed a classification method for developing adaptive systems by incorporating an EEG-based measure of competency as a parameter for adaptation. The study uses different features and a convolutional neural network to validate the method. Classification accuracy of above 80% has been achieved using the raw EEG signals in a reasonable time period, which showed the significance of the proposed method. The proposed method may be used to develop adaptive learning systems for people with learning disabilities and difficulties. The other major application could be EEG-based adaptive games.

1.6 Thesis Organization

The remainder of this thesis is organized as follows:

Chapter 2 (Multimodal Interface Systems) describes multimodal interface systems and modality in detail. It provides a brief overview and history of multimodal systems and discusses the advantages of multimodal systems over unimodal systems. Chapter 2 further explains various modalities, input modes, and information fusion techniques, user modeling, data collection, testing, and evaluation of multimodal systems. Finally, recent applications and challenges are provided at the end. The literature review answers the research questions **RQ 1.1** and **1.2**. The review is submitted to ACM Computing Surveys as a tutorial paper.

Chapter 3 (Human Cognition, Psycho-physiological Analysis & Functional Brain Networks) starts by explaining cognition, including processing, memory, attention, and decision making. The chapter reviews the literature related to psychophysiological analysis, including emotional states and cognitive analysis. The advantages of EEG over other neuroimaging tools are discussed as well as statistical measures used to analyze the connectivity in Functional Brain Networks. Finally, the application of graph theory in studying complex FBNs is provided. The chapter attempts to answer **RQ 2.1** and **2.2** from the literature and the review has been published as a paper, titled as "*A Survey on Psycho-Physiological Analysis & Measurement Methods in Multimodal Systems*" in Multimodal Technologies and Interaction Journal [11].

Chapter 4 (A Multi-Modal Interface System Design (MMIS), Development & Evaluation) explains the development of a multi-modal interface system (xDe-SIGN) that allows the user to design 3D objects in AutoCAD using speech and gesture inputs. The chapter describes the system design methodology and iterative development with implementation details. The chapter lists the improvements in version 2 of xDe-SIGN. Finally, a qualitative evaluation of the system is discussed. In the chapter, we find the answers to research questions **RQ 1.3** and **1.4**. The initial evaluation results have been published in the 9th ACM International Conference on Computer and Automation Engineering (ICCAE 2017) held in Sydney, Australia, in a paper titled " *The Usability of Speech and/or Gestures in Multi-Modal Interface Systems*" *[42]*. The updated version of the MMIS is published in the proceeding of 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), in a paper titled "*Qualitative analysis of a multimodal interface system using speech/gesture*" [43].

Chapter 5 (Experimentation and Instrumentation) explains various aspects of experimental design, the EEG signal acquisition procedure, artifact removal, and EEG pre-processing techniques. The chapter introduces the equipment used in the thesis to record the EEG signals and the software used for pre-processing and analysis.

Chapter 6 (Analysis of Cognitive Activities in a Uni-modal System: using Design Coding Technique) presents the experimental findings of using design techniques to estimate the cognitive activity of a user in a 3D modeling task. A new method for EEG signal segmentation based on design coding technique is presented. The chapter focus is on answering research questions **RQ 3.1** and **3.2**. The findings have been published at the 25th International Conference on Neural Information Processing (ICONIP 2018) held in Siem Reap, Cambodia, in a paper titled "*EEG Signal Analysis in 3D Modelling to Identify Correlations Between Task Completion in Design User's Cognitive Activities*" [46].

Chapter 7 (Analysis of Cognitive Activities in a Uni-modal System: using Transfer Entropy and Functional Brain Networks) explores the use of normalized transfer entropy (NTE) from EEG signals to construct directed FBNs. The aim is to detect and identify patterns of information flow in a cognitive activity (a 3D modeling task). The chapter explains the use of graph-theory to analyze the FBNs and further identifies the changes in connectivity patterns. In the chapter, research questions **RQ 4.1** and **4.2** were investigated. The analysis results have been published in the Brain Sciences Journal titled "*Connectivity Analysis Using Functional Brain Networks* *to Evaluate Cognitive Activity during 3D Modeling*" [47].

Chapter 8 (Comparative Analysis of Cognitive Activity: using Power Spectral Density) contains a comparative analysis of a unimodal and a multimodal system in a design application. The chapter uses the Power Spectral Density (PSD) of various EEG bands for analysis and compares the activity of novice and competent user in using an MMIS. In the chapter, we used PSD analysis to answer research questions **RQ 5.1** and **5.2**. The results have been presented at the 26th International Conference on Systems Engineering (ICSEng) 2018 held in Sydney, Australia, with a paper titled "*Analyzing Novice and Expert User's Cognitive Load in using a Multi-Modal Interface System*" [48].

Chapter 9 (Connectivity Analysis: using Transfer Entropy and Functional Brain Networks) provides a comparative analysis of unimodal and multimodal systems using FBNs. Graph theory measures are used to analyze the network, Data mining algorithms are used to find the regions with intense activity. The chapter uses a different approach (FBN analysis) to answer the research **RQ 5.1** and **5.2**. The results of the study have been accepted for publication at *the 28th International Conference on Information Systems Development (ISD2019)* to be held in Toulon, France.

Chapter 10 (Classification of User's Competency using Convolutional Neural Networks) describes the proposed framework for the classification of user's competency levels using a convolutional neural network (CNN). The aim of this chapter is to design a method that can help researchers to develop adaptive systems that use EEG activity for adaptation. The research questions **RQ 6.1** and **6.2** were explored in the chapter. The results of this study have been under review in the Expert Systems with Applications Journal.

Chapter 11 (Conclusion and Future Work) provides a summary of the thesis and the significant findings. The chapter explains the implications of the findings and limitations of the thesis, along with recommendations for future research.

Chapter 2

Multi-Modal Interface Systems

This chapter starts with an introduction to the Multimodal interface systems (MMIS) with a brief overview and history. It then compares multimodal systems with unimodal systems and presents a detailed description of various modalities, input modes and information fusion techniques, user modeling, data collection, testing, and evaluation of MMIS. The chapter concludes with recent applications of MMIS outlining current challenges in Human-computer interaction (HCI) domain.

2.1 Introduction

The interaction between humans and the world is inherently multi-modal [49]. Humans utilize multiple senses to get an understanding of the environment. All the available senses are employed, both in series and in parallel, to continuously explore the environment and to perceive new information about the environment. The senses used in exploring the outside world can be sight, touch, hear, taste, and smell. These senses when used in a collaborative manner give the humans some useful insights about the surrounding - for example, sight is used to see an object, touch to identify the material, hearing to identify the source of the sound and estimate the location. Multiple modalities support humans to interact with the world and other human beings effectively.

In contrast to human sensing techniques to interact which are inherently multi-modal, the human-computer interaction (HCI) techniques are primarily uni-modal - e.g., writing text using a keyboard. The goal of multimodal interaction research is to design and develop interfaces, technologies, and interaction methods that eliminate the limitations of HCI and unlock its full potential. The introduction of new and sophisticated recognition algorithms enables the use of speech, gesture, facial expression, and other modalities in the development of multi-modal interaction systems [50]. While the chances are highly unlikely that a multimodal interface completely replaces the traditional desktop or Graphical User Interface (GUI), the importance of MMIS is growing with the technological advancements.

An MMIS provides a direct interface for HCI between man and machine. Multi-modal interaction is achieved by combining the contribution of several research areas including signal and image processing, computer vision, artificial intelligence, and many others. The overreaching goal of developing MMIS is to make computing technology more usable by people. To improve the usability of an interface, three things need to be studied: the user, the system and the interaction. We have reviewed the literature to find answers to the following research questions:

RQ 1.1 What modalities are suited the most to the development of an MMIS for 3D modeling?

RQ 1.2 What kind of integration techniques should be used to fuse the multimodal inputs?

2.2 Overview of Multimodal Interaction

A multimodal system aggregates two or more user input modes in an interconnected fashion with the multimedia output. The user input can be speech, pen, touch, manual gestures, gaze, head and body movements [51]. Several research studies have been conducted in the field of multimodal interaction to develop systems that utilize human behavior and a common language to interact [52]. These systems use a recognition algorithm to identify the behavior and language. According to Oviatt [51], " *Multimodal interfaces process two or more combined user input modes (such as speech, pen, touch, manual gesture, gaze, and head and body movements) in a coordinated manner with multimedia system output. They are a new class of interfaces that aim to recognize naturally occurring forms of human language and behavior, which incorporate one or more recognition-based technologies (e.g., speech, pen, vision)*".

The term "multimodal" can be used in many contexts, but in HCI, we use a more human-centered approach to define multimodal interaction. Modality is the mode of communication used as input to activate the computer, and it is a measure of human senses and actions. The input modalities correspond to the human senses are cameras (sight), microphones (hearing), haptic (touching) [53], olfactory (smell) [54] and electronic tongue (taste) [55]. The biofeedback input devices such as galvanic skin response (GSR), Electrocardiogram (ECG), Electroencephalogram (EEG) and many other, used to measure the internal activity of humans, are also considered as input modalities [3].

In HCI, the most common interfaces are perceptual, attentive, and enactive interfaces.

- Perceptual interfaces provide natural, rich, and efficient interaction with the computer using multimodal inputs, and it is highly interactive [56].
- Attentive user interfaces are the ones that rely on a person's attention and use the information gathered from the modalities to estimate the best time for communicating with the user [57]. Many applications of attentive user interfaces involve computer vision as the main component to perform a number of functions such as eye tracking, facial emotion, and gestures [58].

- Enactive interfaces are those interfaces that allow the expression and communication of enactive knowledge to actively utilize the use of hands and body for understanding a task [59].

Speech input has been extensively used in developing commercial products such as the speech-controlled environments as seen in smartphones. Gestures also inspire many researchers to develop systems based on gesture recognition for interaction in several practical applications [60]. However, using multi-modal interfaces rather than a single-user input has been preferred by people for interaction. Multi-modal input improves the handling and reliability of the system. Various studies suggest that it also improves task-completion rate compared to uni-modal systems [61]. Table 2.1 defines some common features to differentiate between a traditional interface and an MMIS [1].

Table 2.1: Differences between traditional systems and MMIS [1]

Traditional Systems	Multimodal Interface System
single input mode	multi input modes
atomistic, deterministic	continuous, probabilistic
sequential processing	parallel processing
centralized architectures	distributed and time-sensitive architectures

2.3 History of Multimodal Interface System

A typical MMIS contains a recognition system that translates human tasks into recognizable computer signals. Once the human input is identified, the next step is to interpret the input and aggregate it to achieve the desired output. There are many examples in the literature using speech and pen input in MMIS [62]. An early example was from Bolt's "Put That There" multi-modal system that combined speech and pointing gestures to move an object [5]. Fig. 2.1 shows an interface of the "Put

That There" experiment. This experiment is considered a groundbreaking demonstration of multimodal interfaces. After "Put That There" experiment, the multimodal

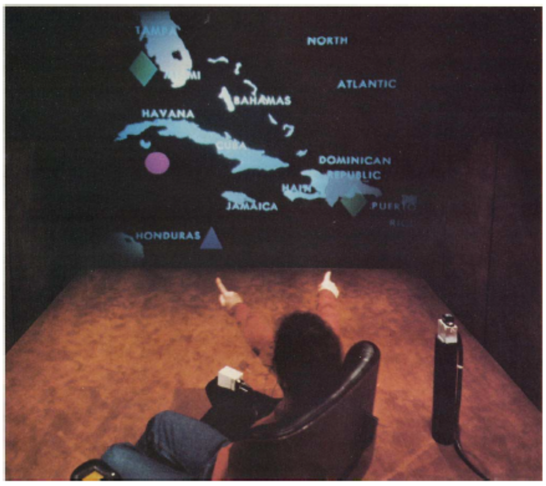


Figure 2.1: Bolt's "Put that there" system [5]

inputs, especially speech and gesture, are used in a range of applications. The early multimodal system was mainly used to perform the spatial task in a map-based environment. Neal et al. developed a multimodal system CUBRICON for tactical mission planning that uses speech, typed text, and gesture as input and displayed the output using a combination of language, maps, and graphics [63]. Koons et al. also developed a multimodal system for a map-based application that uses speech and gesture for interaction [64]. In 1997, Cohen et al. developed Quickset, a multimodal system that uses pen/voice, as a training simulator for US Marine Corps [65]. Some recent applications have also utilized gesture input and combined it with speech to draw sketches [66]. Most of these systems have used Kinect and Leap motion devices to recognize gesture inputs. Speech provides extra assistance to the user in cases such as texture mapping or rotating the object [67].

The multimodal interaction is considered as the expansion of traditional desktop experience, but a considerable amount of research also focused on alternative or "post-WIMP" computing environment which brings new modalities such as haptics and touch interfaces [3]. Post WIMP interfaces are those interfaces that are far away from traditional graphical user interfaces and rely on speech and gesture [68]. These Post WIMP interfaces include a more robust "butler-like" interface that gives life to a new generation of perceptual interfaces [56]. These perceptual user interfaces use a more natural way to interact with the computer by incorporating the natural human capabilities such as communication, cognition, motor skills, and perception into the system.

2.4 Differences Between Unimodal and Multimodal Systems

Usually, MMIS intend to deliver a natural and efficient interaction between a human and computer, as there are some advantages of using a multimodal system over a unimodal system. The literature on the assessment of MMIS state that users prefer multimodal systems over unimodal systems [69]. Multimodal systems also provide flexibility and reliability [70–72] and increase task efficiency. Information processing is better when it is presented in multiple modalities [1,73]. Other possible advantages of multimodal interaction systems explained by [29] include:

- MMIS allow flexible use of input modes, including sequential and parallel use.
- MMIS improve system efficiency, provide greater precision of spatial information, and bring robustness to the interface.
- MMIS give user alternatives in interaction, enhance the error avoidance and correction mechanism.
- MMIS can be made adaptable for a continuously changing environment and accommodate individual differences.

Multimodal interfaces were considered more efficient than unimodal interfaces, but evaluations show that multimodal interfaces improve the task completion rate by only 10% [1]. On the other hand, in the case of error handling and reliability, multimodal interfaces reduce errors by 36% compared to unimodal interfaces. Despite so many advantages, it is difficult to generalize the conclusions as for every combination of task, user, and environment; the required interface is different. Sometimes the use of multiple modalities may be ineffective or disadvantageous [74].

2.5 Modality

In different fields, the terms relevant to multimodal interaction such as modality, devices, multimedia, and multimodal have significantly different meanings [3]. In HCI terms, Modality is the form in which information is displayed or transferred, such as speech, text, visual, and gestures. Each input is transferred to a computer by a specific medium. For example, the text is entered through a keyboard and visual information through a camera. Different modalities have different definitions based on their properties and representation. Modalities are also information-dependent, which means that some type of modalities can be suitable for one type of information but not for other types [75].

2.5.1 Input and Output Modalities

An MMIS can respond to multiple input modes such as speech, gesture, and gaze. in a coordinated way to achieve a particular output. MMIS become a new focus of interest for the future computing generation since they have shifted the paradigm away from the standard keyboard mouse input. The earliest examples of MMIS were probably the ones that were least different from traditional Graphical User Interface (GUI) systems as they only reduced the use of keyboard and mouse as input modes. Since speech and gesture recognition technology has very much matured, the typical GUI systems utilize speech and gesture inputs along with standard keyboard and

Modality	Example
Visual	Face location Gaze Facial expression Lipreading Face-based identity Gesture (head/face, hands, body) Sign language
Auditory	Speech input Non-speech audio
Touch	Pressure Location and selection Gesture
Bio-sensors	Brain-computer interfaces Emotion recognition Cognitive load estimation
Other sensors	Sensor-based motion capture

Table 2.2: Sensor modalities in Multimodal interaction systems [2,3].

mouse interfaces to address user intention [30].

Blattner and Glintert [2] listed some input and output modalities and examples. The list was further updated by Turk et al. [3] given in Table 2.2. We have included some bio-sensor modalities to the list as well because these modalities will be an integral part of future multimodal systems.

A human perceives the world through their five major senses of smell, sight, touch, taste, and hearing. The pathway through which the information from the senses is transmitted or received is called a communication channel [76]. In multimodal interaction, a channel is an interaction technique through which humans transmit or receive information based on user ability and device capability [77]. For example, a keyboard is for text input, mouse for pointing or selecting input and a camera for visual input.

There are several key factors and characteristics that constitute multimodal system architecture and development. These dimensions or characteristics are [51]:

- Input modalities (size and type)

- Communication channel (devices and size)
- Processing modes (series or parallel or both)
- Vocabulary (size and type)
- Sensors and channel fusion
- Application

In a multimodal interaction system, figuring out the correct characteristics for the MMIS architecture is not simple; the designer needs to take the decisions based on intuition or by testing but to find out the best multimodal input for an interface is still an open research question [78]. Most of the work in multimodal interaction is focused on input recognition technologies such as gesture, speech, and facial expression recognition. A few studies focus on the output modality (a channel of sensory output between a human and a computer) which is also a key element of human-computer interaction [79]. The availability of powerful mobile devices (smartphones) and embedded sensors such as the microphone and 3D vision sensor (Kinect, leap-motion, 3D scanners and printers) has opened a plethora of opportunities for multimodal interaction and the HCI world.

2.5.2 Common Multimodal Inputs

With the advances in software and hardware technologies, the multi-modal interface design has emerged as a strong field of research in the last decades. People are more motivated to use natural and human-like ways to interact with computers. It also enables the researchers to integrate inputs in series or in parallel to create new sets of modalities such as speech and pen input [80–82], speech and gestures [83, 84], and speech and lip movement [85, 86]. Nowadays, speech and gestures are the most commonly used modalities in multimodal interaction.

Speech

Speech is one of the most popular inputs for an MMIS. Speech is a tool for communication, comprised of vocabularies. Speech is normally divided into grammar, syntax, semantics, discourse, pragmatics, and prosody. These parts help us understand the language.

- **Grammar** is used to make rules and laws, and the right methods to apply these for speaking and writing.
- **Syntax** links names and actions together in a defined order.
- **Semantics** involves the study of the meanings of individual words and syntactic contexts.
- **discourse** is written and spoken communication that constitutes a sequence of relations to the subject, object and announcements.
- **Pragmatics** is defined as how language is used and investigates the semantic and syntactic uses of language.
- **prosody** is the emotion state-of-utterance [87].

Speech Recognition The process of converting speech into text is known as speech recognition. There are three main types of speech recognition algorithms present in the literature: Hidden Markov models (HMM) [88], Dynamic time warping (DTW) [89], and Neural Networks [90]. In 1994, Microsoft (MS) developed an API for speech recognition and synthesis. The function of this API is to convert speech into text and text into speech in real time. This speech recognition system is available in many languages such as English, French, Spanish, German, Japanese, and Chinese. [91]. The Microsoft Speech API has an approximate recognition rate of 75% [92]. There are some other speech recognition systems available such as Carnegie Mellon University

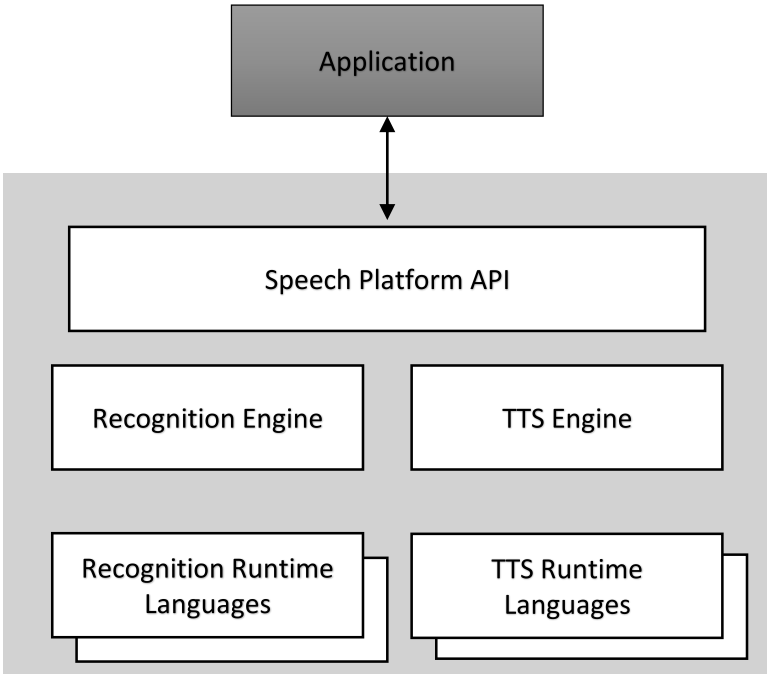


Figure 2.2: Microsoft Speech Recognition Platform

(CMU) Sphinx-4 based on HMM, and Google API based on deep neural networks which are the possible alternatives to MS speech recognition API [93]. In this thesis, we have used MS speech recognition API because it is easy to use and required a minimum training time. Fig. 2.2 shows different modules of Microsoft speech recognition API. The Recognition engine contains the algorithm for recognizing speech to convert into text and the TTS (Text to Speech) Engine converts text input into speech.

Gesture

A gesture is the movement of a body part to express an idea or meaning to someone. Usually, hand and head movements are considered as gestures. There are many gesture-controlled user interfaces available in the literature [94]. The term gesture first appeared in 1979 in a book named "GESTURES, their Origins and Distribution" by Morris et al. [95]. In this book, an analysis of emblem gestures, the action that replaces speech, was given. Rime and Schiaratura explained a gesture taxonomy for

(a) Symbolic gesture "OK"

(b) Deictic gesture "Pointing"

(c) Iconic gesture "Book"

(d) Pantomimic gesture "opening a jar"

Figure 2.3: Commonly used gesture examples

communication with a computer [96]. These gestures are symbolic, deictic, iconic and pantomimic gestures.

- **Symbolic gestures** are those gestures that have a single meaning. Fig. 2.3a shows the symbolic gesture of "OK".
- **Deictic gestures** are the pointing gestures as shown in Fig. 2.3b. These gestures are used to tell the other person about a specific object or event in the surrounding environment.
- **Iconic gestures** are used to give information about size, shape, or orientation of an object. For example, a person doing gesture to describe a rectangular object is shown in Fig. 2.3c.
- **Pantomimic gestures** are used to show the movement of some invisible tool or object. For example, turn a knob as shown in Fig. 2.3d.

McNeill in 1992 added two more gestures to these taxonomies that relate to the process of communication [97]. These gestures are beat and cohesive gestures:

- **Beat gestures** are used to indicate a pace, represent the up and down movement with the rhythm of speech.

- **Cohesive gestures** are used to combine temporally separated but thematically related portions of discourse and these are the variation of iconic, pantomimic or deictic gestures.

Cadoz in 1994 associated three types of functions to the group of gestures; semiotic gestures, ergotic gestures, and epistemic gestures [97].

- **Semiotic gestures** "are used to communicate meaningful information."

- **Ergotic gestures** "are used to manipulate the physical world and create artefacts."

- **Epistemic gestures** "are used to learn the environment through exploration."

Later on in 2005, Karam et al. [98] provide a comprehensive taxonomy of gestures in HCI. They categorizes the gestures in deictic, semaphores, gesticulation, manipulation and sign language gestures. In 2009, Wobbrock et al. presented a taxonomy for surface gestures based on the evaluation of how many users performed a particular gesture for a given task [99]. Ruiz et al. in 2011 proposed a taxonomy for 3D motion gesture and applied it on smartphones [100]. They used the physical characteristic such as kinematic impulse, dimensionality and complexity to create a user-defined gesture set.

Gesture recognition Through the years, gesture recognition has been through very rapid advancement. Gesture-Recognition technology can be divided into three categories. The first category is ***glove-based***, which is further divided into active and passive data gloves. In active data gloves, sensors are embedded on the gloves to measure joint flexing and acceleration. On the other hand, passive gloves use an external device, such as a camera, to define the position and orientation of the hand using some markers and colors for finger detection [101].

The second category is known as ***haptics***. This technology utilizes a tactile experience in the air using haptic technology. The user is not required to wear any device or system. For example, AIREAL extracts tactile sensations in the air and relies on air vortex generation directed by an actuated flexible nozzle to provide effective tactile feedback with a 75-degree field of view, and an 8.5 cm resolution at 1 meter [102].

The third category and most commonly used technology for detecting gestures is ***sensor-based*** motion capture. This method utilizes raw data from a sensor and manipulates the data to generate position and gesture types. The most popular sensors for this purpose are Microsoft Kinect and Leap motion [103]. In this research, we have used a Leap Motion sensor for gesture recognition.

The Leap motion controller is a tiny hardware device that detects hands and fingers. The controller uses two monochromatic IR cameras (IR emitters are inside the device). Leap motion offers a close range. It can detect the fingers even when they are 5 cm away. The advantage of Leap motion is its high-level API. It is very easy to get the position or velocity of the hand, the fingers, and some already defined gestures, such as swipe, circle, point or type. Leap motion offers more facilities for drawing and interaction than Kinect [104]. Fig. 2.4 shows a Leap sensor with its coordinate system.

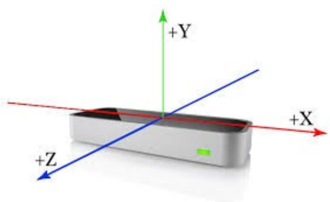


Figure 2.4: Leap motion sensor (http://developer.leapmotion.com/)

2.6 Multimodal Interface System Development

Development of the multimodal system is challenging because standard computing environments usually do not translate into a multimodal environment effectively [3]. In addition to that, each factor or characteristic of the multimodal system may result in different design strategies. Oviatt in 1999 [74] proposed a set of myths about the multimodal systems; "Ten Myths of Multimodal Interaction" have proven to be useful in the MMIS field. These ten myths, as mentioned in [74], are given below:

Myth 1: "If you build a multimodal system, users will interact multimodally."

Myth 2: "Speech and pointing is the dominant multimodal integration pattern."

Myth 3: "Multimodal input involves simultaneous signals."

Myth 4: "Speech is the primary input mode in any multimodal system that includes it."

Myth 5: "Multimodal language does not differ linguistically from a unimodal language."

Myth 6: "Multimodal integration involves redundancy of content between modes."

Myth 7: "Individual error-prone recognition technologies combine multimodally to produce even greater unreliability."

Myth 8: "All users' multimodal commands are integrated in a uniform way."

Myth 9: "Different input modes can transmit comparable content."

Myth 10: "Enhanced efficiency is the main advantage of multimodal systems."

In 2004, Reeves et al. [105] gave some guidelines for multimodal interface design:

- **User Specifications:** The MMIS should be designed for the broadest range of users and contexts of use. They should also address privacy and security issues.

- **Input and Output Specifications:** The design should maximize the human cognitive and physical abilities by considering the information processing abilities and limitations. The system modalities should be harmonious with the user preference, context, and system functionality.

- **Adaptivity:** The multimodal system should adapt to the needs and abilities of users.

- **Consistency:** The system should be consistent in presentation and prompt.

- **Feedback:** The users should be aware of their current connectivity and available modalities.

- **Error Prevention/Handling:** The system should provide error detection, prevention, and handling functionalities.

2.7 Modelling, Fusion, and Data Collection

The principles and techniques used to develop a GUI-based interaction could not be used in MMIS development. This makes the multimodal interface design special [105]. As mentioned in the previous section, special attention should be given to input and output modalities, adaptability, consistency, and error handling issues. The human-related factors such as personality, background, current emotional state need to be considered when designing a multimodal interface [106, 107]. These issues and design decisions further dictate the underlying algorithms and techniques used in interface development.

2.7.1 System Integration Architectures

In the multimodal community, the multi-agent architectures such as Open Agent Architecture [108] and Adaptive Agent Architecture [109] are commonly used. Multi-agent architectures use a distributed approach to implement the complex modules of multimodal processing. The modules and components in a multi-agent architecture can be developed in various programming languages, machines, and operating systems. Inputs in a multi-agent architecture can be in parallel or asynchronous and then be passed to the recognition system. The results from the recognition modules are interpreted and delivered to the user through multimedia feedback [110].

Open Agent Architecture (OAA)

OAA is a multi-agent system architecture that combines the functionalities of those agents that were not designed to work together, thereby facilitating the wider reuse of the expertise embodied by an agent [108]. The key attributes of this architecture are:

- **Open:** The OAA supports agents written in multiple languages and on multiple platforms.
- **Extensible:** Agents can be added to or removed from the system at run-time.
- **Mobile:** OAA-based applications can run from a low-end portable computer.
- **Collaborative:** The user appears to be just another agent to the automated agents. This simplifies creating systems where multiple users and agents collaborate.
- **Multiple Modalities:** The user interface supports multiple modalities in addition to the traditional GUI.
- **Multimodal Interaction:** Users can interact with the system using multiple modalities.

Adaptive Agent Architecture (AAA)

AAA is a robust brokered (or middle agents) architecture that "uses teamwork to recover a multi-agent system broker failure and to maintain a specified minimum number of functional brokers in the system even when some of the brokers become inaccessible [111]." To achieve the functionality, AAA has two mission statements:

- **Mission Statement 1:** "Whenever an agent registers with the broker team, the brokers have a team intention of connecting with that agent, if it ever disconnects, as long as it remains registered with the team."
- **Mission Statement 2:** "The AAA broker team has a team maintenance goal of having at least N brokers in the team at all times where N is specified during the team formation."

In the thesis, we have implemented OAA-based MMIS.

2.7.2 User Modelling and Human Information Processing

Modeling human information processing in HCI and related fields is a challenging job, and there are several studies focusing on this which have received significant attention [107, 112, 113]. In this section, we have mentioned some commonly used models. The most famous model in the HCI literature is the Model Human Processor [114]. It has three parts: the perceptual system, the cognitive system, and the motor system. The perceptual system is responsible for handling the sensory information from the outside world, i.e., the input-output components. The motor system controls all the actions, also known as a processing component. The cognitive system, on the other hand, provides the functionality to connect the other two systems [10]. According to the model human processor, the input-output channels (movement, hearing, vision), memory (both long and short term) and the processing (problem-solving, reasoning) need to be considered when designing an MMIS [115].

GOMS (Goals, Operators, Methods, and Selection rules) is another famous model proposed by Card et al. [114]. There is now a family of many variations of GOMS. The model is suitable for modeling the optimal behavior of the users. Despite so many variations of GOMS, the fundamental concept is the same. **Goals** are the result that a user wants to achieve. An **operator** is an action performed in the service of a goal. A **method** is a sequence of operators to achieve a goal and if more than one method, then a **selection rule** will be used to shortlist one method [116]. All GOMS variations provide useful information, but they also have some drawbacks. For example, these techniques do not incorporate the user fatigue into the model, and the user performance degrades over time because the user performs the same task repetitively. GOMS techniques are less rigid to necessary cognitive actions and most of the time applicable to expert users. There are some other models as well that includes the cognitive architecture models [117] (ACT-R/PM [118], LICAI/CoLiDeS [119]), the psychological models such as human action cycle [120, 121], grammar-based models [122] and application specific models [123].

- **Cognitive Architecture Model** is a theoretical entity based on human cognition using a wide selection of experimental data [117].
- **Psychological Model** uses human psychology to describe different steps to interact with the computer [121].

2.7.3 Adaptability

Adaptability is a vital part of the current and future HCI systems because of the large, diverse types of users. The HCI system should accommodate for the difference in expertise, culture, language, and goals by introducing adaptive and customizable interfaces. The system needs to learn through user behavior and knowledge to predict the user's future behavior and response [124]. With the inclusion of adaptiveness in the system, it will improve the overall user performance and make the interaction exciting [125].

The feedback element in adaptive HCI promises to increase the user ability to handle complex tasks more quickly with greater accuracy and allows the user to learn new techniques quickly. The adaptive HCI interfaces provide an interactive knowledge or agent-based dialog to handle errors, interruptions, understand the current context and situation [126]. The long-term aim of intelligent user interaction is to increase the effectiveness, naturalness, and productivity [127]. The challenging part in developing an adaptive user interface is to find the variables on which the system will adapt. The system needs to analyze input, user behavior, and actions and find those variables that maximize the efficiency, effectiveness of the interface [128].

Adaptive systems need continuous feedback from the user to learn, and for that reason, these systems are also called learning systems. As these systems involve a continues learning cycle, the developer should not consider the adaptive systems a solution for all the problems and should focus on whether the user needs an adaptive system. Some researchers state that adaptive interfaces violate the standard usability principle, but sometimes a static user interface, that does not depend on user state, shows superior performance over an adaptive interface [129, 130]. Despite these studies, adaptive systems can bring undeniable benefits to the interaction system. Nowadays, many researchers work on developing intelligent systems from many different applications including learning and tutoring systems [131–133], medical applications [134], smartphones [135–137] and arts [138].

Although there are many advances in the adaptive interface systems, still there are various research problems that need answering such as usability of input modes, finding variables to adapt, and estimating user behavior. In this thesis, we have aimed at defining ways to create adaptive systems that can adapt to the user's cognitive activity and update the system's response accordingly.

2.7.4 Multimodal Integration

In daily communication, all our input modes are coordinated perfectly, e.g., speech, lip movement, hand gesture, and facial expressions. This is not the case with computers; the inputs in MMIS are not designed to coordinate at all with each other. Unlike humans, computers with an MMIS produce a challenge for integrating complementary modalities to generate a highly cooperative combination. Many fusion techniques and architectures were developed to integrate multi-modal inputs for joint processing [4].

In older multimodal systems, the data collected through various modalities are processed separately and fuse in the end. Modalities with different characteristics such as speech and gesture cannot be connected straight away. Therefore, to integrate inputs with different characteristics, three different levels of integration, also known as fusion levels, are proposed. These levels are signal level, features level, and semantic level [139, 140]. Table 2.3 summarizes these three fusion levels.

Table 2.3: Summary of three fusion levels [4]

Fusion level	**Signal-level fusion**	**Feature-level fusion**	**Semantic-level fusion**
Input type	Raw signal of same type	Closely synchronized	Loosely coupled
Level of information	Highest level of information detail	Moderate level of information detail	Mutual disambiguation by combining data from modes
Noise/failure sensitivity	Highly susceptible to noise or failures	Less sensitive to noise or failures	Highly resistant to noise or failures
Usage	Rarely used for combining multi-modalities	Used for fusion of particular modes	Most widely used type of fusion

Signal level fusion

In signal level data fusion, the raw data from different modalities are combined. This level of fusion is more appropriate for highly synchronized input modalities, such as speech and lip movements, otherwise the performance of the system is affected. The signals are combined in a vector form at a very early stage. The dimensions of the vector are reduced by applying dimensionality reduction techniques (e.g. LDA, PCA) and then the vector is forwarded to the recognition engine for the classification [4].

Feature level fusion

The problem with the signal level fusion is that it fails to model the fluctuations and the asynchrony between the input modalities. One way to solve this issue is to extract features from the signals and use the feature to combine the modalities. Features based on, for example, transformations, probabilistic models, and information measures can be used for fusion in this level [107].

Semantic fusion level

The semantic level fusion of input modalities uses common meaning representation extracted from different modalities into a combined interpretation. Instead of working on raw signals or features, the semantic fusion works on semantic information extracted from individual modalities and combines the interpretation from each mode. This fusion level has the capability to control loosely-coupled input modalities such as speech and gestures [30]. Semantic level fusion has advantages over other fusion techniques such as the multiple inputs do not have to coincide because each modality can be recognized independently [3].

Independent classifiers have been used in semantic fusion level, and the final decision is made by combining the partial outputs of unimodal classifiers. The semantic level fusion provides the advantage of reducing computational complexity by training classifiers separately ($O(2N)$) instead of combined training ($O(N^2)$) which decreases the computational complexity [141]. The semantic fusion approach has been widely used for the coupling asynchronous input modes such as speech and pen input or speech and gestures. It is easy to scale up the input modes and vocabulary size in semantic level fusion because of the late integration of speech and gesture modalities. In this thesis, we have used late integration of modalities for the development of MMIS.

2.7.5 Frameworks for Input Information Fusion

To fuse the multimodal inputs, a framework is needed, and it is the most important requirement in designing a multimodal system. The framework requires a time stamp of individual inputs. In the case of speech and gesture inputs, these can be used in a series or in parallel. A timestamp is an important flag to check the starting and ending of a multimodal input [29]. Fig. 2.5 shows a typical framework

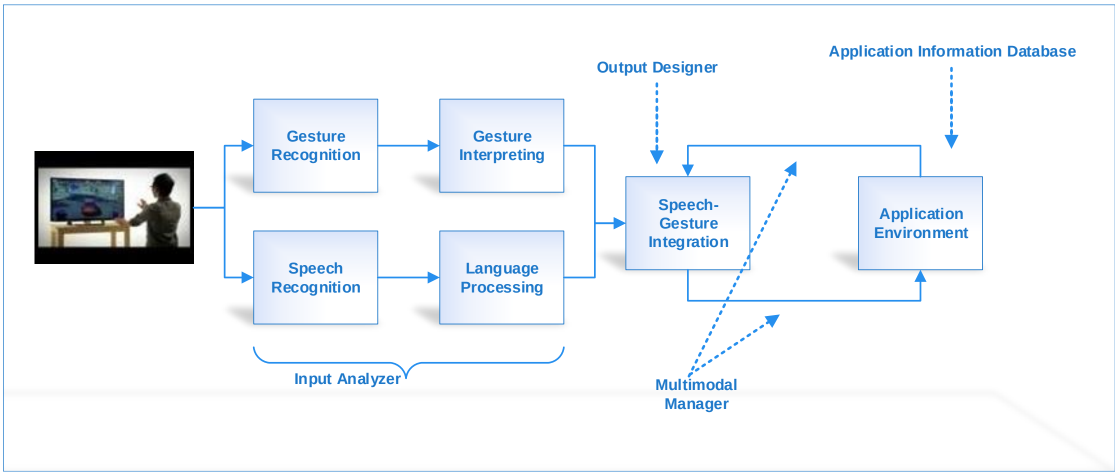


Figure 2.5: Framework for semantic level speech and gestures integration

for speech and hand gesture input at the semantic level. The framework has four main parts: Input Analyzer, Multimodal Manager, Output Designer, and Application Information Database [142]. The input analyzer processes speech and gesture in parallel, and the results are represented in semantics that is given to the output designer block. The task of the multimodal manager is to exchange information between the output designer and application information database for real-time control. The problem with these kinds of architecture is the use of a homogeneous programming language which can be difficult to use in some cases. To overcome this problem, a multi-agent architecture is proposed [4] which uses multiple agents for pooling the pre-semantic information from different sensors and integrates them into a multimodal data structure.

2.7.6 Data Collection and Testing

In an MMIS, one of the most challenging tasks is to obtain the ground truth by collecting and labeling data as these are error-prone and take a lot of effort and time. Usually, in HCI, we rely on self-reported data from the user. For example, in an emotion recognition application, the data used is mostly based on simulated data by actors. Actors imitate certain emotions or perform conversations to generate emotional reactions [143]. Asking someone to generate emotions while watching or listening to something does not create the same authentic feeling. Real emotions, on the other hand, are hard to control and record in a laboratory setting.

Collecting data is a challenging task because of the wide variability of possible inputs. Usually, the researcher uses a small set of data that is available for training and adds unlabeled data into it to make the classifier more robust. Adding of unlabeled data for training needs a lot more care. Probabilistic models are used to deal with this scenario [144]. Recently, deep learning algorithms have been widely used for multimodal input classification [145, 146].

2.7.7 Evaluation Measures

The main issue in multimodal interface design research is the evaluation. Various measures such as efficiency, quality, user satisfaction, and accuracy are available in the literature to evaluate an interface. Usually, computational measures, such as efficiency, are used because the interface will improve the interaction, task completion, and decrease the complexity of the system. Task completion rate with respect to time is one way to measure efficiency. Another measure is the number of actions the user has performed to decide or solve a problem [34].

Another way to measure the effectiveness of an interface is to ask the subjective opinion of the user through questionnaires. The questionnaires can ask the user about user satisfaction, fatigue level, system compatibility, and easiness to use the system [34]. The biofeedback signals such as EEG and ECG can also be used to determine the system effectiveness by measuring the user's emotional level or cognitive state. In this research, we have presented a method to analyze the interface using EEG signals by measuring the cognitive load and activity of the user.

2.8 Current Applications of MMIS

In the previous sections, we have discussed various types of modalities, fusion models, data collection for training and testing, and evaluation of the interface. This section will discuss some of the recent applications of HCI in a wide variety of scenarios, including ambient spaces, mobile and wearable technology, virtual environments, and arts.

Multimodal systems have been used in video indexing for detecting human behavior and expressions [147, 148]. Virtual meeting rooms utilize multimodal systems to record and model user behavior in real-time [149]. Behavioral analysis from multimodal input is utilized in surveillance and intelligence applications [150, 151]. Bradbury et al. [152] proposed an attentive kitchen concept named eyeCOOK, which combines eye-gaze and speech commands to help non-expert users cook a meal. Multimodal systems have been applied to various healthcare applications. Gestonurse is a multimodal robotic scrub nurse that assist the chief surgeon by passing surgical equipment using speech and gesture modalities [6]. Fig. 2.6 shows the gastonurse in a real environment. Nie et al. [153] present a health prediction system based

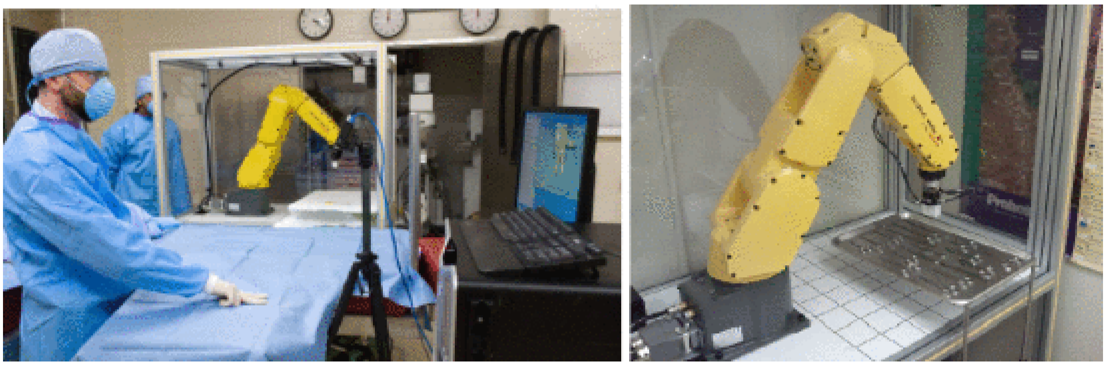


Figure 2.6: A prototype of the interface tested at an OR along with FANUC LR Mate 200iC and Mayo standing with instruments [6]

on multimodal observation. Smart home concepts also utilize multimodal inputs to record various activities of the user [154, 155].

Lv et al. [156] developed an MMIS that utilizes the hand and foot input for interaction with handheld devices. Input modalities are hand and foot, which are processed to generate a coordinated output. Another team of researchers at the University of Michigan developed an MMIS for image editing known as PixelTone [7]. PixelTone uses speech and direct manipulation to edit images, utilizes natural language to express desired changes and sketches for region identification for the corresponding changes. Fig. 2.7 shows a three step procedure to change the color of a shirt in PixelTone. Cohen et al. [144] designed an MMIS titled Sketch-Thru-Plan (STP),

Figure 2.7: Changing the color of a shirt with PixelTone a) select the person's shirt and say "This is a shirt." b) Tell to "Change the color of the shirt," and c) PixelTone offers a slider to change the colors [7]

which is a multimodal interface for command and control. The system recognizes military jargon. The users of STP can give labels, reposition them, and draw symbols on a digital map. Nam and his team [157] developed a human-machine interface for controlling a humanoid robot called GOM-Face. The inputs to the interface are three electric potentials recorded from face: glossokinetic potential (GKP) which is electric potential responses generated by tongue movement, electrooculogram (EOG) which is the potential between the front and the back of the human eye, and electromyogram which is electrical activity of muscles, to help persons with limb motor disabilities.

Smartphones are the most prominent example of multimodal interface systems. Today's smartphones have the capability to interact with speech, touch, gesture, gaze, and facial inputs [158]. Virtual and augmented reality apps open the ways to unlimited possibilities of applications, and with the multimodal inputs, these applications provide a very natural feeling of interaction [159, 160]. A real-time strategy game in which the agents can be instructed with speech and gesture commands is described in [161]. In another application, an augmented reality dialog interface has been present that enables the user to control a robot's verbal and non-verbal behavior accurately [162]. A method of designing an intelligent interface for people with functional disabilities has been presented, which can combine visual, sound, and tactile multimodal inputs in a brain-computer interface to provide highly adaptable and personal services [163].

With the advancement in multimodal interaction and visual technology, the arts industry has produced some of the most exciting applications. Multimodal systems to play music [164–166], generate avatars [167, 168], interacting with objects in museums [169, 170] are some of the applications in the art industry. MMIS has been used to improve the lifestyle of people with disabilities. Speech-activated smart wheelchairs [171, 172], brain-computer interfaces [173, 174] and interfaces that use eye blinks or eyebrow movements [175] are some examples that provide accessibility to disabled people.

MMISs are used in every aspect of computing that involves interaction with humans, objects, and the environment. Although MMISs may not totally replace the traditional interaction systems, they have promising applications that can revolutionize the entertainment, games, art, and health-care industries. In this thesis, we have presented an MMIS for industrial product design that has the capability to model 3D objects using speech and gesture inputs.

2.9 Challenges in Multimodal HCI

Computing field has seen some significant developments in the area of Human-computer interaction in recent decades. The touch-based interfaces have become a vital part in smartphones these days, and with multiple visual sensors on the device, researchers are continuously looking for new ways to interact with the device. Despite the advancements in technology, there are challenges and research problems that need addressing. Each modality itself is an active field of research such as gesture recognition, speech recognition, natural language understanding, activity recognition, haptics, user modeling, and context understanding. Deep learning algorithms have provided substantial support to the recognition algorithms, but much more research is needed in improving performance, personalization, integration, and adaptability.

In addition to the challenges mentioned above, we need to understand the user-dependent issues that affect the performance of the multimodal interfaces. The study of the cognitive load of the user when interacting with a multimodal system and the relationship of cognitive load with multiple modalities need further investigation [176]. In this thesis, we evaluate the MMIS using the traditional questionnaire-based analysis and present a new method to estimate the user's cognitive load and activity using EEG signals.

2.10 Conclusion

In this chapter, the research literature related to the multimodal interface systems has been presented. We presented an overview of multimodal interaction and the history of some of the earliest MMIS. The MMIS provide the advantages of robustness, adaptability, improvement in task completion rate over a unimodal system. The evaluations show that multimodal interfaces improve the task completion rate by only 10% [1], but in the case of error handling and reliability, multimodal interfaces reduce errors by 36% compared to unimodal interfaces.

Speech and gestures are the most widely used input modalities, along with touch and pen-based inputs. Most of the work in multimodal interaction is focused on input recognition technologies such as gesture, speech, and facial expression recognition. A few studies focus on the output modality, channel of sensory output between a human and a computer, which is also a key element of human-computer interaction [79]. The most popular sensors for gesture recognition purpose are Microsoft Kinect and Leap motion [103]. Biofeedback devices are catching the researcher's interest because these devices provide an indirect representation of the user's emotional and cognitive state. The signals such as EEG and ECG can determine the system effectiveness by measuring the user emotional level or cognitive state.

Multimodal inputs are integrated at the signal, feature, and semantic levels. The recent applications have used semantic level integration because it is a late integration process which gives the advantage to update the modalities and vocabulary quickly. Data collection and testing are one of the essential parts, and they require more attention because most of the time, data is collected in a controlled environment. The next important part is the evaluation of the interface, which can be qualitative and quantitative. In the qualitative evaluation, questionnaires are widely used. For quantitative analysis, the measures such as task completion rate, the average time taken to complete a task are used.

To our best knowledge, the literature on quantitative evaluation of an MMIS is limited, especially in 3D modeling applications. We have presented an MMIS for the industrial product design that has the capability to model 3D objects using speech and gesture inputs. The effects of using speech and gesture in modeling a 3D object are the focus of this thesis. In this research, we have used a Leap Motion sensor for gesture recognition. We have presented a method to analyze the interface using EEG signals by measuring the cognitive load and activity of the user. We evaluate the MMIS using the traditional questionnaire-based analysis and present a new method to estimate user cognitive load and activity of the user using EEG signals.

Chapter 3

Human Cognition, Psycho-physiological Analysis & Functional Brain Networks

This chapter discusses human cognition along with a detailed description of processing, memory, attention, and decision making. It further explains the models of cognitive processing and measures used in cognitive processing. In the last section of this chapter, design expertise and novice/ expert differences are discussed. In this chapter, we also review psycho-physiological signal analysis, including a detailed description of emotional and cognitive activities. The chapter introduces measurement methods such as ECG, EEG, and GSR and the advantages of EEG over other neuroimaging tools for cognition research. It further reviews the relevant statistical measures to calculate connectivity between psycho-physiological signals and brain regions and provides a detailed description of the application of graph theory and complex network metrics for the analysis of FBNs. The focus of this review chapter is to find the answers to the following research questions:

RQ 2.1 Can we use psycho-physiological analysis in an HCI system?

RQ 2.2 Which EEG parameters can be used for evaluating the cognitive activity?

The review has been published as a paper, titled as "A Survey on Psycho-Physiological Analysis & Measurement Methods in Multimodal Systems" in Multimodal Technologies and Interaction Journal [11].

- Baig, Muhammad Zeeshan, and Manolya Kavakli. "*A Survey on Psycho-Physiological Analysis & Measurement Methods in Multimodal Systems*." Multimodal Technologies and Interaction 3.2 (2019): 37.

3.1 Cognition

Cognition is a combination of many complex processes which help a human in acquiring knowledge and understanding [36]. Every cognition process contains different levels which can be conscious or non-conscious. For example, in reading, the visual perception comes into action which follows a bottom-up approach at a non-conscious level such as orientation from the vertical alignment of each of the letter parts, followed by pattern matching to make words from letters and then to make sentences. On a higher level of non-conscious processing, these letter shapes are matched with the sound of the speech along with other linguistic processes that run in parallel with memory processes such as syntax and semantics. After all these processes, the meaning is assigned and becomes available to consciousness and further manipulation [177].

In cognition, the processes can be divided into conscious and non-conscious level processes. The conscious processes are those on which humans have a certain degree of freedom or control. The processes that occur naturally and automatically are called non-conscious processes [36]. Some processes are purely non-conscious such as sound perception because it is impossible not to hear the surrounding sound. Actions such as decision making are found to be purely conscious actions with some exceptions [178]. A general understanding is that the non-conscious actions that are performed by humans most of the time can be examined through animal models [179]. On the other hand, the conscious actions are hard to examine because these involve higher-order human cognition and direct experimentation of the human brain, which is prohibited due to ethical considerations. Thus, other non-invasive methods are used to study the basic cognitive processes.

3.1.1 Memory

Memory, in its literal term, means the capacity to retain and retrieve information about the past. Memory is formally defined by Reber [180] as "the mental function of retaining information about stimuli, events, images, ideas, etc. after the original stimuli are no longer present...the hypothesized 'storage system' in the mind/brain that holds this information...the information so retained". There is a general understanding that memory is a combination of functional units or independent sub-types. Some of the sub-types are described below:

Short-term memory (STM)

Short-term memory (STM) stores the recent events. For example, in a stimulus remembering experiment where participants must see and memorize a series of stimuli (digits, pictures, letters, etc.) and after a short amount of time, they need to recall as many stimuli as they can [181], the memory that is in use is STM. The average capacity of STM is seven stimuli, and a normal human can remember 5 to 7 stimuli [182]. The STM could hold information for up to 5 seconds unless rehearsed. It completely degrades in 20 seconds [183].

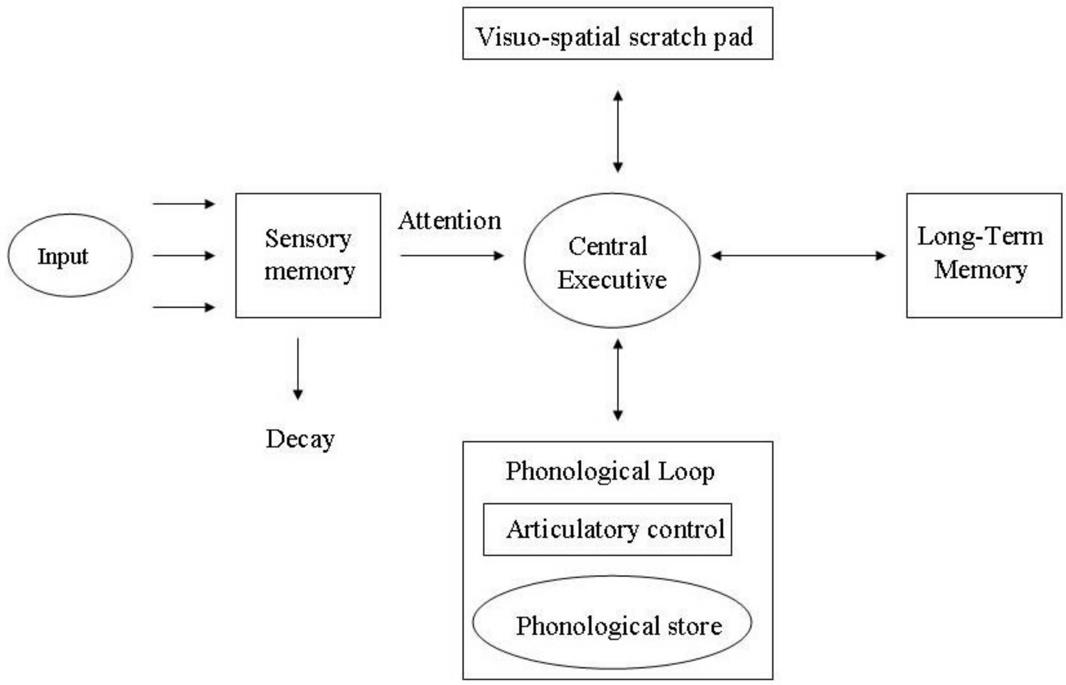


Figure 3.1: The Working Memory Model Components [8]

Working memory (WM)

Working memory (WM) is a temporary storage area, just like STM, which act as a holding station for further processing, and the information may or may not be transferred to long term memory. A model related to working memory was proposed by Baddeley and Hitch [8] that divide the working memory into three sub-components, also known as the tripartite working memory model:

1 the visuospatial sketchpad (VSS) for storing temporary visual information

2 the phonological loop (PL) for the temporary storage of auditory information

3 the central executive (CE) to direct attention towards the information that will be processed further by working memory.

Fig. 3.1 shows the working model presented by Baddeley and Hitch in 1974 [8]. There has been several modifications and refining of working memory models that help to understand the cognitive phenomena better [184].

Long term memory (LTM)

When you recall an event from the past, you are using long-term memory (LTM). Although LTM ties together with STM and WM processes, it is considered as a separate process involving different brain areas and cellular processes [185]. LTM is constructed through the process of long-term potentiation. In the process of potentiation, a small neurons network encodes a memory event. The connection between neurons become increasingly stronger through the unconscious process of consolidation, which means more neurons firing together.

There are some other subcomponents of memory defined in the literature. One of these components is episodic memory, which is the memory of a specific event or episode. For example, if a person has attended a party and some unusual event happened at that party, that unusual event is an example of episodic memory. According to researchers, the episodic memory is not encoded in totality, it is tagged with a salient signpost and reconstructed on the run with these signposts [181]. The problem with this memory is that it has a deconstructive/re-creative nature and is susceptible to both failure and external modulation of thought suggestions. Another component of memory is procedural/ motor memory, which is for motor or procedural actions such as driving a car and riding a bike. These actions are hard to master at first but eventually become over-learned skills, which then can be entirely at non-conscious level [186].

3.1.2 Attention and Decision Making

Attention typically means the focus of neural resources to a specific feature in a scene that becomes the center of cognitive processing. In contrast, distraction occurs when attention loses focus due to internal forces (such as fatigue or boredom) or due to non-task relevant information. Attention is considered a higher order cognitive action but most of the time it is affected by low order processes. Decision making is also a higher order cognitive action that requires both memory and attention for successfully making a choice, forming a conclusion, or reaching a conclusion. Executive processes integrate and organize all low-stream information into a cohesive mental model capable of conscious manipulation [36]. These processes are not purely conscious processes and often operate at an unconscious level with functional manifestations [181]. These processes can influence judgment and decision making [187].

3.1.3 Theories of Human Cognition

All modalities such as speech, gesture, and facial expressions are sources of acquiring and producing information for humans. The human brain has the capability to process sensory data in parallel that comes from the central nervous system, other parts of the brain, muscles, and glands. The brain has a highly dedicated interconnected structure for integration and diffusion of multi-modal sensory data. The argument to support this statement is that the neural processing in language comprehension does not come from verbal semantics only, but also from a more general domain of cognitive processes. The same integration results have appeared for speech and gesture interconnection [188].

Dick et al. [189] studied the influence of gesture activity on semantic information processing by examining functional Magnetic Resonance Imaging (fMRI) of associated brain regions. They found a distributed pattern of neural activity in discourse comprehension as well as in motor perception. These distributed patterns are generated when a user perceives hand movements while someone is speaking. This result confirms that the multi-modal perception and cognitive structure in the brain can process multi-modal information collaboratively.

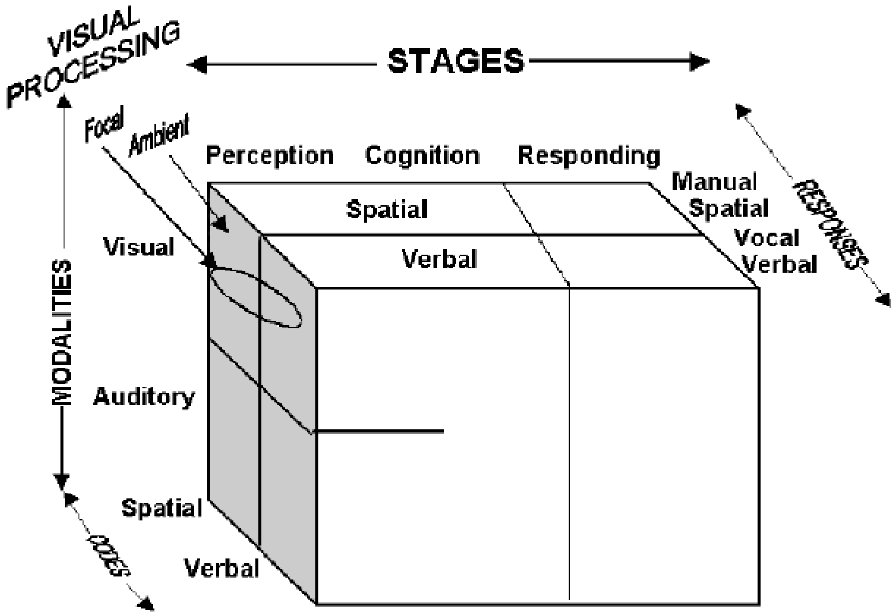


Figure 3.2: The 4-D multiple resource model [9]

Tripartite Working Memory Model

A Tripartite working memory model for MMIS has been proposed by Baddeley et al. [190]. The model has an independent processor called the central executive which works together with three other slave systems (the phonological loop, the visual/spatial sketchpad, and the episodic buffer) in a coordinated and synchronous manner. Baddeley's working memory model gives a general understanding of how multi-modal information is processed in human cognition. It also suggests that multiple resources can be used for one production. For example, gesture input is produced by visual/spatial components, whereas speech is generated by phonological-loop components. The drawback of Baddeley's model is that, for integrated cognition processes, it is hard to explain, as both visual/spatial and auditory verbal processes are involved in a multi-modal production, such as speech and gesture, to explain an object.

Multiple Resource Theory

It is comparatively easy to explain multi-modal tasks with Multiple Resource Theory (MRT) [9]. The MRT works on the principle that there are multiple limited resources available for real-time information processing. In MRT, when a task described by modally-organized central finishes, then the system is taxed. The central resources can be utilized by multiple tasks at the same moment, but they have a limited capacity, and tasks can interfere with each other when using the same resources. To overcome performance degradation when multiple tasks used the same resources, a four-dimensional MRT model has been presented by Wickens and is shown in Fig. 3.2. This model analyzes the resources necessary for completing multiple tasks. If a cognitive process with multiple tasks requires the same the resource usage, then the interference is predicted in advance and there is no interference in a single task because of the sequential processing within a single task [9]. Although MRT provides a useful insight into interference prediction, there is less information regarding the shared resources.

Human Process Model

Another model that has been built upon the Wicken's model is known as the human processor model (shown in Fig. 3.3) and incorporates cognitive, perceptual, and motor processors along with other processors [10]. It basically divides the information processing into three subsystems (cognitive, perceptual, and motor processors) and each subsystem has its own functionality. The main reason for using the human-processor model is that it can calculate the cycle and decay time for each sub-processor. This estimated value allows the designer to estimate a performance with respect to the time taken by a user in performing a task.

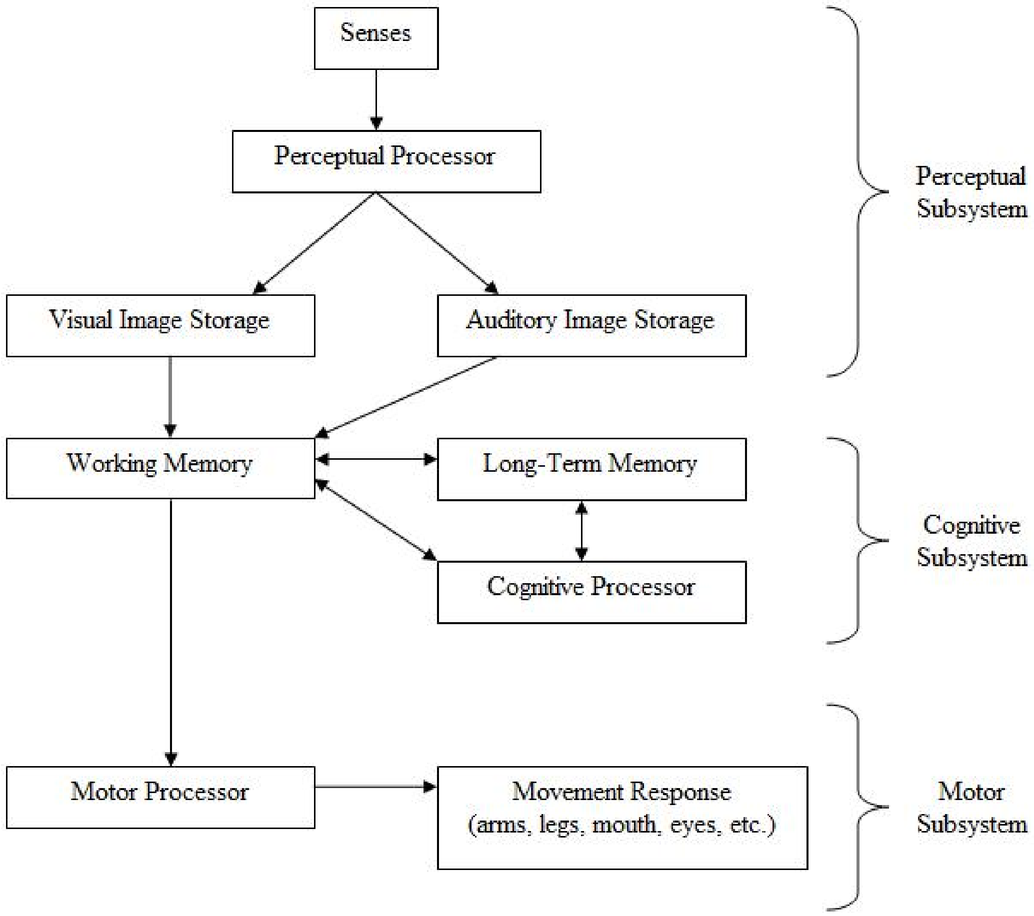


Figure 3.3: Model of the Human Processor [10]

Cognitive Load Theory

Another well-known theory that has been used in the literature to explain multi-modal processing for task performance improvement is cognitive load theory, which tries to decode the mental effort along with the assumption that the brain's working memory has a limited capacity [191]. Multiple modalities not only provide performance superiority to users compared to a single modality, but users also prefer to use multiple modalities, if present. Cognitive load theory interprets multi-modal processing to a set of modality-specific working memory resources. It has also been seen that the user adjusts modalities in a complicated task. For example, the user prefers to interact in multiple modalities when a difficult task is presented, and it has been observed that this change of modality allows the users to handle their cognitive load [192].

A study by Oviatt et al. [69] suggests that an MMIS design that minimizes the cognitive load can free mental resources and improve student performance in educational applications. Morsella and Krauss [193] performed a study in which a user must describe a previously seen object from memory. The analysis shows that the subjects use gestures more than verbal communication in describing the object. When the object is difficult to describe, and gestures are limited, then users describe the object using a non-fluent speech. This suggests that gesturing, while speaking can reduce the cognitive load of the speaker in various practical scenarios [194]. Cook et al. [195] found that gesturing is quite common when speaking even for a fluent speaker. Their analysis suggests that gesturing in meaningful ways can reduce the working memory load, but not if the gestures are in rhythmic synchrony with speech (like beat gestures).

The human brain accepts inputs through multiple modalities and processes the information in parallel and in a coordinated manner. The same kind of architecture can be applied to design a multi-modal interface system that allows the users to use various modalities (speech, gesture, facial expressions, and gaze) in a natural and efficient way for communicating with the machine. However, there are individual differences in cognitive processing. Some studies state gender differences in verbal, quantitative, and visual cognitive activity [196]. For example, women are better in the verbal fluency test, whereas men are good at visual/spatial ability tasks. Males are also proven to be better in arithmetical computations, reasoning, and spatial cognition tasks [196]. Same is the case with handedness and expertise.

3.1.4 Models of Cognitive Processing

Functions of the human brain based on localization were proposed in the late 1700s, but the relationship between human activity and the neurons was established after the development of neural activity measuring techniques, such as EEG [36] which was developed in 1924. With the EEG, an association between activity and specific brain regions became possible, and a variety of theories of cognitive processing were proposed. Some of these theories are explained in the following subsections:

Modularity Theory (MT)

The modularity concept was explained by Chomsky who stated that the only way to acquire language was having infants born with a language acquisition device, such as an innate genetically determined module with a function of learning to speak [177]. The best explanation of the theory of modularity was given by Fodor [197] who stated that the cognitive processes such as perception and language are the result of functions of specific brain modules independent of the global central executive. The central executive organizes the interactions of each module through a mathematical approach governed by syntax.

However, these theories lack empirical support, since none of them has associated a specific brain structure with a specific module [36, 198].

Information Theory (IT)

After Fodor's theory [197] the information theory [181] (IT), was proposed. IT is a way of critical thinking and reasoning processes in the brain that involve input, computation, and output. The IT models in contemporary cognition research are considered oversimplified and do not account for individual differences [199]. Despite their simplicity, a number of models have been proposed accounting for human abilities such as Executive Process-Interactive Control [200], ACT-R [118], and PRODIGY [201].

Network Theory (NT)

Network Theories, also known as the Connectionist theories, are a more recent set of theories compared to Modularity Theory (MT) and Information Theory (IT). The MT lacks ecological validity, whilst IT is an oversimplified approach. Generally, the network theories propose a new function for cognition. The new function is the result of the connection made by individual neurons to form small assemblies which then interact with other neuronal assemblies [202]. The network theory presents a mathematically complex system in terms of graphs and networks that consist of vertices and edges. The edges show the connection between the vertices. The graphs or networks are generated using connectivity/ synchronization measures such as correlation, coherence, mutual information, and Granger causality/prediction [41].

The researchers usually utilize EEG or fMRI to construct the connectivity matrix for showing the relationships between the brain regions [203]. In NT, individual neurons do not need to be function specific as modular theories rely on. Instead, neurons can perform a general function such as depth perception or sound localization. Individual neurons could contribute to as many functional networks as it has connections to other neurons. In the action of neurotransmitters, individual neurons could be contributing to as many as 100 different networks [204], which can increase to thousands when dendrites and spines play an active role [205].

Among NT, Small world network theory (SWNT) is quite popular in describing the human brain function. In 'Small-world' networks, every network contains many short paths through which every vertex can be reached from every other vertex, as can be seen in many real-world networks [41, 206]. The matrices obtained from neuroanatomical databases show **that the functional brain networks in humans exhibit large clustering coefficients and short path lengths like small-world networks [207].** In another factorial study using fMRI, 90 functional brain networks were constructed in cortical and sub-cortical regions, which showed that these networks were small-world and economically efficient in providing information about human information processing [206].

Measures of Cognitive Processing/Function

Measuring cognitive processing or behavioral outcome from a neural activity is not a straight forward task. The neuroimaging techniques such as EEG and fMRI can provide an objective measure of neural activity. There are several tasks that can be used in experimental settings which have a vast literature to support their reliability and validity. The most common tasks are Stroop task [208], memory task [209], N-back task [210], distractor task [211], and Go/No go task [212].

In Stroop task, participants are shown a series of word stimuli; each word is a color name with some words displayed in the same ink as the color name, and some are not. The participants can be asked about color meaning irrespective of color ink or color ink irrespective of color meaning [208].

In memory tasks, participants are required to remember something about the presented stimuli. This may be instructions before the task commencement [209]. For example, in a driving simulator, a participant can be presented with complex auditory stimuli, and the participants are asked to identify whether the audio instruction is true or false. The task can be made more difficult by introducing other parameters to force the participant to use their memory.

The N-back task is a specific type of retrospective memory task in which participants are required to hold a specific type of stimulus in the working memory until a response is required after the presentation of "N" other stimuli [210]. For example, a participant is asked to respond only once in a one-back task wherein 5-back task; participants are required to give the response on the basis of the stimulus presented before five stimuli.

Distractor tasks are used in addition to some other tasks to make it more challenging using distraction [211]. The Go/No go tasks are about reaction time, in which participants need to press buttons when a target stimulus is presented and ignore all other distractor stimuli [212].

3.2 Psycho-physiological Analysis

Psycho-physiology is a branch of physiology that deals with the relationship between psychological and physical phenomenon. For psycho-physiological signals recording, three kinds of measure; reports, reading, and behavior are used. The reports evaluate participants introspection and self-rating about the psychological and physiological states [35]. Questionnaires are most commonly used to record the self-rating. The merits of the report are that it is a representation of user's subjective experience; however, the demerits include human errors such as bias response, misunderstanding of question or scale. [213]

Reading corresponds to the physiological responses that are measured via an instrument to read bodily events such as heart rate, body temperature, muscle tension, brain signals, and skin conductance [35]. The benefit of using these measures is that they provide an accurate and subject independent response; however, they are very prone to physical activity and situation [214]. The behavior measure involves the recording of observations and actions such as facial expressions and eye movements [35]. These responses are easy to measure and are mostly used in attention and emotions related experiments [215, 216].

In psychophysiology, a complex and interactive analysis of bio-signals is usually required. The application of psycho-physiological analysis ranges from stress to lie detection. Often, researchers use it to monitor the effect of an experiment on the user by measuring the short-term affective responses (e.g., feeling, mood, and disposition) [217]. Affective responses are considered as an instinctive state of mind based on circumstances and mood. These responses are spontaneous and last for a few minutes, which makes them hard to recognize. The classical affective states used in the psycho-physiological analysis are anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise.

The researchers have used the psycho-physiological signals to estimate the cognitive state of the participants. These signals have been used to analyze low order (e.g., simple visual inspection) and high order cognitive processes (e.g., attention, memory, language, problem-solving) [218]. Different signal sources are used in the literature for psycho-physiological analysis such as Electrocardiogram (ECG), Skin conductance (GSR), Electroencephalography (EEG), Electromyography (EMG), respiration rate (RR), Electrooculogram (EOG), Skin temperature (ST), and facial expression. Some of these measures are mentioned in the next section.

Table 3.1: Commercially available psycho-physiological signal measuring devices

Biofeedback signals	Commercially available devices
ECG	Alivecor System [219], Biopac [220], EPI mini [221], Omron ECG [222], Ambulatory ECG [223], Quasar sensors [224]
EEG	Mindwave headset [225], Flex sensor [226], Emotiv Headset [227], Neurosky Headset [228], Muse headband [229]
EMG	Neuronode [230], Sx230 [231], Trigno mini sensor [232]
EOG	Google glass [233], SMI eye tracking glassess [234], ASL eye tracking glasses [216]
GSR	Empatica [235], Shimmer 3 [236], Grove-GSR [237]
ST	YSI 400 series temperature probe [238], TIDA-00824 by Texas Instrument
RR	SA9311M [239], TMSI respiration sensor [240]

Various novel technologies have been used in the past to design electrodes for recording the above-mentioned psycho-physiological signals. These technologies have been upgraded from wet to dry electrodes with silver/silver chloride as the most commonly used plating material for these biofeedback sensors. Apart from silver/silver chloride, gold, aluminum, stainless steel, a mixture of some other metals such as nickel and titanium are also utilized [241] in sensors. The wet electrodes require an electrolytic gel to increase the conduction but cause discomfort to the participants. Thus, for applications that involve real-time recording preference was given to the dry electrodes [242]. A list of some commercially available measurement devices for recording biofeedback signals has been given in Table 3.1.

3.3 Measurement Methods

3.3.1 Electrocardiogram

The electrocardiogram (ECG) signal is a measure of electric potential recorded from the skin. The rise and fall of the signal identify different polarization levels of the heart over each heartbeat. The heart rate is measured by calculating the distance from R to R point (peak to peak), as shown in Fig. 3.4. The distance increases with a decrease in heart rate. One drawback of using ECG to find a heart rate is that sometimes it becomes uncomfortable because electrodes are in direct contact with the skin [243].

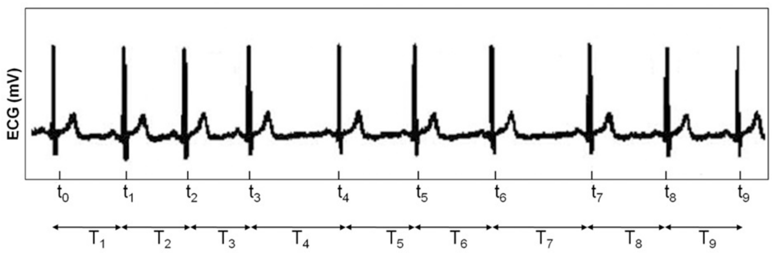


Figure 3.4: An R-R interval time series example [11]

3.3.2 Photoplethysmography

Photoplethysmography (PPG) is a low-cost optical device used to detect changes in blood volume in the microvascular bed of tissue. It is a non-invasive way of measuring blood volume changes. The PPG signal is comprised of two parts; pulsating (AC) signal that measures the changes in blood volume and it is synchronous with cardiac activity, and a slowly varying (DC) signal which contains various low frequencies used to measure the respiration and thermoregulation. These days, it is the most common way of measuring heart-rate, oxygen saturation, and blood pressure [244]. PPG has been used in HCI and Human-robot interaction (HRI) for measuring user experience in terms of emotion and stress [245].

3.3.3 Heart Rate Variability

Heart rate and Heart Rate Variability (HRV) are among the widely used features in detecting emotion states [246]. Autonomic Nervous System (ANS) activity can be effectively derived from the heart rate because the sympathetic and parasympathetic nervous systems govern ANS activity. Stress or activation can be related to ANS because, in a state of stress, the Sympathetic Nervous System (SNS) accelerates the heart rate. In the case of relaxation or rest, the heart rate returns to normal because of the Parasympathetic Nervous System (PNS) [246]. Heart rate is the number of heartbeats per min (bpm). On the other hand, HRV is the sequence of time intervals between heartbeats.

SNS activity is directly related to heart rate; an increase in heart rate is due to an increase in SNS activity. The opposite is the case with PNS; a decrease in heart rate triggers PNS activity which corresponds to the rest or relaxation states. There are some other features that can be derived from the acceleration and deceleration periods including the magnitude and slope of that period, the amount of time taken by these periods, and the mean difference over the baseline [246].

On the other hand, HRV is also sometimes useful in calculating the emotional state. HRV can be used to explain both time- and frequency-domain metrics. This metric can be simple, such as the standard deviation of successive heartbeats, to some relatively complicated metric, such as short-term power spectral density [247].

A simple, robust metric such as standard deviation is sometimes preferred with a short time window because of the limited information [248]. Other metrics can be the maximum and minimum difference between regular R-R wave time interval in a defined window, the successive normal R-R interval difference percentage that is greater than 50 msec (pNN50) and root mean square difference between consecutive R-R interval [249].

With the advancement in signal recording and processing algorithms, complex features such as short-time Fourier Transform (FT) or Power Spectral Density (PSD) of heart rate are becoming the more useful tools for analyzing HRV. The PNS activity can modulate the HRV in frequencies of 0.04 to 0.5 Hz. On the other hand, the SNS activity has functional gain below 0.1 Hz [247, 250]. The spectral domain can function as best in discriminating the SNS and PNS activity influence on HRV, and this is often known as sympathovagal balance.

One easy step to calculate the sympathovagal ratio of all heart rate activity is to measure the ratio of the energy of the lower frequency range (0.04- 0.1Hz) with the total energy in the band (0.04-0.5Hz). Some research suggests that it can also be measured by comparing the energy of the low-frequency band with a variety of combinations of low, medium, and high-frequency bands energy. [249].

Every HRV dependent measure is robust to artifacts such as noise, outliers and abnormal beats and difference in SNS vs. PNS activity. In accession to selecting a suitable metric, scientists and researchers ought to also select the acceptable time frame for heart rate series over which the metric needs to be calculated. The quality of heart rate series and the variable of interest will define the selection of a suitable metric from the cardiac signal. Generally, 5 minutes time window is recommended for an average heart rate of 60 bpm (beats per minutes) [251].

3.3.4 Skin Conductance

Skin conductance is the measure of a person's sweat level in glands. Normally, the skin is an insulator but its conductance changes when there is sweat in the sweat glands. Skin conductance is sometimes referred to as Galvanic Skin Response (GSR). Skin conductivity is a non-invasive method to detect sympathetic activation, which is sweat-gland activity [252]. Karl Jung used GSR for the first time to measure "negative complexes" in a word-connection experiment [253] which was further used as a key component in "lie detector" tests [254]. Skin conductance has been found to have a linearly varying property with respect to emotional arousal. It has been used to classify different states, such as anger and fear. It is also utilized in detecting stress level in experiments that are performed on anticipatory anxiety, and stress, while performing a task [255].

Skin conductance or galvanic skin response can be measured at any place on the skin. However, the highly active sweat glands for emotions are available on the hand's palm and the foot's sole [255]. In experimental studies, the middle and index finger's lower portion is the typical placement for skin conductance electrodes. Usually, a conductive gel is placed on the skin to ensure good conductivity of electrical signals. For measuring skin conductance, the voltage change is measured, while injecting a small amount of current into the skin [256]. By continually monitoring the change in potential difference across the electrodes, the skin conductance can be measured continuously.

For studies that involve movements, alternate electrodes locations are used because hand placement is sometimes found to be inconvenient, and the placement also distorts the signal when a person is moving. Some researchers have measured conductivity even through clothes and jewelry [257, 258].

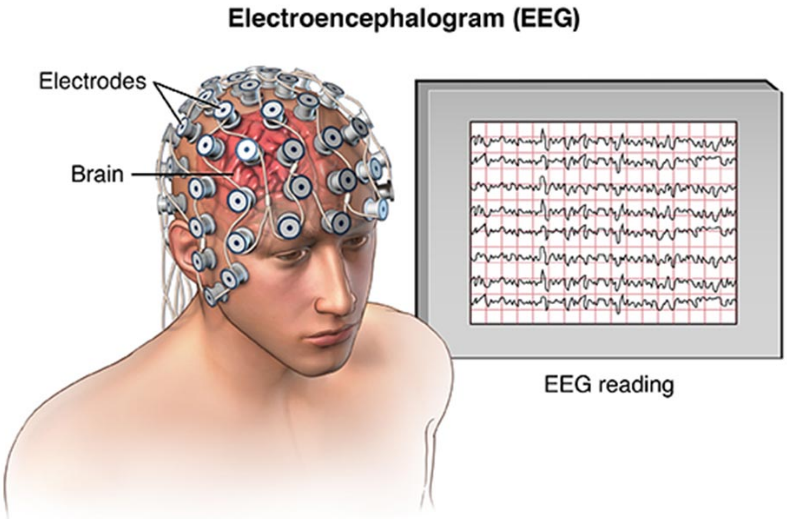


Figure 3.5: Electroencephalography (EEG) [12]

3.3.5 Electroencephalography

An electroencephalogram (EEG) corresponds to the electrical activity of the brain and is observed by measuring the electrical voltage generated by neurons (in μV). Electrodes are placed on the surface of the skull to record an EEG signal. The activity is either transmitted through a wired medium or wirelessly to a computer where it can be seen in the form of graphs for further analysis, as shown in Fig. 3.5. Analysis of EEG signals is a vast field with extensive research going on in the fields of neuroscience and psychology. The first EEG signal was recorded in 1924, almost 100 years ago, by Hans Berger [259] and now it is one of the important diagnostic tool for confirming epilepsy [260], distinguishing between coma and brain death [261], analyzing sleep patterns/ disorders [262], and identifying other brain diseases such as Alzheimer [263] and Brain hemorrhage [264].

Despite the recent advancement in the field of brain imaging such as fMRI, EEG is still an important tool in investigating brain functions and disorder. EEG is preferred over fMRI when the temporal resolution is in the order of seconds [36]. The other reasons for the popularity of EEG are the portability, effective artifacts removal techniques, no noise, no invasive radioactive tracers, non-invasive portable electrodes and cost (just few hundred dollars for modern systems) [36]. A full EEG headset comprises more than 128 channels/electrodes; however, some experiments use fewer electrodes in neurofeedback practice. Experimental studies have shown that the EEG has the potential to differentiate positive emotional valence from negative emotional valence, measure cognitive activity, and identify motor imagery movements.

The EEG signals can also identify different arousal levels. During an experiment that involved walking, the EEG can be considered as only a raw estimate of arousal level, but new advancements have the ability to change this concept [37]. The pre-frontal cortex (PFC) region of the brain seems to represent emotions such as anger [265]. James and Cannon [265] gave a model of the combined working of the mind and body in processing emotion for the first time.

EEG Frequency Bands

Most of the time, it is hard to identify abnormal activity by simply looking into the EEG signals. The raw EEG signal is decomposed into various components based on frequency bands to get a deeper understanding of the underlying activity. These frequency bands are delta, theta, alpha, beta, and gamma frequency bands [266].

Delta band: Delta band is a low-frequency band which ranges from 0.1 - 4 Hz and displays the maximum amplitude among the five bands. Delta band is normally associated with sleep activity, i.e., different stages of sleep [267]. In the literature, delta band activity has been used to distinguish between coma and brain death [268] and anesthesia depth [269].

Theta band: Theta band ranges from 3.5 - 8 Hz and normally relates to drowsiness (sleep or focus state) in adults [270]. The other cognitive functions that are also associated with theta-band are response inhibition [271], memory performance [270], attention deficit hyperactivity disorder (ADHD) [272], and forms of schizophrenia

[273].

Alpha band: The frequency range of 7.5 - 13 Hz lies under alpha band waves. The alpha band activity is seen in the occipital lobe in resting state with eyes closed, attenuated when eyes open [274]. There are many cognitive functions associated with the alpha band activity such as language comprehension [275], error processing [276], motor cognition and interaction [277], working and long-term memory [278], and task performance [279].

Beta band: Beta band activity is a low amplitude and normally lies between 13 - 30 Hz. Beta band activity usually relates to alertness or thinking state [270]. It is commonly used in many emotional processing algorithms [280] as well as sensorimotor movements [281]. Other associations of beta band includes Dementia and Parkinson's diseases [282], drug use/abuse [283], obsessive-compulsive disorder [284] and ADHD [272].

Gamma band: All frequencies above 30 Hz are considered in the gamma band. Gamma band activity is mostly observed in full consciousness and dream phase sleep [285]. Gamma band activity is thought to represent the coordination of various sub-order cognitive functions. Language processing, visual perception, and attention, short-term memory processing, multimodal processing are the basic cognitive processes where gamma band activity can be observed [286].

3.3.6 Facial Input

Some researchers have worked on recognizing emotions by considering facial features. One of the most widely used device for this purpose is MS Kinect. Kinect lets us track facial emotions by taking advantage of Kinect SDK API. Indeed, Kinect can detect and track the face orientation and position; it also detects eyebrow positions and mouth shape in real-time. The Face Tracking SDK can be used to identify facial expressions [287].

3.4 Latest Research in Psycho-physiological Analysis

Psycho-physiological signal analysis has shown promising techniques for measuring valence and arousal level for capturing the emotional and mental state. Self-reported data interrupt the flow of interaction and does not necessarily show the actual state of the user. The psycho-physiological measures help uncover the ground truth. The main problems with the psychological measures are complex equipment setting, signal analysis, and controlled environment which restrict the participant's experience of the interaction in many ways. Nevertheless, the advantages are far more than the disadvantages of psycho-physiological analysis [288]. Psycho-physiological analysis has been used in the literature to recognize emotions or affective states as well as cognitive activity, but most of the research is focused on affect recognition. In the remaining section, we will give an overview of recent trends in the use of psychophysiology in HCI.

3.4.1 Emotion/ Affect Recognition in HCI

The research in affective phenomena focuses on detecting emotions, feelings, mood, attitude, and temperament. A range of algorithms and techniques are available in the literature to detect emotions using different modalities. The first stage in these techniques is to generate the affective signals. This can be done in several ways, such as by watching videos, looking at images, listening to songs, and performing a number of tasks.

Our thoughts, feelings, and behavior are linked with emotions and therefore have a direct effect on decision making and thinking [289]. There are many definitions to describe primary and secondary affective states, but there is no uniform set. Six basic emotions used by many researchers are anger, joy, sadness, disgust, fear, and surprise, as recommended by Ekman [215]. Another model that has been used widely to define emotions is the wheel of emotion proposed by Plutchik [290]. In the wheel of emotion, there are eight emotions. Six of those emotions are the same as defined in [215]. The other two emotions are anticipation and acceptance.

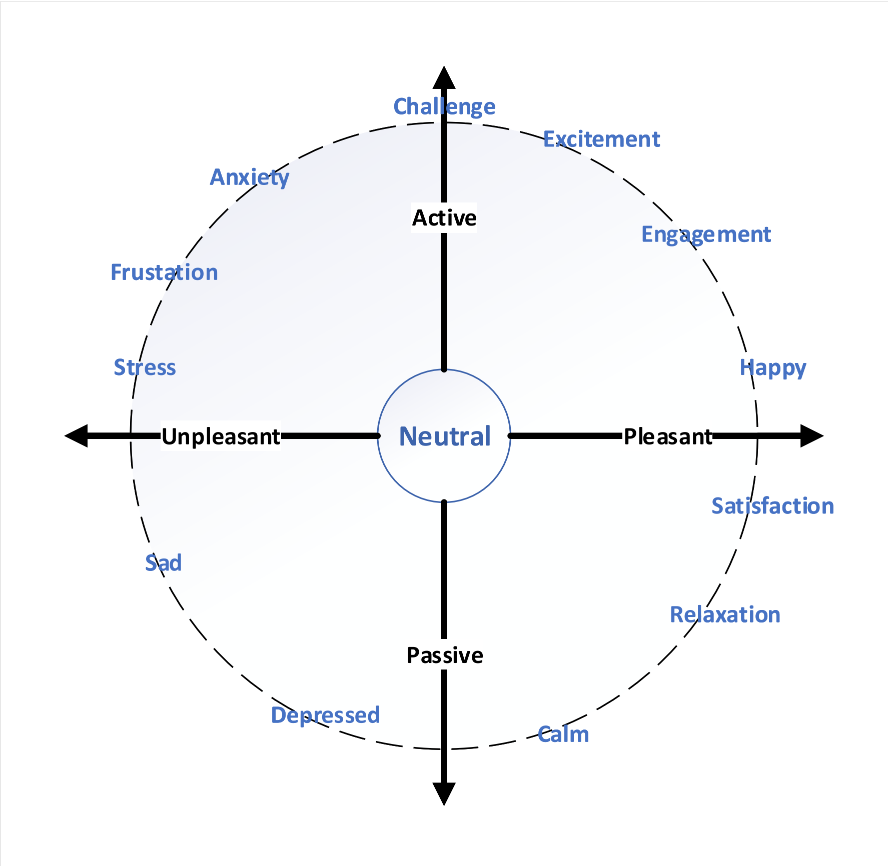


Figure 3.6: Arousal-Valence Space as described by Russel et al. [13]

Arousal and valence dimensions have been used by psychological researchers to model emotions in 2D, as shown in Fig 3.6. In an arousal-valance model, the arousal can be "active", or "passive" and valance can be "positive" or "negative" [13]. Lang [291] labels individual pictures based on an arousal valence space which is further converted into a non-verbal picture assessment called Manikin SAM [213]. Their self-assessment is used widely by advertising agencies and product designers to record affective experiences. The 2D arousal-valance model to define emotion is undoubtedly the most common model. A database named the International Affective Picture System (IAPS) is formed based on this model [292].

Emotion/affect recognition is a fundamental tool for the evaluation of HCIs, and the research is mainly focused on recognizing, interpreting, processing, and simulating human behavior and feelings [293]. Different research studies show that a variation in physiology is highly correlated with a variation in emotions [294]. Table 3.2 shows a comparison of some emotion recognition techniques along with stimulus and evaluation methods. For instance, a person's smile is mapped in positive valence; on the other side, displeasure relates to the negative valence.

Scheirer et al. [295] recognize frustration by classifying galvanic skin response and blood pressure. Klein et al. [296] also experiment with frustration by forcefully frustrating the subject using a game that involves text-based assistance for the user. The results of this experiment show that **the interaction time increases significantly when textual assistance is provided, in contrast to when no assistance is given.** Research studies also support the hypothesis that different stimuli can be used to generate different emotions [297], but these emotions are evoked by seeing a picture/video or listening to audio stimulus and this makes it hard to apply these procedures in real-world applications.

Extensive research has been done in recognizing emotions from face and voice with very high accuracy in cases where the experimental environment is controlled. The accuracy will be lower if the experiment is conducted in normal circumstances. Some researchers believe that emotions are generated due to physiological arousal, while others consider it to be a part of the emotional process [298]. In gaming research, a fuzzy approach has been used by Mandryk et al. [299] to recognize emotions using facial expressions and skin conductance, while playing NHL2003 on a Sony PS2. To record facial expressions, four electrodes have been used. Smiling and frowning are the two emotions that are recognized. The assumption is that smiling is related to positive valence and frowning is related to negative valence, but these assumptions are not enough for strong claims as it does not map the emotions to the valence scale effectively [215].

In experiments on a first-person shooter game, Juma [300] worked on secondary emotions by developing a game in which the primary emotion is combined with a secondary emotion to generate an affective component. The key finding of this experiment is that secondary emotion can be of vital importance in selecting an action in an HCI environment. Emotional films have been used by Costa et al. [301] to evoke five primary emotions of participants. To estimate the valence value of emotions, a synchronization index has been calculated. Li et al. [302] used pictures to generate happiness and sadness in a subject and record 90% classification accuracy. However, Horlings et al. [289] commented that the recognition rate would be low if the arousal and valence values are not extreme. A user-independent emotion-recognition system has been developed by Nie et al. [303]. The emotions in their experiment are generated by movies, and all four emotions are extreme emotions. Frequency-domain features of EEG signals have been extracted, and classification has been performed using a support vector machine (SVM).

Emotion recognition through EEG signals in brain-computer interfaces (BCI) and neuroimaging are usually carried out in a constrained environment. A small tolerance range is allowed for motor movement, which is vital in object manipulation activity. Nowadays, many researchers work on using the psycho-physiological signal analysis in real-life situations such as evaluating the performance of sportsmen and game environments. A review of current research in evaluating the peak performance of sportsmen has been done by Thompson et al. [304]. The study records the finding that the EEG signals are disturbed by motor movement, and it also discusses the techniques that can be used to generate reliable EEG recordings when the subjects are moving.

Nakasone and his team in 2005, presented a model to detect emotions in real time, using EMG and GSR [305] in a gaming scenario between the user and a 3D humanoid agent. Khair et al. published a review paper on human emotions in 2012 [306] in which protocols to generate and analyze human emotions, and an optimal induction method, have been proposed. According to Khair et al., music is considered to be the most popular way of inducing emotions.

The physiological responses to various emotional states are shown in Table 3.2. In another study, they found that different genders relate to different expressions of emotions [307]. Boys induce happiness and anger with faster music and upward movements, unlike girls. A combination of two approaches can be very useful in generating strong emotions such as combining music with a video or a game with strong emotional music.

Table 3.2: Relation of emotions to physiological responses

Emotion	Physiological Response
Pleasure and Sadness	low skin conductance and EMG, high heart rate
Anger	high skin conductance and EMG, flat and fast breathing
Joy	high skin conductance, EMG and heart rate, deep and slow breathing

Zhou et al. present a comprehensive study comparing visual and auditory stimuli to affect generation [308]. The study aims to answer the question: Can auditory stimuli be used effectively to elicit emotions instead of visual stimuli? They found that both stimuli were equally effective in inducing emotions. They also conducted a culture-specific analysis between India and China, but the accuracy was more or less the same. The reason for this may be the strictly controlled experimental environment. Based on their results, we think that visual stimuli strongly backed and synced with auditory data will be much more effective as emotional elicitors in practical HCI applications.

The psycho-physiological data must be sufficient to provide enough evidence in support of recognizing various factors affecting performance and to thoroughly test the developed techniques. Table 3.3 shows the summary of the available datasets for psycho-physiological analysis accessible publicly. A substantial amount of research has been carried out in recognizing emotion from facial input. An excellent review paper on facial emotion recognition on real-world user experience and mixed reality has been written by Mehta et al. [309]. Classification accuracy of almost 90% was seen in the literature using facial input, which indicates that there is still room for improvement.

Table 3.3: Summary of Publicly available datasets for emotion recognition (EDA: Electrodermal activity, GT: Gaze tracking, MEG: Magnetoencephalogram, EOG: Electrooculogram, EMG: Electromyography, RM: Respiration measurement, FT: Facial Tracking, ST: Skin temperature)

Database	Year	No. of Subjects	Psycho-Physiological Signal	Task/Experiment
MAHNOB [310]	2012	27	EEG, ECG, EDA, GT, RM, FT, ST	1st session: Emotional Videos, 2nd Session: Short Videos and Images
DECAF [311]	2015	30	EEG, ECG, MEG, EOG, EMG, FT	Affective Multimedia content (Movies and Music)
DEAP [312]	2012	32	EEG, FT	Watching Video
SEED [313]	2015	15	EEG, FT	Watching Video
Multi-modal Dataset [314]	2015	20	EEG, ECG, RM	Immersive Multimedia
AV communication [315]	2016	20	EEG, ECG, RM	Audiovisual Stimuli

3.4.2 Cognitive State Assessment in HCI

Another major research area in the psycho-physiological analysis is cognitive assessment. The literature is quite limited to cognitive assessment for multi-modal human-computer interface systems. Most of the literature is focused on the assessment of user cognition in games experience. The assessment of human-robot interaction is also popular among many researchers as well as some other HCI's evaluation through psycho-physiological signals.

Gaming Systems

The video gaming industry is one of the biggest industries in World [316]. Still, the assessment of user-game interaction and experience is primarily done by self-reported techniques [317]. With the development in measurement techniques and methods for psycho-physiological system, more and more research has been carried out in measuring user experience using psycho-physiological signals. A game user experience focused survey book written by Bernhaupt defines various user experience and evaluation methods [318]. In a review paper, the use of the psycho-physiological measure in video-games was investigated and listed the pros and cons of using psycho-physiological techniques [319]. They highlighted that the field lacks useful and widely accepted game-specific theory background, research and integrated knowledge.

Drachen et al. presented a study to find a correlation between self-reported data (In-Game Experience Questionnaire (iGEQ)) and psycho-physiological measures and found a direct correlation of iGEQ with heart rate [320]. **Some researcher studied the correlation of psycho-physiological measure and violent games and found an increase in cardiovascular activity when compared to non-violent games [321, 322]. The researchers reported that the psycho-physiological measures, especially heart rate, showed a strong correlation with self-reported data in both positive and negative experiences [321, 323, 324].**

The relationship between level design parameters, user experience, and player characteristics was explored by Pedersen et al., and **found a correlation between gameplay features and three emotions: Fun, challenge and frustration with an average performance of above 70% [325].** In a study [326], McMahan et al. assessed various stimulus modalities and gaming events using an Emotive EEG device.

They found a significant difference between various stimulus modalities that have increasingly difficult cognitive demands. **The power of the β and γ bands of EEG signals was increased during high-intensity events.** They also suggest that an Emotiv EEG headset can be used to differentiate between various cognitive processes.

Nacke et al. [327] studied the user experience in a fast-paced first-person shooter game with and without sound effects. EDA and facial EMG were recorded in addition to the questionnaire to evaluate the game experience. A significant effect of sound was observed in questionnaire results related to tension and flow, and these results correlate with EMG/EDA activity. The EDA, EMG, and ECG data were used to classify two different gaming event with 80% accuracy which showed that the psycho-physiological signal has the capability to differentiate between different user experience [328].

Stein at al. presented a method to adjust the game difficulty using EEG signals [329]. They estimated the long-term excitement of the participant to trigger the dynamic difficulty adjustment and found a correlation between excitement patterns and game events. In the literature, machine learning and evolutionary algorithms are used for clustering various gaming events [330], design new levels [331], difficulty adjustment [329, 332], modeling user experience [325, 333], and feedback to personalized game elements [334]. Despite these advancements, the investigation in modeling and estimating user experience for the improvement of the HCI system is still in its preliminary stages.

Human Robot Interaction (HRI) studies

Psycho-physiological analysis has been applied in HRI studies that involve interacting with actual robots to evaluate the user experience [335]. The main problem with the psycho-physiological analysis is to verify the accuracy and significance of the results. A research conducted by Itoh et al. used ECG, skin response, EDA, blood pressure and upper body movement to estimate the participant stress level and based on the stress level, modify robot action [336]. They found that **the user stress level decreased when the robot shook their hand.** Other researchers have found the same observation when modifying the robot's behaviors based on participants psycho-physiological state [337, 338]. Kulic and Croft evaluated the feasibility of psycho-physiological measures for user experience evaluation [339]. Results showed a relationship between anxiety, calmness, and the speed of the robotic arm. A stronger response was seen in EDA, EMG, and ECG signals. Dehais et al. study showed the same result when they evaluated the human response to different types of robot motion [340].

Researchers have used human gaze analysis to measure situation awareness in real-time in HRI [341]. The model was able to predict a standard measure of situation awareness. Podevijn et al. [342] study the psycho-physiological state of the participants when they interact with a swarm of robots. A direct relationship was found between user state and number of robots which the user is exposed to, and an increase in arousal value was observed when the user was exposed to 24 robots. The visual features of a robot such as an appearance and vocal properties had an affect on the cognitive state of the patient who is receiving some treatment [343].

In a study conducted to record the response of elderly people suffering from mild cognitive impairment interacting with a telepresence robot showed no adverse effect in cardiovascular activity [344]. Psycho-physiological measures have been used in evaluating haptic robot interaction for stroke patients in a multi-modal virtual environment [345] and a weak psycho-physiological response compared to healthy patients was observed. Ting et al. [346] proposed a framework of an adaptive automation system based on the operator's mental state calculated through heart-rate variability and task load index. Munih and Mihelj presented a very interesting article that summarizes the psycho-physiological response in robot-assisted rehabilitation including multi-modal challenges and physical activity [347].

Other HCI systems

Psycho-physiological measures are used as a tool to objectively investigate user experience in many other systems. Zhang et al. [348] studied the cognitive load measurement of a virtual reality driving system with multi-modal information fusion techniques. They found that a hybrid fusion of modalities is best suited for these kinds of challenging tasks, probably because of Dual Coding Theory. Yao et al. [349] uses psycho-physiological signals to evaluate the user experience of mobile applications. Participant's physiological responses, task performance, and self-reported data were collected, and they **found a correlation between self-reported data and skin response, an increase in skin response in failed tasks compared to successful tasks.**

Various frequency bands of EEG signal have been used to study the cognitive load of the user. Kumar and Kumar used EEG to measure cognitive load in an HCI environment and found a significant difference in spectral power between low level and high-level cognitive tasks [350]. Puma et al. used theta and alpha band power of EEG to estimate the cognitive workload in a multitasking environment [351]. The results showed an increase in alpha and theta band powers when there was an increase in the involvement of cognitive resources for completing the sub-tasks.

Significant differences were found in skin response, HR and blood volume pulse (BVP) in response to a video conferencing tool [352]. **An increase in GSR, HR, and decrease in BVP was observed for videos at 5 frames per second compared to 25 frames per second.** Most of the subjects didn't notice the difference in video quality, which indicates that psycho-physiological measure has the capability to mine the underlying fact that cannot be found using traditional methods of measuring user experience [349]. In a comparison study between well- and ill-designed web pages,

Ward et al. [353] found a decrease in GSR and HR in well-designed web pages compared to ill-designed web pages which result in an increase of stress level.

Anders Bruun presents a study where non-specialists analyze GSR data to detect user experience related events and found an accuracy of 60-80% [354]. Lin et al. present an investigation study to find the relationship between physiological measure and traditional usability index and found evidence that physiological data correlates with task performance and subjective reports assessing stress levels [355]. To study the experience in virtual reality, Meehan et al. conducted a study where **they compared the participant's physiological response to a non-threatening virtual height simulation and found a change in heart rate and skin conductance [356].**

The human brain responds differently to text and multimedia stimuli; to investigate this statement, Gere et al. [357] present a study in which they investigate cognitive processes that take place in learning information presented in a visual or text format. They use EEG signals to measure cognitive activity and found higher α-band power, corresponding to less mental activity in the brain, for text presentation. They also concluded that video and picture input gives a spark to visualization strategies, whereas text-induced activity is related to verbal processing. No gender-related differences were observed during this experiment. The same kind of work has been done by Madi and Khan [358]; they focused on analyzing cognitive activity and learning performance in text and multimedia comprehension. Cognitive load and emotions were monitored during the study. They found differences in α-and β-band power. Their study revealed that **multimedia presentation, such as video and image, elicit positive emotions more than a text presentation, which induces a higher cognitive load.**

To study the differences between single-task and dual-task multi-modal human-computer interaction, Novak et al. found significant changes from baseline to single and dual-task in psycho-physiological signals, but no differences were found between single and dual mental arithmetic task [359]. Their results suggested that different task results show different response in the psycho-physiological measure, and it is not compulsory that the response correlates with the participant's subjective feelings. Researchers had found significant differences in respiratory response when the participants were given a high-level cognitive task. Grassmann et al. presented a systematic review of respiratory changes with respect to cognitive load [360].

Psycho-physiological analysis has also been used in the study of cognitive skills and information processing in programmers. Lee et al. [361] present a study in which they examine the differences between novices and experts in programming comprehension. They used EEG to record the neural activity and found apparent differences between novices and experts. The results showed that experts have superior programming comprehension abilities and excel at digit encoding, solving simple programs in a short time, and the ability to recall program functions after an extended period of time compared to a novice. Psycho-physiological analysis has been used for assessing real-time cognitive load for younger and older adults in the situation of divided attention and task interruption with an average cognitive load assessment of 73% for younger and 70% for older adults [362].

Liu et al. [363] analyzed the psycho-physiological signals to detect affective states of engineers in CAD activities and found that the EEG results correlate with the emotions described by the engineers during that activity. In another paper, Nguyen and Zeng used heart rate and EEG signals to find the relationship between the designer's mental efforts and stress levels [364]. They found that mental effort was the lowest at high-stress levels, and no variations in the mental effort were seen in medium and low-stress level tasks. In another research work, Nguyen and Zeng found a strong association between self-rated effort and beta band power. They demonstrated that self-rating itself contributes towards mental activity [365].

Based on the literature review, we found that GSR/EDA is best suited to record arousal and mental efforts, HR is equipped to measure the arousal in emotion, likeability, and attention. HRV, EMG, and respiration are mostly used for emotional state estimation. BVP is used for evaluating relaxation, and facial input is applied to recognize emotions from facial expression. EEG signal is widely used to detect emotions, frustration, and mental effort. In this research, we will be analyzing the cognitive activity of the participants through EEG signals while using a multi-modal interface system. Since the literature focusing on 3D object manipulation is inadequate, this study will serve as a starting point in this research direction.

3.5 Functional Brain Networks

The human brain is a complex network of millions of neurons and the links between these neurons. Therefore, complex network theory can open a whole new portal towards understanding human cognition [38]. However, understanding how information is processed in the brain and how decisions are made can affect the results. Nodes play an important role in the functional brain network (FBN), and there is no clear consensus among neuroscientist on the definition of a node. Despite the considerable differences, the literature is encouraging about the studies of functional brain networks. The commonly used input to generate the functional brain networks are functional magnetic resonance imaging (fMRI), EEG, magnetoencephalography (MEG) and multielectrode array (MEA) [38]. In this research, we have used EEG signals for FBN analysis.

3.5.1 Connectivity Measures

An integrated structure of neurons generating a neural activity from different sources are recorded by the EEG signal, and to analyze the dynamics of this kind of system we need a multivariate analysis. There are many linear and non-linear connectivity measures that have been used in the literature to construct the FBN such as mutual information, Entropy, correlations, and Granger causality [39]. Linear connectivity measures usually fail to identify the non-linear behavior of the brain. Therefore, to analyze a highly non-linear EEG signal, non-linear measures are adopted by the researchers for the construction of FBNs [40]. In this section, we will discuss some of the commonly used linear and non-linear measures to establish connectivity between the nodes of a network.

Pearson's Correlation Coefficient

The basic function to define a linear correlation between two variables is cross-correlation and can be calculated between two signals $x(t)$ and $y(t)$ by the following formula:

$$C_{xy} = \frac{1}{N-\tau} \sum_{k=1}^{N-\tau} x(k+\tau) y(k) \qquad \textit{Eq(3.1)}$$

where τ is the time lag between two signals and N is the total number of samples in the signal. Pearson's correlation coefficient is calculated at zero lag between two time signals. For a signal with zero mean and unit variance, Pearson's correlation coefficient is calculated by putting $\tau = 0$ as:

$$C_{xy} = \frac{1}{N} \sum_{k=1}^{N} x(k) y(k) \qquad \textit{Eq(3.2)}$$

$C_{xy}$is in the range of $-1 \leq P_{xy} \geq 1$:

- 1 : complete direct correlation between $x(t)$ and $y(t)$
- 0 : no linear interdependence between $x(t)$ and $y(t)$
- -1 : complete inverse correlation between $x(t)$ and $y(t)$

The main advantage of using Pearson's coefficient is that it is well-known and fast to compute but only detects linear dependencies [366].

Magnitude squared coherence

The magnitude squared coherence is used to measure the linear correlation between two signals as a function of frequency. The coherence function is calculated as:

$$K_{xy}(f) = \frac{S_{xy}(f)}{\sqrt{S_{xx}(f) S_{yy}(f)}} \qquad Eq(3.3)$$

where $S_{xy}(f)$ is the power spectral density between $x(t)$ and $y(t)$. $S_{xx}(f)$ and $S_{yy}(f)$ are the individual power spectral densities. Thus, the magnitude squared coherence is given by:

$$Coh_{xy}(f) = |K_{xy}(f)|^2 = \frac{|S_{xy}(f)|^2}{S_{xx}(f) S_{yy}(f)} \qquad Eq(3.4)$$

The value of $Coh_{xy}(f)$ ranges between 0 and 1, where 0 means no dependency and 1 means maximum correlation for the corresponding frequency. The coherence is a well-known connectivity measure like correlation coefficient and is employed in many cognitive and clinical EEG applications [367]. The problem with coherence is, like Pearson's correlation coefficient, that it is unable to detect non-linear or causal relationships between signals.

Mutual Information (MI)

Mutual information is a measure to find the dependency between two variables [368]. It quantifies the amount of information obtained by one random variable by observing another [367]. The Mutual Information is calculated as:

$$I(X,Y) = H(Y) - H(Y|X) \qquad Eq(3.5)$$

where I is the mutual information, $H(Y)$ is the entropy of Y and $H(Y|X)$ is the entropy of variable Y observing a variable X. By substituting the expression of entropy in eq.

3.5:

$$I(X,Y)=\sum_{y\epsilon Y}\sum_{x\epsilon X}p(x,y)log\frac{p(x,y)}{p(x)p(y)} \qquad Eq(3.6)$$

where $p(x,y)$ is the joint probability distribution function (PDF) of X and Y. $p(x)$ and $p(y)$ are the marginal PDFs of X and Y respectively. Mutual information will be zero if X and Y are independent and greater than zero if they are dependent. Some researchers also use Kullback-Leibler divergence [369] formula to measure mutual information between two densities. After measuring the mutual information, the next step is to rank the features through a threshold. The problem with mutual information is that it ignores the inter-feature mutual information. Another common variation of mutual information used in the literature is conditional mutual information [370]. MI detects high order correlation and it can detect non-linear dependencies because it uses probability distribution [371]. However, MI cannot detect causal relationships as it doesn't have the directional information [372]. MI has been used in the literature to construct non-directional FBNs from EEG data during resting and cognitive load states [373].

Granger Causality (GC)

Granger Causality (GC), also known as Granger Prediction, is the prediction of the first signal by considering the past information from the second signal instead of only using the information from the first signal, then the second signal can be called causal to the first signal [374]. This concept was originated from econometrics. Granger proposed the mathematical formulation that when x is determining y, then by adding past values of x to the regression of y will provide an improved prediction [374]. Thus, the uni-variate auto-regressive model for x and y can be given as:

$$x(t)=\sum_{k=1}^{P}a_{xk}x(t-k)+u_x(t) \qquad Eq(3.7)$$

$$y(t) = \sum_{k=1}^{P} a_{yk} y(t-k) + u_y(t) \qquad Eq(3.8)$$

where a_{xk} and a_{yk} are the model parameters, P is the model order, u_x and u_y are the uncertainties associated with the model. In the above-mentioned equation, the prediction is estimated by its own past components. The variance of the residual or uncertainties are denoted by:

$$V_{x|\bar{x}} = var(u_x) \qquad Eq(3.9)$$

$$V_{y|\bar{y}} = var(u_y) \qquad Eq(3.10)$$

For bi-variate auto-regressive model:

$$x(t) = \sum_{k=1}^{P} a_{xyk} x(t-k) + \sum_{k=1}^{P} b_{xyk} y(t-k) + u_{xy}(t) \qquad Eq(3.11)$$

$$y(t) = \sum_{k=1}^{P} a_{yxk} y(t-k) + \sum_{k=1}^{P} b_{yxk} y(t-k) + u_{yx}(t) \qquad Eq(3.12)$$

Now, the uncertainties or residuals u depend on the past values of both signals and their variance can be calculated as:

$$V_{x|\bar{x},\bar{y}} = var(u_{xy}) \qquad Eq(3.13)$$

$$V_{y|\bar{x},\bar{y}} = var(u_{yx}) \qquad Eq(3.14)$$

where $x|\bar{x},\bar{y}$ is the prediction of x by the past values of x and y. Therefore, GC from y to x is:

$$GC_{y\rightarrow x} = ln(\frac{V_{x|\bar{x}}}{V_{x|\bar{x},\bar{y}}}) \qquad Eq(3.15)$$

$GC_{y\rightarrow x}$ is evaluated between 0 and ∞, where 0 means the past $y(t)$ does not improve the prediction of $x(t) : V_{x|\bar{x}} \approx V_{x|\bar{x},\bar{y}}$. If the value is greater than zero then $y(t)$ improves the prediction of $x(t) : V_{x|\bar{x}} >> V_{x|\bar{x},\bar{y}}$.

Granger causality has been used in the field of neuroscience for the past 15 years because of its capability to detect causal relationships along with the direction of information flow. GC has been used to analyze brain dynamics during foot movements using EEG signals [375], to find directed interaction from parietal to the frontal lobe in a Stroop task [376]. As most of the interaction is non-linear in nature, the linear GC cannot be applied to measure the causality; instead a non-linear extension of GC is used such as kernel-based methods [377], non-parametric [378] and parametric methods [379].

Partial Directed Coherence (PDC)

Partial directed coherence (PDC) is another popular frequency domain connectivity measure based on GC [380]. The PDC model is based on time series by multivariate auto-regressive (MAR) processes. Consider the following MAR processes of order p with M dimensions:

$$\begin{pmatrix} x_1(k) \\ . \\ . \\ . \\ x_M(k) \end{pmatrix} = \sum_{r=1}^{p} A_r \begin{pmatrix} x_1(k-r) \\ . \\ . \\ . \\ x_M(k-r) \end{pmatrix} + \begin{pmatrix} \varepsilon_1(k) \\ . \\ . \\ . \\ \varepsilon_M(k) \end{pmatrix} \qquad Eq(3.16)$$

where $A_1, A_2 \dots A_p$ are $M \times M$ coefficient matrices, and $\varepsilon_i(k)$ are independent Gaussian white noise with co-variance matrix $\sum$. To calculate the frequency version, we can compute the power spectral density matrix of the above equation:

$$S(f) = H(f) \sum H^H(f) \qquad Eq(3.17)$$

where $(.)^H$ is the Hermitian transpose of the transfer function H: $H(f) = \overline{A^{-1}}(f) = [I - A(f)]^{-1}$, $A(f)$ is the Fourier transform of the coefficient and $\overline{A}(f) = [\overline{a_1}(f)\overline{a_2}(f) \dots \overline{a_M}(f)]$, with $\overline{a_{ij}}(f)$ are the i, jth elements of $\overline{A}(f)$. Then, the PDC from signal j to i is given by:

$$PDC(f) = \pi_{ij}(f) = \frac{\overline{a_{ij}}(f)}{\sqrt{\overline{a_{ij}^H}(f)\overline{a_j}(f)}} \qquad Eq(3.18)$$

The above-mentioned expression evaluates between 0 and 1: 0 means no coupling, 1 means complete coupling.

Some researchers prefer PDC over GC because it doesn't involve matrix inversion which makes PDC computationally efficient [381]. PDC has been used to analyse EEG and fMRI data for estimating directed information flow [382, 383]. Non-linear PDC measure has been proposed along with its application on EEG data [381].

Transfer Entropy (TE)

The concept of entropy was introduced by Shannon in Information Theory [384] to quantify the information in a variable in number of bits required to optimally encode discrete variable X based on probability distribution $p(x)$. The Entropy or Shannon Entropy can be calculated using the following equation:

$$H(X) = -\sum_x p(x) log_2 p(x) \qquad Eq(3.19)$$

For two variables, the Entropy can be calculated as:

$$H(X,Y) = -\sum_{x,y} p(x,y) log_2 p(x,y) \qquad Eq(3.20)$$

where the sum is for all the possible states of x and y. For two time series $X = x_t$ and $Y = y_t$, Schreiber [40] proposed the concept called Transfer Entropy (TE) to compute the deviation from the Markov condition as:

$$p(y_{t+1}|y_t^n, x_t^m) = p(y_{t+1}|y_t^n) \qquad Eq(3.21)$$

where $x_t^m = (x_t, ..., x_{t-m+1})$ and $y_t^m = (y_t, ..., y_{t-n+1})$, t denotes the time step, m and n are the order of Markov Processes X and Y respectively. TE is an information measure to determine quantity and direction of information transfer between two processes [40]. To describe the TE from Y to X, Schreiber proposed the following equation:

$$TE_{y\to x} = \sum_{x_{n+1},x_n,y_n} p(x_{n+1}, x_n, y_n) log\left(\frac{p(x_{n+1}, x_n, y_n).p(x_n)}{p(x_n, y_n).p(x_{n+1}, x_n)}\right) \qquad Eq(3.22)$$

where x_n is the value of signal x at time n, y_n is the value of signal y at time n, and $p(.)$ is the probability distribution. TE is inherently asymmetric and ranges from 0 to ∞. To improve the calculation accuracy, two more steps are mentioned in the literature [385]. EEG is highly non-stationary data of finite length which can induce a large amount of noise, average shuffle TE from Y to X has been subtracted from the estimated TE. Then, normalized TE is calculated from Y to X with respect to total information in sequence X to represent a relative amount of information transfer. The equation for calculating NTE from a vector $Y \to X$ is given in Equation 3.23 as [386]:

$$NTE_{y\to x} = \frac{TE_{y\to x} - < TE_{y_{shuffle}\to x} >}{H(x_{n+1}|x_n)} \qquad Eq(3.23)$$

where $TE_{y \to x}$ is the transfer entropy from Y to X and can be calculated using Equation 3.22, $< TE_{y_{shuffle} \to x} >$ is shuffled TE from Y to X using shuffled version of Y and $H(x_{n+1}|x_n)$ is the conditional entropy of X at time $n+1$ given its value at time n and calculated as given in Equation 3.24. In equation 3.23, $y_{shuffle}$ contains the symbols that are rearranged and shuffled in random order.

$$H(x_{n+1}|x_n) = -\sum_{x_{n+1},x_n} p(x_{n+1},x_n) log\left(\frac{p(x_{n+1},x_n)}{p(x_n)}\right) \qquad Eq(3.24)$$

The NTE from $y \to x$ is not equal to NTE from $x \to y$ and NTE is in the range of 0 and 1. If the value of NTE is 0 that means no transfer of information and if the value is 1 than the information transfer is maximum [387]. Functional Brain Networks (FBNs) can be constructed by computing the NTE from EEG signals.

TE is preferred over GC because GC fails to identify the interaction between highly non-linear processes such as the human brain and TE doesn't require a model of the interaction [388]. TE is a popular measure to quantify information between two non-linear processes. It can also determine the direction of information transfer between two processes [389] and therefore, is ideal for investigating information flow. TE uses the past activity of both variables to estimate the amount of activity of a system irrespective of the interaction model. This property of TE allows the researchers to apply it to various applications such as identifying information transfer between auditory cortical data [390], localization of epileptic patients focus [391], the effect of heart rate on breath rate [392], and information flow patterns in various driving states [387].

3.5.2 Summary of Functional Brain Connectivity Measures

In the literature, both linear and non-linear connectivity measures have been used to study the FBNs. The linear measure such as Pearson's correlation coefficient and

Coherence provide a degree of correlation between signals (or pair of electrodes) and only produce linear brain connectivity, although the analysis of any FBN is restricted by the linearity of the measure. The non-linear connectivity measure provides a much more realistic analysis. Almost all non-linear measure such as GC, MI, and TE can be used to construct FBNs, but only GC and TE can provide directional information. Directionality is essential information that can help to understand how one brain region influences another. Thus, the directionality limits the applicability of MI along with other non-linear measures such as generalized, likelihood, and phase synchronization [367]. GC and TE are very popular in establishing connectivity between electrodes from EEG and MEG signals; thus in this thesis, TE has been used to construct the FBNs because it does not require the interaction model.

3.5.3 Graph Theory and Complex Brain Networks

A graph is a mathematical model that consists of nodes and edges. Nodes are the vertices and edges are the links between each pair of nodes. So, a graph $G = (V, E)$ contains two sets, vertices V and links E, such that $V = \phi$ [393]. An e in set E is identified by the unique pair of nodes $[u, v]$ in V, denoted by $e = [u, v]$. In the context of brain networks, a graph can be considered as a non-linear model of neural activity where each node corresponds to a brain region and the connections between the regions are represented as links [41].

Types of Graphs

The edge or link between graphs nodes constitutes towards different types of graphs such as directed, undirected, and weighted graphs. A brief explanation about these types is given below:

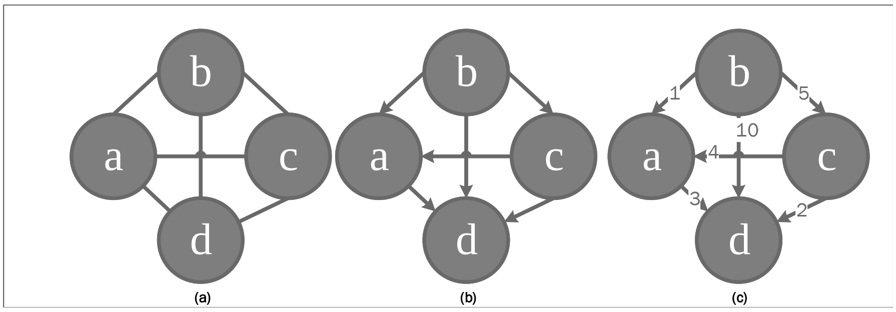


Figure 3.7: Example of (a) undirected (b) directed and (c) weighed directed graphs

Undirected graph: A graph where edges have no direction or orientation is known as an undirected graph. Undirected graphs are created with a straight line between the vertices, as shown in Fig. 3.7.

Directed graph: In a directed graph, the edges are assigned a direction. The direction is shown by adding an arrow to the edges, as shown in Fig. 3.7. The directed graphs have both an in-degree and out-degree by counting the number of edges coming in and going out of a node [41].

Weighted graph: In a weighted graph, each edge is assigned a weight such as a cost, capacity, or length of the edge. A weighted graph can be directed or undirected. A directed weighted graph has been shown in Fig. 3.7.

Representation of Graph

The common ways of representing a graph are the adjacency matrix and the adjacency list.

Adjacency list: In an adjacency list, a linked list representation is used to represent a graph. In the adjacency list, the space needed depends on the number of elements in the graph.

Adjacency Matrix: The adjacency matrix is a 2D array of size $n \times n$, where n is the total number of vertices/nodes in the graph. The adjacency matrix $A = (a_{ij})$ of a graph can be defined as:

$$a_{i,j} = \begin{cases} 1 & \text{if } v_i \text{ and } v_j \text{ are adjacent,} \\ 0 & \text{otherwise.} \end{cases}$$

The adjacency matrix of a weighted graph can be defined as:

$$a_{i,j} = \begin{cases} 1 & \text{if } v_i \text{ and } v_j \text{ are adjacent with weight } w, \\ 0 & \text{otherwise.} \end{cases}$$

The adjacency matrix of an undirected graph is symmetrical, whereas the adjacency matrix of a directed graph is generally asymmetrical.

Types of Connectivity of Complex Brain Network

The connectivity of neuronal networks can be categorized into three types: anatomical/structural, functional, and effective connectivity.

Anatomical connectivity: Anatomical connectivity is a combination of physical or structural connection between the neuronal elements by the synapses or axonal projections at a given time. The connectivity data can range over multiple spatial scales. The patterns are relatively static at short time scales, but can be dynamic for long time scales [394]. The anatomical connectivity can give insight into the neural activity in the spatial domain but cannot reveal when in real time.

Functional connectivity: Functional connectivity is the most popular type of connectivity that helps to capture the patterns of the symmetrical statistical association between neuronal elements. Functional connectivity is time-dependent and modal free and can provide information on both where and when a neural activity occurs [394].

Effective connectivity: The effective connectivity describes the causal effects of one neural system over another. Unlike functional connectivity, it is not model-free and can be inferred through perturbations or observation of temporal ordering of neural events [394]. Fig. 3.8 shows a work-flow to construct the brain network [14].

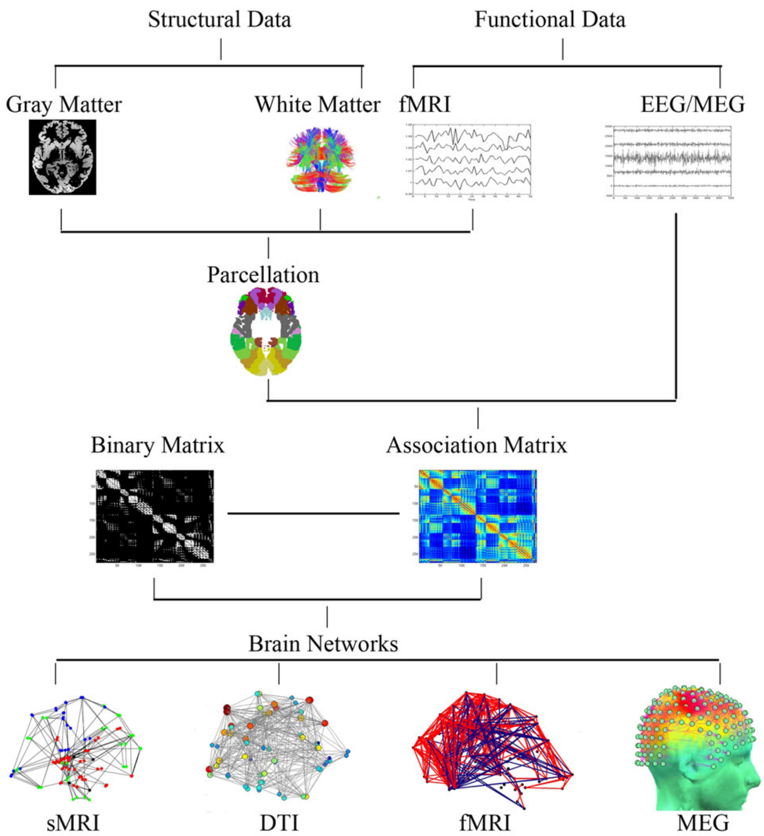


Figure 3.8: Structural and Functional network construction from various types of data [14]

The structural connectivity of the brain is usually measured using Structural Magnetic Resonance Imaging (sMRI) or Diffusion Tensor Imaging (DTI) techniques. The functional connectivity of the brain is measured from fMRI, EEG, or MEG techniques, and can be used to construct functional brain networks. The association between different regions can be estimated through statistical measures.

3.5.4 EEG Based Functional Brain Networks

EEG data can be used to construct functional brain networks (FBNs). Each electrode of the EEG becomes a node, and the connections between the electrodes becomes edges [38]. EEG signals can be recorded from various electrodes positions on the scalp during various states, including resting and cognitive load states. In the literature, FBN is constructed using a short period of EEG data such as 2 or 5 seconds [395]. Both linear and non-linear information measures could be used to construct the connectivity matrices for FBNs. In this thesis, we have used Normalized Transfer Entropy (NTE) to construct the FBNs.

Complex Network Metrics

To analyze complex FBNs, both local or global connectivity measure can be used [396]. In local graph-based connectivity measures, each node or link measures the connectivity profiles associated with them. In global connectivity measure, the global description of the networks is provided and is measured with the help of all elements [41]. The most commonly used metrics for measuring connectivity in the FBNs are motif count, connectivity density, characteristic path length, clustering coefficient, degree centrality.

Connectivity Density: Connectivity density (CD) is the ratio of the actual number of edges to the total number of possible edges [397]. The value of the CD is in the range of 0 and 1. If a graph has complete connectivity, then it has a CD of 1. The connectivity density of a directed network is given in Equation 3.25,

$$CD = \frac{t}{N(N-1)} \qquad Eq(3.25)$$

where t is the number of edges in the network, and N is the number of nodes.

Motif: The motif is used to describe the local structure of a graph. It is the patterns of interconnections that can occur in a complex network [398]. It is further characterized by the number of times a subgraph appears in a complex network [399]. In the case of 3 nodes, a total of 13 classes of a subgraph can be generated, with each class of motif referred to by motif id.

Modularity: There can be several smaller sub-modules in a complex network which can be identified because of the dense interconnection between nodes of the sub-modules, but few connections with other modules [38]. One way of finding the sub-modules in a directed network is given in Equation 3.26 as:

$$Q(p) = \frac{1}{m}\sum_{ij}\left[a_{ij} - \frac{k_i^{out}k_j^{in}}{m}\right]\delta_{ci,cj} \qquad Eq(3.26)$$

where Q(p) is the modularity of a given participation p, a_{ij} is an element of the adjacency matrix, k_i^{out} and k_j^{in} are the out- and in-degrees of the vertices, m is the total number of edges, δ is the Kronecker delta symbol and c_i is the label of the module to which each vertex is assigned [400]. The modularity algorithm divides the complex network into sub-modules in a way that the modules gives maximum Q over a possible division that is considered as the best estimate.

Degree centrality and degree distribution: The degree of a node is defined by the number of connections the node has, and it shows the importance of a node in a network. The degree of a node i is:

$$k_i = \sum_{j\epsilon N} a_{ij} \qquad Eq(3.27)$$

The degree distribution depends on the degree of all the nodes in a network. In a directed graph, the total degree is the sum of in- and out-degrees of a node. There are three possible degree distributions $P(k)$ that can be used to fit the in- and out-degree distribution in a directed FBN:

$$P(k) \approx \begin{cases} k^{-\alpha} & \text{a power law} \\ e^{-\alpha k} & \text{an exponential} \\ k^{\alpha-1}e^{-k/k_c} & \text{an exponentially truncated power law} \end{cases}$$

where k is the in- or out-degree, α is the estimated exponent, and k_c is the cut-off degree. Previous studies in the literature on resting state fMRI data showed that the exponentially truncated power law is the best fitting model for directed and undirected FBNs [401].

Node Strength: Node strength is used to measure the centrality of weighted directed networks. It represents the sum of all incoming and outgoing edge weights [41]. Node strength helps to find the involvement of a particular region in an FBN.

Small-worldness: The small-world network exists in-between regular lattice and completely random network and shows the properties of high clustering and short path length, which means that information traverses mostly between nodes with a small number of edges not between neighbours [402]. It is also considered as a network with both high local and global efficiency [403] which is a representation of effective information propagation over a network.

Clustering Coefficient: The clustering coefficient is the ratio between all the directed triangles formed by node i and the number of all possible edges a node i can form, so the clustering coefficient measures how well the cluster of nodes are communicating and a high value of clustering coefficient relates to the high local efficiency of information transfer [38]. The directed clustering coefficient of a network is calculated using equation 3.28,

$$C_d = \frac{1}{N}\sum_{i=1}^{N} = \frac{1}{N}\sum_{i=1}^{N} \frac{\frac{1}{2}\sum_{j=1}^{N}\sum_{h=1}^{N}(a_{ij}+a_{ji})(a_{ih}+a_{hi})(a_{jh}+a_{hj})}{(k_i^{out}+k_i^{in})(k_i^{out}+k_i^{in}-1)-2\sum_{j=1}^{N}a_{ij}a_{ji}} \qquad Eq(3.28)$$

where C_i is the clustering coefficient of node i, N is the number of nodes, and a_{ij} is the directed connection from node i to node j,

Characteristic Path Length: The characteristic path length is the average shortest path length between all pairs of nodes. It indicates the high global efficiency of information transfer. The shortest path between node i and j is the minimum number of nodes to traverse from node i to reach to node j [38] and is given by:

$$d_{ij} = \sum_{a_{ij} \epsilon g_{i \to j}} a_{ij} \qquad \text{Eq(3.29)}$$

where $g_{i \to j}$ is the directed shortest path from node i to node j. So the characteristic path length of an FBN can be calculated as:

$$L_d = \frac{1}{N} \sum_{i \epsilon N} L_i = \frac{1}{N} \sum_{i \epsilon N} \frac{\sum_{j \epsilon N j \neq 1} d_{ij}}{N-1} \qquad \text{Eq(3.30)}$$

where L_i is the average distance between node i and all the other nodes, and N is the number of nodes.

Small-world Index: The small-world index gives us the comparison of a complex network to a random network. A random network has a low clustering coefficient and typically a short path length compared to a complex network [404]. The small-world networks have high global and local efficiency and are categorized as those networks whose small-world index $\sigma > 1$ [405]. Small-world index can be measured with the following equation:

$$\sigma = \frac{\frac{C_d}{C_{rand}}}{\frac{L_d}{L_{rand}}} \qquad \text{Eq(3.31)}$$

where C_{rand} and L_{rand} are the mean directed clustering coefficient and characteristic path length of a 100 matched random complex network.

Local Information (LI) Measure: The LI measures the amount of information passing through each node in a weighted directed FBN. LI is the difference between outgoing and incoming information of a particular node [406]. Information coming into a node i can be calculated as:

$$k_i^{in} = \sum_{j \epsilon V} w_{ij} \qquad Eq(3.32)$$

The outgoing information is represented as:

$$k_i^{out} = \sum_{j \epsilon V} w_{ji} \qquad Eq(3.33)$$

where w_{ij} is the weight of the edge from node i to node j. Now, the LI of node i can be calculated as:

$$LI[i] = k_i^{out} - k_i^{in} = \sum_{j \epsilon V} w_{ji} - \sum_{j \epsilon V} w_{ij} \qquad Eq(3.34)$$

The value of LI ranges from $-\infty$ to ∞ and can be used to identify the source and sink node in a directed network. A node with positive LI value emits more information than it receives and vice versa. The emitter node is called the source node, and the receiver node is called the sink node. Thus, LI value can help find the total information flow along with its direction.

3.5.5 Analysis of FBNs by Complex Network Metrics

The use of small-world properties of complex network metrics to understand brain networks has rapidly gained interest. The small-world properties have been found in games, control systems, and neural networks. **The results from various studies reviewed in this chapter indicate that the FBNs constructed from neuroimaging data such as EEG, MEG, and fMRI have demonstrated small-world properties [407, 408].** A detailed analysis of the use of graph theory to understand cognitive activities was done by Bullmore and Sporns [38] and they proposed a systematical method to construct FBN from raw EEG data. In another review by Rubinov and

Sporns, applications of complex network measures were reviewed along with the comparison of structural and functional brain networks. They also provide an open-access toolbox for MATLAB to calculate the complex network measures.

Langer et al. [409] investigated the relationship between the intelligence and functional brain networks constructed from the EEG data and found a strong correlation between intelligence, clustering coefficient, and characteristic path length. The results suggested that distortion in such behavior may be relevant to the diagnosis of psychological disorders such as schizophrenia, Alzheimer's disease, and ADHD. Rubinov et al. [410] used weighted FBNs constructed from EEG data to see the differences in clinical and healthy samples during resting state and found significant differences in clustering coefficient, characteristic path length, and degree centrality. The same kind of significant differences was found by Jalili and Knyazeva [411]. In another study to examine the Alzheimer's Disease and other dementia from EEG data found a significant loss of small-world network properties in alpha, beta, and gamma bands [412]. ADHD and depression analysis also showed significant variations in small-world properties when compared to healthy persons [413, 414].

In another application to study music perception, Fallani et al. [375] found that simple foot movement can alter connectivity patterns dramatically. By using graph theoretical measure, Wu et al. [415] showed that during music perception, the clustering coefficient increased, and the characteristic path length decreased. They further demonstrated that the small-world network properties related to the music perception were not sound. Transfer entropy was used to construct FBNs from EEG data [395]. Three different states were evaluated: resting, driving, and driving with audio distraction. **The results showed significant differences in connectivity density, motif, clustering coefficient and degree distribution across all three states.** In a web search-based task analysis through FBN, maximum connectivity was observed in query formulation task [406].

3.6 Conclusion

The chapter has started with the concept of cognition and discussed the conscious and non-conscious processes. The concept of memory has been defined, and a brief introduction of different types of memory has been presented. The task at hand defines which type of memory a subject chooses to use. The main part of cognition such as attention, decision making, and problem-solving have been presented briefly. The pros and cons of various theories of human cognition have been discussed, and a number of models of cognitive processing have been analyzed, such as the human processing model, multiple resource theory and cognitive load theory.

In this chapter, we also discussed the literature related to psycho-physiological analysis and functional brain networks. Concept of psycho-physiology, the methods used in the literature to record psycho-physiological behavior, and the analysis have been presented. The relationships of psycho-physiological signals such as EEG, GSR, and ECG. with the human's emotional and cognitive states have been reviewed in detail. In the later sections, the concept of functional brain network has been introduced and the linear and non-linear connectivity measures used to construct FBNs have been presented.The importance of graph theory-based measure used in the literature to understand the brain dynamics and cognitive activity has been discussed. The key findings from the literature review are mentioned below:

- The interaction time increases significantly when textual assistance is provided, in contrast to no assistance.
- Auditory and visual stimuli are the best ways to elicit emotions in a controlled experimental setting.
- Violent games increase cardiovascular activity compared to non-violent games.
- Psycho-physiological measures show a strong correlation with the self-reported data.

- An increase in β and γ-bands of EEG signals were observed during high-intensity events.
- A decrease in stress level was found while interacting with a social robot.
- In a mobile application evaluation task, an increase in skin response was observed when the user failed the task.
- Psycho-physiological measure has the capability to mine the underlying fact that cannot be found using traditional methods.
- Ill-designed web pages increase the stress level of the user.
- Virtual reality simulations can be used to study brain responses and stress levels.
- Multimedia presentations such as video and image elicit positive emotions more than text presentation, which induces a higher cognitive load.
- Various studies showed that FBNs constructed from neuroimaging data demonstrated small world properties.
- In music perception, drastic change in FBN connectivity patterns was observed with simple foot movement.
- Significant differences were seen in the graph theory-based measures such as connectivity density, motif count, clustering coefficient in tasks with different levels of cognition.
- Non-linear classifier such as Granger causality and transfer entropy showed best results in FBN analysis.

From the analysis, we concluded that the non-linear connectivity measures such as TE and GC, are better equipped to record the causal behavior of highly non-linear EEG signals.

Chapter 4

A Multi-Modal Interface System Design (MMIS), Development & Evaluation

This chapter explains the development of a multi-modal interface system that allows the user to design 3D objects in AutoCAD using speech and gesture inputs. The initial development and evaluation have been published in 9th ACM International Conference on Computer and Automation Engineering (ICCAE 2017) held in Sydney, Australia, in a paper titled " The Usability of Speech and/or Gestures in Multi-Modal Interface Systems" [42]. The updated version of the MMIS published in the proceeding of the 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), in a paper titled "Qualitative analysis of a multimodal interface system using speech/gesture" [43].

- Alibay, Farzana, Manolya Kavakli, Jean-Rémy Chardonnet, and Muhammad Zeeshan Baig. "The usability of speech and/or gestures in multi-modal interface systems." *In Proceedings of the 9th International Conference on Computer and Automation Engineering*, pp. 73-77. ACM, 2017.

- Baig, Muhammad Zeeshan, and Manolya Kavakli. "Qualitative analysis of a multimodal interface system using speech/gesture." *In 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pp. 2811-2816. IEEE, 2018.

4.1 Introduction

In the last decades, many efforts have been made to improve the performance of uni-modal and multi-modal interpreters. One of the main problems in the field of MMIS is to develop systems that utilize human behavior and language to interact with computers. Speech input has been extensively used in smartphones, especially for developing commercial products. Another popular input mode is gestures, inspiring many researchers to develop gesture recognition systems and algorithms for human-computer interaction with practical applications [60]. There is some evidence suggesting that an MMIS not only improves handling and reliability of the system but also task completion rates compared to uni-modal systems [61]. However, the need for a multi-modal interface instead of a single input interface is relatively less explored. In this chapter, we describe the development of an MMIS.

A typical MMIS design consists of a recognition system that translates human tasks into recognizable computer signals. Once the human input has been identified, the next step is to interpret the inputs and aggregate them to achieve the desired output. Most examples in the literature use speech and pen input in MMIS design [142]. Some recent applications have also utilized gesture input combining it with speech to draw and compare digital sketches to hand-drawn sketches [62]. Most of these systems have used Kinect and Leap motion to recognize gesture input. In these examples, speech provided an extra dimension for information required to interact with the computer in cases such as coloring or rotating the object [67].

While combining two input sets is beneficial for some applications, it may not be so beneficial or preferable in some others. For example, in modeling software, many complex words are used to draw a 3D object. The users must be familiar with the vocabulary and have to learn how to navigate in the 3D space. This research investigates the effectiveness of using simple words and gestures to design and navigate in the 3D space by combining speech and gesture inputs to perform design tasks and to facilitate the design process. We focus on how easy it is to use speech and gesture recognition systems instead of a keyboard and mouse, and what the ideal type of communication channel is for designer-computer interaction.

The goal of this chapter is to develop an MMIS which uses speech and gesture inputs to model objects in 3D. The research questions targeted in this chapter are as follows:

RQ 1.3 Is it possible to develop a multi-modal 3D object manipulation system xDe-SIGN v2 using speech and gestures?

RQ 1.4 What are the limitations of using speech and gestures in MMIS?

To answer the above-mentioned research questions, we investigate the following research problems:

1 The effectiveness of using simple words and gestures to design or navigate in the 3D space.

2 The combination of speech and gesture inputs to perform design tasks and facilitate the design process.

3 The easiness for a user to use speech and gesture recognition systems instead of a keyboard and mouse.

4 The ideal type of communication channel for designer-computer interaction.

However, there are a lot of limitations that degrade the performance of the system. The main reason for this is the complexity of the vocabulary used to draw a 3D object.

To create a simple 3D object is difficult for even a skillful user (e.g. a competent user of CAD) using speech and gesture modes of input. The system must be able to accommodate the communication mode desired by the user and adapt to the user.

The research aims to analyze users 3D modeling experience, with a multi-modal interface to create a 3D object. The system also incorporates help throughout the drawing process and identifies simple words and gestures to accomplish a range of (simple to complex) modeling tasks.

4.2 Research Methodology

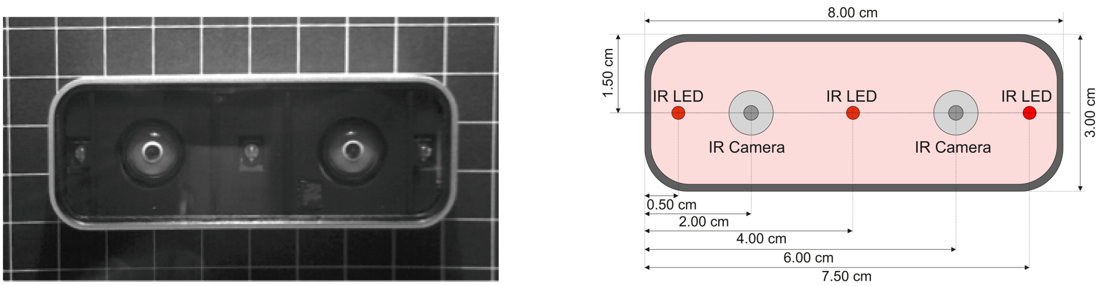
In this section, we will describe the system design and architecture. As a design concept, we developed a model to convert speech and gesture actions into commands given in AutoCAD.

4.2.1 System Specifications

We have used an Intel Core i7 desktop PC, with a Microsoft Windows 10 operating system. For gesture recognition, we have used a Leap Motion sensor, instead of Kinect, since our pilot experiments showed that Kinect 1.0 doesn't allow recognition of users' fingers [416]. Therefore, Leap Motion and its API have been chosen for gesture input, since Leap Motion offers facilities for finger recognition [417]. For speech recognition and synthesis, we have used a typical microphone and the Microsoft Speech Recognition API. We have chosen to use the 3D modeling software, AutoCAD 2017 for the users to design an object. To create an AutoCAD plugin, ObjectARX 2017 SDK was installed [418]. For the implementation of the system, we used C# with the Microsoft Visual Studio 2015 Environment.

Leap motion controller

The Leap Motion controller is a device that translates hand movements/gestures into computer commands. The device tracks gesture and position with high accuracy (usually sub-millimeter). The device uses realistic 3D infrared sensors for capturing the image and detecting hand and fingers in the image [416]. The Leap Motion controller API (Application Programmer Interface) gives the positions in Cartesian space of finger, hands, etc., which are relative to leap motion center point, as shown in Fig. 4.1. The device consists of three Infrared (IR) light emitters and two IR cameras making it a stereo vision based optical tracking system.



(a) (b)

Figure 4.1: A leap motion controller (a) Real view of leap motion sensors (b) Semantic view of leap motion sensors

Microsoft Speech Recognition API

Microsoft has developed the Speech API (SAPI) since 1993 and has continued to develop the powerful speech API. Microsoft has used context dependent deep neural network hidden Markov model (CD-DNN-HMM) to improve the speech recognition engine [93]. Recently Microsoft announced that they had reached human parity in conversational speech recognition and called it an "Historic Achievement" [93]. In this thesis, we have used C# .net framework for speech recognition module. To use the speech recognition API, we just need a microphone and perform the following steps to set up the speech recognition engine:

- Initialize the speech recognizer
- Set the input for speech recognition
- Create a grammar for speech recognition
- Load the grammar into the speech recognizer
- Register for speech recognition event notification
- Create speech recognition event handler
- Start the recognition process

The essential part is to define the grammar for speech recognition. There are two common ways to build grammar. First, to use the Speech Recognition Grammar Specification (SRGS) defined EXtensible Markup Language (XML) format. The second is to use *Microsoft.Speech.Recognition.SrgsGrammar* directly to generate grammar. The structure to define grammar is as follows:

- Define the vocabulary to recognize sentence
- Build the grammar structure
  - In vocabulary building, we define all the scenarios of a sentence we want to recognize. It could be done in one or multiple grammar builders.
  - We need to define the structure of the sentence. For example, if we want to recognize the sentence "The square is blue" and the color could be different then we need to define a variable to store all the colors we want to recognize and define the sentence structure as: The sentence starts with "The square" followed by verb "is" and then the variable that contains the colors.
- The API will recognize and convert it into text based on the defined rules.

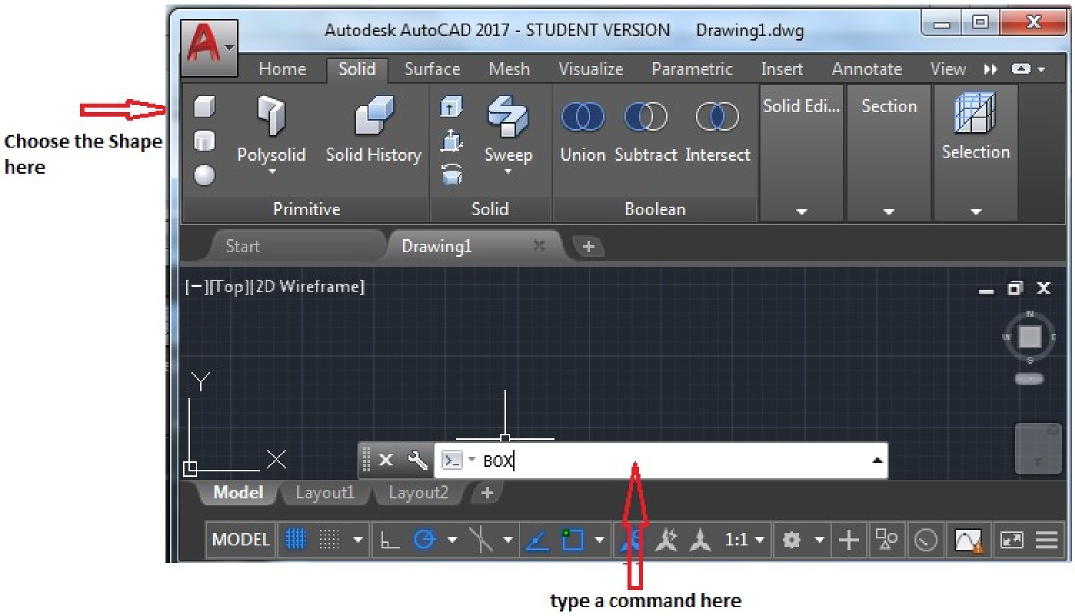


Figure 4.2: AutoCAD 2017 interface

AutoCAD

AutoCAD is a Computer-aided Design (CAD) tool developed by Autodesk. AutoCAD is used by many professionals, including architects, project managers, construction professionals, and engineers. AutoCAD is mostly used to create, draft and edit 2D geometries and 3D models with solids, surfaces and mesh objects [418]. Fig. 4.2 shows a snapshot of the AutoCAD interface. In this thesis, we have used AutoCAD for a 3D modeling task, and ObjectARX 2017 SDK was used to load our program into AutoCAD to draw using speech and gestures.

ObjecARX:

ObjectARX is an API for using run-time extensions in AutoCAD. It consists of C++ headers and libraries that are used to build the Dynamic Link Library (dll) files that can be loaded into AutoCAD. ObjectARX API allows us to use all the functionalities of AutoCAD to use in a customized program. In this thesis, we used ObjectARX 2017 to write the AutoCAD command module that uses speech and gesture decoded output and executes the corresponding AutoCAD command.

4.2.2 Design Concept

For experimental purposes, we have identified a classic chair example to draw and manipulate using multi-modal input. We have analyzed the necessary processes for this design concept. Fig. 4.3 shows an example of a 3D chair.

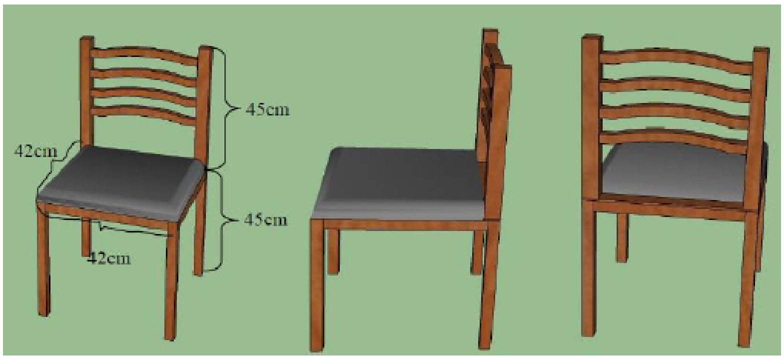


Figure 4.3: A sample 3D chair design

Manipulation and Object Identification

The classical manipulation processes to draw an object involve functions such as select, move, rotate, delete, copy, and scale. We have defined possible actions using speech and gesture inputs to apply the above-mentioned manipulation functions. For example, to rotate an object, the user has to select it first. The object can be selected using speech or gesture. To select an object using gestures, the user needs to navigate the cursor to it and perform a clicking gesture. To perform the same actions using speech and gestures, first, the user needs to navigate the cursor to the object and then articulate the keyword "select" to select the object. Once the object is selected, to rotate the object with a gesture, the user needs to hang on to the clicking gesture and rotate it with the hand position. If the user wants to perform rotation with speech, the keyword "rotate to" followed by the direction of rotation, which should be 90, 45 or 180 degrees, is used. The same set of AutoCAD commands have been used for all other manipulation actions: first, select the object and then use keywords to manipulate it. In summary, to perform a task using both gestures and speech, the process is much more complicated.

3D modeling process and object manipulation

There are two approaches to 3D modeling using either speech or gesture. Using xDe-SIGN, it is possible to mix the order of these two processes in a multimodal 3D modeling task. To draw a chair,

1 We need to draw shapes in AutoCAD such as rectangle, cylinder, and arc.

2 We need to have the ability to manipulate and edit the objects, for example, round the shapes and give some height and thickness to a surface.

3 Finally, we also incorporate the functionality of applying texture, material, and color.

Case 1: Speech For example, if we need to draw a box or cylinder using speech.

1 First we need to **say**, "**I want to draw a box**"; the system will look for the word "**Draw**" in the speech and find the shape, which in this case is a box.

2 After the object has been selected, the next step is to **specify the position**, which can be defined using the command "**the position is x, y, z**", where x, y, z are coordinates in 3D.

3 The third step is to give the object size or dimensions; to achieve this task the user needs to say "**the size is x, y**", where x and y are the lengths and the width. We also have to mention the height of the object by saying "**the height is z**", where z is the height of the specified object. For a circular object, the user needs to mention the radius instead of the size.

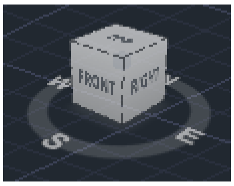


Figure 4.4: The orbit in AutoCAD environment

4 To **assign a color** or **material** to the object, after selecting the object, a speech command "**the material or color is**" can be used. In this project, only wooden material and grey color can be assigned to an object.

Case 2: Gesture If the user wants to perform the same tasks using gestures, all they need to do is:

1 **Use the hands** to **locate position**.

2 **Click on the specific icon** to draw a shape, using the **clicking gesture**.

3 With the help of **click and hold function**, **the size and height of the object** should be adjusted.

4 The material and color of the object can be assigned by using the hand to locate the **menu options** and perform the **clicking gesture** to open the menu dialog box.

Camera Manipulation Usually, in the modeling software, there are **two possible ways** to **manipulate the camera view**:

1 Using a mouse

2 Using the orbit

The orbit is the easiest way to move the camera by clicking directly on the cube (top, right, left, back, down, front or the corner right/back or right/front), as shown in Fig. 4.4.

Case 1: With Speech

1 We can move the camera using classical directions combined with speech commands such as **'move the camera vertically and horizontally'** and **'zoom in and out'**.

2 We also **orientate** the camera by specifying the number of degrees and the direction, stating that **'orientate the camera to 45 degrees on the right'**. Using speech, if no number is specified with direction, then the default value is applied (1 degree, or 1 cm).

Case 2: Using Gestures

1 First, the **camera manipulation option** needs to be activated using a **closed left hand** followed by an **open left hand**.

2 The **right hand** can be use to move the camera **horizontally and vertically**.

3 To **rotate the camera angle**, the user needs to **rotate the right hand** in the corresponding direction.

4 To **deactivate the camera navigation** option, a **closed hand followed by an open hand gesture** should be performed.

Table 4.1 shows a detailed description of words used in xDe-SIGN speech vocabulary and corresponding AutoCAD commands. For **gestures**, the user needs to utilize the clicking gesture on the icon or a tool to enable that command.

User support: During the experiment, we have implemented a user-assistance system. The system also aids while performing an action. To enable help, the following steps need to be performed:

1 **User enables the assistance** by saying, "Help me, please".

2 This **instantiates** the help sequence.

Table 4.1: Speech vocabulary and corresponding AutoCAD commands in xDe-SIGN

Speech	**AutoCAD commands**
Box, rectangle, square, bars, layer	BOX
Cylinder, tube	CYLINDER
Cone	CONE
Wedge	WEDGE
Sphere	SPHERE
Torus, donut	TORUS
Arc	ARC 3 points
Extrude	EXTRUDE
Fillet, FilletEdge, round	FILLETEDGE
Thicken	THICKEN
Move, Displacement	MOVE
Copy, duplicate, clone	COPY
Remove, Delete	DELETE
Scale	SCALE
Rotation, rotate	ROTATE
Undo	UNDO
Finish	ENTER (to finish an action)
Select all	SELECTALL
Select last	SELECTLAST

3 Offers a way to perform an action.

For example, if the user chooses to draw a box, the system will ask them to choose the position. Once the user chooses the position either by speech or gesture, the system asks them to choose the size and height. The system **anticipates** the next step for the current action and **guides the user** to perform it. The implementation of these concepts will be discussed in the next section.

4.3 Implementation of MMIS (xDe-SIGN v1)

To implement the design concept, we have created a dll plugin for AutoCAD containing 4 main classes: *MySpeech, LeapListener, MyMain,* and *MyDrawAutocad*. The *MySpeech* class starts the speech recognition function and sends an event when a speech is recognized. *LeapListener* first initializes the gesture recognizer and defines all the gestures to be recognized and sends an event when a gesture is recognized. *MyMain* receives the speech and gesture events, interprets them and sends the right command to AutoCAD. *MyDrawAutoCad* contains all the functions to draw or to manipulate the object or the camera. Fig. 4.5 shows the C# implemented MMIS structure.

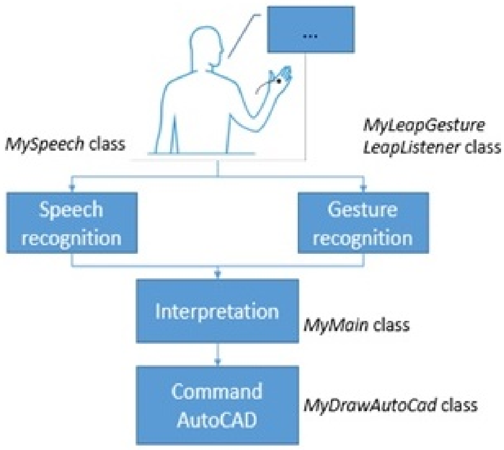


Figure 4.5: Structure of implemented MMIS

4.3.1 Speech Recognition Module (*MySpeech* class)

The speech recognition block starts with the initialization of the Microsoft speech recognition and synthesizer API. Once the initialization is done, the next step is to generate grammar for the speech recognition engine. For grammar building, almost all the main functionalities of AutoCAD mentioned in Table 4.1 have been incorporated in the grammar along with a simple sentence structure. After the successful initialization of the speech recognition API and the detection of the audio input device, the speech recognition process starts. When a speech is detected with a reasonable confidence level, the event is sent to the main block for further interpretation. To utilize the above-mentioned functionalities, the following *dll* need to be included in the program.

- *using System.Speech.Synthesis*: to activate speechSynthesizer that gives information to the user by speech. A method speak(MySpeak) has been created for this purpose.
- *using System.Globalization*: to define language of recognition such as US English (CultureInfo ci = new CultureInfo("en-us");
- *using System.Speech.Recognition*: to use SpeechRecognitionEngine Class for accessing and managing speech recognition engine.

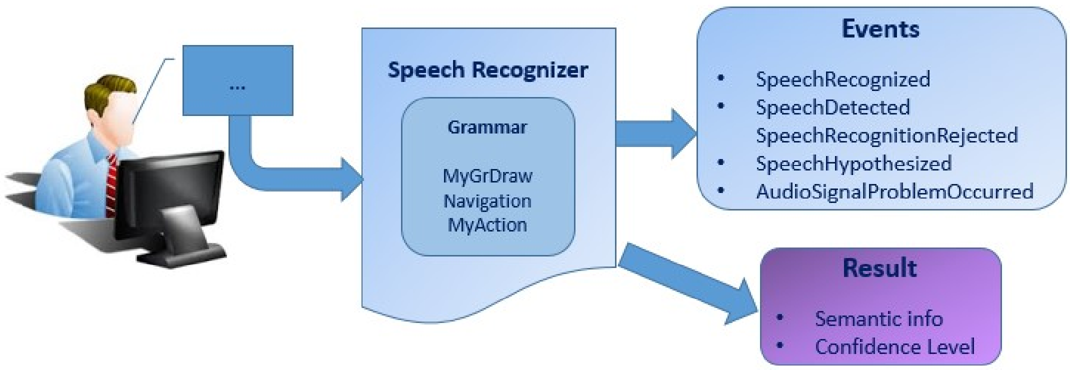


Figure 4.6: Speech recognition module block diagram

Fig. 4.6 shows a block structure of the speech recognition process with corresponding events. The speech recognition module has some limitations that must be considered to effectively translate speech into text, such as noise, syllable lengths, and clarity of speech by the user.

4.3.2 Gesture recognition module (*MyLeapGesture* and *LeapListener* classes)

In this project, **the right hand has been used to control the mouse cursor**. In xDe-SIGN v1 [42], we used the **fist gesture**; when the right hand is closed and then opened, to simulate the click of the left mouse button. Later, we replaced the fist gesture with the **pointing gesture** in xDe-SIGN v2 [43]. The **left hand** has been used to **manage the camera view**.

An event will be sent when the system recognizes:

- Horizontal or vertical hand movement
- Rotation of the hand
- Open or close hand

The opening and closing of the left hand have been used to activate or deactivate the movement of the camera view in AutoCAD. The leap motion sensor and its SDK have been used to recognize gestures. The leap motion sensor can recognize gestures, identify hand, number of fingers, grasping strength, velocity, and direction of movement along with yaw, roll, and pitch. The velocity and movement are manipulated to control the cursor in the X and Y directions. The Main block receives the speech and gesture events and interprets them to apply correct AutoCAD commands.

Two classes (*MyLeapGesture* and *LeapListener*) have been created to acheive the above-mentioned functionalities. *MyLeapGesture* class sent an event to the main class and have all the data required for gesture recognition such as:

- Action : for gesture recognition
- Hand: to identify which hand is detected
- NBfingers: the number of fingers detected
- Strength: the grasping strength
- Directions: define the direction (Left, Right, Up, Down, etc.)
- Position: give hand's current position
- Velocity: give hand's velocity

- Pitch: give the angle of the orientation
- Appx: give the X position of the screen.
- AppY: give the Y position of the screen.

The *LeapListener* is used to access frames by creating an instance of the controller class. These frames are used to generate tracking data and configuration information. The controller class is the main interface to the Leap Motion Controller.

4.3.3 AutoCAD Commands Module (*MyDrawAutocad* class)

In AutoCAD, a user can use speech and gesture inputs to execute a command. For speech, the user needs to speak the desired command, such as "I want to draw a box". With gestures, the user is required to navigate the cursor towards the desired icon and make the clicking gesture. To load the plugin in AutoCAD, "*netload*" functionality of AutoCAD has been utilized which allow the loading of the "*.dll*" file generated by the C# program. Almost all the main functions of AutoCAD have been used such as shapes, color, material, copy, rotate, delete and camera manipulation.

The AutoCAD application arranges the objects in a hierarchical structure, and AutoCAD .net API allow us to access these object and its properties directly. We can create new objects, select a previously created object, and change the properties of the object such as position, size, and orientation. The following ***dll*** needs to be included in the project to use the functionalities mentioned above.

- *using Autodesk.AutoCAD.ApplicationServices* : to control application window and create applications.
- *using Autodesk.AutoCAD.DatabaseServices*: to read or write an object in the database

- *using Autodesk.AutoCAD.EditorInput*: to access current editor for performing the AutoCAD commands.

- *using Autodesk.AutoCAD.Geometry*: to apply geometry designs using, for example, the displacement and rotation.

In this class, we have defined the function *SendCommand(MyCommand)* to send the command to AutoCAD to draw or manipulate the object. All important drawing and manipulating functions have been included in this file:

- Drawing: box, cylinder, cone, etc.

- Object manipulation: sclae, rotate, move, delete, etc.

- Camera manipulation: direction, zoom, orientation, etc.

- Object properties: change color, material, shadow mode etc.

The *MyMain* class receives the speech and gesture events and decodes them into corresponding AutoCAD commands. *Netload* loads the *dll* file generated by the C# code and *launch()* function in AutoCAD command line starts the speech and gesture recognition modules.

4.4 Qualitative Evaluation of MMIS (xDe-SIGN v1)

We evaluated the usability of the system, using both quantitative and qualitative assessments. Every user performed two experiments, first using the keyboard and mouse input, and second using speech and gesture inputs. After the experiment, the user feedback was collected through a questionnaire. This questionnaire includes questions to measure perceived user performance, fatigue, and cognitive load. The time required for completing the whole experiment is 60-90 minutes per user, depending on how familiar the user is with AutoCAD.

4.4.1 Description of Experimentation

The first experiment is to draw a chair using the keyboard and mouse in AutoCAD. For those participants who are not competent user of AutoCAD, a step-by-step guide was provided to draw the chair. Normally, they have to manipulate a camera view to be able to draw the chair correctly. For the second experiment, the participants were asked to familiarize themselves with how to manage both hands: the right hand to simulate the mouse and the left one to control the camera view. Then, they could start to draw the chair. For this experiment, written speech or gesture instructions are also provided.

4.4.2 Testing the Application

To evaluate the system performance and find out if it is easier to draw a chair using speech and gestures than using a keyboard and mouse, a log has been created to store the experimental data, which contains the history of commands. Thus, we can extract the following information:

- Number of speech commands detected
- Number of speech commands recognized
- Number of low-confidence speech commands recognized
- Number of speech commands hypothesized
- Number of speech commands rejected
- Number of audio signal issues

Eight individuals (6 men and 2 women) were selected to evaluate the xDe-SIGN v1 system, and none were competent users of AutoCAD. All the participants are computer science students and academics between the ages of 20 and 50. All are not native English speakers but speak English fluently. At the end of the experiment, the participants were expected to fill in a questionnaire for assessing the qualitative performance of the system. In the questionnaire, each question is asked twice, to compare their status when performing the task with keyboard and mouse and with gesture and speech. In the questionnaire, there are several questions related to the performance of the commands, user's fatigue, and the user perception of interaction.

4.4.3 Log File Evaluation of xDe-SIGN v1

The quantitative analysis has been performed using the data recorded in the log file of each set. Half of the users completed the task in 30 minutes and the other half took 45 minutes. No users finished drawing the chair by using gesture and speech because they gave up. Those who completed in 45 minutes drew with high precision, and the other group drew the chair without obeying the guidelines.

With the log file, we were able to extract information about speech recognition. Fig. 4.7 illustrates different audio signal issues identified during the experiment. The x-axis represents the issues with signal and the y-axis represents the total number of issues detected. 88% of the audio signal issues originated from the signal being too soft - that means that a soft voice is caused by too much attenuation on the signal. 9% of the audio signal issues originated from the signal being too noisy. The rest of the audio signal issues originated from being either too loud, too slow or too fast.

Fig. 4.8 shows the comparison between recognized, rejected, and hypothesized words. Almost 77% of words were hypothesized (detected with low certainty), 14% were rejected, and only 7% of words were accepted. The main problem in speech recognition was the hesitation in participants' voice while speaking. The system expected a complete sentence, but the participant expressed only a portion of the sentence, such as a number to define the size. Therefore, the system misunderstood the words and even recognized the word "zoom", when the user didn't say anything.

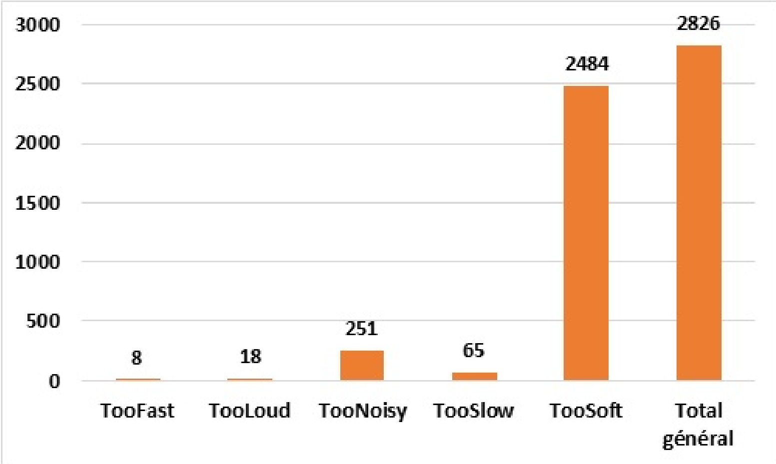


Figure 4.7: Audio signal issues during speech recognition in xDe-SIGN v1

Fig. 4.9 and Fig. 4.10 show the percentage of sentences recognized and rejected for

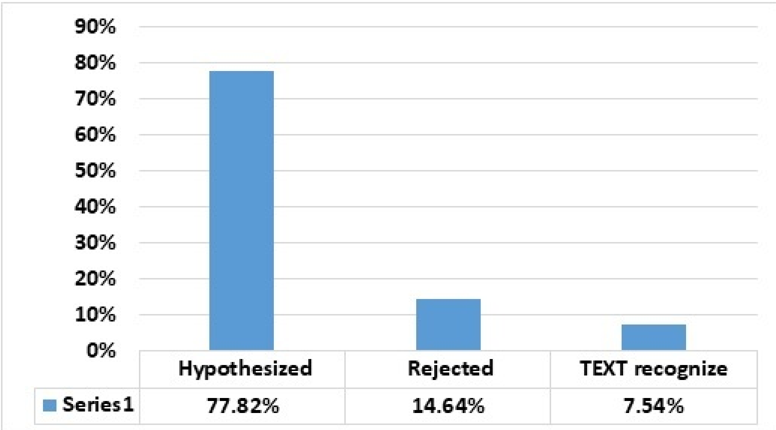


Figure 4.8: Percentage of hypothesized, rejected, and recognized words

drawing or manipulation of the object. In general, the speech signals were highly hypothesized.

It has been noticed that users found it hard to use speech recognition. The system recognized the words "Draw" (27%), "Copy" (33%) and "Material"(29%) rather well, but "Size", "Height", "Depth", "Scale", "Color", "Radius" have less than 10% recognition rate, as shown in Fig. 4.9, 4.10. It was not possible to exploit the data collected in the log file for gesture recognition. However, we analyzed the video records to identify how comfortable the users were through observation. We found that almost no one was comfortable with camera manipulation; sometimes, the system recognized the

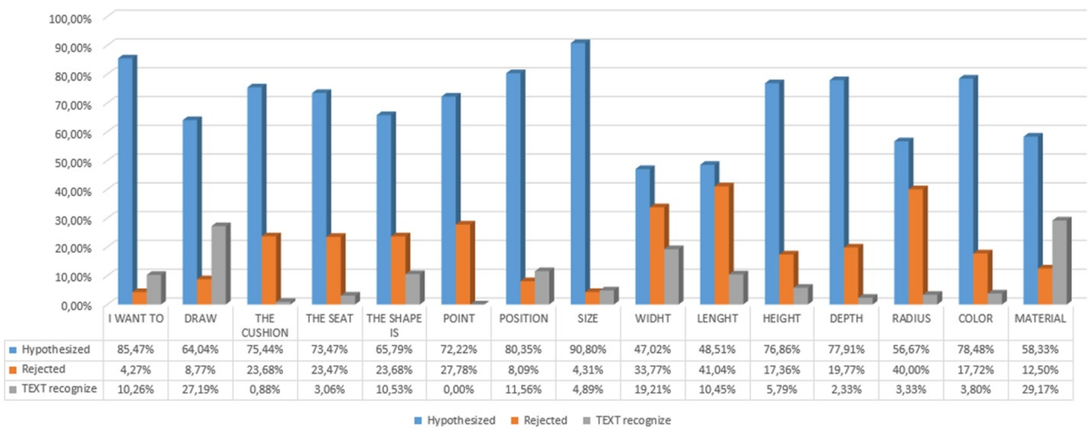


Figure 4.9: Drawing words detected in xDe-SIGN v1

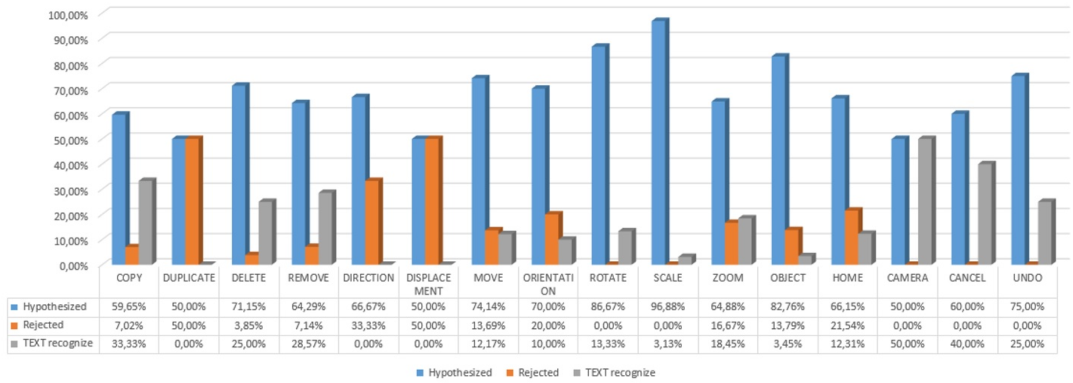


Figure 4.10: Manipulation words detected in xDe-SIGN v1

hand closed as hand open, even if the hand was partially closed. On the contrary, the system did not recognize the hand closed for the clicking gesture. It was difficult for the user to specify the size by using gestures, and for the user who is left-handed, it was challenging to control the mouse.

The questionnaires' results have been shown in 4.11. The x-axis represents various aspects of the system and the y-axis represents the average response of the user given on a scale of 1 (highly disagreed) to 7 (highly agreed). Through the questionnaires results, shown in Fig. 4.11, we concluded that, in general, it was not easy to draw the chair and manipulate the camera in AutoCAD, but it was easier to perform these actions using keyboard and mouse rather than gesture and speech. The users felt

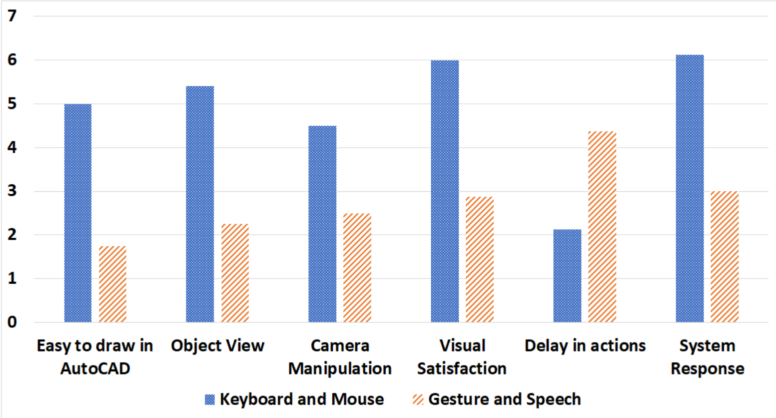


Figure 4.11: Questionnaires Results xDe-SIGN v1

more exhausted while using gesture and speech than using a keyboard and mouse. For the users, it was more natural to use the keyboard and mouse than speech and gesture. They were more satisfied with the response of the computer and more engaged. They felt more frustrated by gesture and speech input but appreciated the help and assistance during the drawing sessions. The user felt frustrated when the system did not respond when they wanted to give a specific position or size.

4.5 Improvements in the xDe-SIGN v1 (xDe-SIGN v2)

We analyzed the experimental data of xDe-SIGN v1 MMIS to define the requirements for the next iteration of system development. Based on the evaluation of MMIS xDe-SIGN v1 [42], we updated the system and performed the following improvements in MMIS xDe-SIGN v2 [43]:

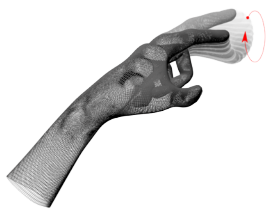
4.5.1 Modification of Clicking Gestures

The gesture used for clicking has been changed from fist/grasp to index-pointing gesture. The reason for this modification is the recognition rate for grasping gestures being low compared to pointing gestures which have a recognition rate of above 90%. We implemented five different gestures for clicking and selected the gesture for clicking with the maximum recognition rate as listed in Table 4.2. Figure 4.13a shows the pointing gesture used for the left-clicking functionality of the mouse.

Table 4.2: Recognition rate of five implemented gestures

Gestures	Recognition rate
Grasp	80%
Pointing	93%
Pinch	70%
Key tap	40%
Screen tap	45%

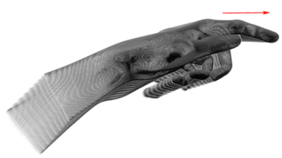
The gesture recognition algorithm has been implemented in C# using the Leap Motion SDK. The Leap Motion SDK has four pre-defined gestures: **circle, swipe, key tap** and **screen tap**, as shown in Fig. 4.12. Two gestures from leap motion SDK were



(a) A circle gesture with index finger

(b) A swipe gesture with index finger

(c) A key tap gesture with index finger



(d) A screen tap gesture with index finger

Figure 4.12: The pre-defined gestures in leap motion SDK [15]

selected as they are most related to the task. The gestures were key and screen tap. For other gestures, we have used fuzzy inference algorithm based on the position of fingers to recognize the gesture. The fuzzy inference based algorithm for gesture recognition is given below:

```
Initialization;
while Right-Hand do
    if Index finger is straight & others are not then
        Pointing Gesture;
    else
        if All fingers are bend & in a circle of 7cm radius then
            Grasp Gesture;
        else
            if Thumb and Index finger tip position in a circle of 2cm radius then
                Pinch Gesture;
            else
                No Gesture;
            end
        end
    end
end
```

Algorithm 1: Fuzzy inference algorithm to detect gestures

4.5.2 Addition of a Filter for Smoothing the Transition

It is challenging to hold the hand still in free space; any unintentional movement changes the hand position and results in changing the cursor position. To make the movement of the cursor smoother and more realistic, a **moving average filter** has been implemented. The filter generated a stable palm position and made the overall interaction as effective as mouse interaction by **taking an average of 15 palm position samples** using the equation:

$$\mathbf{Y}(t) = (1/15)\sum_{n=1}^{15}\mathbf{X}(t+n) \qquad Eq(4.1)$$

where **X**(t) is a vector that consists of palm position in x,y,z, **Y**(t) is the output of moving average filter.

4.5.3 Addition of Right-Click Functionality

The right click functionality of mouse to explore the options related to an object has been implemented in xDe-SIGN v2, which was missing in the previous version

(a) Index-pointing gesture

(b) Thumb-pointing gesture

Figure 4.13: Pointing gestures for left and right click mouse functionality

(xDe-SIGN v1). When a user uses thumb-pointing gesture, the right-click mouse function will be executed. Figure 4.13b shows the gesture used for right-clicking mouse functionality.

4.5.4 Improvement in Camera Manipulation

In the previous version (xDe-SIGN v1), we identified that the users were facing difficulty in-camera manipulation. In this iteration (xDe-SIGN v2), the difficulty level in camera manipulation has been reduced significantly by incorporating simple gestures and moving average filter. The user now manipulates the camera horizontally and vertically. For zoom, pinch gesture of left hand has been used. Pinch gesture with an index finger will zoom in and pinch gesture with a middle finger will zoom out of the current drawing view.

4.5.5 Optimization of Words to use in Speech Recognition

To improve speech recognition, the vocabulary of the system has been optimized by discarding the words that have similar pronunciation such as the word "the" and "draw". The words are found by analyzing the transcript log file generated by the speech recognition module.

4.6 Qualitative Evaluation of MMIS (xDeSIGN v2)

For evaluation, we have used the same questionnaires used in the previous section, along with video recordings and log files. The user had to perform two sets of experiments, one with a keyboard and mouse, and one with speech and gesture. The questionnaires are filled by the user at the end of each experiment. The responses are in Likert scale from 1 to 7, 1 as bad or strong disagreement to 7 as excellent or strong agreement. The experiment lasts for approximately an hour. In both experiments, the user must draw a simple 3D table using keyboard/mouse and speech/gestures input, as shown in Fig 4.14.

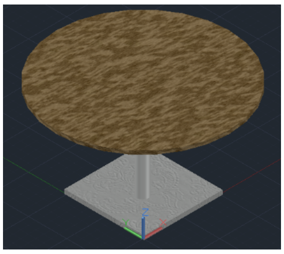


Figure 4.14: A 3D table drawn by a participant

A 15 minute tutorial was given to the participant to introduce them to the options of AutoCAD and speech/gestures input. The participants also had a document of written instructions for both sets of experiments (see appendix D). A total of 12 participants were selected for the experiment. All of them had a scientific background and spoke English fluently. Four out of 12 participants had used AutoCAD previously.

We also recorded a log file based on speech and gesture input to test the system. All the participants were able to draw the table with high precision using the keyboard and mouse. With gesture and speech, the completion rate was above 90% with only one participant who couldn't complete the drawing. The speech recognition rate also increased after simplifying the vocabulary.

Based on the evaluation results, the system has improved a lot from its previous version [42]. The task completion rate has increased to 90% as well as the responsiveness of the system. From the questionnaires, we evaluated the command performance, fatigue level, and user perception in interacting with the computer. We have divided the response into seven stages, with 1 being the lowest or very bad to 7 with the most positive and very good. We have found improvements in almost all parts of the system. It was still difficult for the user to draw in AutoCAD with speech and gestures. The user also faced some difficulties in manipulating the camera and viewing objects from different perspectives. The users did not feel any delay in system response. Questionnaires were used to record the response and to evaluate the performance of the commands. Fig. 4.15 shows the response averaged across all users.

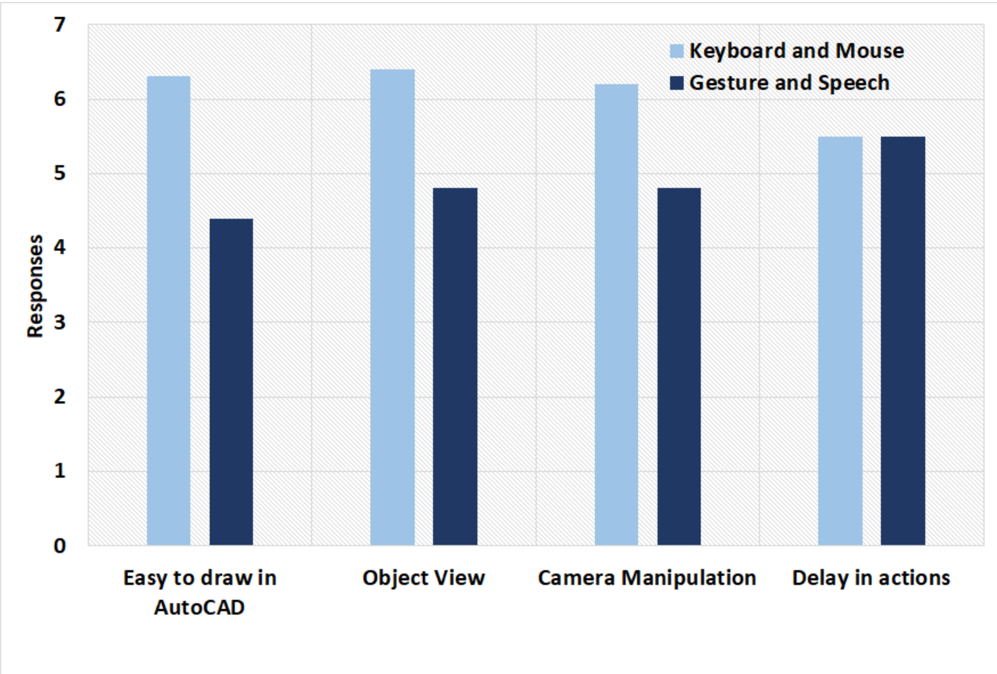


Figure 4.15: Command performance for both keyboard/mouse and speech/gesture with xDeSIGN v2

The results also show that using gestures demands more effort compared to keyboard and mouse while drawing, and fatigue level also increase in the case using gestures. The user feels rather relaxed with the addition of speech for 3D drawing. Figure 4.16 shows the results from questionnaires for the assessment of fatigue. The response is the average of all the user's questionnaire responses.

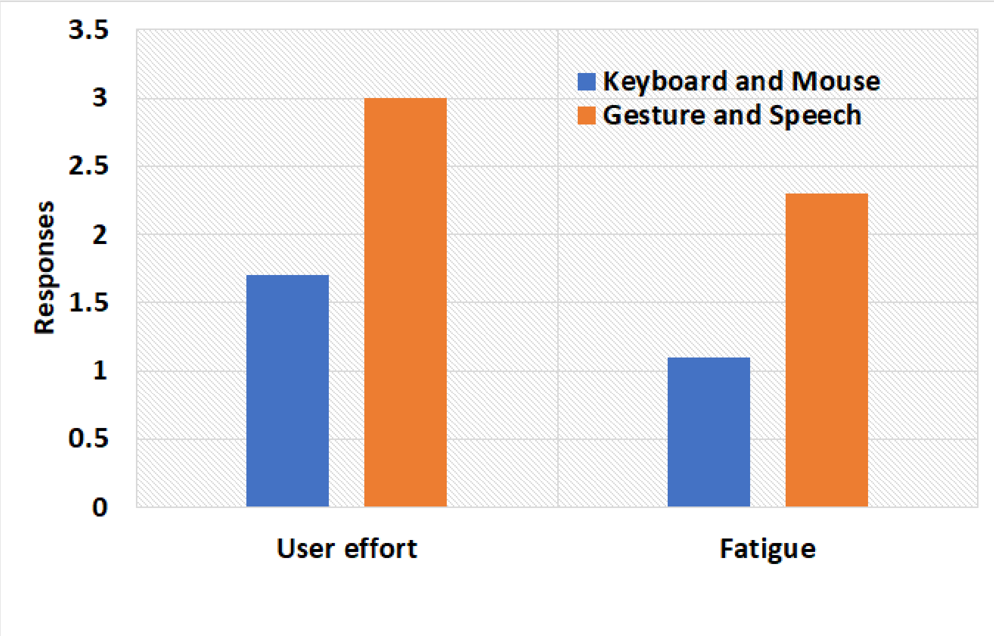


Figure 4.16: Average fatigue response for keyboard/mouse vs speech/gesture

In the case of user perception in interaction, users are more in control of the system compared to the previous version [42] but compared to keyboard/mouse input the control is still lagging and there is a lack of stability and support. The interaction with the objects has also been improved, and it is perceived as more natural compared to the previous version of the system, but there is still room for improvement. Figure 4.17 shows the results of user perception in interacting with the 3D modeling system. The user involvement and frustration levels are the same for both input modalities. Figure 4.17 shows the average response of all users.

For gestures, the main problem is that gesture recognition requires the user's hand in the air, and after some time, it becomes challenging to work with gestures without supporting the arm. Another problem is that some users find it hard to make pointing gestures continuously and when the hand is in space, the position of the hand is not stable which also changes the position of the cursor. For speech, the user finds it easy to work with speech, but the number recognition for coordinates and dimensions of the objects were difficult. The recognition rate for keywords such as drawing shapes

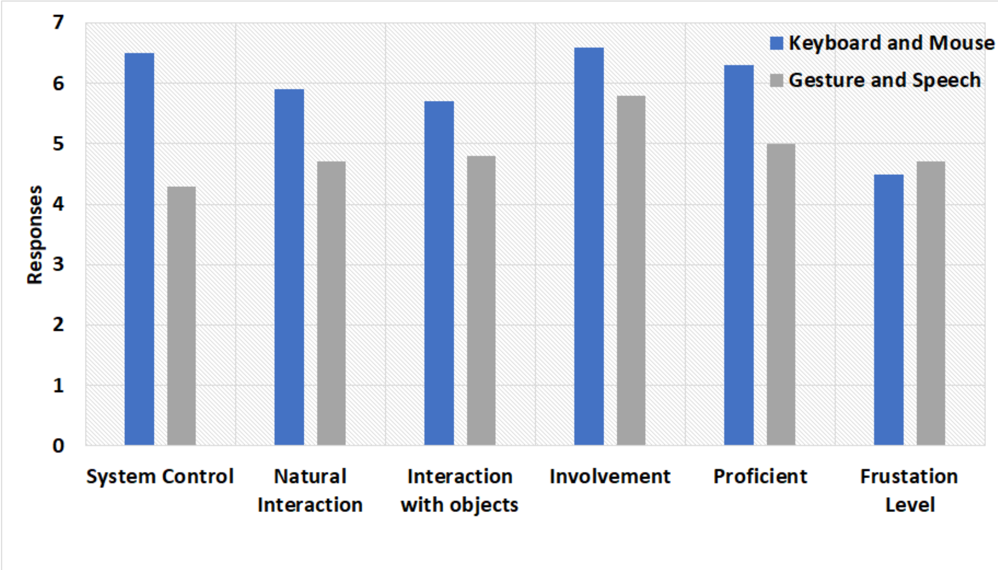


Figure 4.17: Evaluation of user perception in interacting with the multimodal interface system

and actions was around 95%, but for numerical digits, it was approximately 75%.

4.7 Conclusion

In this chapter, a multi-modal system has been presented that utilizes speech and gesture inputs to draw a 3D object in AutoCAD, using Leap motion and a microphone. In section 4.5, an iteration of the multi-modal interface system (xDe-SIGN v2) has been presented. The system is an updated version of the previous system developed by Alibay et al. (xDe-SIGN v1) [42]. With this updated version, a large ratio of participants, more than 90%, were able to carry out the tasks with appropriate precision.

After experimentation and evaluation of the xDe-SIGN v1 MMIS, we found that speech and gestures are well-coordinated in human to human communication. Our results indicate that performing a task using speech is perceived exhausting when there is no shared vocabulary between man and machine, and the usability of traditional input devices supersedes the usability of speech and gestures. Only a small ratio of participants, less than 7% in our experiments, were able to carry out the tasks with appropriate precision using xDe-SIGN v1.

Drawing with precision in 3D modeling software is more complicated than expected. The speech recognition process is exhausting when the system works slowly and does not respond appropriately. Speech recognition requires simple grammar and a quiet environment to reduce noise. Gestures seem to be more natural and less tiring to use in human-computer communication, if the users can use both hands, instead of one hand only. The system has to offer several gestures for the same action to satisfy most users. Even though the system was functional, we still noticed that it losses track of gestures from time to time. People would prefer more natural interaction such as gesture and speech if the performance of the equipment for the interaction could satisfy a standard level of operation in a reasonable amount of time and effort.

Chapter 5

Methodology: Experimentation and Instrumentation

This chapter explains different aspects of experimental design, the EEG signal acquisition procedure, and EEG pre-processing techniques. The chapter introduces the equipment used in this thesis to record the EEG signals and the software used for pre-processing and analysis.

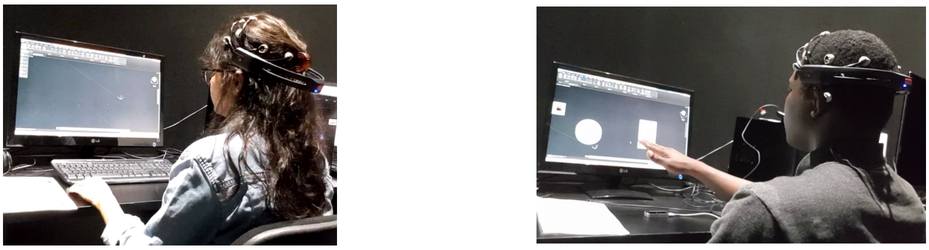
5.1 Experimental Setup

Designing the experiment to stimulate cognitive activity is a complicated process. Real-world tasks are complex and involve many underlying cognitive processes. In the contemporary research literature, it is possible to see examples of sophisticated means to simulate and stimulate cognitive activities in flight simulators [419], driving simulators [420], game-play environments [318], and design simulators [365]. The main aim of the thesis is to study the cognitive activity of both novice and competent users in 3D modeling. To study the behavior of novice and competent users, we have designed a simple experiment so that a novice can also complete the experiment without a high level of difficulty.

The experiment was to design a 3D table with three parts: a base, a pillar, and a top in AutoCAD with two different sets of inputs, i.e. keyboard, mouse and speech and gesture using the MMIS developed in Chapter 4. A total of twelve participants volunteered for the experiment. In most EEG studies, the number of participants varies from 5-20 [365, 395]. In this study, we used 12 participants. All of them were computer-science students at Macquarie University. The ages of the participants range from 21 to 30 years. Four of the participants were competent users of AutoCAD, and 8 were novices. The reason for using mixed competency is to see if there are any differences in novice and competent user's cognitive activity when they are given a totally new set of modalities to draw 3D objects. The experiment was approved (Approval no. 5201700784) by the Faculty of Science and Engineering Human Research Ethics Sub-Committee, Macquarie University. Only right-handed participants were selected as the left-handed participants were likely to have slightly different brain wiring [421]. All participants reported normal hearing and normal or corrected-to-normal vision; no participant reported any history of psychological, neurological, or psychiatric disorders. Each subject was given a tutorial of 10 minutes before the experiment. A video log was maintained for each subject, and EEG signals were also recorded.

Experimental Apparatus:

For the acquisition of EEG signals, we used an off-the-shelf research edition of the Emotiv EEG headset, which has 14 channels: frontal and front-central: AF3, AF4, F3, F4, F7, F8, FC5, FC6; temporal: T7, T8; and the occipital and occipital-parietal: O1, O2, P7, P8. The device has an internal sampling rate of 2048 Hz, which is down-sampled to 128 Hz after the cleaning of artifacts. The electrode placement is based on the international 10-20 system. The Emotiv Headset has a limited number of electrodes. There are no electrodes in the central lobe (Pz, Cz, Fz), and therefore, it is thought that the system has limited applicability in research. The manufacturer states that the signals from the neighboring electrodes are good enough to perform experiments, and researchers have proved the capability of the headset in many applications [422]. The subjects were given an open-ended task to depict the real-world setting, which results in complex cognitive processing and analysis strategies. A picture of the experimental setup is shown in Fig. 5.1. The process can be divided



(a) User using keyboard/mouse to draw 3D object **(b)** User using speech/gesture to draw 3D object

Figure 5.1: Participants using AutoCAD and wearing EEG headset in real experimental setting

into four experimental procedures:

1 Information about the experiment was given to each subject along with the consent form. After reading and signing the consent form, the experimenter gives a walk-through of the experiment and some instructions to minimize the body and head movements.

2 The EEG headset was placed on the head of the subject, and the experimenter made sure that all electrodes were in good contact with the scalp. The subjects were asked to rest for two minutes with their eyes open, with hands on their laps, and after that, the subjects were asked to start drawing using keyboard and mouse.

3 After the completion of the modeling task with keyboard and mouse, the participants were given instruction about the multi-modal interaction system.

4 After completing the task with speech and gesture inputs, subjects were asked to fill in a questionnaire about the experiment and the MMIS.

Figure 5.2: Emotiv EPOC headset [16]

5.2 EEG Data Collection

EEG data collection is an important task and needs a careful preparation of both the participants and equipment. There are various types of equipment available in the market that can be used to record EEG signals. The differences between the equipment is mostly in the number of channels/electrodes, amplifier, electrodes material, and portability. The hardware and software used in this thesis for EEG data preparation and processing are discussed in the following subsections.

5.2.1 Hardware and Software Tools

Emotiv EEG Headset

Electrodes/channels are small metallic discs that are placed on the scalp for EEG data collection. The placement of the electrodes is based on the international 10/20 system [423]. In this thesis, we have used Emotiv EPOC headset [227]. Emotiv EPOC is a multi-channel EEG system developed by Emotiv Inc. to record the EEG signals from the scalp, as shown in Fig. 5.2. The main benefit of Emotiv EPOC headset is that it is a wireless portable EEG device which allows the user certain level of flexibility in performing an experiment. The EPOC headset has 14 channels belonging to various brain regions: frontal and front-central: AF3, AF4, F3, F4, F7, F8, FC5, FC6; temporal: T7, T8; and the occipital and occipital-parietal: O1, O2, P7, P8. The electrode placement is shown in Fig. 5.3.

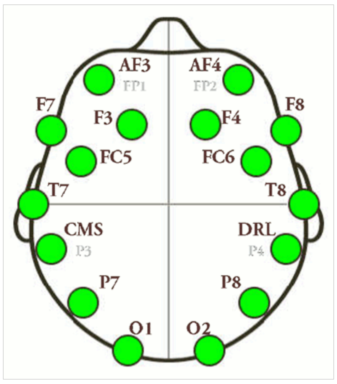


Figure 5.3: Electrode placement of Emotiv Epoch Headset

The Emotiv EPOC provides whole brain sensing with wireless connectivity through Bluetooth. The setup time for the headset is less than 5 minutes, and it has a battery life of up to 12 hours [227]. It has saline-based wet electrode sensors. The internal sampling rate of the headset is 2048 Hz, which is down-sampled to 128 Hz. The resolution is 14-bits with a step size of $0.51\mu V$.

Emotiv TestBench

Emotiv TestBench is the software (Fig. 5.4) used to record EEG signals from the EPOC headset. It allows you to see the raw data acquired from the 14 electrodes in real-time. The user can change the channel spacing and min/max amplitude. It also shows the connectivity of electrodes to the scalp. The software has the option of showing the Fast Fourier Transform (FFT) response with filtering options. The user can send an automatic marker from the serial port or send the markers manually [424]. In this thesis, Emotiv TestBench is used for data acquisition. The other required processing can be conducted in MATLAB and EEGLAB.

MATLAB

MATLAB stands for matrix laboratory; it is a multi-paradigm interactive numerical computing language developed by MathWorks used in a variety of disciplines and fields including signal and image processing, machine learning, communications, and

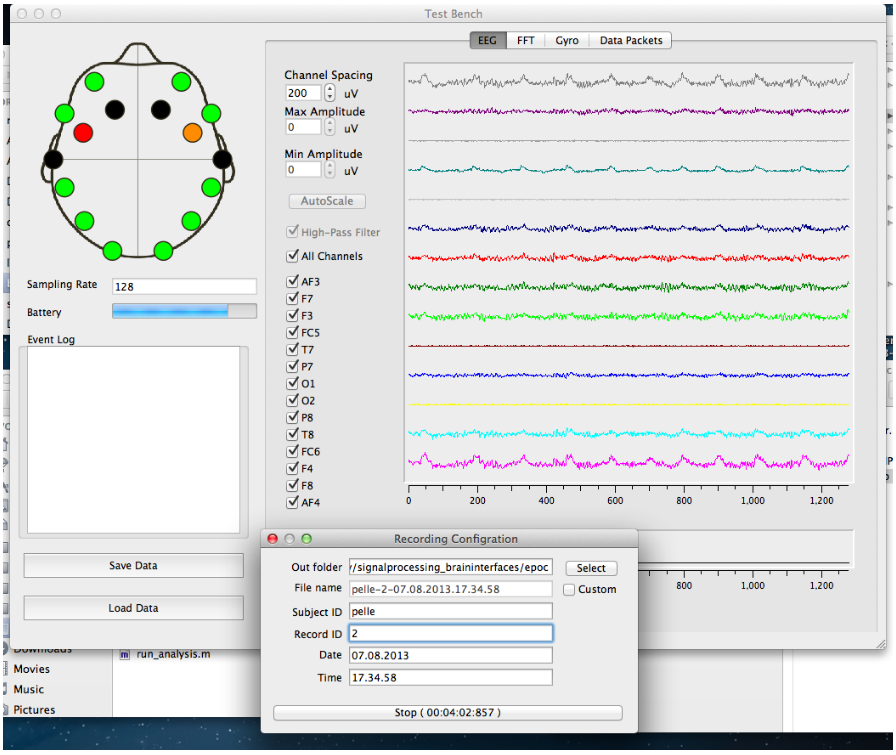


Figure 5.4: Emotiv TestBench Software

control systems. [425]. MATLAB is a matrix-based tool that allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages including C, C++, C#, Java, FORTRAN, and Python. In addition to numerical computing, it also supports symbolic computing and Simulink; a graphical user interface (GUI) for multi-domain simulation and model-based design for dynamic and embedded systems. In this thesis, MATLAB has been used for EEG signal analysis such as filtering, transformation, connectivity matrix generation and plotting of results. EEGLAB (description given below) toolbox has been added to MATLAB to extend the functionality.

EEGLAB

EEGLAB is an open source MATLAB toolbox for processing data from EEG, MEG, and other electrophysiological signals. In addition to all the basic data processing options, EEGLAB implements independent component analysis (ICA), time/frequency analysis, artifact rejection, and several modes of data visualization [426]. It allows importing data from various formats and devices. EEGLAB also provides a graphical user interface (GUI), allowing users to process EEG data and visualize the results interactively. For scripting users, EEGLAB offers a structured programming environment for storing, accessing, measuring, manipulating, and visualizing EEG data. For creative research programs and methods developers, EEGLAB offers an extensible open-source platform through which new methods can be shared with other users. **Topoplot** and **Headplot** are the two strong visualization function that plot the activity on a topographic map (looking down at the top of the head) and a semi-realistic head map respectively. In topoplot, the plots are made in 2-D circular view using interpolation on a Cartesian grid [427].

5.2.2 EEG Data Acquisition

EEG is a powerful non-invasive tool for recording the electrical activity of the brain. The portable EEG devices such as Emotiv EPOC allow participants to perform a task in a more natural and relaxed manner compared to other neuroimaging techniques such as fMRI. In this thesis, EEG data is collected from different brain regions for various states, including resting state with eyes open, other cognitive states such as 3D object modeling with keyboard and mouse and speech and gesture. EEG data were acquired at a sampling rate of 128 Hz through 14 channels (layout shown in Fig. 5.3) of Emotiv EPOC. All tasks were recorded using a camera for all participants to record the modeling actions of the participants.

5.3 EEG Signal Pre-processing

EEG signals are very prone to noise. Any signal that is not generated by the brain and present in the EEG data is called noise or artifact. These artifacts can be eye blinks, participants speaking, muscle movements, and jaw clenching. In addition to these artifacts, other external sources of noise such as electrical line interference also contaminate EEG data. Thus, the major component of EEG signal pre-processing is based on artifacts and noise removal or reduction. In artifact removal, the section contaminated data can be removed from further analysis. On the other hand, artifact reduction is a process of keeping the section of the data and reducing the artifacts.

5.3.1 Eye Blink Artifact

The most common EEG artifact is the eye blink artifact. The blinking artifact is largely observed at frontal sites, and the amplitude decreases as it moves towards posterior electrodes. The electrode FP1 and FP2 are seen to be most sensitive to eye blink artifacts due to their location immediately above the eye. An eye blink artifact lasts from 200-400 ms and occurs several times a minute. The amplitude of eye blink artifact is 10 times more than the surrounding brain activity. Fig. 5.5 shows a typical eye blink artifact.

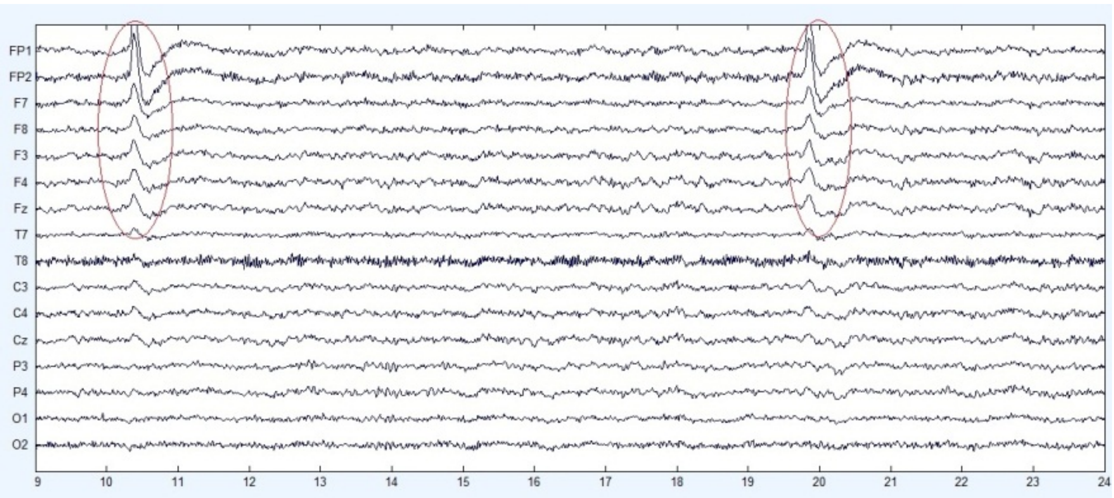


Figure 5.5: Eye blink artifact shown in red circle [17]

Removing eye blinking artifact is relatively easy for those participants who blink the eye only a few times during the experiment, the section containing the artifact can be removed with the minimal loss of data. The feasibility of this approach came into question when there are many eye blinks. In this case, the artifacts can be reduced to minimize the influence on the data. There are many algorithms to reduce the eye blink artifact such as template matching [428], principal component analysis (PCA) [429], and Independent Component Analysis (ICA) [428]. These methods can be implemented using EEGLAB, especially ICA in which multi-channel EEG data is decomposed into independent components, and eye blink component is identified. A minimum of 5-8 channels is required to perform ICA [430].

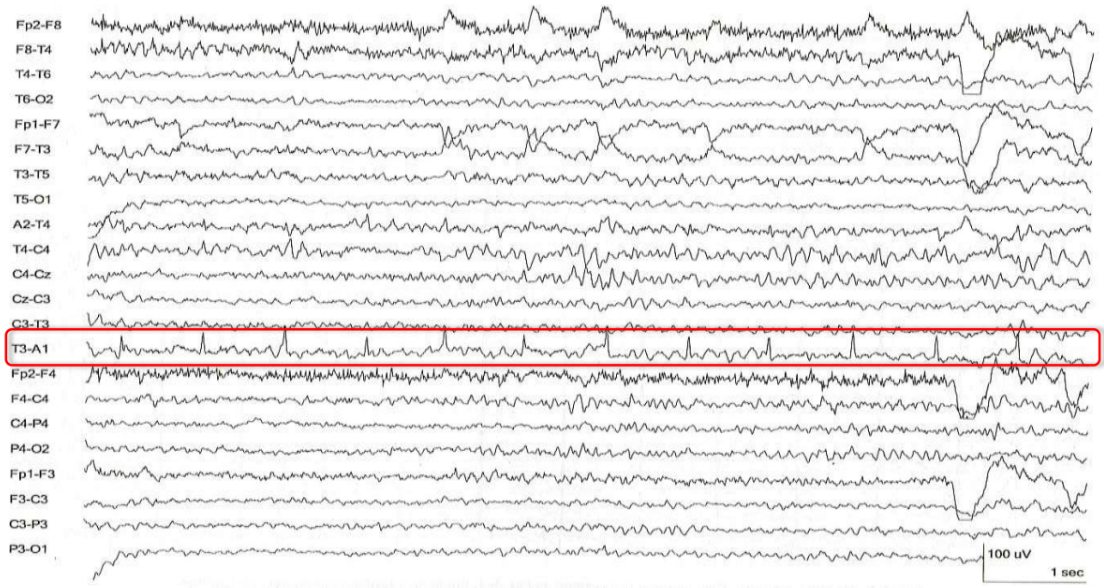


Figure 5.6: Cardiac artifact indicated by the red box [17]

5.3.2 Cardiac Artifact

Cardiac or ECG artifacts are generated with the subtle movement of head and body associated with cardiac contractions. In contrast to eye blink artifacts, cardiac artifacts relate to the field of the heart potential over the scalp. Cardiac artifacts are easily recognized by their periodic nature and coincide with ECG trace as shown in Fig. 5.6. The ECG artifacts can be reduced by adaptive filtering and Independent Component Analysis.

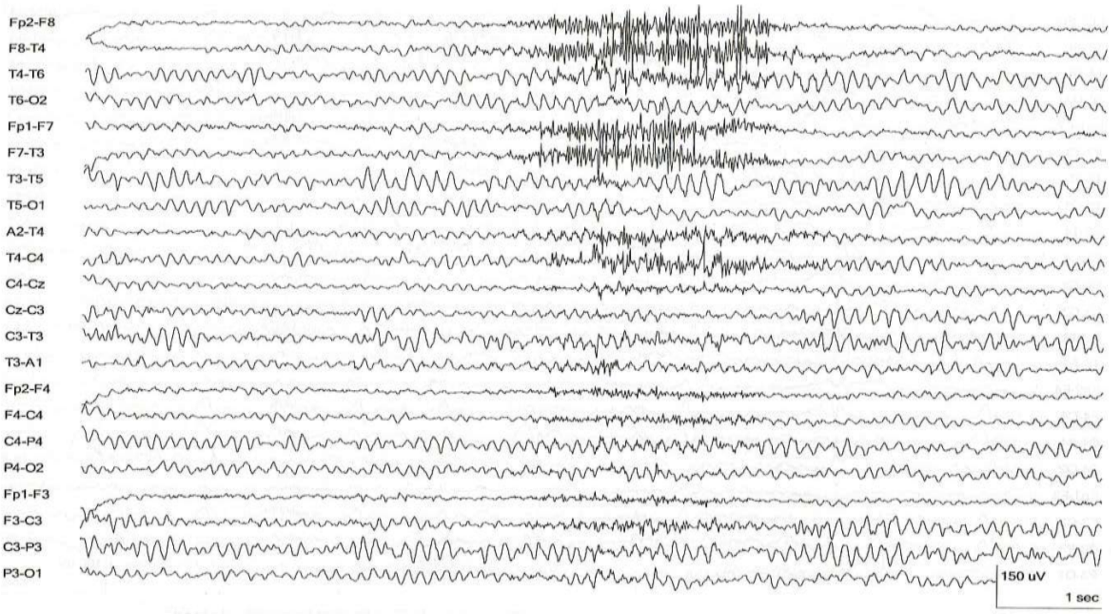


Figure 5.7: EEG signals with muscle artifact [17]

5.3.3 Muscle Artifact

Muscle artifacts are the electric field generated by muscle and through a movement effect on the electrode contacts. They are characterized by surges in high-frequency activity and are readily identified by their outlying high values relative to the local background activity. Muscle activity artifacts are particularly obvious in the gamma range > 20 Hz and cause very short potentials, as shown in Fig. 5.7. Such artifacts are common at electrode sites F3, F4, T3, T4, P3, P4 due to their location near the masseter and temporalis muscles, although the effects can be minimized by properly positioning the participant's head. Any noise related to muscle artifacts can be removed using ICA.

5.3.4 Electrical Interference

Power lines and electric equipment also cause an artifact in the EEG data. These artifacts have a distinct frequency i.e. 50 Hz (Australia) or 60 Hz (North America). Electrical interference artifacts are particularly damaging to EEG data as they over-

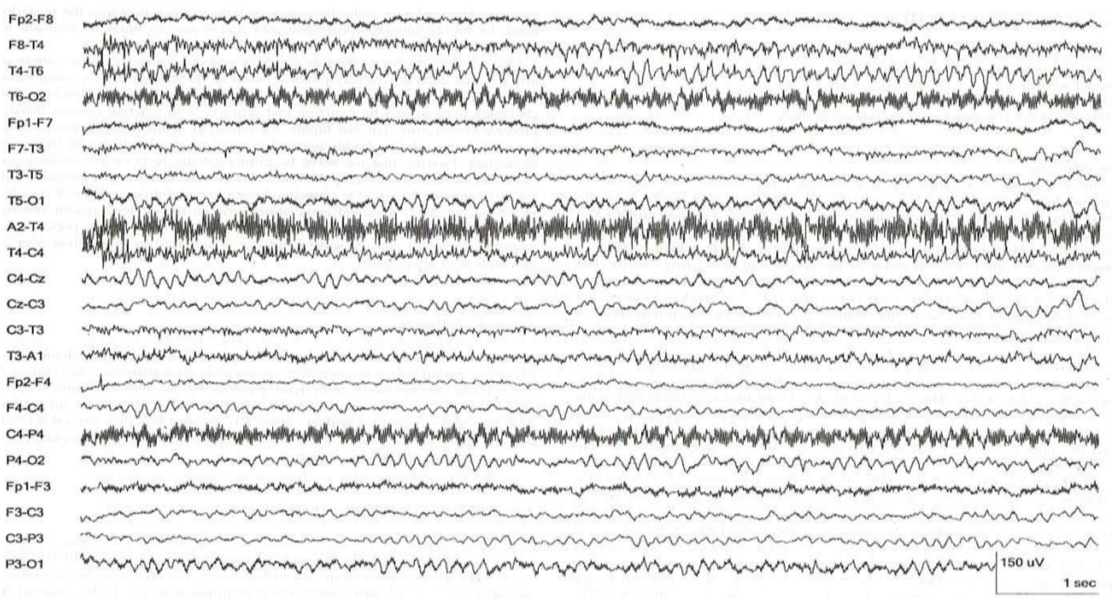


Figure 5.8: EEG signals with electrical interference [17]

whelm the underlying neural activity of interest; the typical value for true EEG ranges from 10 to 100 μV, whereas external electrical interference can range from 10 mV to 1 V (see Fig. 5.8). Such noise contamination can be avoided with properly grounded electrical equipment and 50 or 60Hz notch filter.

5.3.5 Pre-processing and Artifact Reduction in this Study

We have used MATLAB 2017b and EEGLAB to pre-process the EEG data. The baseline was removed from the EEG signal, and low-pass filtering at a cut-off frequency of 45 Hz was performed using a linear-phase FIR filter. EEG signals were then high-pass filtered at a cut-off frequency of 0.16 Hz and notch filtered at 50 and 60 Hz using a linear phase FIR filter. The order of the filter in all cases was 300.

To detect eye blinks, we calculated the ICA without the EoG electrodes. ICA decomposes EEG signals into independent components such that each component can be plotted and then identified through visual inspection. The component can then be removed to produce artifact-free EEG data. The component with the eye blink artifact is identified visually and removed manually. Other bad blocks that contain noise and corrupt data are also removed from the EEG signal. Cardiac and muscle artifacts are also removed using the ICA in EEGLAB. To further minimize the noise component and amplify the EEG activity related to the task, we performed back-to-back epoching which are averaged into one single epoch. In back-to-back epoching, the continuous EEG data is cut into the consecutive epoch of specified duration typically one or two seconds.

5.4 Conclusion

In this chapter, a description of the experimental setup was given that is used in this thesis. The EEG data collection technique has been explained along with the hardware and software used. The Emotiv EPOC headset was used to record the EEG signals. The EEG signal pre-processing was performed in MATLAB and EEGLAB. Eyeblink and muscle artifacts were removed using ICA in EEG. The electrical noise was removed with the help of a notch filter at 50 and 60 Hz. More details related to cognitive activity, pre-processing and post-processing analysis will be discussed in the following chapters.

Chapter 6

Analysis of Cognitive Activities in a Uni-modal System: using Design Coding Technique

This chapter presents the experimental findings of using design coding techniques to estimate the cognitive activity of the user in a 3D modeling application. A new method to segment EEG signals based on design coding technique is presented. The results have been published at the 25th International Conference on Neural Information Processing (ICONIP 2018) held in Siem Reap, Cambodia, in a paper titled "EEG Signal Analysis in 3D Modelling to Identify Correlations Between Task Completion in Design User's Cognitive Activities" [46].

- Baig, Muhammad Zeeshan, and Manolya Kavakli. "EEG Signal Analysis in 3D Modelling to Identify Correlations Between Task Completion in Design User's Cognitive Activities." *In International Conference on Neural Information Processing*, pp. 340-352. Springer, Cham, 2018.

6.1 Introduction

Skills and competencies are developed after learning basic techniques and practicing those techniques over time, but to define how competency is obtained through practice in a design task is quite difficult. Even expert users cannot articulate what kind of techniques are involved in performing a certain task and how they are using these techniques. Our aim is to understand the behavior of a competent designer and to use this understanding to develop the next-generation design and modeling systems to guide a novice. The current study contributes to this understanding by investigating the designer's cognitive activity through EEG signals.

3D modeling or CAD/CAM tools have a great impact on design efficiency [431], but they require a specific set of skills, training, and experience to master these tools. Competent users can capitalize on their design skills at the early stages compared to novice users, but little research has been done on analyzing the stages in the conceptual design process. One way to address this problem is to study the cognitive activities behind the designer's actions. Protocol analysis is a well known method to examine the cognitive activity of a designer, but it mainly focuses on the designer's actions and does not incorporate the mental or emotional state of the designer.

In the last few years, psycho-physiological methods have been used to analyze and understand the science behind a designer's actions. This increase in psychophysiological research is due to the widespread use of non-invasive, inexpensive and easy to use psychophysiological equipment. Researchers have used electrocardiograms (ECG) [432], galvanic skin response (GSR) [365], eye-tracking [433], gesture analysis [434], and electroencephalography (EEG) [364] to study the behavior of design protocols and processes. The most popular technique reported in the literature to analyze the design process is verbal protocol analysis, but it has some limitations [435]. These limitations are apparent when analyzing non-reportable processes such as creativity, judgment or task insights [436], so other techniques for analyzing a designer's actions must be investigated.

Due to the limitations of verbal protocol analysis, alternative techniques to study a designer's actions have been introduced including sketching [437], gesture analysis [438] and eye-tracking [433]. Some researchers use modeling tools and techniques by analyzing the activity through psycho-physiological signals [363]. In this chapter, we present a new approach to use EEG signals segmentation in the analysis of the designer's cognitive activity. The aim of this chapter is to investigate whether the content-oriented approach for analyzing a designer's activities can benefit from overlying EEG signals to understand the cognitive behavior of the designer. The research questions addressed in this chapter are:

RQ 3.1 Why do some novice users perform better than others?

- **RQ 3.1.1** What is the relationship between task completion and mental effort?
- **RQ 3.1.2** What are the factors affecting the task completion of a designer in 3D modeling?

RQ 3.2 What are the factors that affect novice users performance?

- **RQ 3.2.1** What are the relationships between alpha, beta, theta, and gamma bands activities and task completion?

These questions are addressed by the following research tasks:

1 To monitor the cognitive states of designers as they perform a certain 3D modeling task.

2 To investigate the individual modeling behavior involved in 3D object modeling and develop a methodology to understand the behavior.

3 To validate the modeling behavior through EEG signal analysis.

4 To determine if there is a correlation between the designer's performance and psychological signals.

6.2 Designers Cognitive States and Cognitive Analysis

A common empirical method used to estimate the designer's behavior is protocol analysis. Basically, it decodes the actions of the designer to uncover their thought processes. It has been used to study the designer's techniques to solve a problem by using direct-video or audio-recording through observations. The recordings are transcribed, segmented and coded to generate and validate a hypothesis and investigate certain phenomena.

Ullman et al. [439] applied protocol analysis to the mechanical design process and defined a task-episode-accumulation model based on the results of protocol analysis. In another study related to the mechanical design process, Waldron and Waldron [440] observed that experts use an opportunistic approach to quickly identify and focus on the important parts of a design. Many protocol analysis studies also use analysis of sketches to study the underlying cognitive actions. Suwa et al. [441] found that sketches illuminate ideas in the early-stage of design and support the designers marshal their thoughts in creative processes.

Protocol analysis has also been used in the literature to study novice-expert differences. Kavakli and Gero [442] found cognitive differences between novices and experts in the architectural design process. Kavakli et al. [443] also found that the cognitive actions of experts were well organized compared to those of novices, whose actions were highly concurrent. Ho [26] found that novices ignore the problem that they failed to handle, whereas experts directly approach to a goal and then work backward. Some studies also reported that experts' solutions were of high quality; they spent more time in the problem-solving phase and looked for more alternative solutions than novices [444].

Identifying cognitive states of a user is a very complex assignment and has been studied extensively in the field of psychology and cognitive science. With the advancement in the field of clinical psychology, devices such as an EEG headset, ECG and GSR have become more accessible and easier to use. With these devices, we can analyze user behavior using quantitative techniques. In this chapter, we use a coding scheme to segment the designer actions and use EEG analysis to study the user's mental states.

6.3 Methodology

6.3.1 Experimental Setup

The experimental task is to design a 3D table with three parts: a base, a pillar, and a top. The data of eight participants was selected for this experiment. All the participants selected were novices, and we used task completion time to divide the participants. The participants who completed the task quickly were called **low completion time(low-CT)** participants, and took a long time to finish the task were called **high completion time (high-CT)** participants. After cleaning the EEG signal from noise and artifacts, the mean power for the different EEG bands was calculated using Welch power spectral density. The details of the experiment are given in Chapter 5.

6.3.2 Coding Scheme

To examine the differences between the subjects's mental states and their task **completion time (CT)**, we used a coding scheme that allows us to assign codes to the cognitive actions of the designers using the video recordings. This coding scheme is an extension of the coding scheme used by Kavakli et al. [437].

6.3.3 Codes for Modeling Actions

To analyze a certain set of cognitive actions, there are two approaches: the process-oriented approach and the content-oriented approach. We have used the retrospective protocol analysis method that lies under the content-oriented approach, like the one used by Suwa and Tversky [445]. We have categorized the actions of the subjects into three groups: Physical, Perceptual and Conceptual. There is also a functional category, but for this experiment, the subjects were already given a function (i.e. "Table") therefore, this category was not used in the analysis. The perceptual and physical actions present the visual information and the conceptual and functional actions present the non-visual information. The modeling actions of each participant were coded for each cognitive segment.

Physical Actions

Physical actions are all the actions involved in drawing new objects, tracing over the sheet and copying previously drawn elements, and paying attention to previously drawn elements. We have defined three groups of physical actions: D-actions (Drawing, coping), L-actions (paying attention to previous design) and M-actions (movements on design depictions). The details of the Physical actions are given in Table 6.1.

Table 6.1: Sub-codes for D-actions, L-actions and M-actions

D-Actions	**L-actions**	**M-Actions**
Pd: Drawing depictions	Pl: Viewing (Camera Manipulation) Pl: Manipulating depictions	Pm: Moving over depictions

Perceptual Actions

The actions that are related to the visual features of the objects and spatial relations among them are known as perceptual actions (P-Actions). Perceptual actions have a further eight categories, but in this research, we have defined only three perceptual actions as shown in Table 6.2.

Table 6.2: Sub-codes for Perceptual actions (P-actions)

P-Actions related to implicit space	**P-actions related to features**
Ps: Selecting depictions	Pc: Colouring depictions Psd: Deleting depictions

Conceptual actions

Conceptual actions are those actions that are used to retrieve knowledge, previous similar cases or set up goals. In this research, we have only examined the retrieval of knowledge and represent these as C-actions. We defined three conceptual actions as shown in Table 6.3.

Table 6.3: Sub-codes for Conceptual C-actions

C-Actions
Ct: Thinking
Cr: Reading
Ci: Idle state

Table 6.4: Design actions with time stamp and codes for user C and E

User C			User E		
Time	**Action**	**Code**	**Time**	**Action**	**Code**
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
0:56	Change pillar color	Pc	0:41	Drawing Top	Pd
1:00	Thinking	Ct	0:46	Thinking	Ct
1:08	Moving	Pm	0:50	Reading	Cr
1:11	Close color option	Ps	0:55	Reading	Cr
1:15	Moving	Pm	0:59	Open color options	Pc
1:18	Open material change option	Ps	1:05	Reading	Cr
1:20	Change top	Pc	1:13	Move	Pm
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

6.4 Performance Analysis using Coding of 3D Modeling Actions

We examined the actions of the participants who completed the task in different time windows.

Finding 1: We found that the participants with a low task completion time (low-CT) (Users A, B, C) have an average action rate of 20 action/minute, which is 30% higher than for high completion time (high-CT) participants (Users D, E, F). High-CT subjects have an average action rate of 14 actions/minute.

We analyzed the video records and assigned codes to each action performed by participants. The tasks were grouped together to find the ratios between CT levels and cognitive actions. An example of design activity segmentation and codes definition has been shown in Table 6.4. Table 6.5 shows a summary of the modeling actions performed based on task completion time. The major difference is observed in physical and conceptual actions. The high-CT users performed double the conceptual actions compared to low-CT users and 1.5 times more physical actions. Table 6.6

Table 6.5: Summarized action performance comparison based on task completion time

Perceptual	**Low-CT**	**High CT**
Ps	**22**%	17%
Pc	9%	9%
Psd	0%	1%
Total	**31%**	**27%**
Physical		
Pd	**18**%	13%
Pm	**22**%	15%
Pl	**10**%	7%
Total	**51%**	**34%**
Conceptual		
Ct	14%	**18**%
Ci	2%	2%
Cr	3%	**19**%
Total	**19%**	**39%**

shows detailed information about the average time spent by the user at every design stage.

Finding 2: The users with Low-CT have spent less time performing all the three actions. High-CT users have consumed the highest time in performing conceptual actions i.e. 2.2 times more time than low-CT users.

The time spent in performing a corresponding design action has also affected the task completion.

Finding 3: The high-CT users have spent most of the time in doing the conceptual tasks either related to reading the handouts or thinking how to perform the action. On the other hand, low-CT users spent maximum time in performing physical actions which help them in completing the task early.

The statistical results (chi-squared test, $(X^2) < c, p < 0.05$) show that there are significant differences in the modeling actions of low-CT and high-CT users. The maximum difference is observed in conceptual tasks. **Low-CT users performed 1.5 times as many physical actions as high-CT users. High-CT users spend a large amount of time on conceptual actions.** They were relying more on the experimental

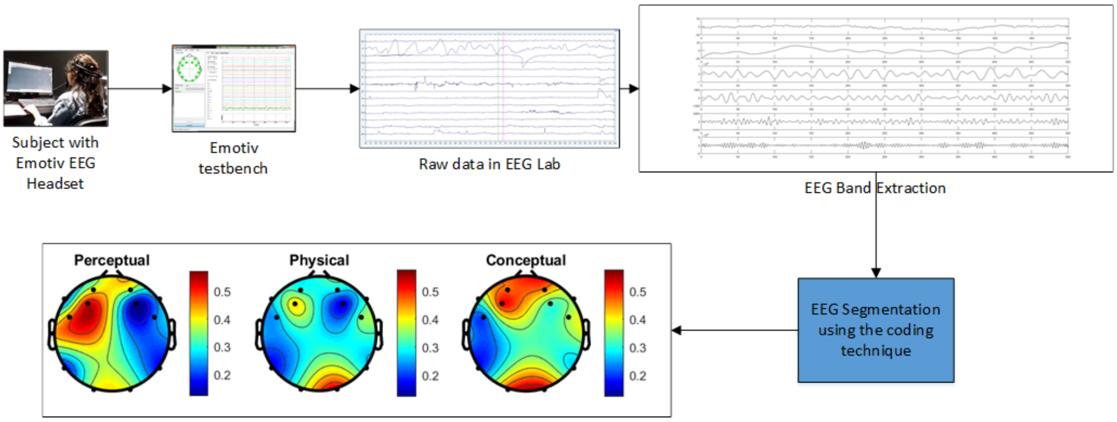


Figure 6.1: EEG Data Processing steps

instructions rather than their short-term memory. The rate of **conceptual actions for high-CT users was twice as high as for low-CT users.** The biggest difference was in reading the experiment instructions from the handouts.

Table 6.6: Average time taken in seconds in performing various design actions

Actions	Perceptual	Physical	Conceptual
User A	5.2	**9.9**	3.7
User B	**6.0**	7.8	7.0
User C	5.2	9.9	**7.2**
Low CT	**5.5**	**9.2**	**6.0**
User D	8.6	**16.9**	9.7
User E	8.4	7.7	**14.7**
User F	**9.4**	11.3	14.4
High CT	**8.8**	**12.0**	**12.9**
Ratio	**1.6**	**1.3**	**2.2**

6.5 EEG Data Analysis

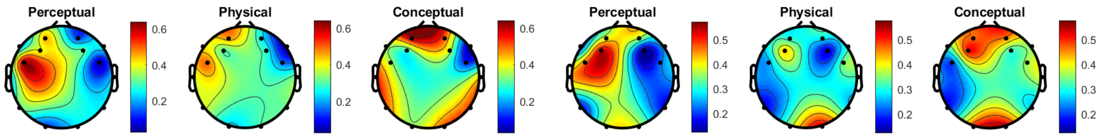
The EEG data of six participants were used for EEG analysis, the data of two participants were rejected because the data was corrupted with noise. The preprocessing was done in MATLAB 2017 using the EEGLAB toolbox [426] as mentioned in Chapter 5. Once the data was clean enough, we calculated the Welch power spectral density

(PSD) (see Section 8.2.1) with a window size of 128 samples without overlapping. The EEG segmentation has been carried out using the proposed coding scheme and the stored video log. Codes were assigned to each action performed by the user and EEG signals were segmented using these codes manually. Data from all electrodes were incorporated into the analysis. A block diagram of all the EEG data processing steps is shown in Fig. 6.1. After calculating the PSD, the mean power for each band has been extracted as shown in Figures 6.2-6.5. The scale in the topoplots shows the average PSD, normalized between minimum and maximum value of each user for comparison.

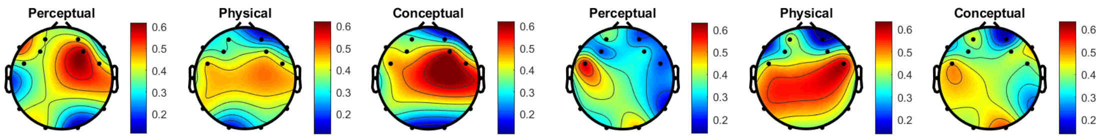
6.5.1 Alpha Band Activity

We have divided the dataset based on the completion time (CT) of the users for comparison. The threshold was set at 190 seconds by calculating the mean and standard deviation of the completion time. Fig. 6.2 shows the alpha-band energy in perceptual, physical and conceptual actions. The subjects on the left side of the figure have a low task CT, whereas the subjects on the right side of the figure have a high task CT. As alpha-band activity is associated with cognitive functions such as task performance preparation [279], language comprehension, and memory [270], so a change in alpha activity indicates a change in the cognitive activity. In the literature, researchers have established that **task complexity is inversely related to alpha-band activity [446].**

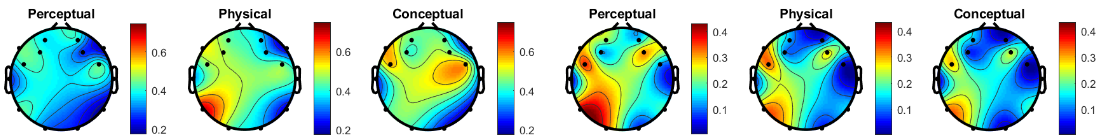
In Fig. 6.2, **the alpha-band activity in the conceptual actions is higher than in the physical actions for low-CT subjects**, which means that the **subjects are more relaxed in performing the conceptual tasks or they have spent less time in conceptual tasks than in other segments.** Subjects A and B have performed more physical tasks and the corresponding alpha-band activity is less in that segment, meaning that their attention highly focuses on physical actions. **For the users with**



(a) User A (low-CT) **(b)** User D (high-CT)



(c) User B (low-CT) **(d)** User E (high-CT)

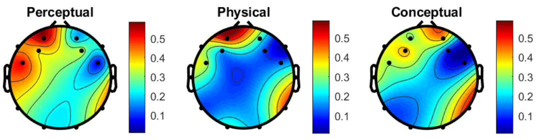


(e) User C (low-CT) **(f)** User F (high-CT)

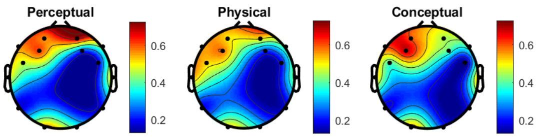
Figure 6.2: Mean alpha-band activity at different segments

a high-CT, the alpha activity is less in the conceptual actions (excluding User D) which is a sign that they were more relaxed or comfortable when they were thinking or reading handouts compared to others while performing modeling actions. **The variation in alpha activity is higher in the frontal cortex than in other regions for low-CT users.** The frontal cortex of the brain is responsible for higher mental functions such as concentration, planning, and problem-solving [447], so it is also an indication that **low-CT users were performing more cognitive activity than high-CT users.**

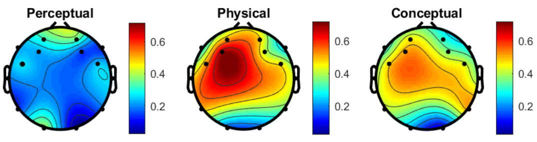
For subjects with high-CT, alpha activity variations were more in the motor, temporal and parietal cortex (electrodes T7, FC6, and P4) than in the frontal cortex. These locations are more associated with voluntary motor functions [447]. The continuous activation on the left frontal cortex (electrodes FC5 and F3) in the perceptual segment of high-CT subjects is possibly due to the continuous eye movements because of the nature of the experiment. We have also observed that **the change in alpha-band activity for users with low-CT is more than for subjects with high-CT. The reason behind this response may be that the low-CT subjects**



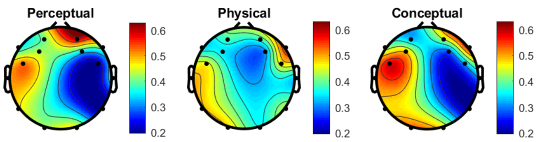
(a) User A (low-CT)



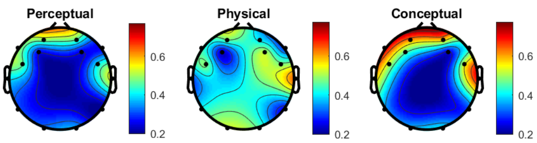
(b) User D (high-CT)



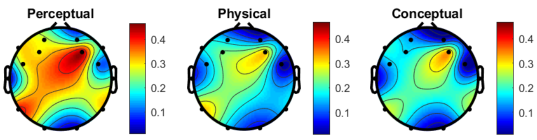
(c) User B (low-CT)



(d) User E (high-CT)



(e) User C (low-CT)



(f) User F (high-CT)

Figure 6.3: Mean beta-band activity at different segments

change the design stages very rapidly compared to the other subjects.

Coding Design actions and using these for EEG analysis are also very beneficial in providing further insight into which sub action caught the attention of the user. For example, in Fig 6.4, the alpha activity of user F has been shown. We can observe that the activity is more in segments drawing depiction (Pd), coloring depiction (Pc), deleting depiction (Psd), idle state (Ci) compared to other segments.

6.5.2 Beta Band Activity

Fig. 6.3 shows the beta activity of the three segments. The beta band usually relates to alertness and has a very low amplitude [280]. By looking at the beta activity response in Fig. 6.3, we can say that on **average the beta band activity is high in physical segments for low-CT users and high in the perceptual segments in high-CT users.**

Finding 4: As we mentioned before, the beta activity directly relates to concentration, the high beta activity is an indication that the concentration is high

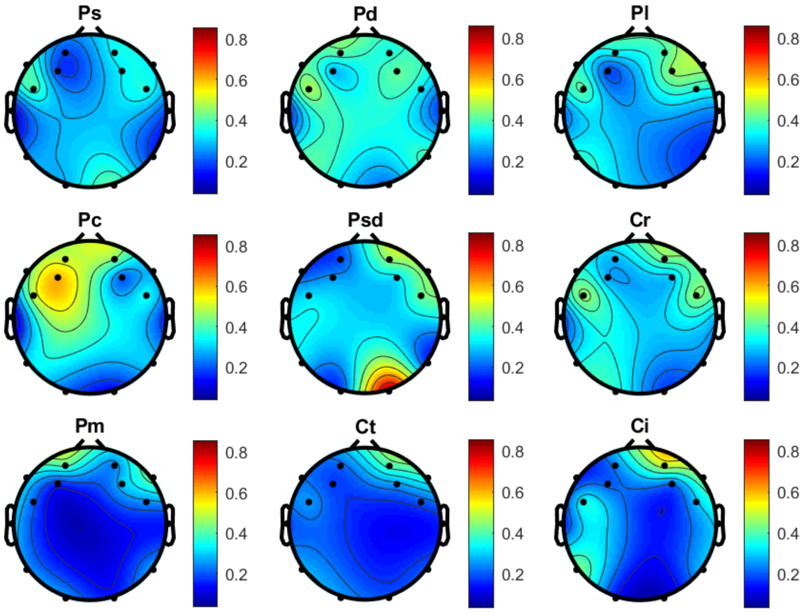


Figure 6.4: Sub design action alpha-band activity of User F

in the physical segments compared to the perceptual and conceptual segments of low-CT users. The higher concentration can be due to the higher number of physical actions performed by low-CT users as seen in Table 6.5.

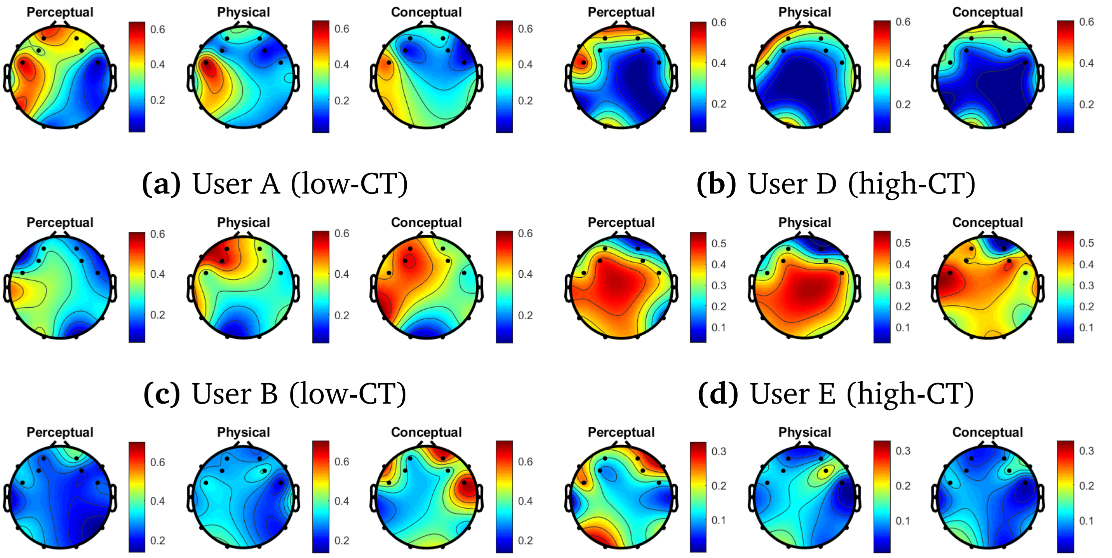
For high-CT users, on average, the beta activity is lower in physical segments than in other segments, and this finding can also relate to the fact that fewer actions are performed in physical sections. Users with low-CT are more attentive in physical segments whereas users with high-CT concentrate more on perceptual and conceptual activities.

6.5.3 Theta Band Activity

The theta-band response is observed in adult individuals who are in a state of focus and is also associated with memory performance and functional processes [270].

Finding 5: We observe that the theta-band activity varies in each segment for low-CT users, especially for users A and B, whereas for other users the relative change is very small as shown in Fig. 6.5.

This is also an indication that the focus or attention level of users with low-CT



(a) User A (low-CT) **(b)** User D (high-CT)
(c) User B (low-CT) **(d)** User E (high-CT)
(e) User C (low-CT) **(f)** User F (high-CT)

Figure 6.5: Mean theta activity at different segments

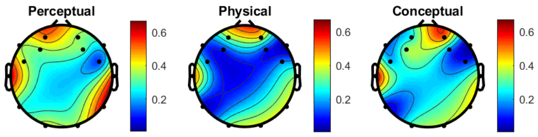
varied based on what actions they were performing.

6.5.4 Gamma Band Activity

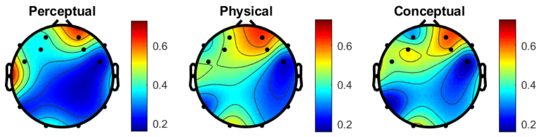
We have also observed that, by looking at the gamma-band activity (Fig. 6.6) and comparing it with the actions performed in each segment, **the average gamma-band activity is low in segments where there are more actions and high in segments with fewer actions.**

6.6 Conclusion

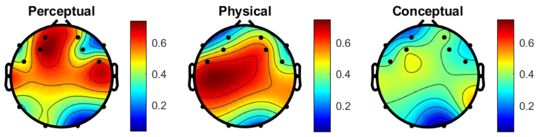
In this chapter, we have presented a new method to segment EEG signals for understanding cognitive actions and their relation to brain activities. For this purpose, we have conducted an experiment in which each user had to draw a 3D table and we have used video recording and EEG signals to analyze the user's cognitive activities. Participants had no prior experience of designing 3D objects in AutoCAD, so we have used task completion time as a measure to differentiate between designers. We have



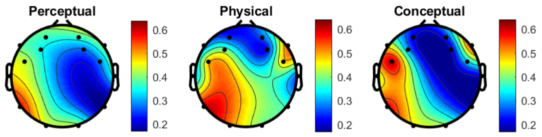
(a) User A (low-CT)



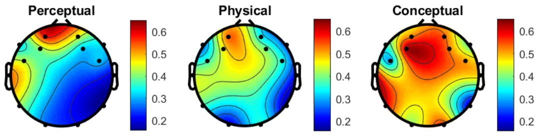
(b) User D (high-CT)



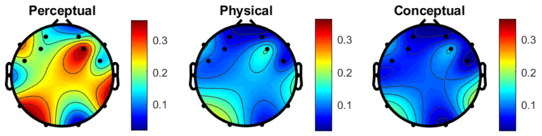
(c) User B (low-CT)



(d) User E (high-CT)



(e) User C (low-CT)



(f) User F (high-CT)

Figure 6.6: Mean gamma activity at different segments

analyzed the reasons for why some users completed the task earlier than the other users.

We have used a coding scheme designed by Suwa et al. [441] to analyze the quantified designer's actions. The coding analysis for EEG segmentation provides two advantages: the first is that we can see the EEG power variation in different segments, and the second is that we can use the results of coding analysis and compare with EEG power to easily track the cause of particular behavior. From video recording, all the actions were decoded and divided into three segments: Perceptional, Physical and Conceptual actions. These segments were used to segment out the EEG data. We have analyzed the alpha, beta, theta and gamma activity of the users. We categorized users in 2 groups: Low-CT and High-CT users. The findings from our data analysis are listed below:

1 Low-CT users performed 1.5 times more physical actions, which gave them the advantage of drawing quickly.

2 The rate of conceptual actions for high-CT users was twice as high as for low-CT users. This slows the overall design process.

3 The action rate per minute for low-CT users is 30% higher than for high-CT users. This is an indication that they are utilizing their short-term memory more efficiently. This result is aligned with previous findings of Kavakli and Gero [442].

4 The alpha band shows that low-CT users were comfortable in performing physical tasks whereas high-CT users were not relaxed in physical segments as their mean alpha-band power was high.

5 High-CT users spent maximum time in performing conceptual tasks compared to low-CT users, who spent most of the time in focusing on physical design actions.

6 The maximum variation in the frontal cortex was found in low-CT users, which indicates that they were using their short-term memory more.

7 From the beta activity, we have found that low-CT users were more attentive to physical segments, whereas the attention of high-CT users was focused on perceptual and conceptual actions.

8 We have found more variation in theta-band activity for low-CT users than for high-CT users, which indicates that the focus of cognitive activity of low-CT users was changing in relation to the action performed.

As low-CT users were performing the task more easily and efficiently compared to high-CT users, so we can deduce the following hypothesis for competency:

- Conceptual C-actions are the key to the performance. One would expect to see a low number of C-actions in competent (low-CT) users.

- Competent (low-CT) users use their short term memory more efficiently and more frequently.
- Competent (low-CT) users utilize more motor processes (physical actions) compared to novices that focus more on perceptual and conceptual actions as evidenced by beta-band activity.
- Variations (changes) in cognitive processing are higher in competent (low-CT) users as stated by theta band activity.

From the above analysis, we have concluded that if a user would utilize short-term memory more, reducing their attention to the conceptual actions and performing more physical actions instead, then their performance would improve.

Chapter 7

Analysis of Cognitive Activities in a Uni-modal System: using Transfer Entropy and Functional Brain Networks

This chapter explains the use of transfer entropy (TE) to measure connectivity between EEG signals during cognitive activity. The matrix generated from transfer entropy is used to construct functional brain networks (FBN). The analysis results have been published in the Brain Sciences Journal titled as "Connectivity Analysis Using Functional Brain Networks to Evaluate Cognitive Activity during 3D Modeling" [47]. Both the weighted and binary FBN analysis have been presented in the chapter, along with the information transfer between left/right hemisphere and lobes of the brain. The classification of novice and competent users from transfer entropy of EEG signal has been published at the 2019 11th International Conference on Computer and Automation Engineering (ICCAE 2019) held in Brisbane, Australia, in a paper titled "Expertise Classification using Functional Brain Networks and Normalized Transfer Entropy of EEG in Design Applications" [448].

- Baig, Muhammad Zeeshan, and Manolya Kavakli. "Connectivity analysis using functional brain networks to evaluate cognitive activity during 3d modelling." *Brain sciences* 9, no. 2 (2019): 24.
- Baig, Muhammad Zeeshan, and Manolya Kavakli. "Expertise Classification using Functional Brain Networks and Normalized Transfer Entropy of EEG in Design Applications." *In Proceedings of the 2019 11th International Conference on Computer and Automation Engineering,* pp. 41-46. ACM, 2019.

7.1 Introduction

Differences in information processing have been observed between males and females, novices and competent and left-handed and right-handed people while describing a simple 3D object [30]. The study of these user-dependent factors makes the interaction robust and enhances system flexibility, efficiency, naturalness. In this chapter, we investigate the differences in information flow patterns by analyzing the electroencephalogram (EEG) signals of various participants to find the relationships between cognitive strategies in information processing and competencies while performing a set of experiments.

In the research literature, TE has not been used to study user dependent differences. In the present work, we apply TE to the analysis of information flow patterns between novice and competent users during a 3D modeling task with keyboard and mouse. We have used normalized TE values to construct both binary and weighted directional FBNs. After constructing an FBN, we have applied graph theory measures and statistical analysis to quantify the information flow patterns. The main objective of this analysis is to identify the topological differences between novice and competent user's FBNs in different design activities. The second objective is to identify the information flow patterns in the design actions of a user. The following research

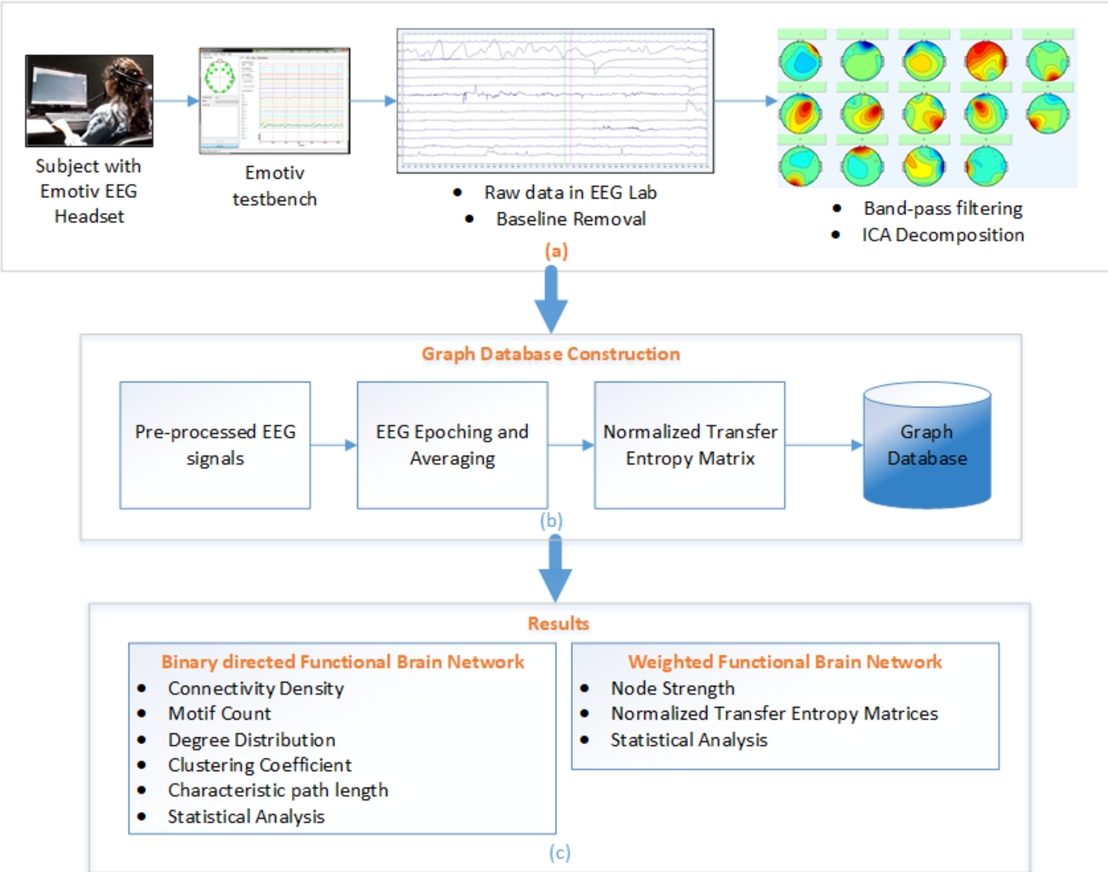


Figure 7.1: Normalized Transfer Entropy Framework (a) EEG data acquisition and pre-processing, (b) Transfer Entropy calculation and graph database construction, and (c) Result and analysis of binary and weighted FBNs

questions are addressed in this chapter:

RQ 4.1 Are there any differences in information processing and cognitive activity between novice and competent users?

RQ 4.2 Can Functional Brain Networks (FBNs) be used to identify the information flow patterns?

7.2 EEG Data Analysis Methodology

Data of eight participants were used for the experiment. The experiment is the same as the one described in Chapter 6. Three of them were competent users and five novice users. EEG data were filtered (described in Chapter 5) and segmented into two segments. The first segment was the one where the participants were drawing the table. In the second segment, the participants were asked to change the appearance of the table, such as materials, etc. using AutoCAD options. For convenience, we named the first segment, "Drawing" and the second segment, "Manipulation".

The preprocessing was done in MATLAB 2017b using the EEGLAB toolbox [426] as described in Chapter 5.

Once the data was clean enough, we extracted two-second epoch averaged data from resting, drawing, and manipulation tasks by performing back-to-back epoching with a 0.5-sec difference between epochs. The response time for a modeling action varies from 0.5 to 4 seconds, and we made sure that a minimum of one modeling action must be performed in an epoch. Table 7.1 shows the time taken by each user in seconds in drawing and manipulation states. Novice-User 2 and Competent-User 2 both deleted some objects when drawing, and these phases were also considered in this experiment. There was no correlation found between competency level and task completion time.

Table 7.1: Time taken by each participant to complete the task

Competency	Drawing (sec)	Manipulation (sec)
Competent-User 1	38	72
Competent-User 2	120	101
Competent-User 3	58	113
Novice-User 1	81	125
Novice-User 2	137	93
Novice-User 3	54	72
Novice-User 4	43	118
Novice-User 5	61	98

7.2.1 Functional Brain Network

The pre-processed EEG signals were used in the construction of NTE connectivity matrices, where each cell denotes the NTE value from one electrode to another. The normalization was done by subtracting a noise matrix (averaged shuffled TE matrix) from the Transfer Entropy (TE) matrix. The NTE matrices were used to create both binary and weighted directed FBN. To analyze the results, we had used the graph analysis measure such as the Connectivity density, clustering coefficient, characteristic path length, motif count, node strength, and small-worldness. Fig. 7.1 illustrates EEG analysis using normalized transfer entropy.

7.2.2 Binary Directed Functional Brain Network

A threshold was applied on NTE matrix to convert them into binary directed functional brain networks (FBN) for calculating complex network parameters. For this experiment, the threshold was set to 0.001, which is an arbitrary value to remove the very insignificant connections.

7.2.3 Weighted Directed Functional Brain Network

The NTE matrices, without applying any threshold (shown in Fig 7.2 and 7.3) were also used to design **weighted directed functional brain networks (WDFBN)**. The NTE matrices show that the transfer entropy has a larger variance in competent users compared to novice users. The information transfer measured using NTE is different for almost every electrode in competent user. Fig 7.12 gives the **node strength** values of novice and competent users during various cognitive states. WDFBN was used to calculate the node strength using equation 7.1.

$$Strength_i = \sum_{j \epsilon V} w_{ij} + \sum_{J \epsilon V} w_{ji} \qquad Eq(7.1)$$

where w_{ij} is an element weight of NTE matrix.

Hemisphere-wise information flow

The hemisphere analysis has been performed to identify the **information flow patterns** within and between the left and right hemispheres. For this analysis, we have

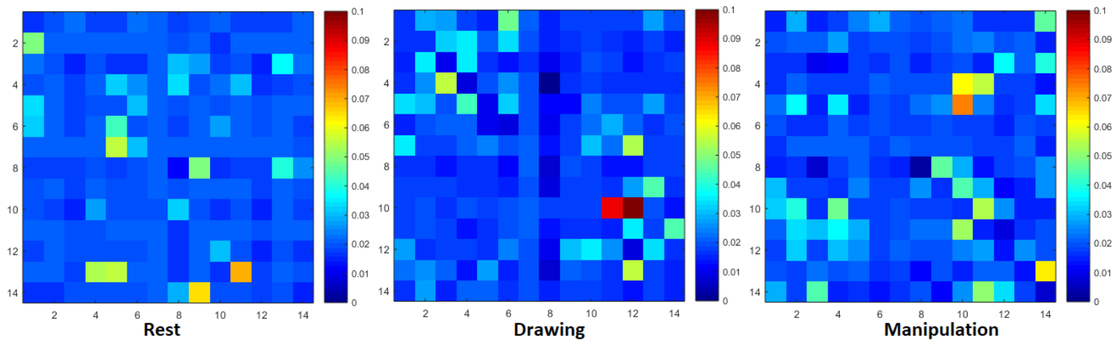


Figure 7.2: Normalized transfer entropy matrix during rest, drawing and manipulation states of novice user

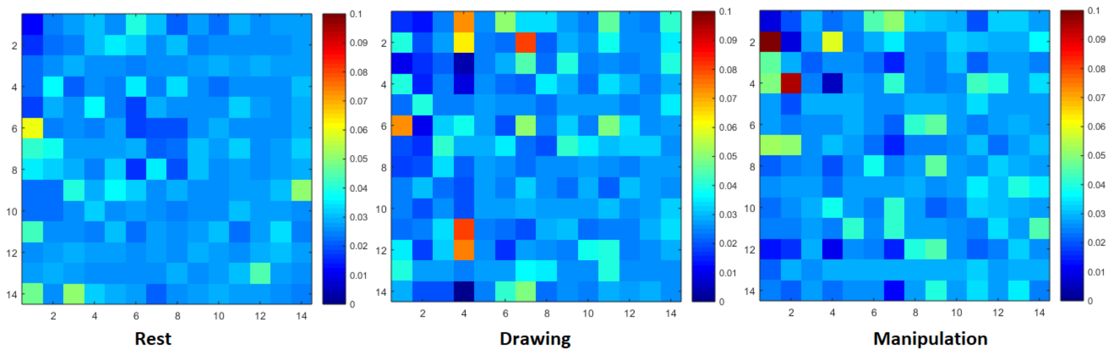


Figure 7.3: Normalized transfer entropy matrix during rest, drawing and manipulation states of competent user

divided 14 EEG electrodes into two sets of 7 electrodes that correspond to each hemisphere. The electrodes which belong to the left hemisphere (LH), and right hemisphere (RH) are shown in Fig. 7.4. A total of four sub-NTE matrices were generated to represent the information flow between electrodes in LH to LH, LH to RH, RH to RH and RH to LH. The size of sub-NTE matrices was 7x7, and the total information flow from one electrode to all other electrodes was calculated by row-wise summation of each sub-NTE matrices. To analyze these sub-NTE matrices, one-way ANOVA was applied, and the results of 2 typical novice and 2 typical competent users have been shown in Fig. 7.15.

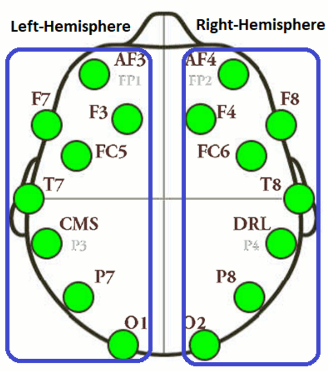


Figure 7.4: Electrodes layout of LH and RH

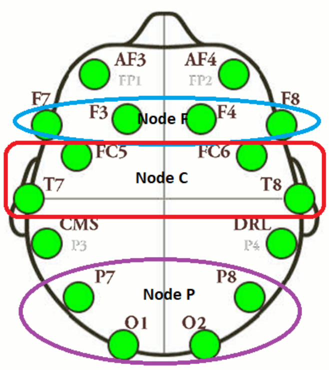


Figure 7.5: Electrodes layout of F, C and P nodes

Region-based information flow

Region-based information flow has been studied and documented by other researchers [387, 406]. To study the information flow (IF) in different regions of the brain, we also performed the region-wise analysis. We divided the brain into three regions and called them nodes. Each of the three nodes had four electrodes. From the NTE matrix, three sub-matrices were constructed, and information flow from and to the node was calculated. The nodes are F, C, and P. The node F contains electrodes F7, F3, F4, and F8 and represent the frontal cortex of the brain. The node C contains electrodes FC5, T7, T8, and FC6 and represent the central and temporal cortex. The node P contains electrodes P7, O1, O2, and P8 and gives the information of parietal and occipital lobes. The information flow from one region to another region was calculated by the summation of information flow from the electrodes of one region to another. For example, the IF from F to C was calculated by summation of all IF from F7, F3, F4, F8 to FC5, T7, T8, FC6. The results of region-wise IF are shown in Table 7.5 for 2 competent users (Competent-User 2 and 3) and 2 novice users (Novice-User 2 and 5) for all three states.

7.2.4 Classification of Novice/Competent

We have used measures such as connectivity density and clustering coefficient mentioned in Section 3.5.3 as features for the classification of novice and competent users. To depict real-time situations and generate a reasonable number of trials, a 20-second window of EEG signal from every second onward was used and considered as one trial. By this method, we extracted a total of 470 trial in which 285 trials correspond to novice users, and 185 trials belong to competent users. The connectivity density, motif count, clustering coefficient, and mean information flow are used as a feature for classification. The complete feature set is the combination of all the measures mentioned above. The actual feature set is given in Fig. 7.6.

Feature selection algorithm was applied to extract the best features for classification. We have used a sequential forward search (SFS) for searching for the best feature. The technique used for feature selection is wrapper technique. In the wrapper technique, the classification algorithm is a part of the feature selection process. In this chapter, classification accuracy has been used as the optimization criteria for feature selection.

Various classification algorithms are used to classify features into different classes. These classification algorithms are categorized into supervised and unsupervised techniques. Support Vector Machines (SVM) [449], K-nearest neighbors (k-NN) [450], Bayesian classifier [451] are some of the most commonly used classifiers for EEG applications. In this Chapter, we have used five different classifiers to classify the feature set. The classifiers used in this chapter are SVM, k-NN, Linear Discriminant

Analysis [452], Naive Bayes [451] , and decision trees [453]. For evaluation we have used classification accuracy, sensitivity, specificity, precision, F-measure and mean squared error.

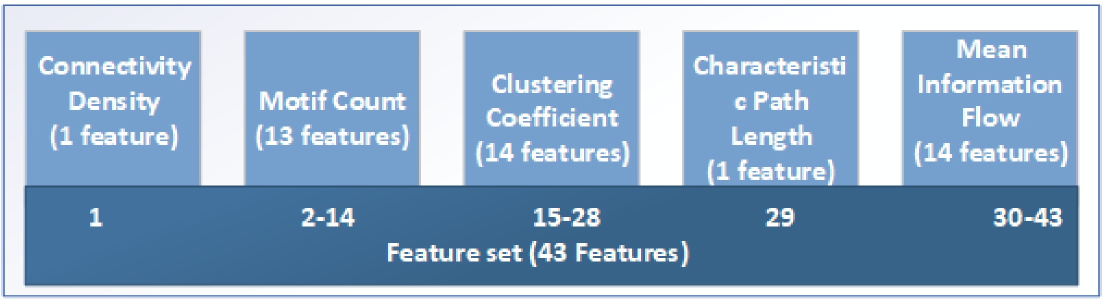


Figure 7.6: The complete feature set

7.3 Results and Discussion

The NTE matrices of size 14x14 were calculated, as shown in Fig. 7.2, 7.3. Fig. 7.2 shows the NTE matrix of a novice user with three different design states and Fig. 7.3 shows the NTE matrix of a competent user. The cluttered and brighter pixel shows the increase in information flow. From the NTE matrices, it can be inferred that the information flow pattern of the user changed and increased from the baseline rest condition. To maintain the readability of the chapter, the results of two novice and two competent users have been shown in this chapter instead of the entire cohort. Most of the figures and tables in this section show the results of the same participants. In addition, the average results for novice and competent users have been shown to show the efficacy of the technique.

7.3.1 Binary Directed FBN Analysis

The results of binary directed FBNs are shown in Fig. 7.7 for a novice user and Fig. 7.8 for a competent user. For the novice, the connections between electrodes increase when participants move from rest to drawing and from drawing to manipulation. In

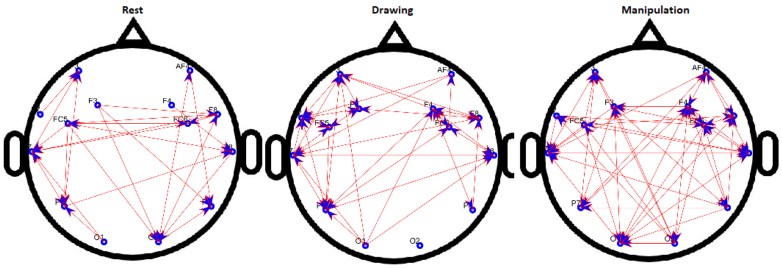


Figure 7.7: Binary directed functional brain network during rest, drawing and manipulation states of novice user

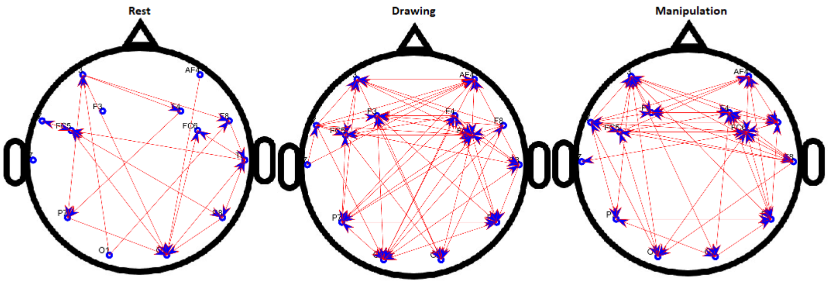


Figure 7.8: Binary directed functional brain network during rest, drawing and manipulation states of competent user

contrast, connections increase from rest to drawing and no significant change has been observed from drawing to manipulation for competent users.

The density of network increased from the baseline for both novice and competent users, but the change in density for the novice user is more, compared to a competent user in drawing and manipulations states. Most of the activity was focused on the frontal cortex, which also indicates the use of short-term memory [454].

The **connectivity density** for all the user is shown in Fig. 7.9. To compare the connectivity density across all users in different states, we normalized the value by dividing the actual connectivity value in one state by the sum of connectivity in all states. As indicated in Fig. 7.9, the connectivity density is higher for all the users in drawing and manipulation state than in the rest condition, which is a control condition. From this figure, we can deduce that the information flow increases in the drawing and manipulation states compared to a rest state by establishing more connections. The main difference between novice and competent users has been

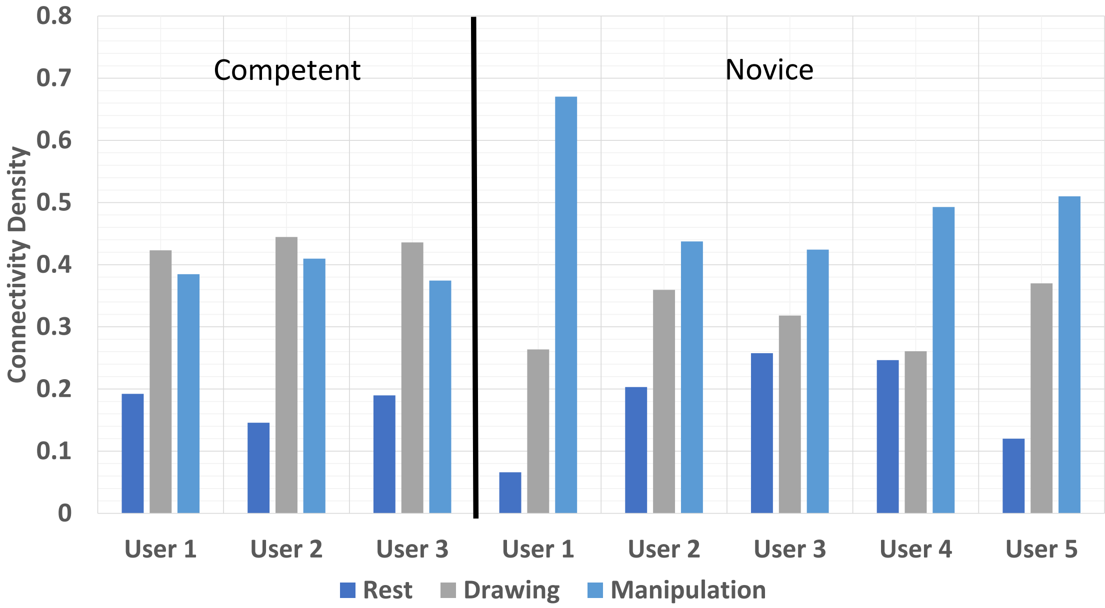
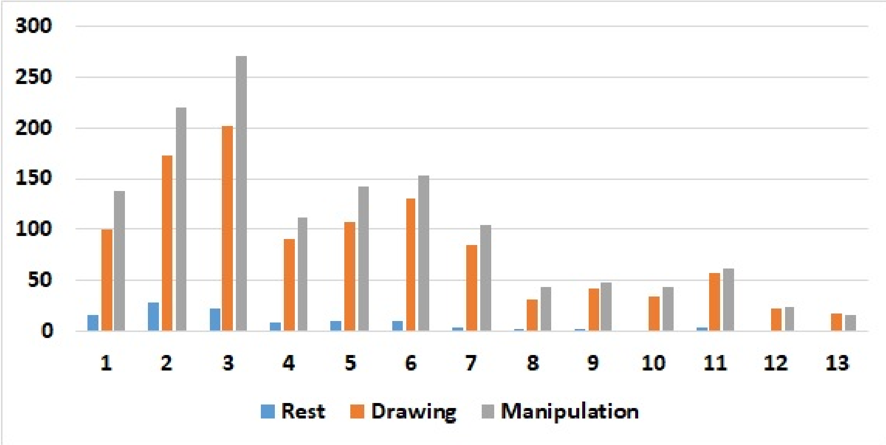


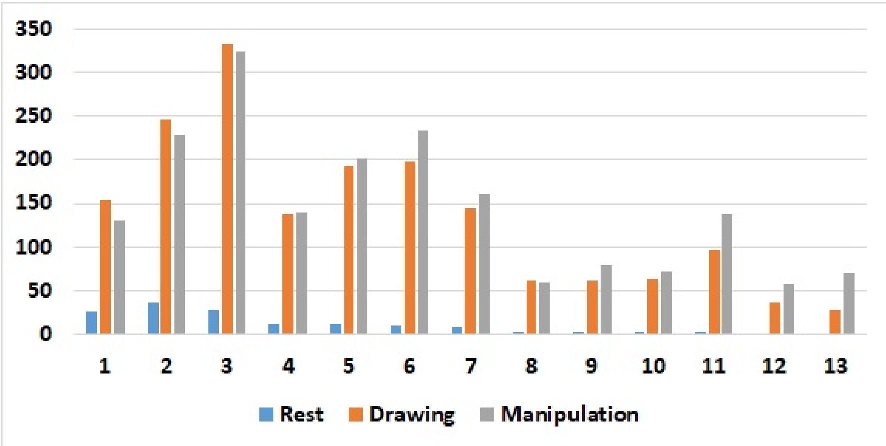
Figure 7.9: Comparison of connectivity density of all users brain activity during rest, drawing and manipulation states

observed in drawing and manipulation states and the connectivity density relatively increased in manipulation states from drawing states for novice users. The competent user's connectivity density difference was much less compared to novices, and a slight decrease in density was seen for competent users, which means that the information flow was more or less the same in drawing and manipulation states. The increase in connectivity for novice users also indicated that the electrodes were trying to establish a mutual connection to facilitate effective information transfer within the FBN as novice users are using the drawing and manipulation functions for the first time.

Fig. 7.10 shows the average number of **motifs** for three nodes during rest, drawing and manipulation state for novice and competent user. The motif count is higher in drawing and manipulation states compared to the rest state. Novice users have an increase in the number of motifs in manipulation states from drawing state, and this pattern exists for almost all the novice users except for Novice-User 3, in which the motif count difference is not significant. Like connectivity density, the motif count for competent users decreases in manipulation state compared to the drawing state.



(a) Average motif count for novice users

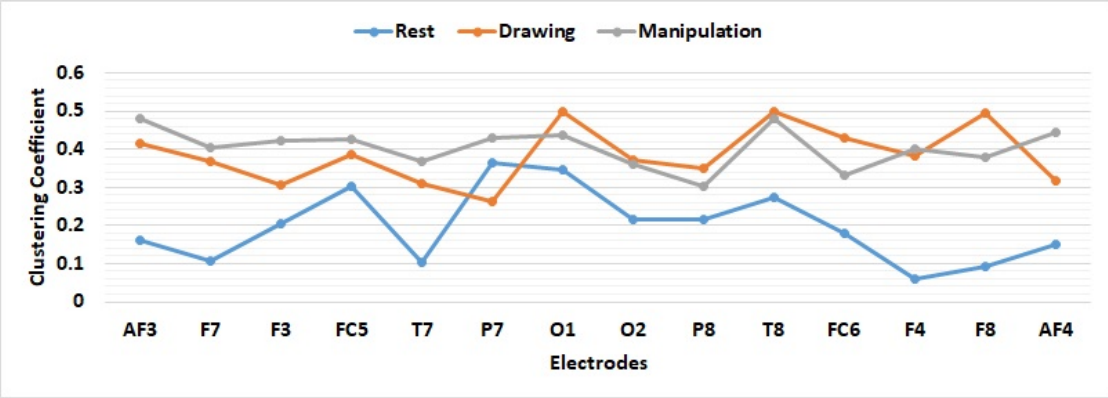


(b) Average motif count for competent users

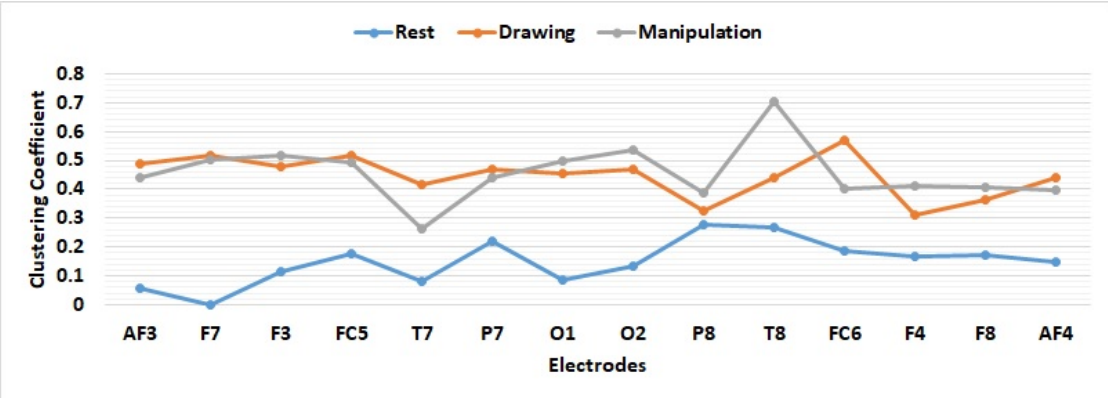
Figure 7.10: Comparison of average number of the motif with three nodes during rest, drawing and manipulation states

As the number of motifs is used to describe the local features of the network, the increase in motifs is related to a substantial increase in information exchange among the neighboring nodes of directed FBN. The information transfer between neighbor electrodes was more for novice users compared to competent users, which means that novice user's information flow pattern changes more rapidly compared to competent users.

Average **clustering coefficient** of novice and competent users has been shown in Fig. 7.11 during rest, drawing and manipulation states. The value increases for almost all the electrodes in drawing and manipulation states compared to baseline



(a) Average clustering coefficient for novice users



(b) Average clustering coefficient for competent users

Figure 7.11: Average clustering coefficient of all electrodes for novice and competent users during rest, drawing and manipulation states

rest state. This is a clear indication that each electrode was communicating directly with its neighboring electrodes and formed clusters. For novice users, the clustering coefficient value in manipulation state is slightly higher than drawing state except for T8 and FC6 electrodes. The trend is opposite for competent users, the value in drawing state is higher than manipulation state except for electrodes O1, O2, P3. These regions belonged to the occipital and parietal cortex and from the literature, we found that these regions were associated with sensations from muscles, visual perception and recognition [455].

The statistical significance of the clustering coefficient across all 14 electrodes has been calculated using a 2-sample t-test with unequal variance at $\alpha = 0.05$ for novice and competent users. The results are shown in Table 7.2 for 2 typical users. The

Table 7.2: Statistical validation of clustering coefficients for typical novice and competent users (D:Drawing, R:Rest, M:Manipulation, MD:Mean Difference)

Users	States		MD	95% CI	DF	t	P
Novice-User 2	D	R	**0.156**	(0.0763 , 0.2357)	20	4.0838	**0.0006**
	M	D	0.031	(-0.0169 , 0.0789)	24	1.3362	0.097
	M	R	**0.187**	(0.1102 , 0.2638)	17	5.1374	**0.00001**
Competent-User 3	D	R	**0.273**	(0.2148 , 0.3312)	15	10	**0.00001**
	M	D	-0.015	(-0.0988 , 0.0688)	17	-0.7895	0.3551
	M	R	**0.258**	(0.1935 , 0.3225)	22	8.2958	**0.00001**

results suggested that mean difference is significant ($p < 0.05$) across rest/drawing and rest/manipulation states for both users and the same trend has been observed for all other users. The difference is not significant across manipulation/drawing state but if we compare the values for novice and competent users the difference across manipulation/drawing state is more for novice users than competent users with a negative mean difference for Competent-User 2 and 3.

The **small-world properties** of directed FBN during rest, drawing and manipulation states for novice and competent users are shown in Table 7.3. In Table 7.3,

Table 7.3: Small-worldness of binary directed FBNs during rest, drawing and manipulation states for 2 typical users

Users	Cognitive State	C_d	C_{rand}	L_d	L_{rand}	$\sigma = \frac{C_d}{C_{rand}} / \frac{L_d}{L_{rand}}$
Novice-User 2	Rest	0.9119	0.8676	1.1939	1.1934	1.0506
	Drawing	0.6675	0.5358	1.5255	1.5043	**1.2286**
	Manipulation	0.7714	0.6582	1.3878	1.3879	1.0809
Competent-User 3	Rest	0.8214	0.6998	1.3469	1.3470	1.1739
	Drawing	0.4777	0.4126	1.7092	1.6537	**1.1202**
	Manipulation	0.6607	0.5267	1.5408	1.5256	**1.2420**

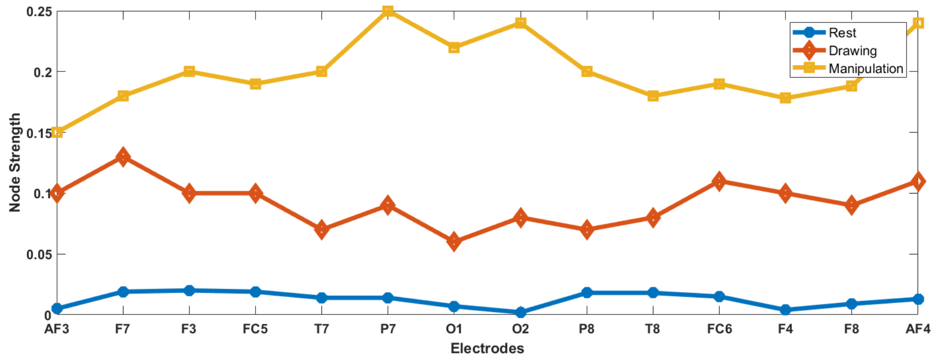
C_d and C_{rand} are the clustering coefficient of an actual and random network respectively; L_d and L_{rand} are the characteristic path lengths of actual network and random network respectively. If the value of $\sigma > 1$ for an FBN, then the network shows the small-world properties, which means that the FBN has both high local and global efficiency. For novice users, the difference between σ in rest, drawing and manipulation states are more than for the competent user, which means that competent users have relatively high global and local efficiency of information transfer than novice users. To calculate the random values for the clustering coefficient and characteristic path length, 100 matched random networks have been generated with 14 nodes. The small-world values are computed with the transitive clustering coefficient and Monte-Carlo realizations [405].

7.3.2 Weighted FBN Analysis

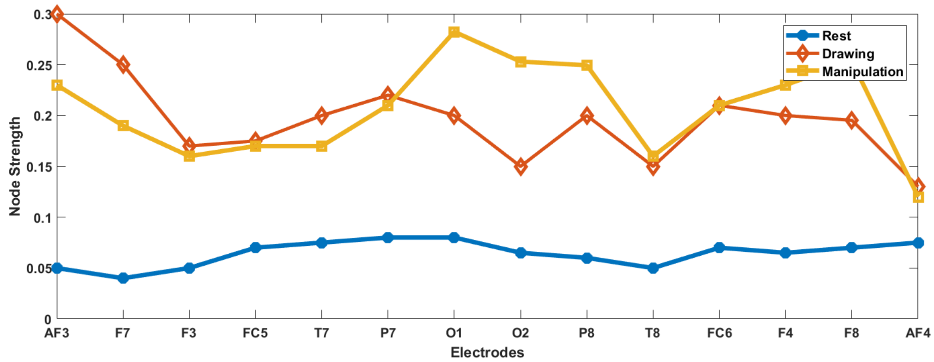
From Fig. 7.12, we can observe that both novice and competent users have more node strength in drawing and manipulation states than in the rest state, which indicates that each node or electrode sends and receives more information during drawing and manipulation states. The competent user's node strength is approximately the same in drawing and manipulation states, which means that their brain is more sending and receiving the same amount of information, and information flow is the same in both states, whereas, for novice users, the node strength increases in manipulation state, which indicates more information transfer in manipulation state than any other state. All novice users showed the same trend of node strength.

Statistical Analysis

To show the significance of information flow during the three states, we have applied the one-way analysis of variance (ANOVA) test. The mean information flow of each state, which is calculated by row-wise summation of each NTE matrix, was used in ANOVA and the multi-comparison procedure was applied, and results are shown in Fig. 7.13. The ANOVA results show that there are significant differences in the mean information flow between rest and drawing/ manipulation states. For novice users, the mean information flow difference is also significant in drawing and manipulation states, but this is not the case with competent users because of the



(a) Node strength averaged across all novice users



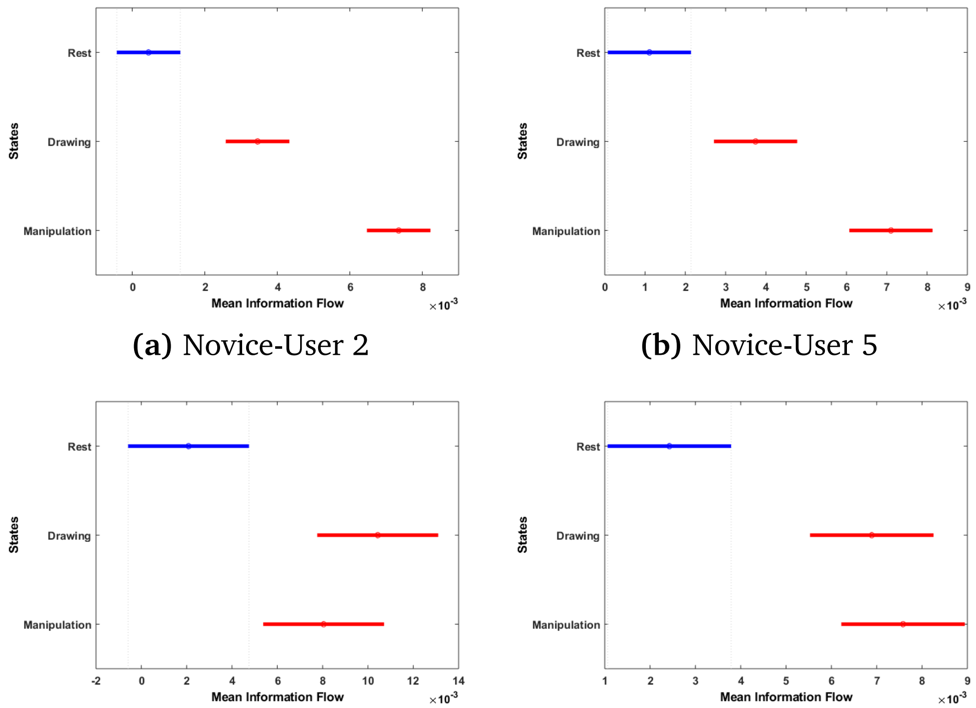
(b) Node strength averaged across all competent users

Figure 7.12: Comparison of node strength across electrodes during rest, drawing and manipulation states

overlap seen in drawing and manipulation states mean information flow. This is a clear indication that the information flow was increased in novice users when they started manipulating the object, whereas competent users were comfortable with both drawing and manipulation. Fig. 7.14 showed a multi-comparison of ANOVA on mean information flow of each electrode to all other electrode averaged across all the novice and competent users during rest, drawing and manipulation states.

Hemisphere-wise Analysis

Fig. 7.15a-7.15d shows the ANOVA test results for hemisphere-wise analysis. The results show that the information transfer from LH to LH is greater than the information transfer in other hemispheres for Competent-Users 2 and 3 in drawing state. In manipulation state, the t-test results showed that there are no significant differ-



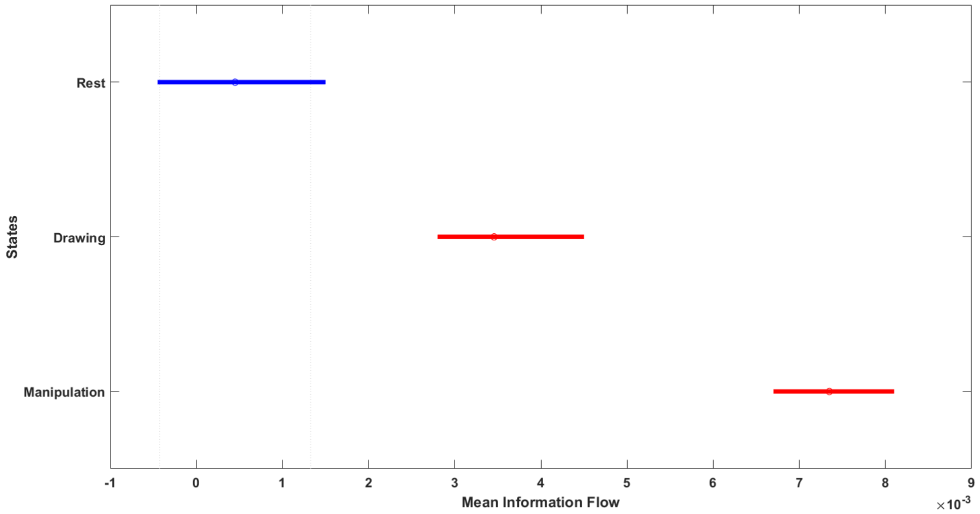
(a) Novice-User 2 **(b)** Novice-User 5

(c) Competent User 2 **(d)** Competent-User 3

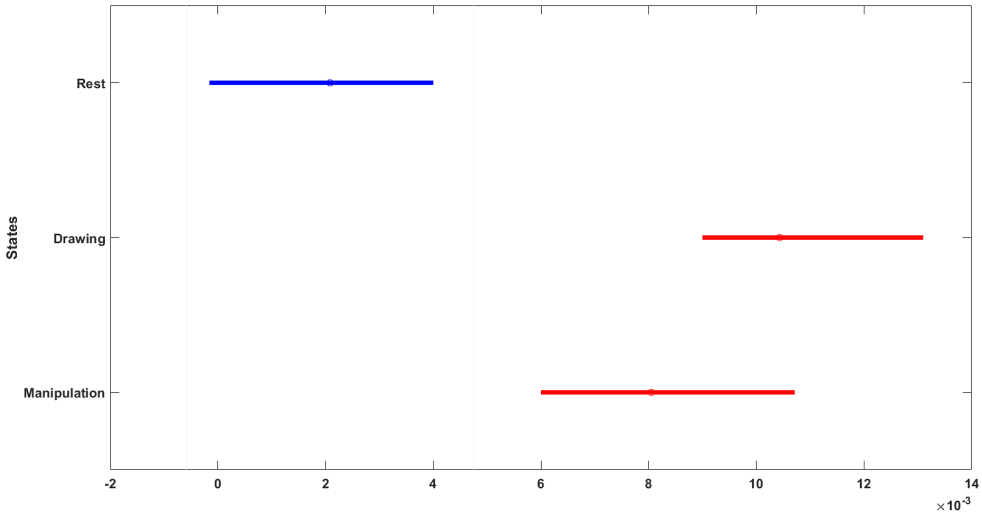
Figure 7.13: Multi-comparison of ANOVA on mean information flow of each electrode to all other electrodes during rest, drawing and manipulation states

ences in mean information transfer value for RH to RH and RH to LH transfer. For Competent-User 2, the values of RH to LH and RH to RH increased in manipulation state compared to drawing state. Competent-User 3 shows the opposite behavior; the mean information flow value from RH to LH and LH to LH decreases in manipulation state, and other information transfer values do not change much. In the case of the novice users, an increase in information flow in all the electrodes has been observed from rest to drawing to the manipulation state. The flow of information is more towards the left hemisphere from the right and left hemisphere electrodes. Thus, it can be said that, in manipulation state, the left region of the brain has received more information from its own electrodes then the right-side electrodes.

The statistical significance of the difference in information transfer between different states has been given in Table 7.4. These results were calculated by applying the two-sample t-test with $\alpha = 0.05$ (two-tailed). The results show that in most of

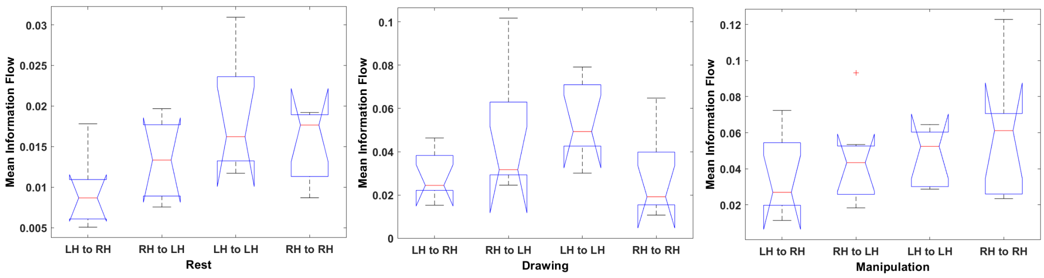


(a) Multi-comparison of ANOVA averaged across all Novice User

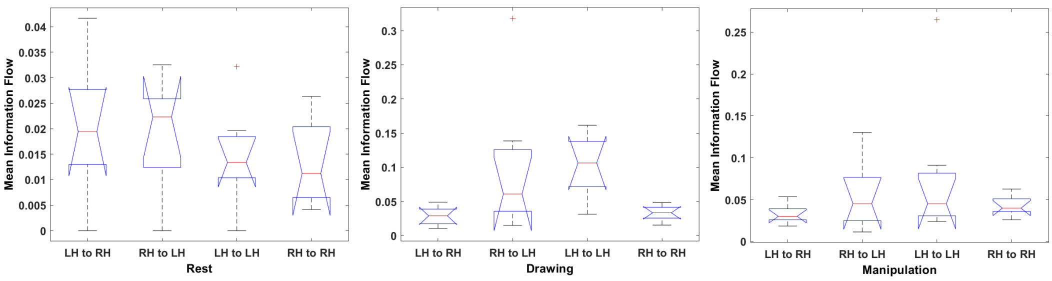


(b) Multi-comparison of ANOVA averaged across all Competent User

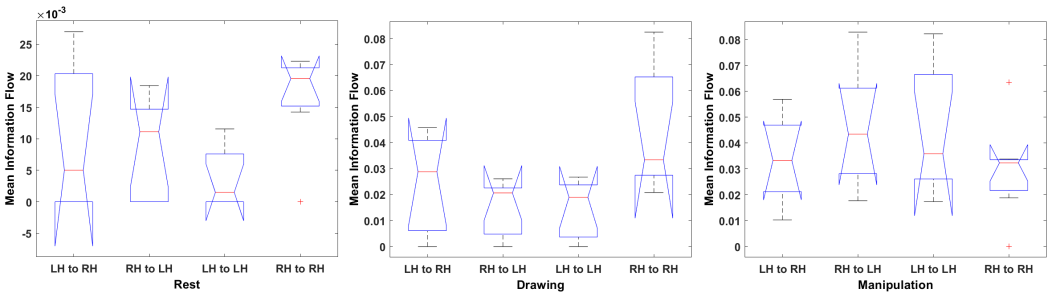
Figure 7.14: Multi-comparison of ANOVA on mean information flow during rest, drawing and manipulation states



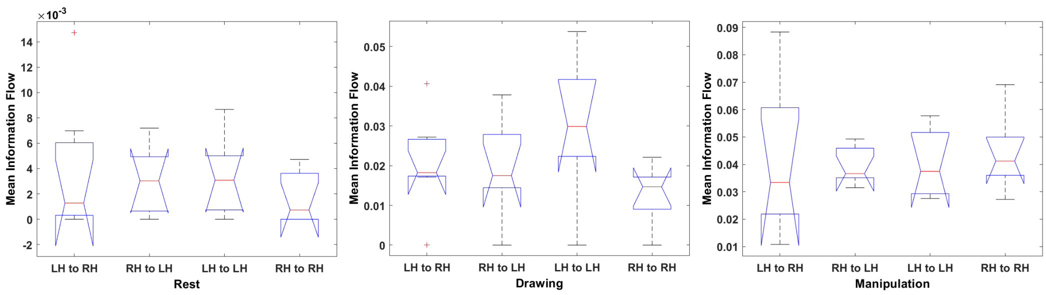
(a) Competent-User 2



(b) Competent-User 3



(c) Novice-User 2



(d) Novice-User 5

Figure 7.15: One-way ANOVA of Hemisphere-wise mean information flow during rest, drawing and manipulation states

Table 7.4: Result of pairwise mean difference using the t-test for hemispheres information flow during rest, drawing and manipulation states(MD:Mean Difference

States		Novice-User 2			Novice-User 5		
Drawing	**Rest**	**MD**	**CI**		**MD**	**CI**	
LH-RH	LH-RH	0.017*	0.005	0.029	0.016	-0.003	0.034
RH-LH	RH-LH	0.017*	0.006	0.028	0.005	-0.004	0.019
LH-LH	LH-LH	0.027*	0.011	0.043	0.012*	0.001	0.022
RH-RH	RH-RH	0.012*	0.005	0.018	0.030*	0.007	0.052
Manipulation	**Drawing**						
LH-RH	LH-RH	0.022	-0.004	0.048	0.008	-0.014	0.029
RH-LH	RH-LH	0.020*	0.008	0.032	0.031*	0.001	0.052
LH-LH	LH-LH	0.010	-0.007	0.028	0.029*	0.006	0.052
RH-RH	RH-RH	0.031*	0.018	0.044	-0.016	-0.042	0.009
Manipulation	**Rest**						
LH-RH	LH-RH	0.039*	0.014	0.064	0.024*	0.002	0.041
RH-LH	RH-LH	0.037*	0.01	0.043	0.038	0.017	0.059
LH-LH	LH-LH	0.037*	0.026	0.048	0.041*	0.018	0.064
RH-RH	RH-RH	0.043*	0.030	0.055	0.014	-0.004	0.032
		Competent-User 2			**Competent-User 3**		
Drawing	**Rest**	**MD**	**CI**		**MD**	**CI**	
LH-RH	LH-RH	0.008	-0.008	0.024	0.020*	0.009	0.030
RH-LH	RH-LH	0.080	-0.017	0.177	0.034*	0.008	0.060
LH-LH	LH-LH	0.088*	0.046	0.130	0.035*	0.018	0.052
RH-RH	RH-RH	0.020*	0.008	0.031	0.013	-0.005	0.031
Manipulation	**Drawing**						
LH-RH	LH-RH	0.005	-0.010	0.019	0.006	-0.016	0.028
RH-LH	RH-LH	-0.042	-0.141	0.057	-0.003	-0.035	0.028
LH-LH	LH-LH	-0.025	-0.108	0.057	-0.007	-0.027	0.013
RH-RH	RH-RH	0.009	-0.005	0.023	0.029	-0.005	0.063
Manipulation	**Rest**						
LH-RH	LH-RH	0.013	-0.002	0.027	0.026*	0.004	0.047
RH-LH	RH-LH	0.038	-0.000	0.076	0.031*	0.008	0.054
LH-LH	LH-LH	0.063	-0.017	0.142	0.028*	0.013	0.043
RH-RH	RH-RH	0.029*	0.016	0.041	0.042*	0.009	0.075

*Mean difference is significant at $p < 0.05$ level.

the cases, the difference in mean information flow is significant, especially for novice users. The mean difference is more for novice users when comparing manipulation and drawing states. For competent users, a small decrease in information flow has been observed, when comparing the RH-LH and LH-LH regions in manipulation and drawing states. This is an indication that the flow of information decreases for competent users in the left hemisphere from the left and right hemisphere electrodes. Overall, Novice users show significant mean differences in all three states compared to competent users. With competent users, the difference is not significant in all cases which indicates that **the competent user's brain is trained in a way that it allows certain regions to be more active than others with respect to tasks.**

Region-based Analysis

The region-based analysis in Table 7.5 shows that the information flow from F-C is high in competent users compare to novice users, and the IF from F-C is more than IF from C-F. The IF from F-C increased in the manipulation state compared to other states for every user, which indicates that more information is transferred from the frontal region. The IF from P-C is more for novice users compared to competent users. The IF from F-P increases from rest to manipulation state for novice users but decreases for competent users. From the analysis in manipulation state, **maximum information was transferred through frontal electrodes. The reason for the maximum activation of the frontal scalp regions can be because this region is associated with reasoning, planning, and problem-solving [456].**

In Table 7.5, the highlighted blue color shows an increase in information flow from drawing state, and red color shows a decrease in IF. The analysis clearly shows that competent users' IF increased in some region and decreased in others whereas, for novice users, the information flow has increased in manipulation state from drawing state with some exception (Novice-User 5). It is also an indication that competent

Table 7.5: Information flow among F, C and P nodes during rest, drawing and manipulation states (NU: Novice-User, CU: Competent-User, D:Drawing, R:Rest, M:Manipulation)

Users	States	F-C	C-F	F-P	P-F	P-C	C-P	Max	Min
NU 2	**R**	0.0061	0.0121	0.0111	0.0037	0.0048	0.0058	C-F	P-F
	D	0.0611	0.0502	0.0709	0.0224	0.0330	0.0637	F-P	P-F
	M	**0.1097**	**0.0919**	**0.1350**	**0.1091**	**0.0987**	**0.1139**	F-P	C-F
NU 5	**R**	0.0212	0.0162	0.0135	0.0313	0.0372	0.0242	P-C	F-P
	D	0.0741	0.0555	0.0434	0.0386	0.0964	0.0730	P-C	P-F
	M	**0.1080**	**0.0797**	**0.0557**	**0.0579**	**0.0431**	**0.0636**	F-C	P-C
CU 2	**R**	0.0242	0.0134	0.0356	0.0406	0.0606	0.0689	C-P	C-F
	D	0.2260	0.0752	0.1570	0.0739	0.0735	0.1027	F-C	P-C
	M	**0.1631**	**0.0577**	**0.0952**	**0.0642**	**0.0782**	**0.0590**	F-C	C-F
CU 3	**R**	0.0465	0.0325	0.0175	0.0254	0.0359	0.0446	F-C	F-P
	D	0.0736	0.0702	0.1154	0.1516	0.0622	0.0606	P-F	C-P
	M	**0.1792**	**0.1215**	**0.1027**	**0.1478**	**0.0526**	**0.1020**	F-C	P-C

users have developed the capability to control the performance of a specific brain region to perform a particular task effectively.

In the rest state most of the IF is through the central region and the least information transfer was in between frontal and parietal lobes for all users. All users show maximum information transfer from the frontal cortex in manipulation state and minimum IF has been seen in the central cortex in manipulation state. **In drawing state maximum IF has been seen through the frontal to the central cortex**. The parietal and occipital lobes have shown minimum activation in sending information except for Competent-User 3.

Table 7.6 shows the comparison of mean information flow between F, C and P nodes in various states averaged across all the novice and competent users. The results clearly shows the difference between novice and competent user information flow between different brain regions.

Table 7.6: Information flow among F, C and P nodes during rest, drawing and manipulation states averaged across all users

User	States	F-C	C-F	F-P	P-F	P-C	C-P
	Rest	0.0136	0.0141	0.0123	0.0175	0.0210	0.0150
Novice	Drawing	0.0676	0.0528	0.0571	0.0305	0.0647	0.0684
	Manipulation	0.1088	0.0858	0.0954	0.0835	0.0709	0.0888
	Rest	0.0353	0.0229	0.0266	0.0330	0.0482	0.0568
Competent	Drawing	0.1498	0.0727	0.1362	0.1127	0.0679	0.0816
	Manipulation	0.1712	0.0896	0.0990	0.1060	0.0654	0.0805

Table 7.7: Classification results using all features

Classifiers	Accuracy	Sensitivity	Specificity	F-measure
LDA	88%	0.94	0.80	0.91
KNN	71%	0.75	0.65	0.76
NB	69%	0.80	0.53	0.76
SVM	82%	0.87	0.74	0.85
Tree	83%	0.88	0.75	0.86

7.3.3 Classification Results

The classification results are shown in Table 7.7 - 7.9. The data from all eight subjects were used in this experiment. The data-set was divided into training and testing data-sets and the results are evaluated with 5-fold cross-validation. Table 7.7 shows the classification results of different classifiers with all features as input. The LDA classifier shows maximum classification accuracy of 88% with an F-measure of 0.91. The NB classifier shows the least classification accuracy. Sequential forward search has been used to select the best features and the results are shown in Table 7.8. The k-NN (k=3) shows the best classification accuracy of 95% with just 11 features. The classification accuracy for all classifiers increased after applying SFS for feature selection except for LDA. NB classifier accuracy also increased from 69% to 78% by selecting only 4 features. Table 7.9 shows the timing analysis to select five features using SFS. The SVM classifier takes the maximum time (818.47 sec)

Table 7.8: Classification results after feature selection by SFS

Classifiers	Accuracy	Sensitivity	Specificity	F-measure	Features
LDA	88%	0.95	0.76	0.90	40.00
KNN	95%	0.96	0.92	0.96	11.00
NB	78%	0.93	0.56	0.84	4.00
SVM	88%	0.95	0.76	0.90	20.00
Tree	89%	0.90	0.88	0.91	16.00

Table 7.9: Time taken by the algorithm to select features using various classifiers

Classifiers	Accuracy	Feature number	Time in sec
LDA	80%	12 32 35 37 41	7.96
KNN	92%	23 29 32 35 41	7.30
NB	78%	29 30 32 35 41	7.52
SVM	81%	19 20 29 32 38	818.47
Tree	86%	9 17 22 29 35	7.71

and k-NN takes the least time (7.30 sec) to select five features. k-NN also gives the maximum classification accuracy among all other algorithms. The results also show some common features selected by different classification algorithms The feature number 29, 32, 35 and 41 selected by k-NN also appears in the features selected by other algorithms. The feature 29 is the characteristic path length and 32, 35, and 41 belong to the mean information flow. The feature 32, 35, and 41 correspond to the mean information flow of electrodes F3, T7 and F8 respectively. It shows that the channel that contributes to maximum variance between novice and competent users are F3, and F4 from frontal lobe and T8 from temporal lobe of the brain. Fig. 7.16 shows the mean square error (MSE) of various classification algorithms with respect to the features selected. The graph shows that the error decreased when the number of features increased. After selecting a certain number of features, the change in MSE was not significant. The least MSE of 8% was observed with k-NN classifier and NB shows the maximum MSE compared to other classifiers.

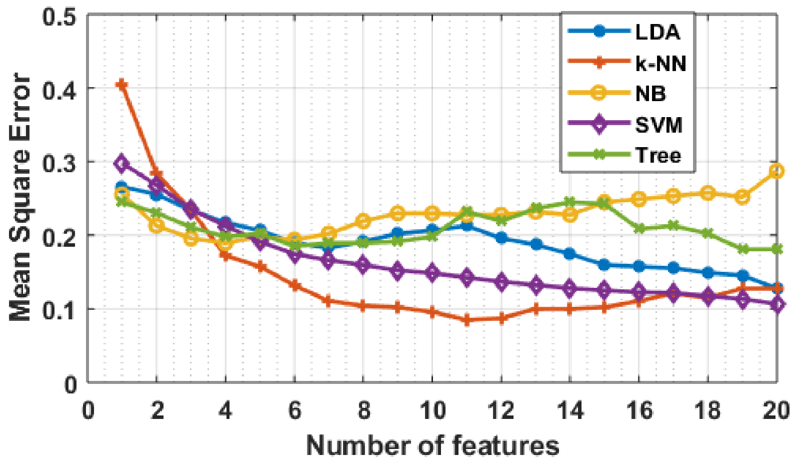


Figure 7.16: Mean square error with respect to features selected

7.4 Conclusion

In this chapter, a novice/competent user study has been presented that uses NTE to construct FBN and estimate the information flow patterns. Both binary and weighted directed FBNs were used for the analysis. Using the techniques of signal and information processing to construct FBNs and applying graph theory and statistical analysis, we have estimated the cognitive activity and information flow pattern for novice and competent users by measuring the Connectivity density, clustering coefficient, characteristic path length, motif count, node strength and small-worldness. The results showed a significant difference in information processing between novice and competent users. The findings from the analysis are listed below:

1 The main difference was observed in the manipulation state, novice users information flow patterns changed in manipulation state compared to competent users. The connectivity density, motif count, clustering coefficient all showed the same trend.

2 The network density increased from the baseline for both novice and competent users, but the change is more for novice users compared to competent users in drawing and manipulation states.

3 Most of the activity was focused in the frontal region, which indicates the use of short-term memory.

4 The small-worldness shows that competent users have relatively high global and local efficiency of information transfer than novices.

5 The hemisphere analysis shows that the information flow has increased in both hemispheres for novice users, but competent users managed to control the information flow according to the task.

6 In the lobe-wise analysis, the frontal lobe was most active in sending and receiving information in drawing and manipulation states for all users.

Competent users have developed the capability which gives them control of different brain regions for the different tasks, unlike novice users where almost all regions became active. The results indicate that competent users were more relaxed in both drawing and manipulation states, whereas novice users put more effort into manipulation state than drawing state. The feature selection algorithm also helps us to find the features that maximize the novice competent difference. These features belong to the frontal and temporal lobes of the brain. The results are aligned with the earlier findings of Kavalki & Gero [443] regarding differences between novice and competent users concurrent cognitive processing.

Chapter 8

Comparative Analysis of Cognitive Activity: using Power Spectral Density

This chapter presents the analysis of power spectral density in estimating cognitive activity of novice and competent users in using a traditional unimodal input system and MMIS. The theta and alpha bands of EEG signals have been used to measure the cognitive activity of the user. The results have been presented at the 26th International Conference on Systems Engineering (ICSEng) 2018 held in Sydney, Australia, with a paper titled "Analyzing Novice and Expert User's Cognitive Load in using a Multi-Modal Interface System" [48].

- Baig, Muhammad Zeeshan, and Manolya Kavakli. "Analyzing Novice and Expert User's Cognitive Load in using a Multi-Modal Interface System." *In 2018 26th International Conference on Systems Engineering (ICSEng)*, pp. 1-7. IEEE, 2018.

8.1 Introduction

Effects of human-factors on Human-Computer Interactions (HCI) has been studied for a while and the centre of interest was the performance-based measures, such as task completion timings and accuracies, to draw the conclusions. These types of performance measures can give us a global understanding but they failed to shed light on the individual variations amongst participants [457].

The individual differences were recorded with the measures of human expertise and experimental task analysis which are subjective and rely heavily on self-reported data and are biased based on culture, age, personal behaviour, and overestimation. In this chapter, we have used a multi-modal interface system for 3D modelling (xDeSIGN) [43] and EEG to replace the traditional self-reported data. We have compared the traditional unimodal input (Keyboard and mouse) and multi-modal input (Speech and gesture). We have also used the qualitative self-reported data along with the quantitative data provided by the EEG analysis.

In the literature, most of the research work is focused on analyzing games and programming expertise [361, 422]. We have used the EEG signals to analyze the relationship between human cognition and CAD competency when the user is designing a 3D object using both unimodal and multimodal inputs. To the best of our knowledge, this is the first study that explores the use of EEG to estimate the cognitive activity associated with CAD competency and the design tasks in multi-modal interaction.

An EEG signal can be divided into different bands based on frequency ranges and each band has its own characteristics as shown in Table 8.1. In this chapter, we have used the EEG Power Spectral Density (PSD) to investigate the user's cognitive activity while the user performs a 3D object modelling task. In the literature, it is stated that the θ band over the parietal lobe is better associated with low-and high-cognitive load tasks compared to β and α bands [462]. A high β-band activity over occipital lobe has been associated with the high-visual-attention tasks [460]. We construct the

EEG Band	Frequency Range	Characteristic
Theta (θ)	4-8 Hz	Emotional Information [458]
Alpha (α)	8-13 Hz	Cognitive Processing [459]
Beta (β)	13-30 Hz	Logical Thinking, Conscious Thought, Stimulating effect [460]
Gamma (γ)	30-50 Hz	Memory, Linguistic Processing, Cognition, Attention [461]

Table 8.1: EEG bands and corresponding characteristics

following hypothesis based on the literature review:

Hypothesis 1: After reviewing the literature, we identified a hypothesis that α-band activity is inversely, and θ- and β-band activity are directly correlated with mental effort.

Hypothesis 2: We hypothesized that the keyboard state activity levels would be lower compared to gesture state cognitive activity levels.

Hypothesis 3: The competent users face difficulty in adjusting to new input modes compared to novice users.

In this chapter, we find the answers to the following research problems:

RQ 5.1 What are the differences in cognitive activity between multimodal and unimodal systems?

RQ 5.2 Does competency play a role when a new set of inputs were used for a predefined task?

8.2 Methodology

In this chapter, we have used a multimodal interface system (MMIS) xDe-SIGN v2 [42, 43] to draw and manipulate 3D objects and to draw comparison with the traditional use of a unimodal AutoCAD system. The MMIS allows the users to design objects in 3D, using AutoCAD commands as well as speech and gesture.

We analyzed the cognitive activity of the participants through EEG signals in an MMIS. Since the literature focusing on 3D object manipulation is (very) limited, more detailed investigation is needed to study the cognitive states of novices and competent users in a 3D-modelling environment. The experimental details and EEG signal pre-processing steps are provided in Chapter 5. The EEG data from 11 participants (3 Competent and 8 Novice) have been used for this study. For convenience, we named the first task that used traditional (mouse and keyboard) input was named "Keyboard" and the second task that used multimodal (speech and gesture) input was named "Gesture" based drawing state.

We extracted 50 epochs from data in the keyboard, and gesture state tasks by performing back-to-back epoching with a 0.5-sec difference between epochs. The response time for a modelling action varies from 0.5 to 4 seconds and we made sure that a minimum of one modelling action must be performed in an epoch through the video log. An epoch size of 4 sec has been selected in this study. 25% of the users were competent users of AutoCAD with over 2 years of experience in 3D modelling. Two users deleted the objects when drawing and these actions were also considered among design actions in this experiment. According to our findings, there was no significant correlation found between competency level and task completion time.

8.2.1 Power Spectral Density

After the pre-processing stage, Power Spectral Density (PSD) estimate was calculated for each epoch of 4-seconds using Welch's method. The Welch's method is prefered over other methods for PSD because Welch's method reduces noise in the estimated PSD in exchange for reducing the frequency resolution. A hamming window of size 64 samples was used without overlapping. The method can be divided into 4 steps [463]:

1 Partition the data into K segments:

Segment 1:$x(0), x(1), ..., x(M-1)$

Segment 2:$x(S), x(S+1), ..., x(S+M-1)$

.

.

.

Segment 1:$x(N-M), x(N-M+1), ..., x(N-1)$

where M is the number of points in each segment, S is the shift or gap between segments and K is the total number of segments.

2 For each segment ($k = 1$ to K), compute a windowed Discrete Fourier Transform (DFT) at some frequency $f = i/M$ with $-(M/2-1) \leq i \leq M/2$:

$$X_k(v) = \sum_m x(m)w(m)exp(-j2\pi vm) \qquad Eq(8.1)$$

where, $m = (k-1)S, ..., M + (k-1)S - 1$ and $w(m)$ is the window function, which is hamming window in this case.

3 For each segment ($k = 1$ to K, from the modified periodogram value, $P_k(f)$, from the DFT:

$$P_k(v) = \frac{1}{W}|X_k(v)|^2 \qquad Eq(8.2)$$

where W is:

$$W = \sum_{m=0}^{M} w^2(m) \qquad Eq(8.3)$$

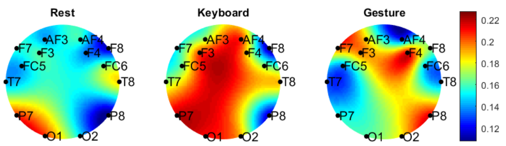
4 Average the periodogram values to obtain Welch's estimate of PSD:

$$S_x(v) = \frac{1}{K}\sum_{k=1}^{K} P_k(v) \qquad Eq(8.4)$$

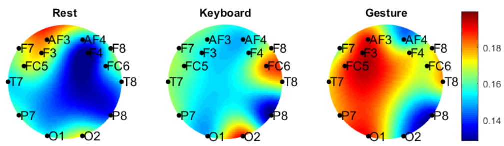
To compare the data of all users, we calculated the relative PSD by dividing the PSD of each band with the sum of PSD of all frequencies.

8.3 Results and Discussion

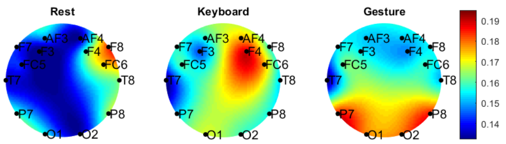
We performed the analysis on all the 50 epochs extracted from the EEG signals of all participants. The PSD estimate averaged over all 50 trials for 6 users (> 50%) have been shown in Fig. 8.1 - 8.2. Three users were competent and all other users were novice users of 3D modelling software (See appendix E.1 for the results of all the users).



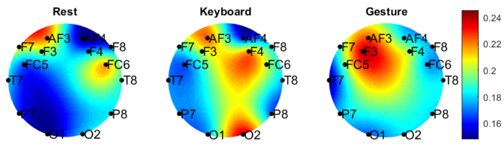
(a) Competent-User 1



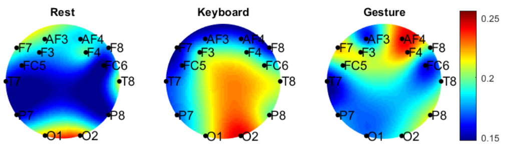
(b) Novice-User 1



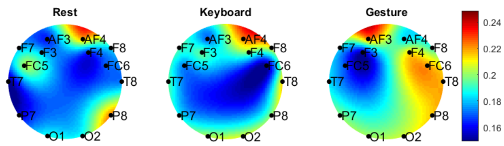
(c) Competent-User 2



(d) Novice-User 2



(e) Competent-User 3



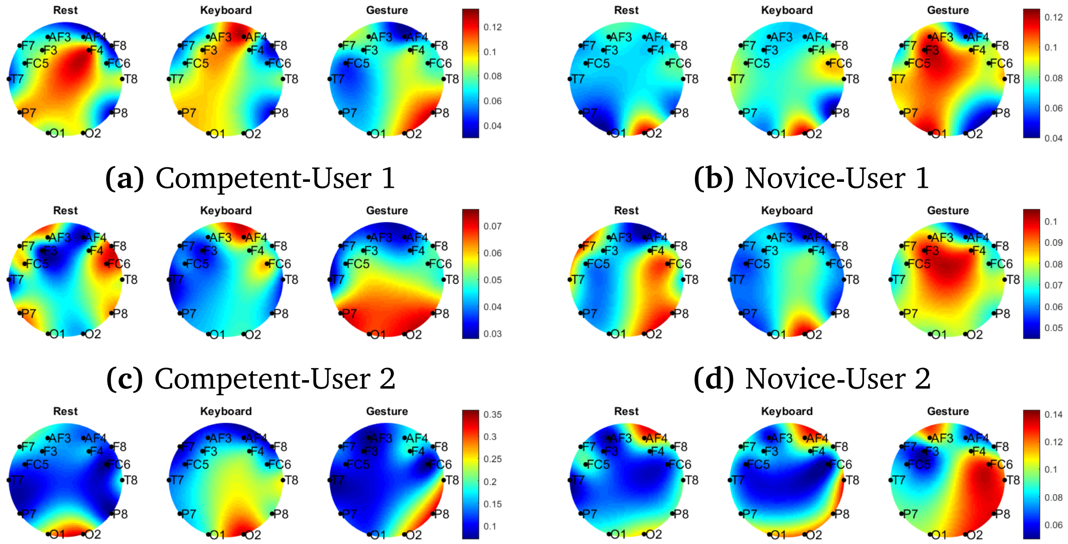
(f) Novice-User 3

Figure 8.1: Averaged theta activity at rest, keyboard and gesture states for three competent and three novice users

Result 1: The average theta band activity for 3 competent and 3 novice users, as shown in Fig. 8.1, demonstrates that the θ-band activity was more intense in the keyboard state compared to gesture state for competent users. However, for novice users, the θ-band activity increased in gesture state rather than keyboard state.

The theta band activity corresponds to the emotional information processing which means that novice users were processing relatively more emotional information and were possibly more stressed in the gesture state compared to competent users. The θ-band activity was mainly observed on the frontal and occipital lobes for most of the users which can be interpreted as the users were using their short term memory more with the motor tasks.

Result 2: The α-band activity of the users is shown in Fig. 8.2. It can be seen from the figure that the α-band activity decreased for almost all the users in the keyboard state compared to the rest state which shows the increase in mental effort in keyboard state from rest state. In gesture state, we observed a very unexpected pattern. On average, the α-band activity of novice users increased more in gesture state compared to keyboard state when compared with competent users. If we apply



(a) Competent-User 1 **(b)** Novice-User 1

(c) Competent-User 2 **(d)** Novice-User 2

(e) Competent-User 3 **(f)** Novice-User 3

Figure 8.2: Averaged alpha activity at rest, keyboard and gesture states for three competent and three novice users

the hypothesis that α-band activity is inversely related to mental effort then the results can be interpreted as the mental effort of competent users increased when they were using gesture and speech to draw the 3D objects unlike novice users.

Result 3: As all of the users were using the xDE-SIGN (multi-modal interface system) for the first time so they are all considered to be novice users of xDE-SIGN, but the competent users of AutoCAD were finding it hard to use the multi-modal input compared to the novice users. Based on the α-band activity, we can say that it was relatively easy for the novice users to use the new set of inputs compared to competent users.

Table 8.2: Statistical validation of alpha band PSD of keyboard and gesture states

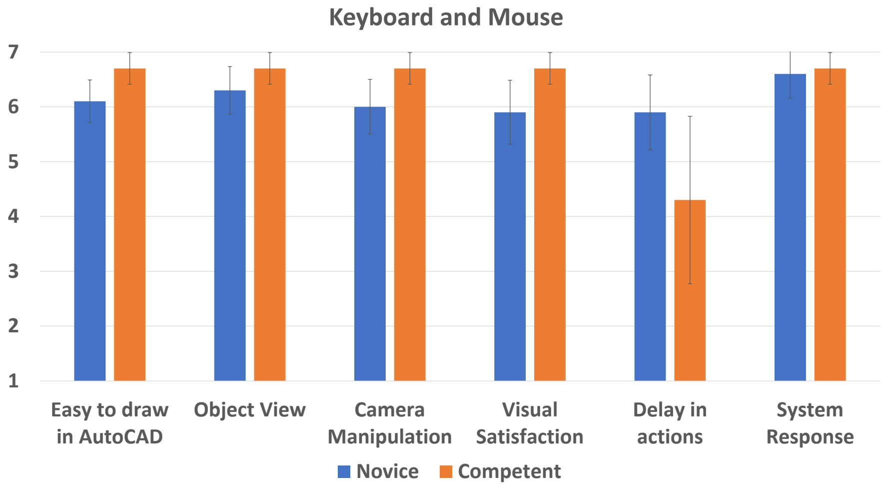
Users	Mean	CI		P
Competent-User 1	-0.04771	-0.0574	-0.03802	2.85E-13
Competent-User 2	0.006986	0.001907	0.012066	0.008025
Competent-User 3	-0.02094	-0.03402	-0.00785	0.002312
Novice-User 1	0.026834	0.011602	0.042067	0.000888
Novice-User 2	-0.00733	-0.01164	-0.00302	0.001273
Novice-User 3	-0.04591	-0.05621	-0.0356	7.02E-12
Novice-User 4	-0.02645	-0.03396	-0.01895	4.97E-09
Novice-User 5	-0.02223	-0.03268	-0.01178	8.8E-05
Novice-User 6	-0.0189	-0.02777	-0.01003	8.62E-05
Novice-User 7	0.013459	0.004327	0.022592	0.004708
Novice-User 8	0.01033	0.001146	0.019514	0.028273

Result 4: Table 8.2 shows the results of 2-sample t-test with unequal variance at $\alpha = 0.05$. The t-test was applied on alpha band PSD values of keyboard and gesture state averaged across the 14 electrodes. The results showed that for 11 users the change between alpha activity in keyboard state and gesture state was significant i.e. $p < 0.05$.

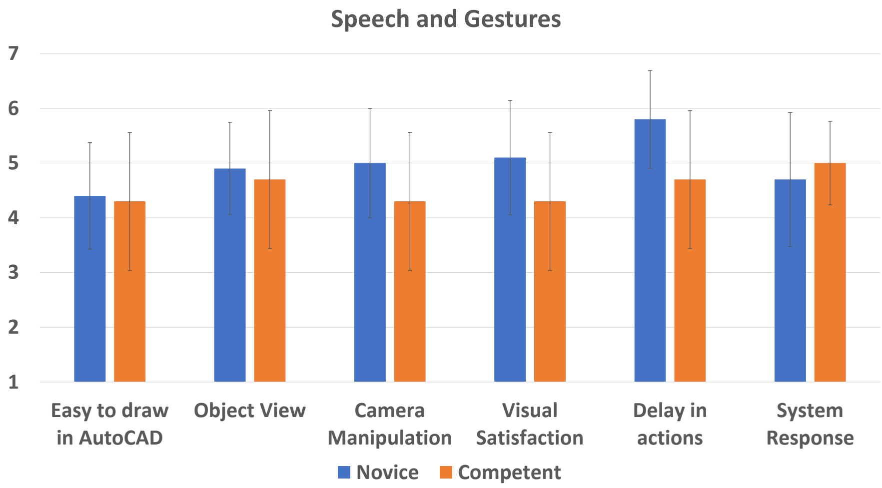
For evaluation of the system, we also collected data from questionnaires. The questionnaires were completed by all 11 users at the end. The recorded response varies from 1 to 7 with 1 being bad or strong disagreement and 7 being excellent or strong agreement. From the answers to questionnaires, we investigated the performance of the system, fatigue level of the user and perception in HCI using unimodal (keyboard/mouse) and multimodal (speech/gesture).

Result 5: The questionnaire results showed that it was difficult for all the users to draw using multimodal (speech/gesture) compared to unimodal (keyboard/mouse).

This is due to the fact that they were using an MMIS for the first time and mouse and keyboard inputs were used by the users traditionally all the time. The responses of the users that correspond to performance evaluation are shown in Fig 8.3.



(a) Keyboard and mouse

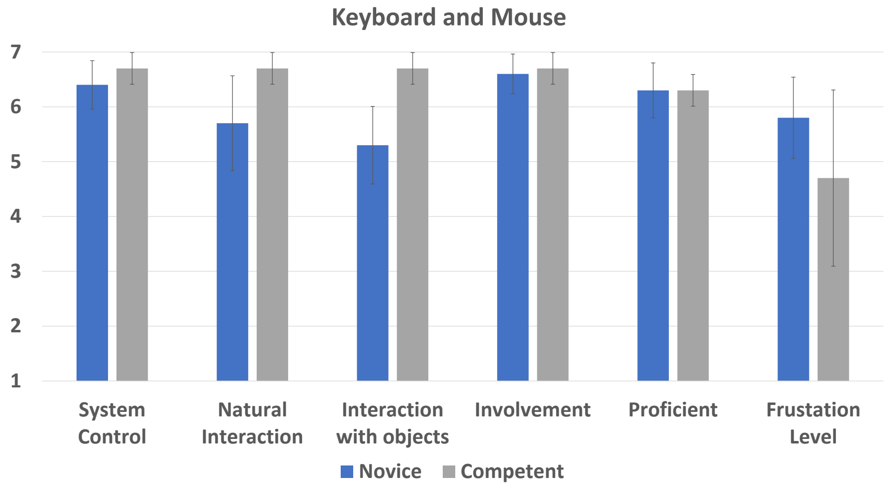


(b) Speech and gesture

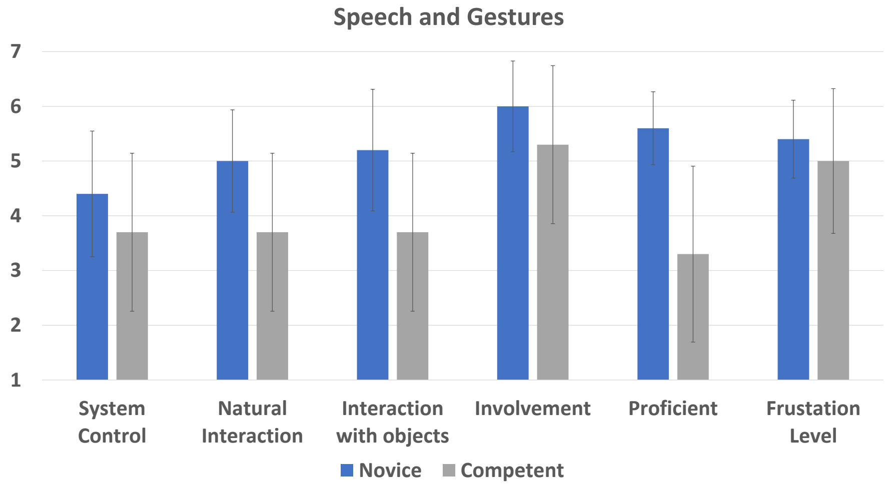
Figure 8.3: Command Performance Evaluation

Result 6: Competent users were more comfortable in using AutoCAD with traditional inputs compared to novice users.

The one exception can be seen in the case of delay in action, where novice users find the system acceptable but competent users were feeling some kind of delay in action. In the case of speech and gesture input, both the novices and competent users find it hard to use compared to keyboard and mouse inputs as shown in the questionnaire responses recorded in Fig. 8.3b. A difference of opinion was observed when we compared the novices and competent users responses about the performance



(a) Keyboard and mouse



(b) Speech and gesture

Figure 8.4: User perception in interacting

of MMIS.

Result 7: Novice users find it relatively easier to use the MMIS compared to competent users.

The same trend between novices and competent users has been observed when asked about user perception in interacting with the system as shown in Fig. 8.4. The competent users were more comfortable than the novice users in interacting with the system with unimodal (keyboard and mouse) input. However, in multimodal (gesture and speech) inputs, the novice users were more relaxed and comfortable compared to competent users.

Result 8: A similar correlation between multi-modal input was seen in the α-band activity response which highlights the finding that a novice user has a good chance to learn and use a new set of inputs in a lesser amount of time compared to competent users.

8.4 Conclusion

In this chapter, we have presented a comparative analysis of novice and competent users in using a multi-modal interface system that uses speech and gesture inputs to draw 3D objects. We have recorded EEG signals to see the cognitive activity in the brain. The participants were asked to draw a simple 3D table using keyboard/mouse and speech/gesture. The results showed that using speech and gesture for drawing and manipulating 3D objects is more difficult compared to using keyboard and mouse. We found that it was more difficult for competent users to use a new set of inputs to draw 3D objects than novice users. When comparing the results with a competent user for using multi-modal inputs, we observed that novice users' alpha band activity increased. This means that their mental effort decreased. The same kind of trend was seen in the questionnaire responses. The novice user finds the speech and gesture input more usable than competent users which is an indication of that it is easier for a novice to use a new set of inputs i.e. speech and gesture for a particular application compared to a competent user. This is an important finding regarding competency and an extensive investigation for underlying reasons is needed to validate our hypothesis that novice users could adapt to new input modes more easily than competent users.

Chapter 9

Connectivity Analysis: using Transfer Entropy and Functional Brain Networks

In this chapter, a comparative analysis of the cognitive activity in using both unimodal and multimodal interface systems is presented. The aim is to find the relationships between a 3D modeling task and the user's cognitive activity. We have performed the connectivity analysis of Functional Brain Networks (FBNs) constructed from transfer entropy. The results of the study are accepted for publication at *the 28th International Conference on Information Systems Development (ISD2019)* to be held in Toulon, France.

9.1 Introduction

The advancements in virtual and augmented reality have presented the users the opportunity to use the multi-modal input rather than the traditional mouse and keyboard input. The multi-modal input is a vital part of smartphones, VR headsets, and game boxes. The applications of multi-modal input are also used in the 3D modeling and Computer-aided design (CAD), Computer-aided manufacturing (CAM) and Computer Aided Engineering (CAE) industries as well.

Modeling systems with multiple inputs are preferred by the users to design a 3D object. These inputs can be speech, touch, facial expressions, gestures, and handwriting [464]. The availability of multiple inputs improves the usability of Human-Computer Interaction (HCI) systems. To develop such systems, the system requires the obtaining of human information effectively. To achieve this, we need to overcome technological, physiological, and psychological barriers. In the case of humans, we utilize all available modalities in which our brain's cognitive and perceptual functions are perfectly synchronized.

In this chapter, we shed light on the user-dependent factors affecting MMIS. We analyzed the cognitive activity and information flow patterns when a user is using the traditional multi-modal input to draw 3D objects. We have used transfer entropy (TE) of the EEG signals to analyze the connectivity between electrodes and then used the TE matrix to generate both binary and weighted functional brain networks (FBNs). We have used the graph theory methods to study characteristics the FBNs.

The study tries to find the human dependent factors that affect the performance of a user in a multimodal interaction. We investigate individual differences in information processing when the user is using traditional unimodal and multimodal input systems. This study aims at finding the differences between competent users and novices cognitive activities in interface modalities as well as proposing quantitative measures to evaluate HCI systems. Other researchers generally provided qualitative evidence on self-reported data, whereas we have provided quantitative evidence in this chapter which is aligned with the previous work.

Our primary goal is to study the fundamental structure of the brain and its response to the user interaction task and input modality and to discover the differences in information flow patterns between various users in using speech and gestures for design and manipulation of an object in 3D space. The secondary goal is to investigate how we can use these information processing differences to improve the overall system performance in Human-Computer Interaction (HCI). In this chapter, we have answered the following research questions. In this chapter, we find the answers to the following research problems:

RQ 5.1 What are the differences in cognitive activity between multimodal and unimodal systems?

- **RQ 5.1.1** Are there any differences in information flow patterns and cognitive activity in HCI using speech and gestures, based on the user's competency in using a CAD system?
- **RQ 5.1.2** Do novices and competent users employ different cognitive processing models in 3D object manipulation?

RQ 5.2 Does competency play a role when a new set of inputs were used for a predefined task?

- **RQ 5.2.1** Do the differences in information flow patterns give any advantage in performance to the user in certain tasks?

9.2 Methodology

In this chapter, we analyze the cognitive activity of the participants through EEG signals while using a multi-modal interface system (xDeSIGN v2) [42, 43]. We have used a connectivity measure to estimate the information flow between electrodes and then used FBNs and graph theory methods to study the behavior. We apply TE to the analysis of information flow patterns between novice and competent users. We have used normalized TE values to construct both binary and weighted directional FBNs.

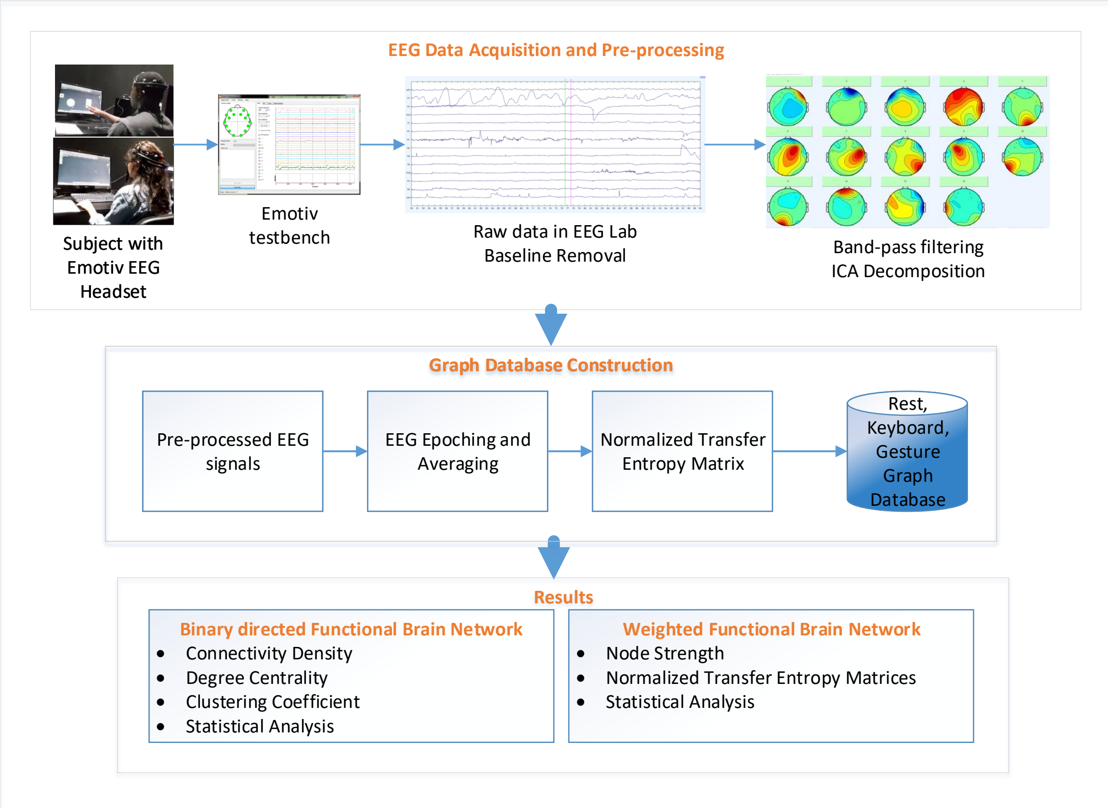
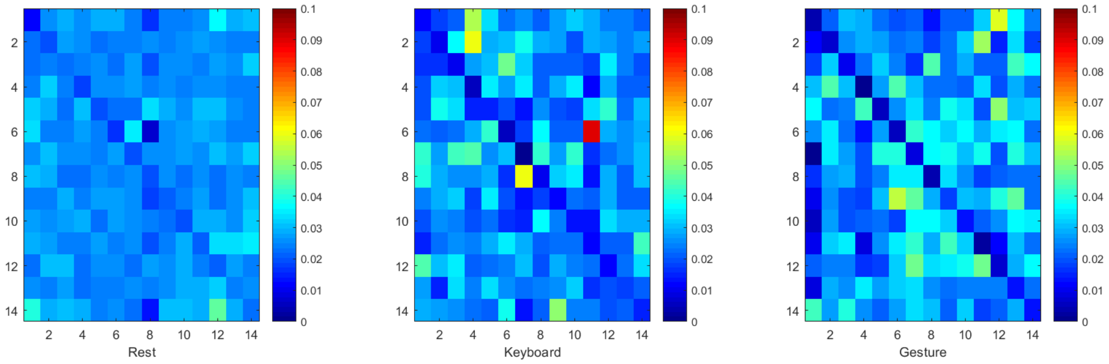


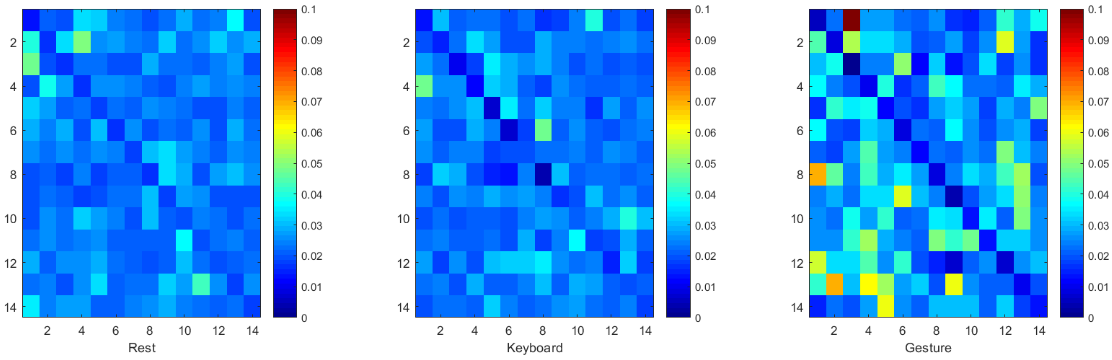
Figure 9.1: Normalized Transfer Entropy Framework

After constructing an FBN, we have applied graph theory measures and statistical analysis to quantify the information flow patterns.

The experimental details, EEG data collection, and preprocessing steps are described in Chapter 5. For convenience, we named the first task (drawing with keyboard and mouse), "Keyboard" and the second task (drawing with speech and gestures) "Gesture" drawing state. To analyze FBNs, we have used graph analysis measures, such as the Connectivity density, clustering coefficient, and node strength. Fig. 9.1 shows is an analysis framework to compute FBN metrics. Data from 12 participants (4 Competent and 8 Novice) were used in this experiment. The data for 8 of the participants is the same as used in previous chapters and 4 additional participant data has been collected for this analysis.



(a) Averaged across all novice users



(b) Averaged across all competent users

Figure 9.2: NTE connectivity matrices during rest, keyboard and gesture drawing states

9.3 Results and Discussion

9.3.1 Analysis of Directed Binary FBNs

This section shows the results of the analysis performed on binary directed FBNs constructed using the connectivity matrix calculated by NTE. The Fig 9.2 shows the NTE connectivity matrices during various states averaged across novice and competent users.

Fig. 9.2 shows that the brain connectivity increased progressively as the cognitive activity increased from rest state to keyboard and gesture drawing states. Fig. 9.3 shows the connectivity density for all 12 users. **It shows that the connectivity density is higher in keyboard and gesture drawing states compared to the baseline resting state,** inferring more connections between electrodes to accommodate more

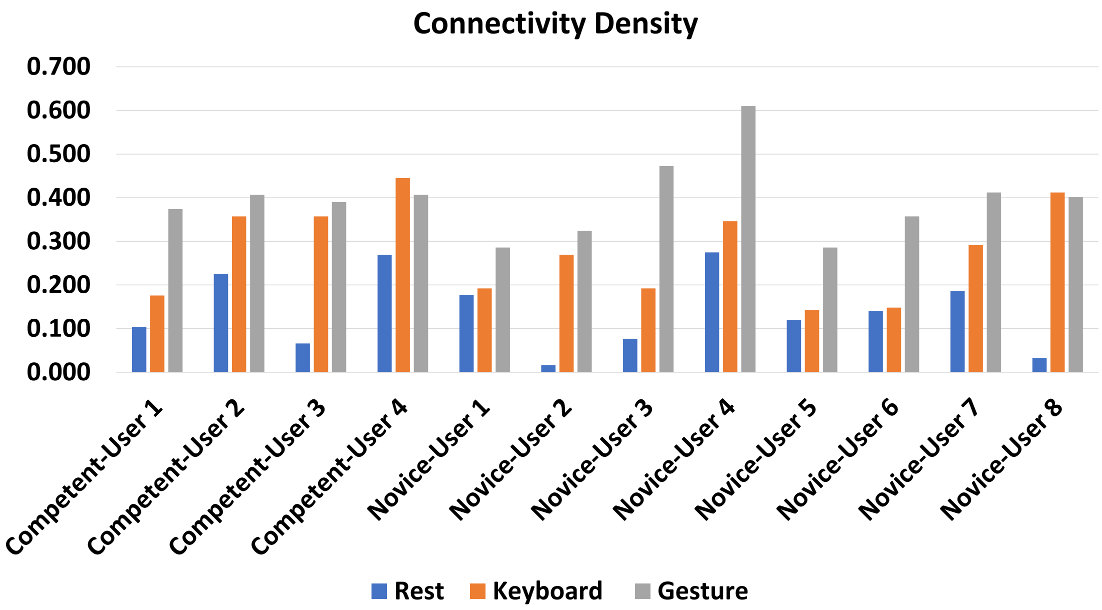


Figure 9.3: Comparison of connected density for all users during rest, keyboard and gesture drawing states

active information flow which means that the brain recruits a higher number of neurons to facilitate high cognitive activity processes in keyboard and gesture drawing states.

If we compare the keyboard and gesture connectivity density, **an increase in connectivity density is seen except for Competent-User 4 and Novice-User 8.** The increase in connectivity density in gesture drawing state from the keyboard drawing state is due to the execution of a number of simultaneous processes (motor cognition, visual processing, designing). Fig. 9.4 shows the statistical significance of the connectivity density of all the users during the three states. The lines extended out from the mean show the confidence intervals, which means that the group means are significantly different because there is no overlap.

The clustering coefficient was calculated from the binary directed FBNs across the electrodes for all participants, and the results of two participants (Novice-User 3 and 8) have been shown in Fig. 9.5. It can be seen that **the clustering coefficient value increases for almost all the electrodes during the cognitive activity compared**

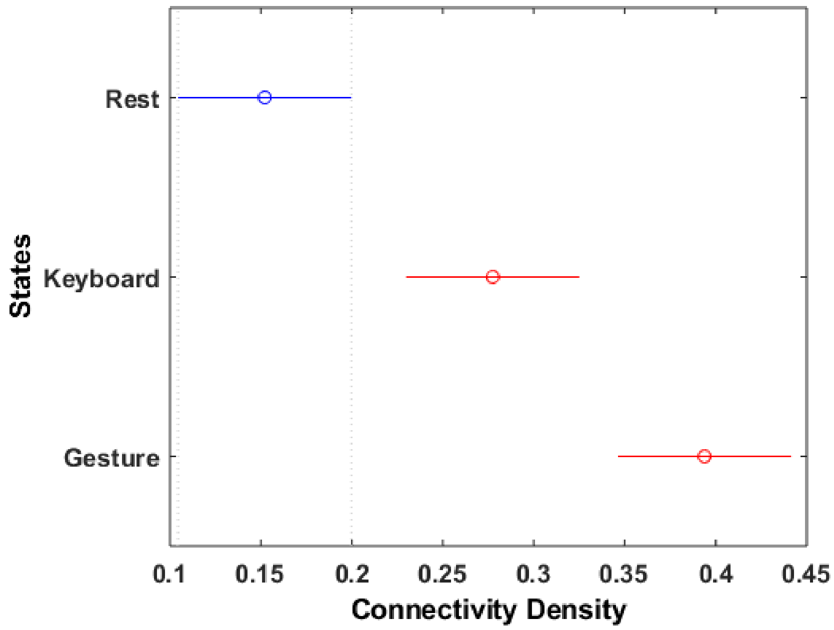


Figure 9.4: Multiple comparison test of connectivity density group mean for 12 users during rest, keyboard and gesture drawing states

to the rest state. This shows that each electrode is communicating effectively with its neighboring electrodes to form clusters which represent an increase in local efficiency of the information transfer between electrodes. The clustering coefficient values also show that the electrodes from right side frontal lobe and left side occipital and parietal lobes are communicating more with the nearest neighbors.

Table 9.1 shows the statistical significance of the difference between means of clustering coefficient across electrodes during keyboard and gesture drawing states for all participants computed using a two-tailed t-test at $\alpha = 0.05$. The results show that **the mean difference is significantly different in keyboard and gesture drawing states for almost all the users except for Competent-User 1, 2, and 3.** The mean difference is significant for all the users in rest and keyboard and rest and gesture drawing states.

We have also calculated the degree centrality for all participants from the NTE matrix. Degree centrality shows the importance of a node in the network. Fig. 9.6 shows the topographic map of degree centrality data of a novice and a competent user

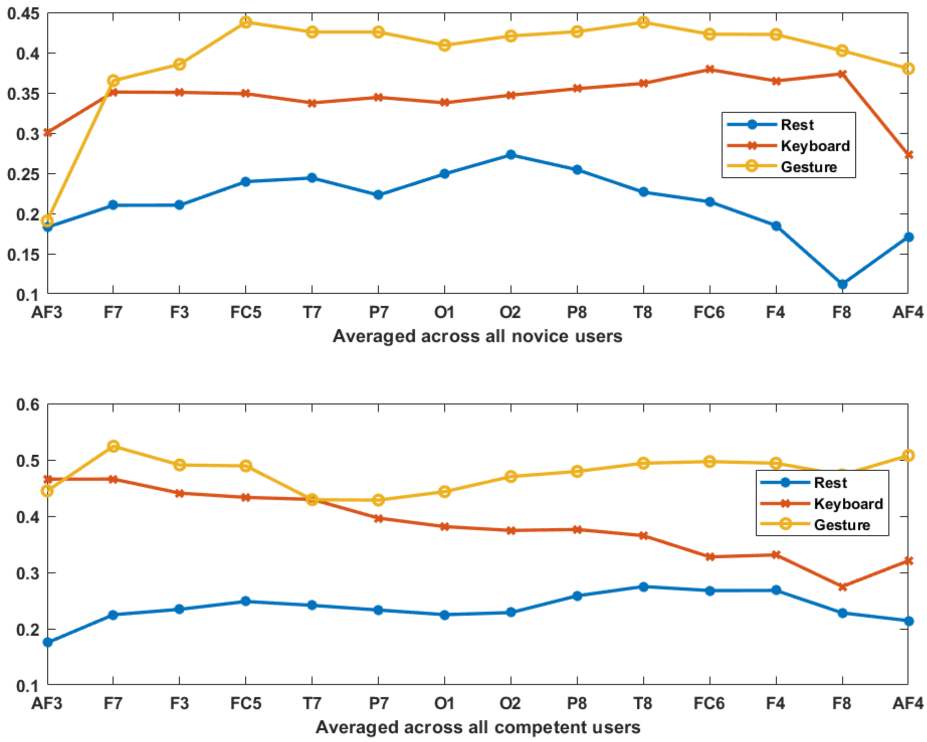
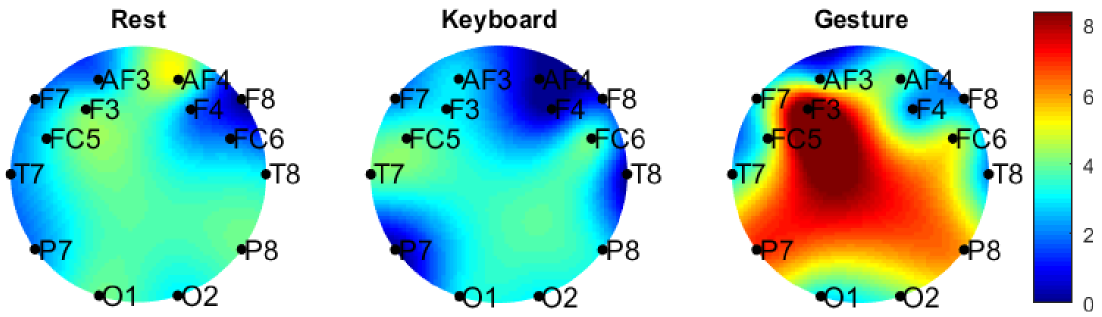


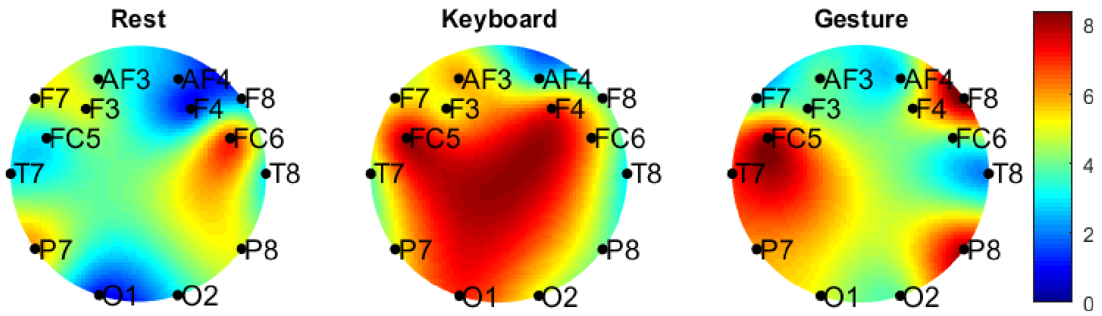
Figure 9.5: Clustering coefficient across electrodes for during rest, keyboard and gesture drawing states

Table 9.1: Statistical validation of clustering coefficient values for all user during keyboard and gesture drawing states

Users	Mean Diff.	95% CI		P
Competent-User 1	0.0744	-0.0110	0.1597	0.0824
Competent-User 2	-0.0180	-0.0609	0.0250	0.3828
Competent-User 3	0.0173	-0.0425	0.0771	0.5421
Competent-User 4	0.1395	0.1036	0.1753	0.0000
Novice-User 1	0.0655	0.0396	0.0913	0.0001
Novice-User 2	0.1728	0.0575	0.2881	0.0065
Novice-User 3	0.0189	-0.0364	0.0741	0.0474
Novice-User 4	0.1855	0.1280	0.2431	0.0000
Novice-User 5	0.1403	0.0591	0.2214	0.0025
Novice-User 6	0.3033	0.2722	0.3344	0.0000
Novice-User 7	0.2020	0.1652	0.2388	0.0000
Novice-User 8	-0.0512	-0.1061	0.0037	0.050



(a) Novice-User 6



(b) Competent-User 4

Figure 9.6: Degree centrality topographical plot of two users during rest, keyboard and gesture drawing states

in a 2D circular view of all states. The topographical map was customized such that the color map scales from minimum degree centrality value to maximum degree centrality value to visualize subtle changes. The blue color represents the minimum value, and the red represents the maximum value of degree centrality. The user in the drawing state shows greater engagement than in the resting state. The major focus is seen in the frontal area of the brain. Both drawing with keyboard/mouse and speech/gesture requires intense visual attention, which shows elicited central and frontal areas in the brain. **Some competent users of AutoCAD show a small variation in degree centrality across electrodes in keyboard and gesture drawing state compared to other users.** (Degree centrality of all the users is given in Appendix E.2 Fig. E.3.)

Table 9.2 shows the results of t-test at $\alpha = 0.05$ (two-tailed) between means of average degree centrality across electrodes in all three states for all the user and the results show that **the mean difference is significantly different across**

Table 9.2: Statistical validation of mean degree centrality across electrodes for all participants during rest, keyboard, and gesture drawing states

States		Mean Diff.	95% CI		P
Keyboard	Rest	1.5000	0.4778	2.5222	0.0080
Gesture	Rest	2.9762	1.9589	3.9935	0.0000
Gesture	Keyboard	1.4762	0.5625	2.3899	0.0045

all cognitive states. The maximum mean difference was observed in gesture drawing and rest states.

Table 9.3: Electrodes with maximum variance in keyboard and gesture drawing states using LDA

Graph Measure	Electrodes				
Clustering coefficient	AF4	O2	P8	T8	FC6
Degree Centrality	F7	FC5	T7	O2	P8

Linear Discriminant Analysis (LDA) was used to find the electrodes that showed the maximum variance between keyboard and gesture drawing task. The results are shown in Table 9.3. The results showed that **for the clustering coefficient, the maximum variation is seen in the electrodes that are on the right side of the hemisphere compared to the left hemisphere electrodes.** It can also be interpreted as the electrodes on the right-hemisphere form clusters more often than the left hemisphere electrodes. **In the case of degree centrality, the maximum difference is seen in the left frontal hemisphere (F7, FC5, T7) and back right hemisphere (O2, P8). This finding indicates that the receiving and transmitting of information are from these electrodes more than the other electrodes.**

9.3.2 Analysis of Directed Weighted FBNs

We have used the weighted FBNs to find the mean information flow and node strength for all the users. Fig. 9.7 shows the multiple comparison test of the average node strength of all the users across all electrodes in the three states. **The results clearly**

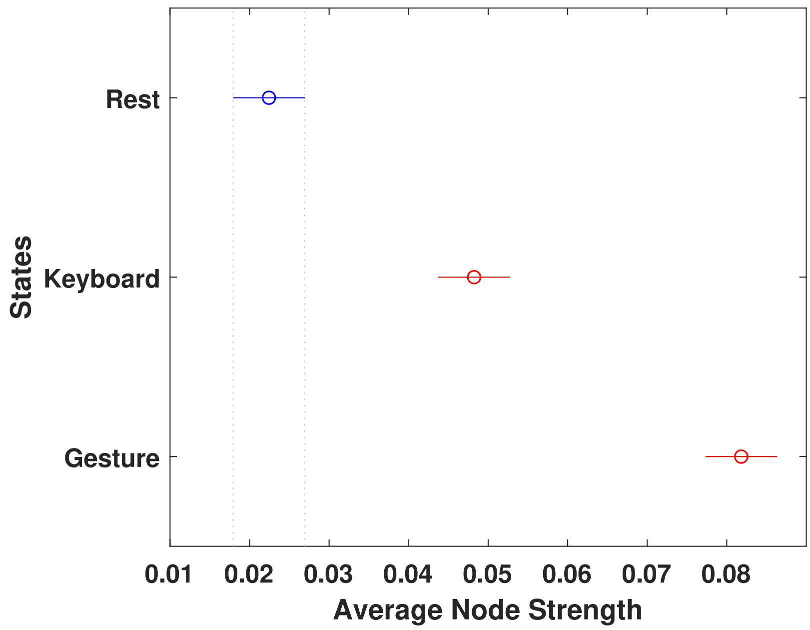


Figure 9.7: Multiple comparison test of node strength group mean for 12 users during rest, keyboard and gesture drawing states

show that the node strength in gesture drawing state is more than the keyboard drawing and rest state, and the results are statistically significant as seen in Fig. 9.7. The electrodes in the frontal lobes send and receive more information while drawing with gestures compared to the keyboard state. The total information flow from each electrode to all the other electrodes has been calculated by the row-wise summation of the NTE connectivity matrix.

Fig. 9.8 shows one-way analysis of variance (ANOVA) results on the mean information flow across electrodes for all participants. **The mean information flow increased from approximately 0.005 to 0.02 when the user was using multi-modal inputs instead of keyboard and mouse for drawing.** Table 9.4 shows the statistical significance of the difference in means of total information flow for all users using t-test with $\alpha = 0.05$ (two-tailed). The bold values of mean difference showed that the difference was not statistically significant i.e. $p > 0.05$. In the case of gesture and keyboard drawing analysis, the Competent-User 2, 3, 4 and Novice-User 2's mean information flow difference was not significant. The possible reason can be

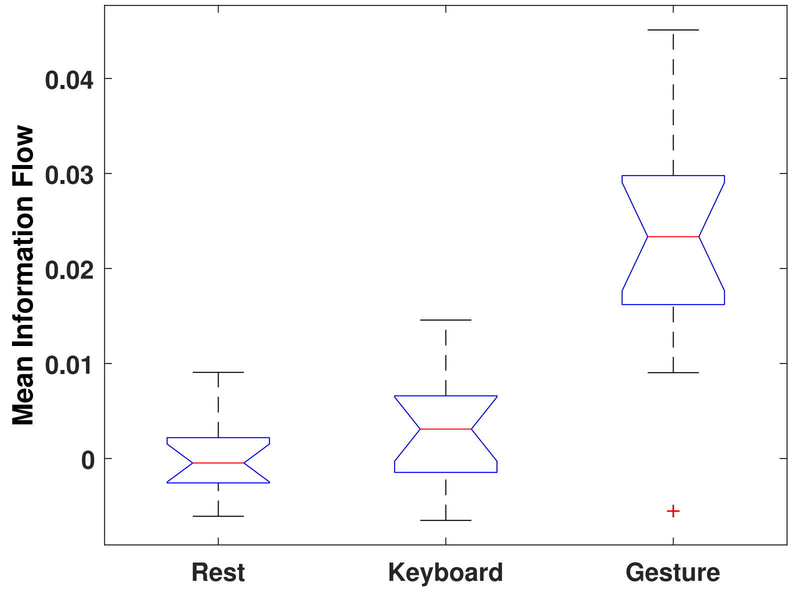


Figure 9.8: Multiple comparison test of node strength group mean for 12 users during rest, keyboard and gesture drawing states

that the Competent-User 2, 3, 4 have some previous knowledge of AutoCAD but this can't be generalized as other users did not show the same trend.

We also applied LDA to information flow and node strength values to find out the electrodes that provide maximum variance during keyboard and gesture drawing states, and the results are shown in Table 9.5. In the case of node strength, the electrodes in the frontal lobes are sending and receiving more information than other electrodes. **The electrodes in the frontal right hemisphere of the brains (F4, F8, AF4) have shown more variation in node strength compared to other electrodes.** The frontal and parietal lobes electrodes mean information flow variations are more than the other regions when using gesture drawing states. We have also tried to find the correlations between the participants experiment completion time and mean information flow patterns using LDA and found that **the information flow from right hemisphere was more compared to left-hemisphere specifically in the frontal and central cortex for participants who completed the experiment within 3 minutes.**

The results show that **the cognitive activity of the users increased in the ges-**

Table 9.4: Statistical validation of total information flow for all users during rest, keyboard, and gesture drawing states. (MD: Mean difference, CI: Confidence interval, CU: Competent-User, NU: Novice-User)

Users	Keyboard-Rest			Gesture-Rest			Gesture-Keyboard		
	MD	**95% CI**		**MD**	**95% CI**		**MD**	**95% CI**	
CU 1	0.0217	0.0092	0.0341	0.0674	0.0471	0.0877	0.0457	0.0174	0.0740
CU 2	**0.0076**	-0.0019	0.0170	0.0149	0.0017	0.0281	**0.0073**	-0.0060	0.0206
CU 3	0.0402	0.0328	0.0476	0.0369	0.0279	0.0459	**-0.0033**	-0.0118	0.0053
CU 4	0.0385	0.0195	0.0575	0.0632	0.0345	0.0920	**0.0248**	-0.0118	0.0613
NU 1	**0.0059**	-0.0015	0.0133	0.0255	0.0154	0.0356	0.0196	0.0065	0.0327
NU 2	0.0379	0.0143	0.0615	0.0387	0.0157	0.0618	**0.0009**	-0.0276	0.0293
NU 3	0.0153	0.0012	0.0294	0.0570	0.0374	0.0767	0.0417	0.0259	0.0576
NU 4	0.0130	0.0025	0.0235	0.1064	0.0738	0.1389	0.0933	0.0618	0.1249
NU 5	0.0139	-0.0019	0.0297	0.0370	0.0170	0.0571	0.0231	-0.0015	0.0477
NU 6	**-0.0005**	-0.0148	0.0139	0.0397	0.0245	0.0549	0.0402	0.0207	0.0596
NU 7	0.0447	0.0360	0.0534	0.0738	0.0555	0.0920	0.0291	0.0089	0.0492
NU 8	0.0693	0.0527	0.0860	0.1425	0.1170	0.1680	0.0731	0.0474	0.0989

Table 9.5: Node strength and mean information flow electrodes with maximum variance in keyboard and gesture drawing states

Graph Measure	Electrodes				
Node strength	FC5	T7	F4	F8	AF4
Mean Information Flow	FC5	P7	P8	FC6	F8

ture drawing state, but the increase was less for some Competent-Users (2, 3, 4) and Novice-User (user 2). It is an indication that with a little training, the users will be able to demonstrate the same cognitive level as they show in keyboard drawing state. The novice's efficiency can be increased if we can stimulate frontal and central lobes of the brain more often than the other lobes. One way to achieve this is to ask the users to use their short-term memory by displaying some hints or other related information for the experiment.

9.4 Conclusion

In this chapter, NTE has been applied to construct functional brain networks from EEG signals to study the user's cognitive activity in rest, traditional input (keyboard and mouse) and multi-modal input (Speech and gesture). After pre-processing of the EEG signal, we have extracted 2 seconds epochs to construct the connectivity matrix using normalized transfer entropy. The averaging of epochs was done to remove the noise from the signals and to extract more global properties of the drawing phenomena. Functional brain networks were constructed using the NTE matrix. Graph theory-based measures were used to analyze the FBNs. Both binary and weighted FBNs were used to study different cognitive states. The list below shows the major findings of this chapter:

1 The results show that the cognitive activity of the user increased when they were using multi-modal input for drawing 3D objects in AutoCAD.

2 The connectivity density, clustering coefficient, and degree centrality results demonstrate that the information transfer between electrodes increases in gesture drawing state from keyboard drawing state.

3 The mean information flow and node strength show that the maximum variation in sending and receiving information is seen in frontal and central lobes because drawing with multi-modal input requires intense attention and motor cognition.

4 Some competent users of AutoCAD show a small variation in degree centrality across electrodes in keyboard and gesture drawing state compared to other users. The maximum difference is seen in the left frontal hemisphere (F7, FC5, T7) and back right hemisphere (O2, P8).

5 The electrodes in the frontal right hemisphere of the brains (F4, F8, AF4) show more variation in node strength compared to other electrodes.

6 The mean information flow from right hemisphere is more compared to left-hemisphere specifically in the frontal and central cortex for participants who completed the experiment within 3 minutes.

Although 3 out of 4 competent users showed that the variation in cognitive activity was less when they moved from keyboard to gesture drawing state, due to very few competent participants, we cannot generalize this hypothesis. The results could be used as a real-time metric for cognitive activity measure. Usability designers can benefit from the insights into the mental processing using the presented method for newly designed systems and classification of cognitive activity.

Chapter 10

Classification of User's Competency using Convolutional Neural Networks

In this chapter, we have proposed a method to classify the user's competency level using convolutional neural networks. To develop a system that can accommodate the lack of competency, it needs to adapt to the competency level of the user. To solve this problem, we have presented a deep convolutional neural network (CNN) model that uses the Electroencephalography (EEG) of the user to classify the level of competency in a 3D modeling task. The five competency levels were defined based on the task completion time, final 3D model rating, and previous modeling competency. The results are compared with other commonly used feature set results. To the best of our knowledge, this is the first study to classify competency levels in HCI. The results of this study have been under review in Expert Systems with Applications Journal.

10.1 Introduction

The 3D graphics industry is one of the biggest industries that incorporates many sub-disciplines, including graphics in games, special effects in movies, and creativity in visual arts. The target audience of the 3D modeling industry has grown from a narrow customer base to a more broader audience in recent years. Modeling in a 3D designing tool is a challenging job and require a unique skill set. In order to achieve a particular competency level, a novice user undergoes rigorous training to become an expert [465]. The system needs to be flexible and adaptable to the user's skill-set to attract more customers. The user-dependent factor must be studied to make the 3D modeling tool adaptive [43]. The purpose of this research is to classify the competency of the user into five levels using Electroencephalography (EEG) signals.

To develop the state of the art tools for any application, the system needs to be adaptive to accommodate both novice and competent users, but to define the criteria on which the system will adapt itself is difficult. One way is to use the user competency level and skills on which the system adapts. The problem is that it is hard to predict the competency of a user in real-time. If a robust technique is developed that can classify the user into various skill levels, then it will be easier to make the system adaptive and accessible by both novice and competent users. By doing that we can reduce the training time of the user. **This application can be used in the gaming industry to develop adaptive games in which difficulty level can be changed based on the user's mental states and cognitive loads.** The development in the Virtual and Augmented Reality (VR/AR) industry and the availability of the VR headset such as looxidlabs VR device that has a built-in EEG sensor [466] can be the gateway to next-generation games, systems, and visual arts.

For the classification of user competency level, we have to analyze the user behavior and actions. A little research has been carried out in analyzing and classifying novice and competent users behavior and the majority of the literature is focused on analyzing the behavior in video games. Researchers proposed various methods to estimate the cognitive response of the users in video games [467, 468]. In the field of 3D modeling, researchers analyze the stress level and cognitive load of the user [365, 469] but there is no research currently focusing on predicting the user skill and competency level with EEG signals. With this research, we want to explore the answers to the following research questions:

RQ 6.1 Can we effectively classify a user's competency into different levels with the EEG signals?

RQ 6.2 Which features contribute the most in the classification of the user' competency?

In this chapter, we have presented a pilot study for the classification of the user's competency level. We have defined five different skill levels of users in a 3D object design and manipulation task. We have recorded the EEG signals of the users with a portable EEG headset (Emotiv EEG headset) while they were drawing a 3D object in AutoCAD. After noise reduction and artifacts removal, features were extracted, including power spectral density, normalized transfer entropy (NTE), common spatial patterns (CSP). A deep CNN model was used to classify the features in five skill levels. We have implemented a CNN model of 14 layers and validate the results with 5-fold cross-validation.

10.2 Review of Related Methods

CNNs have been used in classifying driver's cognitive activity with >70% accuracy using EEG signals [470]. EEG decoding and visualization algorithms also use CNNs [471]. Most of the studies that utilize CNNs for EEG classification focus on healthcare applications. Only a few studies used CNN and EEG signals to detect fatigue and cognitive activity [472, 473]. In this chapter, we have proposed a method to quantify the user skill level into five stages using the deep convolutional neural network. This study will be the first one for the classification of user competency into different levels according to the best knowledge of the authors.

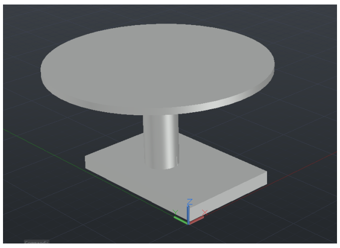
In the literature, deep neural networks have been used to classify EEG signals in several applications. Lawhern et al. presented EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces (BCI) [474]. CNN has been used in biometrics for the identification of individuals using resting state EEG signals with an identification rate of 88% [475]. Researchers have used CNNs to perform feature extraction and classification of motor imagery EEG signals and achieved a classification rate of above 85% [476]. Acharya et al. used CNN for the detection and diagnosis of seizure using EEG signals [477]. In their study, a 13 layer deep neural network has been used to classify normal, preictal, and seizure EEG signals and obtained a classification accuracy of >88%.

CNN is a type of deep neural network that has been used in many applications related to image recognition. In recent years, CNN has been successfully used in various fields including biomedical image processing [478–480], object recognition [481], face detection [482], and sentence modelling [483], but a small amount of research has been carried out in studying psychophysiological signals. For psychophysiological signals, a CNN was used to segment the intervals of tachycardia ECG with an accuracy of > 90% [484]; it was used for automated diagnosis of myocardial infarction [485] and seizure [477].

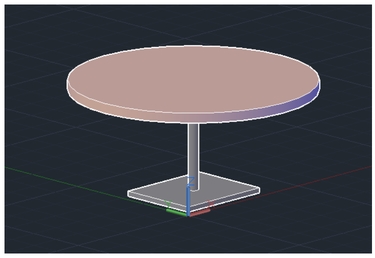
10.3 Methodology

In this chapter, we have used EEG signals to classify competency levels. We have extracted various features from the EEG signal for classification purposes such as Power Spectral Density (PSD), Normalized Transfer Entropy (NTE), and Common Spatial Patterns. We have used the original trials extracted from raw EEG data for training and classification. A deep Convolutional Neural Network (CNN) is used for training and classification of the features. The experimental details on collecting and processing EEG signals are presented in Chapter 5. In this experiment, to prove the

(a) The experimental setting



(b) Novice drawing



(c) Competent drawing

Figure 10.1: Experimental setup with novice and competent users drawing results

concept, we only consider the data of the experiment in which the user draws the 3D model using the keyboard and mouse. Later, we will evaluate the same task using gestures and speech. A picture of the experimental setup along with sample drawings of novice and competent users are shown in Fig. 10.1.

Once the EEG data was clean enough, we split the EEG data of a user into trials of four-second trials. The four-second intervals were selected using the video log data of the users. The four-second interval is the average period in which a user had performed at least one design action such as selecting a shape, changing the dimension, or manipulating the object. By splitting the EEG data into 4-second trials, we extracted a total of 586 trials from 12 users. The overall framework has been given in Fig. 10.2.

10.3.1 Level of Competency

As the users have different skill levels, so a criterion needs to be defined to identify the level of competency. To define a measure for competency level, we have taken into account three different modeling characteristics, i.e., previous knowledge of 3D modeling, task completion time, and rating of the final design by a competent user. We have used these three factors to define five levels of competency. If the users have previous knowledge of 3D modeling, then the users were given five points and for no experience, one point. The task completion time was mapped linearly to a scale of 1

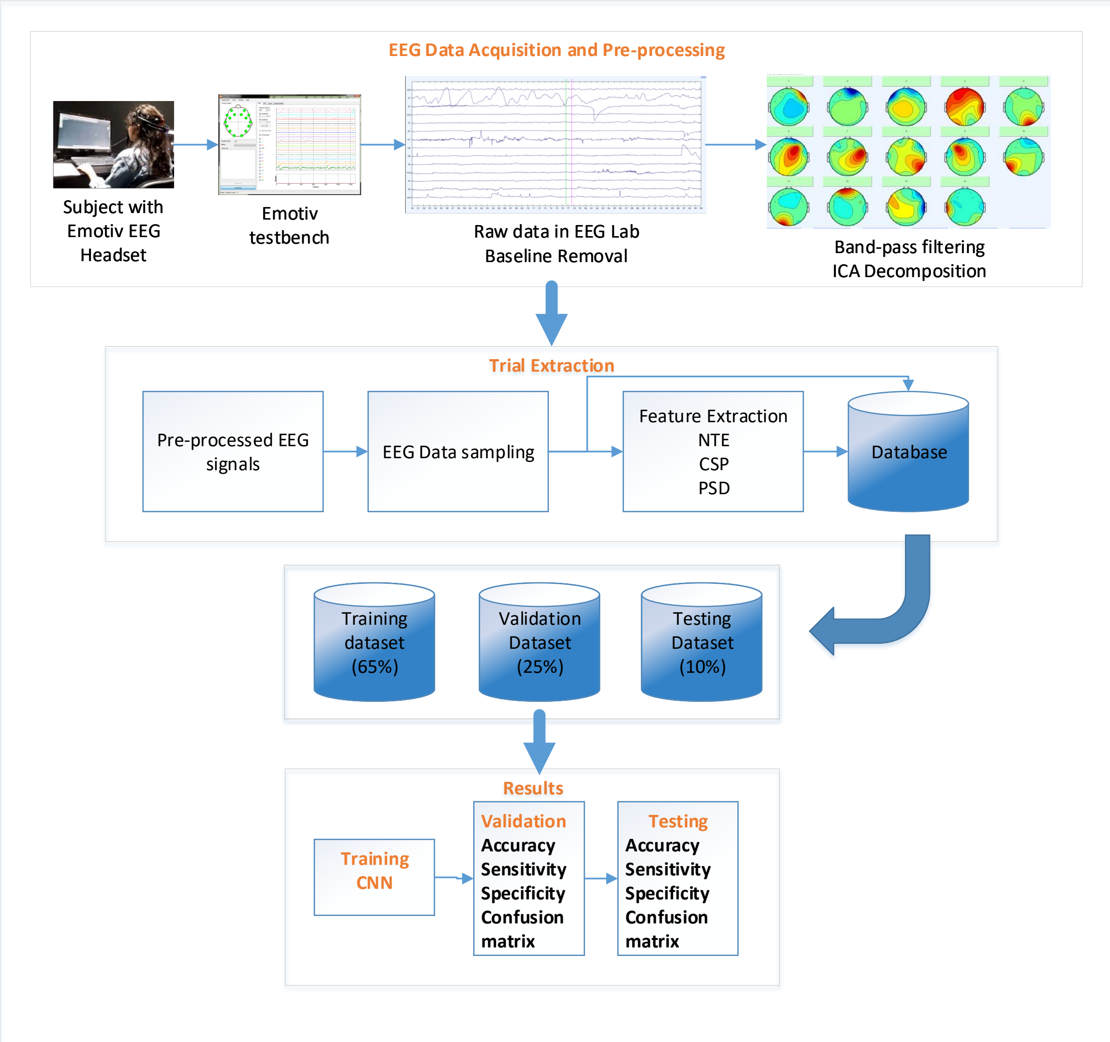


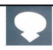
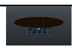
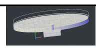
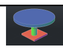
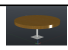
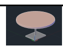
Figure 10.2: Experimental Framework

to 5. A competent user of 3D modeling gave the ratings to the user's final designs on a scale of 1 (bad) to 5 (good). The competency measures of users, along with the number of trials extracted from the EEG signals and final designs of the users, have been given in Table 10.1. Once the user has been awarded points on previous knowledge, completion time, and drawing rating, the total score was calculated by linearly mapping the sum of the three measure on a scale of 1 to 5 with the following competency attribute:

1 Fundamental

2 Novice

Table 10.1: Competency level of all 12 users along with task completion time, number of trials and final drawings

Users	Novice/ Competent	Completion time (s)	Weight	Rating	Sum	Competency level	No. of trials	Final Drawings
User 1	5	58.5	5.00	3.5	13.5	4.0	14	
User 2	1	68.0	4.90	2.0	7.9	2.0	17	
User 3	1	93.1	4.63	3.0	8.6	3.0	23	
User 4	5	110.0	4.45	5.0	14.4	5.0	27	
User 5	1	136.3	4.17	5.0	10.2	4.0	34	
User 6	1	157.8	3.94	1.0	5.9	1.0	39	
User 7	1	164.4	3.87	2.0	6.9	2.0	41	
User 8	5	173.3	3.77	5.0	13.8	5.0	43	
User 9	1	205.1	3.43	2.0	6.4	2.0	51	
User 10	1	213.2	3.34	3.5	7.8	2.0	53	
User 11	1	263.3	2.81	5.0	8.8	3.0	65	
User 12	5	290.1	2.52	5.0	12.5	4.0	72	

3 Intermediate

4 Advanced

5 Competent

10.3.2 Feature Extraction

To classify the EEG signals, we have used the trial data extracted from EEG signals without any feature extraction as well as the extracted features from the trial data. We extracted some widely used EEG features for comparison. We have used Power Spectral Density (PSD), Normalized Transfer Entropy (NTE), and Common Spatial Patterns (CSP) filters. The reason for selecting these three features is that PSD provides the frequency information of the EEG signals, and there is an inverse relationship between mental activity and alpha frequency band of EEG signal. NTE provides directional information patterns between various brain regions. CSP is the most common feature extraction technique used in motor imagery-based BCIs.

Power Spectral Density

Power Spectral density (PSD) is used to extract frequency vs. power spreading information. PSD is the auto-correlation of Fourier transform (FT) which is considered stationary in a wide range [486]. The Welch PSD estimate is used, and all frequency bands ($\delta, \alpha, \beta, \gamma$, and θ) of EEG signals are used in this experiment.

Normalized Transfer Entropy

Normalized transfer entropy (NTE) is used to estimate the information transfer between two variable (see section 3.5.1). The NTE from $y \rightarrow x$ is not equal to NTE from $x \rightarrow y$ and NTE is in the range of 0 and 1. If the value of NTE is 0, that means no transfer of information and if the value is one then the information transfer is a maximum. NTE is used to study the cognitive load of the user in different applications [387]. In design application, novice and competent users should have different cognitive activity; for this reason, NTE is selected as a feature.

Common Spatial Pattern

Common Spatial Pattern (CSP) has been widely used because it can maximize the difference in variance between the two classes [487]. It has been used in many EEG applications and is probably the first choice of feature extraction when designing a Brain Computer Interface (BCI) [488].

10.4 Classification

After trials and features extraction, the next step was to classify the features. In this chapter, we have used deep Convolutional Neural Networks (CNNs). Deep CNNs are a very loose simulation of neurons in the brain. The method uses several levels and layers of data to learn automatically by using a deep structure of neural networks made of many hidden layers of neurons. The advantage of using a CNN is that it automatically mines features that contribute more towards classification [489].

10.4.1 Convolutional Neural Network Architecture

The CNN architecture consists of a convolutional layer, pooling layer, and fully connected layer [490]. The convolutional layer consists of filters (also known as kernels), which is a matrix that convolves with the input signal (EEG Signal). The output of this layer is also called a feature map. The following equation performs the convolutional operation:

$$y_k = \sum_{n=0}^{N-1} x_n h_{k-n} \qquad \textit{Eq(10.1)}$$

where x is the signal, h is the filter coefficient, N is the number of elements in x, and y is the output of convolution.

The pooling layer (down-sampling layer) reduces the dimension of output neurons from the convolutional layer. The reason is to reduce the computational intensity and avoid overfitting. In this chapter, max pooling operation is used as it selects only the maximum value in each feature map to reduce the number of output neurons.

The fully connected layer has all the connection to the activation in the previous layer. Softmax activation function has been used after the fully connected layer in this work. The softmax function computes the probability distribution of the k output classes to predict the actual class corresponds to the EEG signal. The probability is calculated using the following equation:

$$p_j = \frac{\exp^{x_j}}{\sum_1^k \exp^{x_k}} for j = 1, \ldots k \qquad \textit{Eq(10.2)}$$

where p is the output value between 0 and 1, x is the net input and k is the output classes.

After every convolutional layer, batch normalization layer has been used to normalize the activation of each feature map by subtracting the means and dividing by the standard deviation of mini-batch. After normalization, the layer shifts the input by a learnable offset β and scales it by a factor γ. It is a common practice to use batch normalization layer between the convolutional and nonlinear layers such as Rectified Linear Unit (ReLU) layer. The purpose is to speed up the training of the CNN. The Rectified Linear Unit (ReLU) layer used as an activation function performs a threshold operation to the input element, where any negative value is set to zero.

The output of a CNN model is based on weights and biases of the previous layers, and the weights and biases of each layer are updated with Equation 10.3 and Equation 10.4 respectively.

$$\Delta W_l(t+1) = -\frac{x\lambda}{r} W_l - \frac{x}{n}\frac{\delta C}{\delta W_l} + m\Delta W_l(t) \qquad Eq(10.3)$$

$$\Delta B_l(t+1) = -\frac{x}{n}\frac{\delta C}{\delta B_l} + m\Delta B_l(t) \qquad Eq(10.4)$$

where W is the weight, B is bias, l is layer number, λ is regularization parameter, x learning rate, n is total number of samples, m is momentum, t is updating step, and C cost function. The parameters that are used to train CNN are regularization parameter, learning rate, and momentum. These parameters are tuned according to application and data-set to achieve maximum performance. The regularization parameter is used to avoid overfitting, learning rate to control the pace of learning during training, and momentum is used for convergence of the data. In this work, the regularization parameter λ is set to $1*10^{-3}$, the learning rate is 0.01, and momentum is set to 0.9. These parameter values are obtained by trial and error method. Table 10.2 shows the details of the CNN model used in this work.

Table 10.2: A 14-layers CNN structure.

Layer	Type	Options
0	Input	zerocenter normalization
1	Convolution	8 kernels, 3x3 convolutions with 2 stride
2	Batch Normalization	Batch normalization with $\epsilon = 10^{-5}$
3	ReLU	-
4	Max Pooling	pool size 2 with 2 stride
5	Convolution	16 kernels, 3x3 convolutions with 2 stride
6	Batch Normalization	Batch normalization with $\epsilon = 10^{-5}$
7	ReLU	-
8	Max Pooling	pool size 2 with 2 stride
9	Convolution	32 kernels, 3x3 convolutions with 2 stride
10	Batch Normalization	Batch normalization with $\epsilon = 10^{-5}$
11	ReLU	-
12	Fully Connected	Output size 5
13	Softmax	-
14	Classification Output	cross entropy loss error function

The CNN structure contains three convolutional, 2 max-pooling, and one fully connected layers. The stride is set at 2 for convolution and max-pooling operations.

10.4.2 Training of CNN

We have used stochastic gradient descent with momentum (SGDM) algorithm to train the CNN network. A mini-batch size of 128 is selected in this work to train CNN. SGDM is a method that updates the weights and biases to minimize the loss function by taking small steps in the direction of the negative gradient of the loss. A total of 150 epochs were used to train the CNN model in this work. The number of trials required to train CNN for a small number of class labels is not necessarily large. In most of the cases, approximately 60 trials per class are considered enough to train a CNN for a 5 class problem.

10.4.3 Testing of CNN model

We divided the data into three sets, training (65%), validation (25%), and testing (10%). The CNN model used training data (65%) for training, and 25% data is used for validation of the CNN model. The 10% testing data was used to test the CNN model and validate the classification accuracy. A total of 150 epochs (iteration of one training set) were used to train the CNN. After each iteration, the method validates the CNN model by using 25% of the validation dataset. The reason for using this structure is to avoid overfitting of the CNN model.

A 5-fold cross-validation approach was used to test the CNN model. EEG data (90%) is divided into five random portions. Four portions are used for training the CNN, and the remaining one is used to validate the CNN model in each epoch. One-tenth of the EEG data (testing dataset) is used to test the performance of the model. This mechanism was repeated five times by shifting the training and testing datasets. The average values of this shifted evaluation have been used to report accuracy, sensitivity, specificity, and F-measure.

10.5 Results and Discussion

The experiment was performed on an Intel(R) Core(TM) i7-6700 CPU @ 3.40GHz (8 CPUs), with 16GB RAM. The programming environment was MATLAB, and all simulations run on a single CPU. MATLAB was used for the pre-processing of EEG signals and implementation of CNN (training and testing). Table 10.3 shows the classification accuracy of the baseline performance, i.e., the resting state. The label class used for the classification of the resting-state dataset is the same described in Table 10.1 under the heading competency level. The results showed that the CNN model is perfectly equipped to classify EEG signals of various competency level when they are not performing any design task. Classification accuracy of greater than 80% was achieved with the CNN model for baseline dataset.

Table 10.3: Validation accuracy of baseline dataset (Resting state)

Epochs	Validation Accuracy (%) Trial Size (Electrodes x No. of samples)				
	14 x 320	14 x 256	14 x 192	14 x 128	14 x 64
1	25.00%	15.19%	39.64%	34.68%	35.64%
10	51.67%	53.16%	49.55%	50.87%	50.83%
20	63.33%	58.23%	51.35%	57.80%	52.21%
30	65.00%	63.29%	57.66%	65.90%	53.59%
40	66.67%	62.03%	59.46%	65.90%	61.33%
50	66.67%	59.49%	65.77%	67.63%	66.85%
60	71.67%	59.49%	70.27%	71.10%	70.44%
70	73.33%	63.29%	72.07%	69.94%	71.55%
80	73.33%	67.09%	77.48%	73.99%	75.14%
90	76.67%	68.35%	80.18%	76.88%	76.24%
100	78.33%	68.35%	82.88%	78.61%	78.18%
110	80.00%	68.35%	84.68%	82.66%	82.04%
120	80.00%	70.89%	84.68%	83.82%	83.15%
130	80.00%	72.15%	84.68%	85.55%	85.91%
140	80.00%	72.15%	84.91%	86.01%	86.15%
150	80.00%	70.89%	84.91%	86.01%	86.33%

We have used three features sets along with the original trial data. Table 10.4 shows the details about the datasets used in this experiment. The time taken in

Table 10.4: Features with size and computational time to extract all features

Features	Size	Time (s)	Total time(s)
Original data trials	14x512	-	-
PSD of all bands	14x257	0.2537	148.6682
Normalized TE	14x14	4.6529	2726.599
CSP	14x73	0.0468	27.4248

extracting a single CSP feature was 0.04 sec which is minimum among all three features extracted from the original data. The confusion matrix and accuracy is presented in Table 10.5. It is observed that EEG signals without any feature extraction process show an accuracy of >80%. The reason is that the CNN model mines the data to select the optimal elements for classification, so in most cases, a deep neural network does not need any features. The original trial set achieved more than 80%

Table 10.5: The confusion matrix averaged across five iteration of five-fold cross validation

Feature	Classes	Predicted					Accuracy (%)
		1	2	3	4	5	
Original	**1**	10.0	0.5	0.0	1.5	0.5	80.00
	2	0.5	15.0	0.8	3.5	2.0	68.97
	3	0.0	2.8	11.3	2.0	1.8	63.38
	4	0.8	2.8	1.0	32.5	2.3	82.80
	5	0.8	1.0	0.5	0.8	35.8	92.26
PSD	**1**	8.6	1.8	0.8	0	1	70.49
	2	0	21.8	0.2	0	1.8	91.60
	3	0	1	15.8	0	1.4	86.81
	4	0	7.2	3.2	23.2	4.4	61.05
	5	0	2.6	0	0	35.2	93.12
NTE	**1**	0.75	2.25	2.5	5.5	2.25	5.66
	2	1	5.25	2.25	9	4.5	23.86
	3	0.75	4.25	3.25	4.75	4.75	18.31
	4	2.75	7	4.25	21	5	52.50
	5	0.75	9.25	2.5	6.5	18	48.65
CSP	**1**	11.75	1	0	0.25	0.5	87.04
	2	0.25	15.3	1	6.25	0	67.03
	3	0	0.75	12.25	2	2.25	71.01
	4	0	4.5	0.25	34.3	0	87.82
	5	1.25	0	1.75	0	34.5	92.00

accuracy in classifying class 1, 4. The best classification accuracy is shown by the CSP features. The worst performance is shown by NTE features. Table 10.6 shows the average classification results of all feature sets across all five-folds. The results showed that CSP feature classify >80% of the trials correctly. Both the original and PSD features correctly classify the EEG trials 8 out of 10 times. The maximum training time is taken by the original trial set because of the trial size of 14 x 512 samples. The

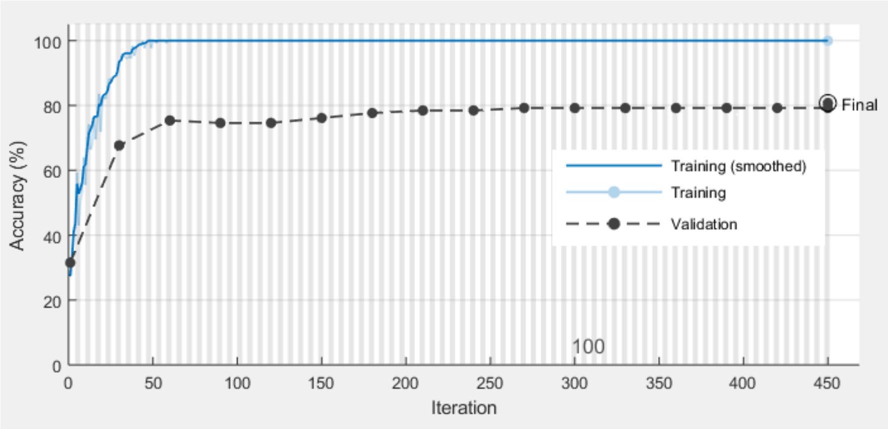


Figure 10.3: CNN training process for original trial set

least training time is for the NTE features because they only have 14 x 14 elements in one trial.

Table 10.6: The overall classification result of all feature sets averaged across all five-folds.

Features	Validation accuracy%	Testing accuracy %	Senstivity	Specificity	F-measure	Train time (s)
Original	80.38	78.79	0.80	0.80	0.44	182.61
PSD	80.46	77.27	0.70	0.82	0.40	96.67
NTE	39.23	37.88	0.06	0.41	-	8.16
CSP	83.65	86.36	0.87	0.83	0.53	31.09

If we consider training time and classification accuracy as factors to choose the best feature, then the CSP features surpass the performance of other features. Fig 10.3 shows the training process of the CNN model along with all 150 epochs on original trial data. It can be seen clearly that the CNN model achieved a validation accuracy of almost 80%. Fig. 10.4 shows the training process for CSP features along with validation. The original trial size was 14 x 512 samples (i.e., four-second sample) which was taking approximately 180 sec for CNN to train. To reduce the training time, we shortened the trial size of original data by half a second repeatedly and then trained and tested the CNN model on these shortened datasets. The trial size was reduced by extracting a sub-trial from the original trial. The number of trials

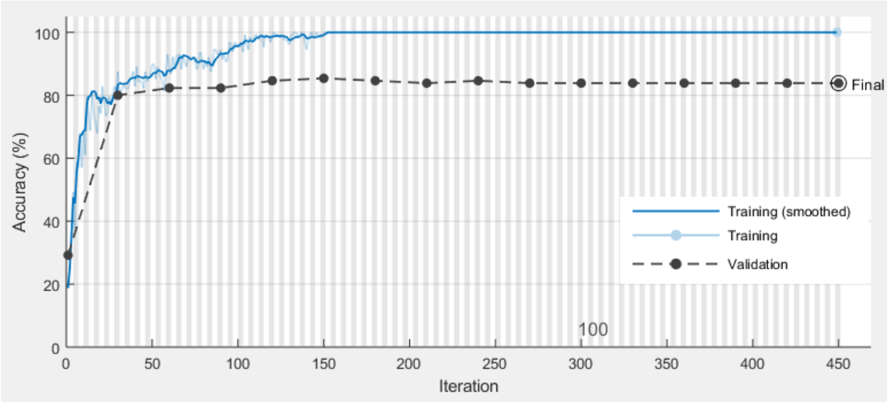


Figure 10.4: CNN training process for CSP feature set.

remains the same. The results are listed in Table 10.7. It is shown in Table 10.7 that reducing the original trial size increased the classification accuracy of the model and training time is also reduced. The best validation, and testing accuracy of 88.27% and 90.91% respectively was achieved by a trial size of 14 x 64 samples.

Table 10.7: The classification results with different trial size of original signal averaged across all five-folds.

Feature size	Validation accuracy%	Testing accuracy %	Senstivity	Specificity	F-measure	Train time (s)
14 x 512	80.38	78.79	0.80	0.80	0.44	182.61
14 x 448	80.79	79.58	0.70	0.81	0.42	174.86
14 x 384	84.81	80.30	0.81	0.85	0.51	155.87
14 x 320	85.58	80.30	0.87	0.85	0.55	115.00
14 x 256	84.62	80.30	0.83	0.85	0.52	94.13
14 x 192	86.92	90.91	0.84	0.87	0.56	71.67
14 x 128	86.54	83.33	0.83	0.87	0.54	50.93
14 x 64	88.27	90.91	0.74	0.90	0.56	17.86

We also tested the extracted features from the reduced trial data but the classification results became worse. To test whether the results of the reduced trial set was reliable, we divided the trials into intervals of 64 samples from the four seconds interval and applied the CNN model. The classification results were consistent. The 14 layer CNN model is selected using trial and error. We have tried different layers of CNN and selected 14 layer CNN because of it produces the efficient results. We have applied the traditional classification algorithms, including SVM, LDA, QDA, kNN, and regression trees, but no one has achieved a classification accuracy of more than 50%. The only classifier with a classification accuracy of 58.4% was kNN with 10 neighbors using cosine distance and PCA components (95% variance).

10.6 Conclusion

In this work, the major contribution is the classification of user competency levels using a deep CNN model. A 14-layer CNN is proposed for the prediction of user skill levels in a design application. This is the first attempt to classify the user's competency into five different levels for design application. The results shown in Table 10.7 are quite promising. To the best of our knowledge, this work is the first work in the field of competency classification using EEG for any application.

A novel method for classifying user's competency into five different levels has been presented in this chapter. A deep CNN model is used for classification. The EEG signal data of every user is divided into trial samples of four seconds and the trials data, as well as the features extracted from the data, were used for training and testing of the CNN model. The features extracted were PSD, NTE, and CSP; these are some of the most commonly used features for EEG applications. The results showed the effectiveness of the proposed method. Maximum classification accuracy of >88%, a specificity of >90% and sensitivity of >70% are recorded by the original reduced trial data of dimension 14 x 64. CSP features performed far better than the other two features sets, i.e., NTE and PSD. The results showed that the method could be used to design a real-time futuristic system that can adapt its functionality according to user skill level. This study is the first study in the field of competency classification using EEG signals.

The significance of the study is that it can be used in conjunction with the new Virtual reality (VR) devices (such as looxidlab VR headset [466]) that have EEG electrodes embedded in the headset. It can allow developers or programmer to develop software, games, and even visual arts that can adapt themselves based on user skill level, behavior, or emotional level.

Chapter 11

Conclusion and Future Work

The primary objective of this thesis is to identify the user's competency by analyzing their cognitive activities using EEG signals in a design task. We tested the usability of the multimodal system in 2 sets of experiments with 12 participants. We used EEG signals to record users' mental states and estimate their cognitive activity. First, we analyzed users' cognitive activity in a unimodal system (keyboard and mouse inputs), and second, in a multimodal system (speech and gesture inputs). The thesis is divided into four major parts: a) the design, development of an MMIS (Chapter 4) b) qualitative evaluation of a multimodal interface system that uses speech and gestures for 3D modeling (Chapter 4) c) quantitative evaluation of the interface using EEG signals (Chapter 6-9) d) classification of user's competency level for adaptive systems design (Chapter 10). The key findings of the thesis are given below:

11.1 Key Findings

The thesis began with an extensive literature review on MMIS, cognition, psychophysiological signals analysis, and FBNs. Chapter 2 discusses the MMIS in detail, along with the differences between unimodal and multimodal systems, various modalities, system integration architectures, fusion techniques, data collection, and evaluation measures. This chapter attempts to answer the following research questions:

11.1.1 Input Modalities

RQ 1.1: What modalities are suited the most to the development of an MMIS for 3D modeling?

Finding 1.1: Based on the literature review presented in Chapter 2, we deduced that speech and gestures are the most widely used input modalities along with the pen and touch-based inputs. Most of the studies in the literature focus on input recognition methods and very few studies focus on output modalities. The MMIS xDe-SIGN uses speech and gestures for 3D modeling, and the system evaluation showed that performing tasks using speech and gesture were perceived as exhausting, but with proper preprocessing techniques and optimization, speech and gesture can become well-coordinated multimodal inputs in an HCI system. In addition to speech, gestures, and pen inputs, biofeedback devices were used in the literature to estimate the emotional and cognitive state of the participant in an MMIS, but the applications were quite limited.

RQ 1.2: What kind of integration techniques should be used to fuse the multimodal inputs?

An MMIS is more robust, adaptable, and provides better results in task completion rate compared to a unimodal system. The literature showed that multimodal interfaces improve the task completion rate by only 10%, but in the case of error handling and reliability, multimodal interfaces reduce errors by 36% compared to unimodal interfaces.

Finding 1.2: The late integration (semantic level) in multimodal input systems was preferred over other integration techniques because late integration (semantic level) gives the advantage to update the modalities and vocabulary quickly.

The literature justifies the use of an MMIS over a unimodal system when speech and gestures are used as the input modalities. The MMIS development and evaluation using qualitative methods have been presented in Chapter 4. The results showed that the speech and gestures were well-coordinated in human to human communication but not in human-computer interaction (HCI). The findings of Chapter 4 gave the answer to the following research questions:

RQ 1.3: Is it possible to develop a multi-modal 3D object manipulation system xDe-SIGN using speech and gestures?

Finding 1.3.1: The results of xDe-SIGN v2 evaluation indicate that performing a task using speech and gestures is exhausting when there is no shared vocabulary between human and machine, and the usability of traditional input devices exceeds the usability of speech and gestures.
The gap between traditional and multimodal input systems will be minimized with the advancements in technology in the near future.

Finding 1.3.2: With the updated xDe-SIGN v2, a large ratio of participants, more than 90%, were able to carry out the tasks with appropriate precision.

RQ 1.4: What are the limitations of using speech and gestures in MMIS?

Finding 1.4.1: The speech recognition systems do not work well and respond appropriately as they require a simple grammar and a quiet environment to reduce noise. On the other hand, gestures seem to be more natural and cognitively less tiring to use in human-computer communication compared to the speech recognition systems. However, gesture recognition systems have to offer several gestures for the same action to address user preferences, as mentioned by Jahani and Kavakli [491]. Even though the xDe-SIGN v2 system was functional, we still noticed that it had lost track of gestures from time to time. People would prefer more natural interaction such as gesture and speech if the performance of the HCI system could satisfy a standard level of operation.

Finding 1.4.2: With proper preprocessing techniques and optimization, speech and gestures can become well-coordinated inputs in HCI.

11.1.2 Psycho-physiological Analysis

In Chapter 3, literature review relevant to cognition, psychophysiological signal analysis, and FBNs is presented. The methods and measurement techniques that utilize psychophysiological analysis and FBNs in analyzing user's cognitive and affective states are discussed. This review helps us find some of the answers to the following research questions:

RQ 2.1: Can we use psycho-physiological analysis to evaluate the HCI?

There are various studies in the literature that use psycho-physiological analysis to evaluate the user's emotional and cognitive responses. In games, violent games increased cardiovascular activity compared to non-violent games, and psychophysiological measures show a strong correlation with the self-reported data. When interacting with a social robot, an increase in β and γ-bands of EEG signals were observed during high-intensity events, and a decrease in stress level was found. Researchers observed an increase in stress level when the user interacts with ill-designed web pages.

Finding 2.1: These studies in the literature showed that the psycho-physiological analysis such as power analysis of EEG-bands could be used to evaluate the HCI systems.

RQ 2.2: Which EEG parameters can be used for evaluating the cognitive activity?

We concluded from the literature that the EEG power bands can be used to evaluate the cognitive activity of the user. The power analysis of EEG signals showed a correlation with task complexities.

Finding 2.2.1: In the literature, researchers have established that the task complexity is inversely related to α-band and directly related to β and θ-band activity.

The other technique to evaluate cognitive activity is to analyse Functional Brain Networks (FBNs). Researchers have observed significant differences in the graph theory-based measures such as connectivity density, motif count, and clustering coefficient in FBNs related to various tasks with different levels of cognition. Non-linear classifier such as Granger causality and transfer entropy showed the best results in FBN analysis.

The literature review showed the applicability of EEG analysis to study user's behavior and cognitive activities.

Finding 2.2.2: The EEG power band and connectivity analyses are the most common techniques used in the literature to study cognitive activities and information processing. The analysis also justified the superiority of nonlinear information measure transfer entropy for the construction of FBNs.

11.1.3 Cognitive Activity

In Chapter 6, we have presented a new empirical method to segment the EEG signals. The aim is to understand cognitive actions and their relation to brain activities in a design application. The participants were divided into two groups: low completion time (low-CT) and high completion time (high-CT) participants. Low-CT users are those who completed the task in a short amount of time and high-CT users completed the task in a long time period. This helped us to answer the following research questions:

RQ 3.1: Why do some novice users perform better than others?

We used a coding scheme to analyze the designer's actions. All actions were divided into three categories: Perceptual, Physical, and Conceptual actions. For analysis of the EEG signals, we used average power of alpha, beta, theta, and gamma bands.

Finding 3.1: The overall results demonstrated that low-CT users performed 1.5

times more physical actions, which gave them the advantage of drawing quickly. The rate of conceptual actions in high-CT users was twice as high as in low-CT users, which slows down the performance of these users in the overall design process.

The action rate per minute for low-CT users is 30% higher than for high-CT users. This is an indication that low-CT users utilize their short-term memory more efficiently.

RQ 3.2: What are the factors that affect novice users performance?

The alpha band shows that low-CT users were comfortable in performing physical actions, whereas high-CT users were not. High-CT users' mean alpha-band power was high. High-CT users spent maximum time in performing conceptual tasks compared to low-CT users, who spent most of the time in focusing on physical design actions.

Finding 3.2.1: The maximum variation in the frontal cortex was found in low-CT users, which indicates that they were using their short-term memory more. From the beta activity, we have found that low-CT users were more attentive to physical actions, whereas the attention of high-CT users was focused on perceptual and conceptual actions. We have found significant variation in theta-band activity for low-CT users than for high-CT users, which indicates that the focus of low-CT users was changing in relation to the actions performed.

Finding 3.2.2: The results clearly showed that performing physical actions with focused attention can decrease task completion time significantly.

These findings suggest that if the interface is designed in a manner that it allows the user to perform more physical actions than conceptual actions, the user performance, and learning rate may improve.

11.1.4 Information Processing

Chapter 7 showed the results of FBNs constructed using NTE and estimated the information flow between various brain regions. Both binary and weighted directed

FBNs were used for the analysis. The research goal in this chapter was to estimate the cognitive activity and information flow patterns for novice and competent users using graph theory. In this chapter, the 3D modeling was performed using unimodal input. The results showed a significant difference between novice and competent user's FBNs. This answers the research questions **RQ 4.1** and **RQ 4.2**.

RQ 4.1: Are there any differences in information processing and cognitive activity between novice and competent users?

The main difference was observed in the object manipulation state. Novice users information flow patterns changed more significantly in the manipulation state compared to competent users. The connectivity density, motif count, clustering coefficient all showed the same trend. The network density increased from the baseline for both novice and competent users, but the change was more significant for novice users compared to competent users in drawing and manipulation states.

Finding 4.1: Most of the activity was focused on the frontal region, which indicates the use of short-term memory. The small-worldness (analysis of clustering coefficient and characteristic path length) shows that competent users have relatively high global and local efficiency of information transfer than novices which means efficient information propagation over the FBN.

RQ 4.2: Can Functional Brain Networks (FBNs) be used to identify the information flow patterns?

The hemisphere analysis shows that the information flow has increased in both hemispheres for novice users, but competent users managed to control the information flow according to the task.

Finding 4.2.1: In the lobe-wise analysis, the frontal lobe was most active in sending and receiving information in drawing and manipulation states for all users. Classification accuracy of more than 90% was achieved with the proposed technique using a simple k-NN classifier in classifying novice and competent users.

The feature selection algorithm showed that features which belong to the frontal and temporal lobes of the brain contribute the most towards competency classification.

Finding 4.2.2: The findings clearly showed that competent users have developed the capability which enables information processing in different brain regions for the different tasks, unlike novice users where almost all regions became active. The main activity was observed in frontal and temporal lobes, which directly relate to the motor, problem-solving, memory, and language functions.

Chapter 7 demonstrates the application of transfer entropy in estimating the cognitive activity and information flow in an open-ended task.

11.1.5 Modality Comparisons

In chapter 8 and 10, we have presented a comparative analysis of unimodal and multimodal interface systems. We have used both the average power spectral density and FBN analysis to compare the unimodal and multimodal systems. The goal was to see whether there are any differences in the cognitive activity of different users with various competencies.

RQ 5.1: What are the differences in cognitive activity between multimodal and unimodal systems?

Finding 5.1.1: For competent users, the average theta band activity demonstrates that the θ-band activity was more intense when the users were using keyboard and mouse for 3D modeling, but for novice users, the θ-band activity increased in gesture state rather than keyboard state.

Finding 5.1.2: The α-band activity decreased for almost all the users in keyboard state compared to the rest state, which shows the increase in mental effort in keyboard state from the rest state.

A very unexpected trend was observed in the gesture state.

Finding 5.1.3: On average, the α-band activity of novice users increased more in gesture state compared to keyboard state when compared with competent users.

The mean information flow and node strength showed that the maximum variation in sending and receiving information was seen in frontal and central lobes because drawing with multi-modal input requires intense attention and motor cognition.

RQ 5.2: Does competency play a role when a new set of inputs are used for a predefined task?

As all of the users were using the xDe-SIGN v2 (multi-modal interface system) for the first time, so they were all considered to be novice users of xDe-SIGN v2, but the competent users of AutoCAD were finding it hard to use the multi-modal input compared to the novice users.

Finding 5.2.1: The FBN connectivity analysis showed that the cognitive activity of the users increased when they were using multimodal input for drawing 3D objects in AutoCAD. The connectivity density, clustering coefficient, and degree centrality results demonstrate that the information transfer between electrodes increased in gesture drawing state from keyboard drawing state.

The results demonstrated the usability of speech and gesture in MMIS. The evaluation of MMIS showed that the system could be used for a 3D modeling application, but there are factors that affect the performance. It is evident that EEG analysis and FBNs have the potential to identify the neural activity related to cognition.

Finding 5.2.2: The power spectral density and connectivity analysis showed that they are sensitive to the cognitive activity of the user and changes in this analysis directly related to changes in cognitive activity. The approach mentioned in this thesis can be used in the development of quantitative metrics to measure cognitive activity in an HCI system.

In chapter 10, a novel method for classifying user's competency level has been presented. The method used a deep CNN model for the classification of competency levels using EEG signals.

RQ 6.1: Can EEG signals be used in classifying user's competency level?

Finding 6.1: The users were divided into five levels of competencies, and a convolutional neural network (CNN) was used to classify the users into various competency levels. The results showed a maximum classification accuracy of above 88%. The method can be used to classify a user's competency using EEG signals and to develop competency-based adaptive systems.

RQ 6.2: Which features contribute the most toward classification accuracy?

Finding 6.2: The features extracted were Power Spectral Density, Normalized Transfer Entropy, and Common Spatial Patterns. Maximum classification accuracy of more than 88% was recorded by original reduced data of dimension 16 x 64. In terms of extracted features, CSP features performed far better than the other two feature sets, i.e. NTE and PSD. The results showed that the competency levels can be used as an adaptive parameter to design a real-time futuristic system that can adapt its functionality according to user skill level.

11.2 Evaluation of Results

To the best of our knowledge, this thesis is the first study in the field of competency classification using EEG signals. This thesis has outlined 22 major findings in total including 6 in the area of input modalities, 3 in the area of psycho-physiological analysis, 3 in cognitive activities, 3 in information processing, 5 in modality comparison and 2 in the area of competency classification. This section sheds light on implications and limitations of results along with future recommendation.

11.2.1 Implications of Results

The reported results have significance in many real-world applications as they demonstrate the potential of EEG signal analysis in measuring cognitive activity in HCI.

The results also showed the application of NTE based FBNs to detect the unusual variations in neural activity. The thesis provides experimental evidence that EEG based measures can be used as a quantitative metric to analyze cognitive activity in HCI. These EEG based cognitive metrics may be used in monitoring and diagnosing cognitive impairments/disorders. For example, an EEG-based portable device can be beneficial for people who are in remote areas, working under severe conditions and/or with a high cognitive load.

In the computing domain, the results could be used to develop adaptive learning systems, where the pace of learning could be changed based on the user's mental effort and cognitive load. Adaptive games could be another example in which the game complexity level could be adapted based on the user's competency levels. With the advancements in portable headset devices that embed EEG electrodes such as Looxidlab VR headsets [466], adaptive system applications will soon become a reality.

11.2.2 Limitations of Results and Future Recommendations

Every research work contains limitations irrespective of the discipline of study, and so it is in this thesis. The most obvious limitation is the sample size of the participants. The sample size was less than 20 in most of the experiments which may affect the results reported in this research. This sample size is common in the EEG literature, but a higher number of participants would be much more desirable [365, 387]. Despite the fact that EEG is relatively cheaper than other neuroimaging techniques such as fMRI, it is still expensive compared to traditional surveys and other testing techniques from both the time and financial perspective.

Most of the EEG related research is conducted in a controlled psychology/ medical environment and uses the psychology/medical students for sample population [492]. In these scenarios, a credit point is awarded as an incentive to participants, which in return increase the sample size significantly although sometimes these studies receive criticism for not targeting a more general community [493]. In this thesis, the participants are computer science students, and no incentive was given for participation. As the limited sample size shows, this voluntary participation didn't yield a large number of participants. Because of the small sample size, the results in this thesis lack statistical power, but this does not mean that the results are not valid or reliable. In future research, other ways should be explored to attract the wider community to increase the sample sizes.

Another limitation is the statistical tests used in the thesis to show the significance of results: One-way ANOVA and t-test. Although many researchers proved the robustness of these tests [494, 495], it is arguable that the sample size reported in this thesis makes the testing somewhat problematic. Future research should focus on increasing the sample size to obtain statistical power.

In addition to the limitation mentioned above, another limitation is the averaging of EEG signal, which has both pros and cons. The averaging increases the original signal component and decreases the noise component but at the cost of some degradation in the neural signal. The reason for this is that complex neural activities are not synchronized perfectly and averaging can eliminate the neural activities that have different phase delays.

Another limitation is the potential inference of actual brain functions from the estimated neural activity from EEG signals. Implanting electrodes into the healthy brain for recording the neural activity is prohibited due to ethical reasons. The electrical potential recorded over the brain (such as EEG) could be a combination of more than one brain function. However, the reported results presented in this thesis have been described relative to the available literature. Future research could explore new ethical ways to extract the brain activity to observe the functional changes and impairment directly related to the measured activity for a cause/effect conclusion.

Finally, the last limitation could be related to the adaptive systems that use the

EEG signals for adaptability. In this thesis, we presented a possible solution for the development of adaptive learning and gaming systems that uses EEG signals measured as the variable for adaptation. Although the results showed a classification accuracy of greater than 80%, the sample size used is small. The future research could increase the sample size and use a VR headset that has embedded EEG sensors such as the looxidlab VR headset [466] to test the system or use sensors which are more reliable and sensitive than the Emotive headset.

In summary, our findings reported in this thesis are in line with the existing literature. Various techniques related to computational neuroscience, psychology, and graph theory have been explored in this thesis to achieve the objectives. The findings related to cognitive activity and competency classification could have a significant impact on future MMIS designs. These findings are directly applicable to the development of adaptive systems.

References

[1] B. Dumas, D. Lalanne, and S. Oviatt, “Multimodal interfaces: A survey of principles, models and frameworks,” in *Human Machine Interaction*. Springer, 2009, pp. 3–26.

[2] M. M. Blattner and E. P. Glinert, “Multimodal integration,” *IEEE Multimedia*, vol. 3, no. 4, pp. 14–24, 1996.

[3] M. Turk, “Multimodal interaction: A review,” *Pattern Recognition Letters*, vol. 36, pp. 189–195, 2014.

[4] M. Paleari and C. L. Lisetti, “Toward multimodal fusion of affective cues,” in *Proceedings of the 1st ACM International Workshop on Human-centered Multimedia*. ACM, 2006, pp. 99–108.

[5] R. A. Bolt, *"Put-that-there": Voice and gesture at the graphics interface*. ACM, 1980, vol. 14, no. 3.

[6] M. G. Jacob, Y.-T. Li, and J. P. Wachs, “Gestonurse: a multimodal robotic scrub nurse,” in *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2012, pp. 153–154.

[7] G. P. Laput, M. Dontcheva, G. Wilensky, W. Chang, A. Agarwala, J. Linder, and E. Adar, “Pixeltone: A multimodal interface for image editing,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’13. New York, NY, USA: ACM, 2013, pp. 2185–2194.

[8] A. D. Baddeley and G. Hitch, “Working memory,” in *Psychology of Learning and Motivation*. Elsevier, 1974, vol. 8, pp. 47–89.

[9] C. D. Wickens, “Multiple resources and performance prediction,” *Theoretical Issues in Ergonomics Science*, vol. 3, no. 2, pp. 159–177, 2002.

[10] T. S. Jastrzembski and N. Charness, “The model human processor and the older adult: parameter estimation and validation within a mobile phone task.” *Journal of Experimental Psychology: Applied*, vol. 13, no. 4, p. 224, 2007.

[11] M. Z. Baig and M. Kavakli, “A survey on psycho-physiological analysis & measurement methods in multimodal systems,” *Multimodal Technologies and Interaction*, vol. 3, no. 2, p. 37, 2019.

[12] D. D. Mridha, “EEG,” http://drmridha.com/services/eeg/, [Online; accessed 06-Jun-2019].

[13] J. A. Russel, A. Weiss, and G. A. Mendelsohn, “Affect grid: A single-item scale of pleasure and arousal,” *Journal of Personality and Social Psychology*, vol. 57, no. 3, pp. 493–502, 1989.

[14] D. S. Bassett and E. T. Bullmore, “Human brain networks in health and disease,” *Current Opinion in Neurology*, vol. 22, no. 4, p. 340, 2009.

[15] Leap-Motion, "Gestures," https://developer-archive.leapmotion.com/documentation/csharp/devguide/Leap_Gestures.html, (Accessed on 06/07/2019).

[16] imotions, "Emotiv epoc EEG | EEG hardware," https://imotions.com/emotiv-epoc-eeg/, [Online; accessed 06-Jun-2019].

[17] S. R. Benbadis and D. Rielo, "EEG artifacts," *Distribution*, vol. 12, pp. 1–23, 2010.

[18] P. F. Drucker, "Knowledge-worker productivity: the biggest challenge," *IEEE Engineering Management Review*, vol. 34, no. 2, pp. 29–29, 2006.

[19] G. Iazzolino and D. Laise, "Knowledge worker productivity: is it really impossible to measure it?" *Measuring Business Excellence*, vol. 22, no. 4, pp. 346–361, 2018.

[20] B. Hayes-Roth and F. Hayes-Roth, "A cognitive model of planning," *Cognitive Science*, vol. 3, no. 4, pp. 275–310, 1979.

[21] R. E. Boyatzis, *The competent manager: A model for effective performance*. John Wiley & Sons, 1982.

[22] N. Cross and A. C. Cross, "Expertise in engineering design," *Research in Engineering Design*, vol. 10, no. 3, pp. 141–149, 1998.

[23] N. Cross, "Expertise in design: an overview," *Design Studies*, vol. 25, no. 5, pp. 427–441, 2004.

[24] M. T. Chi, "Laboratory methods for assessing experts' and novices' knowledge," *The Cambridge Handbook of Expertise and Expert Performance*, pp. 167–184, 2006.

[25] D. Schwartz, J. Bransford, and D. Sears, "Innovation and efficiency in learning and transfer," *Transfer of Learning from a Modern Multidisciplinary Perspective*, pp. 1–51, 2005.

[26] C.-H. Ho, "Some phenomena of problem decomposition strategy for design thinking: differences between novices and experts," *Design Studies*, vol. 22, no. 1, pp. 27–45, 2001.

[27] H. Christiaans and K. H. Dorst, "Cognitive models in industrial design engineering: a protocol study," *Design Theory and Methodology*, vol. 42, no. 1, pp. 131–140, 1992.

[28] M. Kavakli and J. Gero, "Difference between expert and novice designers: An experimental stud," *Human Behaviour in Design*, 2003.

[29] S. Oviatt, P. Cohen, L. Wu, L. Duncan, B. Suhm, J. Bers, T. Holzman, T. Winograd, J. Landay, J. Larson *et al.*, "Designing the user interface for multimodal speech and pen-based gesture applications: state-of-the-art systems and future research directions," *Human-computer Interaction*, vol. 15, no. 4, pp. 263–322, 2000.

[30] J. Liu, "Analysis of gender differences in speech and hand gesture coordination for the design of multimodal interface systems," Ph.D. dissertation, Macquarie University, 2014.

[31] F. Qiao, J. Sherwani, and R. Rosenfeld, "Small-vocabulary speech recognition for resource-scarce languages," in *Proceedings of the First ACM Symposium on Computing for Development*. ACM, 2010, p. 3.

[32] Y. Liu, O. Sourina, and M. K. Nguyen, "Real-time eeg-based human emotion recognition and visualization," in *Cyberworlds (CW), 2010 International Conference on*. IEEE, 2010, pp. 262–269.

[33] A. Boyali, M. Kavakli, and J. Twamley, "Real time six degree of freedom pose estimation using infrared light sources and wiimote IR camera with 3D TV demonstration," in *International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services*. Springer, 2010, pp. 137–148.

[34] C. Remy, O. Bates, A. Dix, V. Thomas, M. Hazas, A. Friday, and E. M. Huang, "Evaluation beyond usability: Validating sustainable hci research," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 2018, p. 216.

[35] J. T. Cacioppo, L. G. Tassinary, and G. Berntson, *Handbook of Psychophysiology*. New York, NY, US: Cambridge University Press, 2007.

[36] B. Kolb and I. Q. Whishaw, *Fundamentals of human neuropsychology*. USA: Worth Publishers Inc., 2017.

[37] C. F. Levinthal, *Introduction to physiological psychology*. Englewood Cliffs, NJ, US: Prentice-Hall, Inc, 1990.

[38] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience*, vol. 10, no. 3, p. 186, 2009.

[39] R. Salvador, A. Martinez, E. Pomarol-Clotet, J. Gomar, F. Vila, S. Sarró, A. Capdevila, and E. Bullmore, "A simple view of the brain through a frequency-specific functional connectivity measure," *Neuroimage*, vol. 39, no. 1, pp. 279–289, 2008.

[40] T. Schreiber, "Measuring information transfer," *Physical Review Letters*, vol. 85, no. 2, p. 461, 2000.

[41] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010.

[42] F. Alibay, M. Kavakli, J.-R. Chardonnet, and M. Z. Baig, "The usability of speech and/or gestures in multi-modal interface systems," in *Proceedings of the 9th International Conference on Computer and Automation Engineering*. ACM, 2017, pp. 73–77.

[43] M. Z. Baig and M. Kavakli, "Qualitative analysis of a multimodal interface system using speech/gesture," in *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, 2018, pp. 2811–2816.

[44] Y. B. Australia, "1301.0," *Australian Bureau of Statistics*, 2012.

[45] ADCET, "Learning difficulty vs learning disability," https://www.adcet.edu.au/disability-practitioner/reasonable-adjustments/disability-specific-adjustments/specific-learning-disability/learning-difficulty-versus-learning-disability/, [Online; accessed 16-July-2019].

[46] M. Z. Baig and M. Kavakli, "EEG signal analysis in 3D modelling to identify correlations between task completion in design user's cognitive activities," in *Neural Information Processing*, L. Cheng, A. C. S. Leung, and S. Ozawa, Eds. Cham: Springer International Publishing, 2018, pp. 340–352.

[47] M. Z. Baig and M. K. , "Connectivity analysis using functional brain networks to evaluate cognitive activity during 3D modelling," *Brain sciences*, vol. 9, no. 2, p. 24, 2019.

[48] M. Z. Baig and M. Kavakli, "Analyzing novice and expert user's cognitive load in using a multi-modal interface system," in *2018 26th International Conference on Systems Engineering (ICSEng)*, Dec 2018, pp. 1–7.

[49] F. Quek, D. McNeill, R. Bryll, S. Duncan, X.-F. Ma, C. Kirbas, K. E. McCullough, and R. Ansari, "Multimodal human discourse: gesture and speech," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 9, no. 3, pp. 171–193, 2002.

[50] F. Cutugno, V. A. Leano, R. Rinaldi, and G. Mignini, "Multimodal framework for mobile interaction," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*. ACM, 2012, pp. 197–203.

[51] S. Oviatt, "Advances in robust multimodal interface design," *IEEE Computer Graphics and Applications*, vol. 23, no. 5, pp. 62–68, 2003.

[52] H. J. Fariman, H. J. Alyamani, M. Kavakli, and L. Hamey, "Designing a user-defined gesture vocabulary for an in-vehicle climate control system," in *Proceedings of the 28th Australian Conference on Computer-Human Interaction*. ACM, 2016, pp. 391–395.

[53] M. Hafez, "Tactile interfaces: technologies, applications and challenges," *The Visual Computer*, vol. 23, no. 4, pp. 267–272, 2007.

[54] H. Liu, Z. Ma, W. Shao, and Z. Niu, "Schedule of bad smell detection and resolution: A new way to save effort," *IEEE Transactions on Software Engineering*, vol. 38, no. 1, pp. 220–235, 2012.

[55] A. Riul Jr, C. A. Dantas, C. M. Miyazaki, and O. N. Oliveira Jr, "Recent advances in electronic tongues," *Analyst*, vol. 135, no. 10, pp. 2481–2495, 2010.

[56] M. Turk and M. Kölsch, "Perceptual interfaces," *Emerging Topics in Computer Vision, Prentice Hall PTR Upper Saddle River, NJ, USA*, 2004.

[57] D. Chen and R. Vertegaal, "Using mental load for managing interruptions in physiologically attentive user interfaces," in *CHI'04 extended abstracts on Human Factors in Computing Systems*. ACM, 2004, pp. 1513–1516.

[58] S. Bostandjiev, J. O'Donovan, and T. Höllerer, "Tasteweights: a visual interactive hybrid recommender system," in *Proceedings of the sixth ACM Conference on Recommender Systems*. ACM, 2012, pp. 35–42.

[59] P. B. S. O'Modhrain, "Towards tangible enactive-interfaces," in *4th International Conference on Enactive Interfaces*, 2007, p. 37.

[60] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly, "Vision-based hand pose estimation: A review," *Computer Vision and Image Understanding*, vol. 108, no. 1, pp. 52–73, 2007.

[61] S. Oviatt, "Multimodal interactive maps: Designing for human performance," *Human-Computer Interaction*, vol. 12, no. 1, pp. 93–129, 1997.

[62] M. Kavakli and K. Nasser, "Impacts of culture on gesture based interfaces: A case study on anglo-celtics and latin americans," *International Journal on Advances in Life Sciences*, vol. 4, 2012.

[63] J. G. Neal, C. Thielman, Z. Dobes, S. M. Haller, and S. C. Shapiro, "Natural language with integrated deictic and graphic gestures," in *Proceedings of the Workshop on Speech and Natural Language*. Association for Computational Linguistics, 1989, pp. 410–423.

[64] D. B. Koons, C. J. Sparrell, and K. R. Thorisson, "Integrating simultaneous input from speech, gaze, and hand gestures," *MIT Press: Menlo Park, CA*, pp. 257–276, 1993.

[65] P. R. Cohen, M. Johnston, D. McGee, S. L. Oviatt, J. Pittman, I. A. Smith, L. Chen, and J. Clow, "Quickset: Multimodal interaction for distributed applications." in *ACM Multimedia*, vol. 97, 1997, pp. 31–40.

[66] M. L. Newman, C. J. Groom, L. D. Handelman, and J. W. Pennebaker, "Gender differences in language use: An analysis of 14,000 text samples," *Discourse Processes*, vol. 45, no. 3, pp. 211–236, 2008.

[67] J. A. Rodger and P. C. Pendharkar, "A field study of the impact of gender and user's technical experience on the performance of voice-activated medical tracking application," *International Journal of Human-Computer Studies*, vol. 60, no. 5, pp. 529–544, 2004.

[68] A. Van Dam, "Post-wimp user interfaces," *Communications of the ACM*, vol. 40, no. 2, pp. 63–68, 1997.

[69] S. Oviatt, "Human-centered design meets cognitive load theory: designing interfaces that help people think," in *Proceedings of the 14th ACM International Conference on Multimedia*. ACM, 2006, pp. 871–880.

[70] B. Xiao, C. Girand, and S. Oviatt, "Multimodal integration patterns in children," in *Seventh International Conference on Spoken Language Processing*, 2002.

[71] S. Oviatt, R. Lunsford, and R. Coulston, "Individual differences in multimodal integration patterns: What are they and why do they exist?" in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2005, pp. 241–249.

[72] D. Bohus and E. Horvitz, "Facilitating multiparty dialog with gaze, gesture, and speech," in *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*. ACM, 2010, p. 5.

[73] V. Van Wassenhove, K. W. Grant, and D. Poeppel, "Visual speech speeds up the neural processing of auditory speech," *Proceedings of the National Academy of Sciences*, vol. 102, no. 4, pp. 1181–1186, 2005.

[74] S. Oviatt, "Ten myths of multimodal interaction," *Communications of the ACM*, vol. 42, no. 11, pp. 74–74, 1999.

[75] Y. Cao, *Multimodal information presentation for high-load human computer interaction*. University of Twente, 2011.

[76] M. E. Dawson, A. M. Schell, and D. L. Filion, "7 the electrodermal system," *Handbook of Psychophysiology*, vol. 159, 2007.

[77] M. Pantic and L. J. Rothkrantz, "Toward an affect-sensitive multimodal human-computer interaction," *Proceedings of the IEEE*, vol. 91, no. 9, pp. 1370–1390, 2003.

[78] J.-H. Kim, S.-W. Lee, D. Kwak, M.-O. Heo, J. Kim, J.-W. Ha, and B.-T. Zhang, "Multimodal residual learning for visual qa," in *Advances in Neural Information Processing Systems*, 2016, pp. 361–369.

[79] Z. Obrenovic and D. Starcevic, "Modeling multimodal human-computer interaction," *Computer*, vol. 37, no. 9, pp. 65–72, 2004.

[80] K. Shinoda, Y. Watanabe, K. Iwata, Y. Liang, R. Nakagawa, and S. Furui, "Semi-synchronous speech and pen input for mobile user interfaces," *Speech Communication*, vol. 53, no. 3, pp. 283–291, 2011.

[81] N. Ruiz, Q. Q. Feng, R. Taib, T. Handke, and F. Chen, "Cognitive skills learning: pen input patterns in computer-based athlete training," in *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*. ACM, 2010, p. 41.

[82] S. Dusan, G. J. Gadbois, and J. Flanagan, "Multimodal interaction on pda's integrating speech and pen inputs," in *Eighth European Conference on Speech Communication and Technology*, 2003.

[83] M. Miki, N. Kitaoka, C. Miyajima, T. Nishino, and K. Takeda, "Improvement of multimodal gesture and speech recognition performance using time intervals between gestures and accompanying speech," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2014, no. 1, p. 2, 2014.

[84] A. W. Ismail and M. S. Sunar, "Multimodal fusion: gesture and speech input in augmented reality environment," in *Computational Intelligence in Information Systems*. Springer, 2015, pp. 245–254.

[85] J. Wang, A. Samal, J. R. Green, and F. Rudzicz, "Sentence recognition from articulatory movements for silent speech interfaces," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 4985–4988.

[86] D. G. Stork and M. E. Hennecke, *Speechreading by humans and machines: models, systems, and applications*. Springer Science & Business Media, 2013, vol. 150.

[87] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[88] J. M. Baker, L. Deng, J. Glass, S. Khudanpur, C.-H. Lee, N. Morgan, and D. O'Shaughnessy, "Developments and directions in speech recognition and understanding, part 1 [dsp education]," *IEEE Signal Processing Magazine*, vol. 26, no. 3, pp. 75–80, 2009.

[89] L. Muda, B. KM, and I. Elamvazuthi, "Voice recognition algorithms using mel frequency cepstral coefficient (mfcc) and dynamic time warping (dtw) techniques," *Journal of Computing*, vol. 2, no. 3, pp. 138–143, 2010.

[90] A. L. Maas, Q. V. Le, T. M. O'Neil, O. Vinyals, P. Nguyen, and A. Y. Ng, "Recurrent neural networks for noise reduction in robust asr," in *Thirteenth Annual Conference of the International Speech Communication Association*, 2012.

[91] S. E. Zuberec, L. Matheson, C. Cyr, H. Li, and S. P. Masters, "Speech recognition system with changing grammars and grammar help command," Oct. 2 2001, US Patent 6,298,324.

[92] S. Burger, Z. A. Sloane, and J. Yang, "Competitive evaluation of commercially available speech recognizers in multiple languages," in *Proc. of Fifth International Conference on Language Resources and Evaluation (LREC), Genoa, Italy*, 2006.

[93] V. Këpuska and G. Bohouta, "Comparing speech recognition systems (Microsoft API, Google API and CMU Sphinx)," *International Journal of Engineering Research and Application*, vol. 7, no. 03, pp. 20–24, 2017.

[94] F. Vafaei, "Taxonomy of gestures in human computer interaction," *Retrieved from North Dakota State University, Fargo: http://library.ndsu.edu*, 2013.

[95] D. Morris, *Gestures, their origins and distribution*. London, UK: Stein & Day Pub, 1979.

[96] B. Rimé and L. Schiaratura, "Gesture and speech." *In R. S. Feldman & B. Rimé (Eds.), Studies in emotion & social interaction. Fundamentals of Nonverbal Behavior*, 1991.

[97] M. Billinghurst and B. Buxton, "Gesture based interaction," *Haptic Input*, vol. 24, 2011.

[98] M. Karam and M. C. Schraefel, *A Taxonomy of Gestures in Human Computer Interactions*. S.N., 2005.

[99] J. O. Wobbrock, M. R. Morris, and A. D. Wilson, "User-defined gestures for surface computing," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2009, pp. 1083–1092.

[100] J. Ruiz, Y. Li, and E. Lank, "User-defined motion gestures for mobile interaction," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011, pp. 197–206.

[101] P. Premaratne, "Historical development of hand gesture recognition," in *Human Computer Interaction Using Hand Gestures*. Springer, 2014, pp. 5–29.

[102] R. Sodhi, I. Poupyrev, M. Glisson, and A. Israr, "Aireal: interactive tactile experiences in free air," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 4, p. 134, 2013.

[103] P. Molchanov, X. Yang, S. Gupta, K. Kim, S. Tyree, and J. Kautz, "Online detection and classification of dynamic hand gestures with recurrent 3d convolutional neural network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4207–4215.

[104] G. Marin, F. Dominio, and P. Zanuttigh, "Hand gesture recognition with leap motion and kinect devices," in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 1565–1569.

[105] L. M. Reeves, J. Lai, J. A. Larson, S. Oviatt, T. Balaji, S. Buisine, P. Collings, P. Cohen, B. Kraal, J.-C. Martin *et al.*, "Guidelines for multimodal user interface design," *Communications of the ACM*, vol. 47, no. 1, pp. 57–59, 2004.

[106] A. Jaimes and N. Dimitrova, "Human-centered multimedia: culture, deployment, and access," *IEEE Multimedia*, vol. 13, no. 1, pp. 12–19, 2006.

[107] N. Sebe, "Multimodal interfaces: Challenges and perspectives," *Journal of Ambient Intelligence and Smart Environments*, vol. 1, no. 1, pp. 23–30, 2009.

[108] D. L. Martin, A. J. Cheyer, and D. B. Moran, "The open agent architecture: A framework for building distributed software systems," *Applied Artificial Intelligence*, vol. 13, no. 1-2, pp. 91–128, 1999.

[109] I. Cohen, F. G. Cozman, N. Sebe, M. C. Cirelo, and T. S. Huang, "Semisupervised learning of classifiers: Theory, algorithms, and their application to human-computer interaction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 12, pp. 1553–1566, 2004.

[110] M. Turunen and J. Hakulinen, "Jaspisˆ 2-an architecture for supporting distributed spoken dialogues," in *Eighth European Conference on Speech Communication and Technology*, 2003.

[111] S. Kumar, P. R. Cohen, and H. J. Levesque, "The adaptive agent architecture: Achieving fault-tolerance using persistent broker teams," in *Proceedings Fourth International Conference on MultiAgent Systems*. IEEE, 2000, pp. 159–166.

[112] H. Yoshikawa, "Modeling humans in human-computer interaction," *Human-computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, pp. 118–146, 2002.

[113] A. Guo, J. Ma, I. Kevin, and K. Wang, "From user models to the cyber-i model: Approaches, progresses and issues," in *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*. IEEE, 2018, pp. 33–40.

[114] S. K. Card, A. Newell, and T. P. Moran, *The Psychology of Human-Computer Interaction*. Hillsdale, NJ, USA: L. Erlbaum Associates Inc., 2000.

[115] A. Dix, J. Finley, G. Abowd, and R. Beale, "Human-computer interaction (2nd ed.)," Upper Saddle River, NJ, USA, 1998.

[116] P. Biswas and P. Robinson, "A brief survey on user modelling in hci," in *Proc. of the International Conference on Intelligent Human Computer Interaction (IHCI)*, vol. 2010, 2010.

[117] M. D. Byrne, "Cognitive architecture," in *The Human-computer Interaction Handbook*. Hillsdale, NJ, USA: L. Erlbaum Associates Inc., 2007, pp. 110–130.

[118] M. Byrne, "ACT-R/PM and menu selection: Applying a cognitive architecture to HCI," *International Journal of Human-Computer Studies*, vol. 55, no. 1, pp. 41–84, 2001.

[119] A. Hijazi, "Cognitive architectures: Potential for application in complex dynamic environments," *University of Colorado*, 2004.

[120] D. A. Norman, *The design of everyday things Doubleday*. New York, NY: Basic Books, 1988.

[121] J. E. Cahn and S. E. Brennan, "A psychological model of grounding and repair in dialog," in *Proc. Fall 1999 AAAI Symposium on Psychological Models of Communication in Collaborative Systems*, 1999.

[122] S. C. Chase, "A model for user interaction in grammar-based design systems," *Automation in Construction*, vol. 11, no. 2, pp. 161–172, 2002.

[123] Y. Motomura, K. Yoshida, and K. Fujimoto, "Generative user models for adaptive information retrieval," in *Smc 2000 Conference Proceedings. 2000 IEEE International Conference on Systems, Man and Cybernetics.'cybernetics evolving to systems, humans, organizations, and their complex interactions'*, vol. 1. IEEE, 2000, pp. 665–670.

[124] D. Browne, P. Totterdell, and M. Norman, Eds., *Adaptive User Interfaces*. London, UK, UK: Academic Press Ltd., 1990.

[125] T. Lavie and J. Meyer, "Benefits and costs of adaptive user interfaces," *International Journal of Human-Computer Studies*, vol. 68, no. 8, pp. 508–524, 2010.

[126] R. Oppermann, *Adaptive user support: ergonomic design of manually and automatically adaptable software*. New York: Routledge, 2017.

[127] E. Ross, "Intelligent user interfaces: Survey and research directions," Bristol, UK, UK, Tech. Rep., 2000.

[128] M. Pantic, N. Sebe, J. F. Cohn, and T. Huang, "Affective multimodal human-computer interaction," in *Proceedings of the 13th annual ACM International Conference on Multimedia*. ACM, 2005, pp. 669–676.

[129] K. Höök, "Steps to take before intelligent user interfaces become real," *Interacting with Computers*, vol. 12, no. 4, pp. 409–426, 2000.

[130] D. N. Chin, "Empirical evaluation of user models and user-adapted systems," *User Modeling and User-adapted Interaction*, vol. 11, no. 1-2, pp. 181–194, 2001.

[131] A. O. Mahdi, M. I. Alhabbash, and S. S. A. Naser, "An intelligent tutoring system for teaching advanced topics in information security," *World Wide Journal of Multidisciplinary Research and Development*, vol. 2, no. 12, pp. 1–9, 2016.

[132] M. W. Alawar and S. S. A. Naser, "Css-tutor: An intelligent tutoring system for css and html," *International Journal of Academic Research and Development*, vol. 2, no. 1, pp. 94–98, 2017.

[133] H. A. Al Rekhawi and S. S. Abu Naser, "An intelligent tutoring system for learning android applications ui development," *International Journal of Engineering and Information Systems (IJEAIS)*, vol. 2, no. 1, pp. 1–14, 2018.

[134] D. Sonntag, S. Zillner, P. van der Smagt, and A. Lörincz, "Overview of the cps for smart factories project: deep learning, knowledge acquisition, anomaly detection and intelligent user interfaces," in *Industrial Internet of Things*. Springer, 2017, pp. 487–504.

[135] H. Lee, Y. S. Choi, and Y.-J. Kim, "An adaptive user interface based on spatiotemporal structure learning," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 118–124, 2011.

[136] C. Zhang, X. Ding, G. Chen, K. Huang, X. Ma, and B. Yan, "Nihao: A predictive smartphone application launcher," in *International Conference on Mobile Computing, Applications, and Services*. Springer, 2012, pp. 294–313.

[137] G. López, L. Quesada, and L. A. Guerrero, "Alexa vs. siri vs. cortana vs. google assistant: a comparison of speech-based natural user interfaces," in *International Conference on Applied Human Factors and Ergonomics*. Springer, 2017, pp. 241–250.

[138] M. S. El-Nasr, "Interaction, narrative, and drama: Creating an adaptive interactive narrative using performance arts theories," *Interaction Studies*, vol. 8, no. 2, pp. 209–240, 2007.

[139] A. Corradini, M. Mehta, N. O. Bernsen, J. Martin, and S. Abrilian, "Multimodal input fusion in human-computer interaction," *NATO Science Series Sub Series III Computer and Systems Sciences*, vol. 198, p. 223, 2005.

[140] D. Lalanne, L. Nigay, P. Robinson, J. Vanderdonckt, J.-F. Ladry *et al.*, "Fusion engines for multimodal input: a survey," in *Proceedings of the 2009 International Conference on Multimodal interfaces*. ACM, 2009, pp. 153–160.

[141] M. Turk, "Multimodal human-computer interaction," in *Real-time Vision for Human-Computer Interaction*. Springer, 2005, pp. 269–283.

[142] J. Liu and M. Kavakli, "A survey of speech-hand gesture recognition for the development of multimodal interfaces in computer games," in *Multimedia and Expo (ICME), 2010 IEEE International Conf. on*. IEEE, 2010, pp. 1564–1569.

[143] M. Mansoorizadeh and N. M. Charkari, "Multimodal information fusion application to human emotion recognition from face and speech," *Multimedia Tools and Applications*, vol. 49, no. 2, pp. 277–297, 2010.

[144] P. R. Cohen, E. C. Kaiser, M. C. Buchanan, S. Lind, M. J. Corrigan, and R. M. Wesson, "Sketch-thru-plan: A multimodal interface for command and control," *Commun. ACM*, vol. 58, no. 4, pp. 56–65, Mar. 2015.

[145] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 2011, pp. 689–696.

[146] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Challenges and applications in multimodal machine learning," in *The Handbook of Multimodal-Multisensor Interfaces*. Association for Computing Machinery and Morgan & Claypool, 2018, pp. 17–48.

[147] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 39–58, 2009.

[148] M. Pantic, A. Pentland, A. Nijholt, and T. S. Huang, "Human computing and machine understanding of human behavior: A survey," in *Artifical Intelligence for Human Computing*. Springer, 2007, pp. 47–71.

[149] D. Reidsma, R. op den Akker, R. Rienks, R. Poppe, A. Nijholt, D. Heylen, and J. Zwiers, "Virtual meeting rooms: from observation to simulation," *AI & Society*, vol. 22, no. 2, pp. 133–144, 2007.

[150] Z. Zhu and T. S. Huang, *Multimodal surveillance: sensors, algorithms, and systems*. Norwood: Artech House, 2007.

[151] P. Roth and T. Pun, "Design and evaluation of multimodal system for the non-visual exploration of digital pictures." in *INTERACT*, 2003.

[152] J. S. Bradbury, J. S. Shell, and C. B. Knowles, "Hands on cooking: towards an attentive kitchen," in *CHI'03 extended abstracts on Human factors in Computing Systems*. ACM, 2003, pp. 996–997.

[153] L. Nie, L. Zhang, Y. Yang, M. Wang, R. Hong, and T.-S. Chua, "Beyond doctors: Future health prediction from multimedia and multimodal observations," in *Proceedings of the 23rd ACM International Conference on Multimedia*. ACM, 2015, pp. 591–600.

[154] A. Fleury, M. Vacher, and N. Noury, "Svm-based multimodal classification of activities of daily living in health smart homes: sensors, algorithms, and first experimental results," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 274–283, 2010.

[155] O. Brdiczka, M. Langet, J. Maisonnasse, and J. L. Crowley, "Detecting human behavior models from multimodal observation in a smart home," *IEEE Transactions on Automation Science and Engineering*, vol. 6, no. 4, pp. 588–597, 2009.

[156] Z. Lv, A. Halawani, S. Feng, H. Li, and S. U. Réhman, "Multimodal hand and foot gesture interaction for handheld devices," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 11, no. 1s, pp. 10:1–10:19, Oct. 2014.

[157] Y. Nam, B. Koo, A. Cichocki, and S. Choi, "Gom-face: Gkp, eog, and emg-based multimodal interface with application to humanoid robot control," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 2, pp. 453–462, 2014.

[158] S. Shaheen, A. Cohen, and E. Martin, "Smartphone app evolution and early understanding from a multimodal app user survey," in *Disrupting Mobility*. Springer, 2017, pp. 149–164.

[159] M. K. Bekele, R. Pierdicca, E. Frontoni, E. S. Malinverni, and J. Gain, "A survey of augmented, virtual, and mixed reality for cultural heritage," *Journal on Computing and Cultural Heritage (JOCCH)*, vol. 11, no. 2, p. 7, 2018.

[160] M. Billinghurst, A. Clark, G. Lee *et al.*, "A survey of augmented reality," *Foundations and Trends® in Human–Computer Interaction*, vol. 8, no. 2-3, pp. 73–272, 2015.

[161] S. Link, B. Barkschat, C. Zimmerer, M. Fischbach, D. Wiebusch, J.-L. Lugrin, and M. E. Latoschik, "An intelligent multimodal mixed reality real-time strategy game," in *2016 IEEE Virtual Reality (VR)*. IEEE, 2016, pp. 223–224.

[162] A. Pereira, E. J. Carter, I. Leite, J. Mars, and J. F. Lehman, "Augmented reality dialog interface for multimodal teleoperation," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2017, pp. 764–771.

[163] P. Gang, J. Hui, S. Stirenko, Y. Gordienko, T. Shemsedinov, O. Alienin, Y. Kochura, N. Gordienko, A. Rojbi, J. L. Benito *et al.*, "User-driven intelligent interface on the basis of multimodal augmented reality and brain-computer interaction for people with functional disabilities," in *Future of Information and Communication Conference*. Springer, 2018, pp. 612–631.

[164] I. Wicaksono and J. A. Paradiso, "Fabrickeyboard: multimodal textile sensate media as an expressive and deformable musical interface." in *NIME*, 2017, pp. 348–353.

[165] K. C. Ng, T. Weyde, O. Larkin, K. Neubarth, T. Koerselman, and B. Ong, "3D augmented mirror: a multimodal interface for string instrument learning and teaching with gesture support," in *Proceedings of the 9th International Conference on Multimodal interfaces*. ACM, 2007, pp. 339–345.

[166] D. Overholt, J. Thompson, L. Putnam, B. Bell, J. Kleban, B. Sturm, and J. Kuchera-Morin, "A multimodal system for gesture recognition in interactive music performance," *Computer Music Journal*, vol. 33, no. 4, pp. 69–82, 2009.

[167] C. L. Lisetti and F. Nasoz, "Maui: a multimodal affective user interface," in *Proceedings of the tenth ACM International Conference on Multimedia*. ACM, 2002, pp. 161–170.

[168] N. A. Shaked, “Avatars and virtual agents–relationship interfaces for the elderly,” *Healthcare Technology Letters*, vol. 4, no. 3, pp. 83–87, 2017.

[169] F. Liarokapis, P. Petridis, D. Andrews, and S. de Freitas, “Multimodal serious games technologies for cultural heritage,” in *Mixed Reality and Gamification for Cultural Heritage*. Springer, 2017, pp. 371–392.

[170] C. Capurro, D. Nollet, and D. Pletinckx, “Tangible interfaces for digital museum applications,” in *2015 Digital Heritage*, vol. 1. IEEE, 2015, pp. 271–276.

[171] J. Leaman and H. M. La, “A comprehensive review of smart wheelchairs: past, present, and future,” *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 4, pp. 486–499, 2017.

[172] C. Munteanu, P. Irani, S. Oviatt, M. Aylett, G. Penn, S. Pan, N. Sharma, F. Rudzicz, R. Gomez, K. Nakamura *et al.*, “Designing speech and multimodal interactions for mobile, wearable, and pervasive applications,” in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2016, pp. 3612–3619.

[173] F. Miralles, E. Vargiu, X. Rafael-Palou, M. Solà, S. Dauwalder, C. Guger, C. Hintermüller, A. Espinosa, H. Lowish, S. Martin *et al.*, “Brain–computer interfaces on track to home: results of the evaluation at disabled end-users’ homes and lessons learnt,” *Frontiers in ICT*, vol. 2, p. 25, 2015.

[174] A. L. S. Ferreira, L. C. de Miranda, E. E. C. de Miranda, and S. G. Sakamoto, “A survey of interactive systems based on brain-computer interfaces,” *SBC Journal on Interactive Systems*, vol. 4, no. 1, pp. 3–13, 2013.

[175] J.-S. Lin and W.-C. Yang, “Wireless brain-computer interface for electric wheelchairs with EEG and eye-blinking signals,” *International Journal of Innovative Computing, Information and Control*, vol. 8, no. 9, pp. 6011–6024, 2012.

[176] F. Chen, N. Ruiz, E. Choi, J. Epps, M. A. Khawaja, R. Taib, B. Yin, and Y. Wang, “Multimodal behavior and interaction as indicators of cognitive load,” *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 2, no. 4, p. 22, 2012.

[177] M. D. Hauser, N. Chomsky, and W. T. Fitch, “The faculty of language: what is it, who has it, and how did it evolve?” *Science*, vol. 298, no. 5598, pp. 1569–1579, 2002.

[178] B. Libet, “Do we have free will?” *Journal of Consciousness Studies*, vol. 6, no. 8-9, pp. 47–57, 1999.

[179] B. S. Reddy, *Diet, Nutrition and Cancer: a Critical Evaluation*. Boca Raton: CRC Press, 2018, ch. 8, pp. 47–66.

[180] A. S. Reber, *The Penguin dictionary of psychology*. New York, NY, US: Penguin Press, 1995.

[181] D. Medin, B. Ross, and A. Markman, *Cognitive Psychology*, 4th ed. NY, US: wiley, 2005.

[182] G. Repovš and A. Baddeley, “The multi-component model of working memory: Explorations in experimental cognitive psychology,” *Neuroscience*, vol. 139, no. 1, pp. 5–21, 2006.

[183] A. Kirkwood, H.-K. Lee, and M. F. Bear, “Co-regulation of long-term potentiation and experience-dependent synaptic plasticity in visual cortex by age and experience,” *Nature*, vol. 375, no. 6529, p. 328, 1995.

[184] A. Baddeley, “Working memory: theories, models, and controversies,” *Annual Review of Psychology*, vol. 63, pp. 1–29, 2012.

[185] R. L. Buckner, "Beyond hera: Contributions of specific prefrontal brain areas to long-term memory retrieval," *Psychonomic Bulletin & Review*, vol. 3, no. 2, pp. 149–158, 1996.

[186] T. Brashers-Krug, R. Shadmehr, and E. Bizzi, "Consolidation in human motor memory," *Nature*, vol. 382, no. 6588, p. 252, 1996.

[187] W. McDougall, *An introduction to social psychology*. London, UK: Psychology Press, 2015.

[188] A. Özyürek, R. M. Willems, S. Kita, and P. Hagoort, "On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials," *Journal of Cognitive Neuroscience*, vol. 19, no. 4, pp. 605–616, 2007.

[189] A. S. Dick, S. Goldin-Meadow, U. Hasson, J. I. Skipper, and S. L. Small, "Co-speech gestures influence neural activity in brain regions associated with processing semantic information," *Human Brain Mapping*, vol. 30, no. 11, pp. 3509–3526, 2009.

[190] A. D. Baddeley, *Essentials of human memory*. Hove, England: Psychology Press, 1999.

[191] G. A. Miller, "The magical number seven, plus or minus two: some limits on our capacity for processing information." *Psychological Review*, vol. 63, no. 2, p. 81, 1956.

[192] J.-P. Thiran, H. Bourlard, and F. Marques, *Multimodal Signal Processing: Methods and Techniques to Build Multimodal Interactive Systems*. Academic Press, 2009, no. LIDIAP-BOOK-2009-003. [Online]. Available: http://infoscience.epfl.ch/record/146300

[193] E. Morsella and R. M. Krauss, "The role of gestures in spatial working memory and speech," *The American Journal of Psychology*, pp. 411–424, 2004.

[194] R. Ping and S. Goldin-Meadow, "Gesturing saves cognitive resources when talking about nonpresent objects," *Cognitive Science*, vol. 34, no. 4, pp. 602–619, 2010.

[195] S. W. Cook, T. K. Yip, and S. Goldin-Meadow, "Gestures, but not meaningless movements, lighten working memory load when explaining math," *Language and Cognitive Processes*, vol. 27, no. 4, pp. 594–610, 2012.

[196] D. C. Geary, S. J. Saults, F. Liu, and M. K. Hoard, "Sex differences in spatial cognition, computational fluency, and arithmetical reasoning," *Journal of Experimental Child Psychology*, vol. 77, no. 4, pp. 337–353, 2000.

[197] J. A. Fodor, "Precis of the modularity of mind," *Behavioral and Brain Sciences*, vol. 8, no. 1, pp. 1–5, 1985.

[198] A. Caramazza, D. Chialant, R. Capasso, and G. Miceli, "Separable processing of consonants and vowels," *Nature*, vol. 403, no. 6768, p. 428, 2000.

[199] J. P. Spencer, M. Clearfield, D. Corbetta, B. Ulrich, P. Buchanan, and G. Schöner, "Moving toward a grand theory of development: In memory of esther thelen," *Child Development*, vol. 77, no. 6, pp. 1521–1538, 2006.

[200] D. E. Meyer, J. M. Glass, S. T. Mueller, T. L. Seymour, and D. E. Kieras, "Executive-process interactive control: A unified computational theory for answering 20 questions (and more) about cognitive ageing," *European Journal of Cognitive Psychology*, vol. 13, no. 1-2, pp. 123–164, 2001.

[201] M. Veloso, J. Carbonell, A. Perez, D. Borrajo, E. Fink, and J. Blythe, "Integrating planning and learning: The prodigy architecture," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 7, no. 1, pp. 81–120, 1995.

[202] Y. Yoshimura, J. L. Dantzker, and E. M. Callaway, "Excitatory cortical neurons form fine-scale functional networks," *Nature*, vol. 433, no. 7028, p. 868, 2005.

[203] F. Musso, J. Brinkmeyer, A. Mobascher, T. Warbrick, and G. Winterer, "Spontaneous brain activity and EEG microstates. a novel EEG/fMRI analysis approach to explore resting-state networks," *Neuroimage*, vol. 52, no. 4, pp. 1149–1161, 2010.

[204] S. H. Snyder and C. D. Ferris, "Novel neurotransmitters and their neuropsychiatric relevance," *American Journal of Psychiatry*, vol. 157, no. 11, pp. 1738–1751, 2000.

[205] K. Sidiropoulou, E. K. Pissadaki, and P. Poirazi, "Inside the brain of a neuron," *EMBO Reports*, vol. 7, no. 9, pp. 886–892, 2006.

[206] S. Achard and E. Bullmore, "Efficiency and cost of economical brain functional networks," *PLoS Computational Biology*, vol. 3, no. 2, p. e17, 2007.

[207] C. J. Stam and J. C. Reijneveld, "Graph theoretical analysis of complex networks in the brain," *Nonlinear Biomedical Physics*, vol. 1, no. 1, p. 3, 2007.

[208] J. M. G. Williams, A. Mathews, and C. MacLeod, "The emotional stroop task and psychopathology." *Psychological Bulletin*, vol. 120, no. 1, p. 3, 1996.

[209] C. Berka, D. J. Levendowski, M. N. Lumicao, A. Yau, G. Davis, V. T. Zivkovic, R. E. Olmstead, P. D. Tremoulet, and P. L. Craven, "EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks," *Aviation, Space, and Environmental Medicine*, vol. 78, no. 5, pp. B231–B244, 2007.

[210] S. M. Jaeggi, M. Buschkuehl, W. J. Perrig, and B. Meier, "The concurrent validity of the n-back task as a working memory measure," *Memory*, vol. 18, no. 4, pp. 394–412, 2010.

[211] Y. Liu and X. Fu, "How does distraction task influence the interaction of working memory and long-term memory?" in *International Conference on Engineering Psychology and Cognitive Ergonomics*. Springer, 2007, pp. 366–374.

[212] G. Georgiou and C. A. Essau, "Go/no-go task," *Encyclopedia of Child Behavior and Development*, pp. 705–706, 2011.

[213] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[214] L. M. Glynn, N. Christenfeld, and W. Gerin, "The role of rumination in recovery from reactivity: Cardiovascular consequences of emotional states," *Psychosomatic Medicine*, vol. 64, no. 5, pp. 714–726, 2002.

[215] P. Ekman, "Facial expression and emotion." *American Psychologist*, vol. 48, no. 4, p. 384, 1993.

[216] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer, *Eye tracking: A comprehensive guide to methods and measures*. Oxford, UK: Oxford University Press, 2011.

[217] M. McKee, "Biofeedback: an overview in the context of heart-brain medicine." *Cleveland Clinic Journal of Medicine*, vol. 75, pp. S31–4, 2008.

[218] K. Sterelny, "Thought in a hostile world: The evolution of human cognition," *Philosophy and Phenomenological Research*, vol. 74, no. 2, pp. 476–497, 2007.

[219] W. Reeves, "System for simultaneously producing and synchronizing spectral patterns of heart sounds and an ecg signal," Aug. 16 1994, US Patent 5,337,752.

[220] I. ECG, "Electrocardiography i," https://www.biopac.com/curriculum/l05-electrocardiography-i/, [Online; accessed 06-Jun-2019].

[221] C.-T. Lin, L.-W. Ko, M.-H. Chang, J.-R. Duann, J.-Y. Chen, T.-P. Su, and T.-P. Jung, "Review of wireless and wearable electroencephalogram systems and brain-computer interfaces–a mini-review," *Gerontology*, vol. 56, no. 1, pp. 112–119, 2010.

[222] D. Scherr, D. Dalal, C. A. Henrikson, D. D. Spragg, R. D. Berger, H. Calkins, and A. Cheng, "Prospective comparison of the diagnostic utility of a standard event monitor versus a "leadless" portable ecg monitor in the evaluation of patients with palpitations," *Journal of Interventional Cardiac Electrophysiology*, vol. 22, no. 1, pp. 39–44, 2008.

[223] D. Jabaudon, J. Sztajzel, K. Sievert, T. Landis, and R. Sztajzel, "Usefulness of ambulatory 7-day ecg monitoring for the detection of atrial fibrillation and flutter after acute stroke and transient ischemic attack," *Stroke*, vol. 35, no. 7, pp. 1647–1651, 2004.

[224] C. Park, P. H. Chou, Y. Bai, R. Matthews, and A. Hibbs, "An ultra-wearable, wireless, low power ecg monitoring system," in *Biomedical Circuits and Systems Conference, 2006. BioCAS 2006. IEEE*. IEEE, 2006, pp. 241–244.

[225] E. Fiolet, "Neurosky mindwave brings brain-computer interface to education," *) ˆ(Eds.):'Book NeuroSky MindWave Brings Brain-Computer Interface to Education'(2011, edn.)*, 2011.

[226] C. Grozea, C. D. Voinescu, and S. Fazli, "Bristle-sensors-low-cost flexible passive dry EEG electrodes for neurofeedback and bci applications," *Journal of Neural Engineering*, vol. 8, no. 2, p. 025008, 2011.

[227] E. Emotiv, "Neuroheadset," https://www.emotiv.com/epoc/, 2012, [Online; accessed 06-Jun-2019].

[228] D. B. Hunter, K. B. McCarraher, K. P. Fratto, R. M. Brueggman, T. J. Stenstrom, and H. I. Gustafson Jr, "Headset for EEG measurements," Nov. 28 2000, US Patent 6,154,669.

[229] J. Amores, X. Benavides, and P. Maes, "Psychicvr: Increasing mindfulness by using virtual reality and brain computer interfaces," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2016, pp. 2–2.

[230] D. Burton, "Vigilance monitoring system," Jun. 10 2003, US Patent 6,575,902.

[231] M. Alam and A. Azad, "Development of biomedical data acquisition system in hard real-time linux environment," in *Biomedical Engineering (ICoBE), 2012 International Conference on*. IEEE, 2012, pp. 436–440.

[232] C. Wong, Z.-Q. Zhang, B. Lo, and G.-Z. Yang, "Wearable sensing for solid biomechanics: A review," *IEEE Sensors Journal*, vol. 15, no. 5, pp. 2747–2760, 2015.

[233] S. Ishimaru, K. Kunze, Y. Uema, K. Kise, M. Inami, and K. Tanaka, "Smarter eyewear: using commercial eog glasses for activity recognition," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 2014, pp. 239–242.

[234] S. Instruments, "Smi eye tracking glasses," https://imotions.com/hardware/smi-eye-tracking-glasses/, 2012, [Online; accessed 06-Jun-2019].

[235] M. Garbarino, M. Lai, D. Bender, R. W. Picard, and S. Tognetti, "Empatica e3-a wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition," in *Wireless Mobile Communication and Healthcare (Mobihealth), 2014 EAI 4th International Conference on*. IEEE, 2014, pp. 39–42.

[236] A. Burns, B. R. Greene, M. J. McGrath, T. J. O'Shea, B. Kuris, S. M. Ayer, F. Stroiescu, and V. Cionca, "Shimmer–a wireless sensor platform for noninvasive biomedical research," *IEEE Sensors Journal*, vol. 10, no. 9, pp. 1527–1534, 2010.

[237] S. K. Panigrahy, S. K. Jena, and A. K. Turuk, "Study and analysis of human stress detection using galvanic skin response (gsr) sensor inwired and wireless environments," *Research Journal of Pharmacy and Technology*, vol. 10, no. 2, pp. 545–550, 2017.

[238] J. McKenzie and D. Osgood, "Validation of a new telemetric core temperature monitor," *Journal of Thermal Biology*, vol. 29, no. 7-8, pp. 605–611, 2004.

[239] A. Bonarini, M. Matteucci, and S. Tognetti, "Chest expansion reconstruction from respiration sound by using artificial neural networks," *IET Conference Proceedings*, pp. 126–126(1), January 2008.

[240] V. Jones, A. Van Halteren, N. Dokovsky, G. Koprinkov, J. Peuscher, R. Bults, D. Konstantas, R. Herzog *et al.*, "Mobihealth: mobile services for health professionals," in *M-Health*. Springer, 2006, pp. 237–246.

[241] A. Albulbul, "Evaluating major electrode types for idle biological signal measurements for modern medical technology," *Bioengineering*, vol. 3, no. 3, p. 20, 2016.

[242] X. Zhang, J. Li, Y. Liu, Z. Zhang, Z. Wang, D. Luo, X. Zhou, M. Zhu, W. Salman, G. Hu *et al.*, "Design of a fatigue detection system for high-speed trains based on driver vigilance using a wireless wearable EEG," *Sensors*, vol. 17, no. 3, p. 486, 2017.

[243] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 3, pp. 289–294, 2003.

[244] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Measurement*, vol. 28, no. 3, p. R1, 2007.

[245] A. Barreto, J. Zhai, and M. Adjouadi, "Non-intrusive physiological monitoring for automated stress detection in human-computer interaction," in *International Workshop on Human-Computer Interaction*. Springer, 2007, pp. 29–38.

[246] J. Healey, "Physiological sensing of emotion," *The Oxford handbook of affective computing*, pp. 204–216, 2014.

[247] G. G. Berntson, "and interpretive caveats," *Psychophysiology*, vol. 34, pp. 623–648, 1997.

[248] C. M. van Ravenswaaij-Arts, L. A. Kollee, J. C. Hopman, G. B. Stoelinga, and H. P. van Geijn, "Heart rate variability," *Annals of Internal Medicine*, vol. 118, no. 6, pp. 436–447, 1993.

[249] D. W. Rowe, J. Sibert, and D. Irwin, "Heart rate variability: Indicator of user state as an aid to human-computer interaction," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM Press/Addison-Wesley Publishing Co., 1998, pp. 480–487.

[250] S. Akselrod, D. Gordon, F. A. Ubel, D. C. Shannon, A. Berger, R. J. Cohen *et al.*, "Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardiovascular control," *Science*, vol. 213, no. 4504, pp. 220–222, 1981.

[251] H. Itoh, K. Takeda, K. Nakamura, H. Fujita, A. Uchida, T. Kuwahara, K. Takenaka, M. Tanaka, T. Nakata, S. Sasaki *et al.*, "Young borderline hypertensives are hyperreactive to mental arithmetic stress: spectral analysis of rr intervals," *Journal of the Autonomic Nervous System*, vol. 54, no. 2, pp. 155–162, 1995.

[252] J. T. Cacioppo, G. G. Berntson, J. T. Larsen, K. M. Poehlmann, and T. A. Ito, "The psychophysiology of emotion," *Handbook of Emotions*, vol. 2, pp. 173–191, 2000.

[253] C. G. Jung, *Studies in word-association*. Taylor & Francis, 1969.

[254] W. M. Marston, *The lie detector test*. Oxford, England: R. R. Smith, 1989.

[255] W. Boucsein, *Electrodermal activity*. Boston, MA: Springer Science & Business Media, 2012.

[256] J. Healey, "Gsr sock: A new e-textile sensor prototype," in *2011 15th Annual International Symposium on Wearable Computers*. IEEE, 2011, pp. 113–114.

[257] R. W. Picard and J. Healey, "Affective wearables," *Personal Technologies*, vol. 1, no. 4, pp. 231–240, 1997.

[258] J. Healey, "Recording affect in the field: towards methods and metrics for improving ground truth labels," in *International Conference on Affective Computing and Intelligent Interaction*. Springer, 2011, pp. 107–116.

[259] L. F. Haas, "Hans berger (1873–1941), richard caton (1842–1926), and electroencephalography," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 74, no. 1, pp. 9–9, 2003.

[260] L. Douw, M. De Groot, E. Van Dellen, J. J. Heimans, H. E. Ronner, C. J. Stam, and J. C. Reijneveld, "'functional connectivity'is a sensitive predictor of epilepsy diagnosis after the first seizure," *PLoS One*, vol. 5, no. 5, p. e10839, 2010.

[261] E. F. Wijdicks, "The diagnosis of brain death," *New England Journal of Medicine*, vol. 344, no. 16, pp. 1215–1221, 2001.

[262] R. M. Benca, W. H. Obermeyer, R. A. Thisted, and J. C. Gillin, "Sleep and psychiatric disorders: a meta-analysis," *Archives of General Psychiatry*, vol. 49, no. 8, pp. 651–668, 1992.

[263] J. Jeong, "EEG dynamics in patients with alzheimer's disease," *Clinical Neurophysiology*, vol. 115, no. 7, pp. 1490–1505, 2004.

[264] G. Greisen, L. Hellström-Westas, H. Lou, I. Rosén, and N. Svenningsen, "EEG depression and germinal layer haemorrhage in the newborn," *Acta Pædiatrica*, vol. 76, no. 3, pp. 519–525, 1987.

[265] J. A. Coan and J. J. Allen, "Frontal EEG asymmetry as a moderator and mediator of emotion," *Biological Psychology*, vol. 67, no. 1, pp. 7–50, 2004.

[266] B. J. Baars and N. M. Gage, *Cognition, brain, and consciousness: Introduction to cognitive neuroscience*. Burlington, MA, USA: Academic Press, 2010.

[267] A. Buchmann, M. Ringli, S. Kurth, M. Schaerer, A. Geiger, O. G. Jenni, and R. Huber, "EEG sleep slow-wave activity as a mirror of cortical maturation," *Cerebral Cortex*, vol. 21, no. 3, pp. 607–615, 2010.

[268] G. B. Young, "The EEG in coma," *Journal of Clinical Neurophysiology*, vol. 17, no. 5, pp. 473–485, 2000.

[269] V. Billard, P. L. Gambus, N. Chamoun, D. R. Stanski, and S. L. Shafer, "A comparison of spectral edge, delta power, and bispectral index as EEG measures of alfentanil, propofol, and midazolam drug effect," *Clinical Pharmacology & Therapeutics*, vol. 61, no. 1, pp. 45–58, 1997.

[270] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," *Brain Research Reviews*, vol. 29, no. 2-3, pp. 169–195, 1999.

[271] R. J. Huster, S. Enriquez-Geppert, C. F. Lavallee, M. Falkenstein, and C. S. Herrmann, "Electroencephalography of response inhibition tasks: functional networks and cognitive contributions," *International Journal of Psychophysiology*, vol. 87, no. 3, pp. 217–233, 2013.

[272] M. Arns, C. K. Conners, and H. C. Kraemer, “A decade of EEG theta/beta ratio research in ADHD: a meta-analysis,” *Journal of Attention Disorders*, vol. 17, no. 5, pp. 374–383, 2013.

[273] D. Begić, L. Hotujac, and N. Jokić-Begić, “Quantitative EEG in ’positive’ and ’negative’ schizophrenia,” *Acta Psychiatrica Scandinavica*, vol. 101, no. 4, pp. 307–311, 2000.

[274] S. K. Lal and A. Craig, “A critical review of the psychophysiology of driver fatigue,” *Biological Psychology*, vol. 55, no. 3, pp. 173–194, 2001.

[275] D. Röhm, W. Klimesch, H. Haider, and M. Doppelmayr, “The role of theta and alpha oscillations for language comprehension in the human electroencephalogram,” *Neuroscience Letters*, vol. 310, no. 2-3, pp. 137–140, 2001.

[276] J. van Driel, K. R. Ridderinkhof, and M. X. Cohen, “Not all errors are alike: theta and alpha EEG dynamics relate to differences in error-processing dynamics,” *Journal of Neuroscience*, vol. 32, no. 47, pp. 16 795–16 806, 2012.

[277] A. Perry, N. F. Troje, and S. Bentin, “Exploring motor system contributions to the perception of social information: Evidence from EEG activity in the mu/alpha frequency range,” *Social Neuroscience*, vol. 5, no. 3, pp. 272–284, 2010.

[278] B. Berger, S. Omer, T. Minarik, A. Sterr, and P. Sauseng, “Interacting memory systems-does EEG alpha activity respond to semantic long-term memory access in a working memory task?” *Biology*, vol. 4, no. 1, pp. 1–16, 2015.

[279] T. A. De Graaf, J. Gross, G. Paterson, T. Rusch, A. T. Sack, and G. Thut, “Alpha-band rhythms in visual task performance: phase-locking by rhythmic sensory stimulation,” *PloS One*, vol. 8, no. 3, p. e60035, 2013.

[280] W. J. Ray and H. W. Cole, “EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes,” *Science*, vol. 228, no. 4700, pp. 750–752, 1985.

[281] G. Pfurtscheller, “Central beta rhythm during sensorimotor activities in man,” *Electroencephalography and Clinical Neurophysiology*, vol. 51, no. 3, pp. 253–264, 1981.

[282] K. J. Stam, D. L. Tavy, B. Jelles, H. A. Achtereekte, J. P. Slaets, and R. W. Keunen, “Non-linear dynamical analysis of multichannel EEG: clinical applications in dementia and parkinson’s disease,” *Brain Topography*, vol. 7, no. 2, pp. 141–150, 1994.

[283] R. I. Herning, R. T. Jones, W. D. Hooker, J. Mendelson, and L. Blackwell, “Cocaine increases EEG beta: a replication and extension of hans berger’s historic experiments,” *Electroencephalography and Clinical Neurophysiology*, vol. 60, no. 6, pp. 470–477, 1985.

[284] S. Velikova, M. Locatelli, C. Insacco, E. Smeraldi, G. Comi, and L. Leocani, “Dysfunctional brain circuitry in obsessive–compulsive disorder: source and coherence analysis of EEG rhythms,” *Neuroimage*, vol. 49, no. 1, pp. 977–983, 2010.

[285] R. Ferri, F. I. Cosentino, M. Elia, S. A. Musumeci, R. Marinig, and P. Bergonzi, “Relationship between delta, sigma, beta, and gamma EEG bands at REM sleep onset and REM sleep end,” *Clinical Neurophysiology*, vol. 112, no. 11, pp. 2046–2052, 2001.

[286] C. S. Herrmann, I. Fründ, and D. Lenz, “Human gamma-band activity: a review on cognitive and behavioral correlates and network models,” *Neuroscience & Biobehavioral Reviews*, vol. 34, no. 7, pp. 981–992, 2010.

[287] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin, “Static and dynamic 3D facial expression recognition: A comprehensive survey,” *Image and Vision Computing*, vol. 30, no. 10, pp. 683–697, 2012.

[288] E. Ganglbauer, J. Schrammel, S. Deutsch, and M. Tscheligi, "Applying psychophysiological methods for measuring user experience: possibilities, challenges and feasibility," in *Workshop on User Experience Evaluation Methods in Product Development*. Citeseer, 2009.

[289] R. Horlings, D. Datcu, and L. J. Rothkrantz, "Emotion recognition using brain activity," in *Proceedings of the 9th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing*. ACM, 2008, p. 6.

[290] P. E. Ekman and R. J. Davidson, *The nature of emotion: Fundamental questions.* Oxford, UK: Oxford University Press, 1994.

[291] P. J. Lang, "The emotion probe: Studies of motivation and attention." *American Psychologist*, vol. 50, no. 5, p. 372, 1995.

[292] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (iaps): Technical manual and affective ratings," *NIMH Center for the Study of Emotion and Attention*, pp. 39–58, 1997.

[293] T. Partala and V. Surakka, "The effects of affective interventions in human–computer interaction," *Interacting with Computers*, vol. 16, no. 2, pp. 295–309, 2004.

[294] D. Norman, "Emotion & design: attractive things work better," *Interactions*, vol. 9, no. 4, pp. 36–42, 2002.

[295] A. Haag, S. Goronzy, P. Schaich, and J. Williams, "Emotion recognition using biosensors: First steps towards an automatic system," in *Tutorial and Research Workshop on Affective Dialogue Systems*. Springer, 2004, pp. 36–48.

[296] J. Klein, Y. Moon, and R. W. Picard, "This computer responds to user frustration:: Theory, design, and results," *Interacting with Computers*, vol. 14, no. 2, pp. 119–140, 2002.

[297] T. Partala and V. Surakka, "Pupil size variation as an indication of affective processing," *International Journal of Human-Computer Studies*, vol. 59, no. 1, pp. 185–198, 2003.

[298] N. H. Frijda, *The emotions: Studies in emotion and social interaction*. New York, NY, US: Cambridge University Press, 1986.

[299] R. L. Mandryk and M. S. Atkins, "A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies," *International Journal of Human-Computer Studies*, vol. 65, no. 4, pp. 329–347, 2007.

[300] L. Juma, "The role of secondary emotions in action selection and its effects on the believability of a character," *Master Programme Final Project, Aalborg University Copenhagen*, 2008.

[301] T. Costa, E. Rognoni, and D. Galati, "EEG phase synchronization during emotional response to positive and negative film stimuli," *Neuroscience Letters*, vol. 406, no. 3, pp. 159–164, 2006.

[302] M. Li and B.-L. Lu, "Emotion classification based on gamma-band EEG," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE, 2009, pp. 1223–1226.

[303] X.-W. Wang, D. Nie, and B.-L. Lu, "EEG-based emotion recognition using frequency domain features and support vector machines," in *International Conference on Neural Information Processing*. Springer, 2011, pp. 734–743.

[304] T. Thompson, T. Steffert, T. Ros, J. Leach, and J. Gruzelier, "EEG applications for sport and performance," *Methods*, vol. 45, no. 4, pp. 279–288, 2008.

[305] A. Nakasone, H. Prendinger, and M. Ishizuka, "Emotion recognition from electromyography and skin conductance," in *Proc. of the 5th International Workshop on Biosignal Interpretation*. Citeseer, 2005, pp. 219–222.

[306] N. M. Khair, S. Yaacob, M. Hariharan, and S. N. Basah, "A study of human emotional: Review," in *2012 International Conference on Biomedical Engineering (ICoBE)*, Feb 2012, pp. 393–399.

[307] R. T. Boone and J. G. Cunningham, "Children's expression of emotional meaning in music through expressive body movement," *Journal of Nonverbal Behavior*, vol. 25, no. 1, pp. 21–41, 2001.

[308] F. Zhou, X. Qu, J. R. Jiao, and M. G. Helander, "Emotion prediction from physiological signals: A comparison study between visual and auditory elicitors," *Interacting with Computers*, vol. 26, no. 3, pp. 285–302, 2014.

[309] D. Mehta, M. Siddiqui, and A. Javaid, "Facial emotion recognition: A survey and real-world user experiences in mixed reality," *Sensors*, vol. 18, no. 2, p. 416, 2018.

[310] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42–55, 2012.

[311] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe, "Decaf: Meg-based multimodal database for decoding affective physiological responses," *IEEE Transactions on Affective Computing*, vol. 6, no. 3, pp. 209–222, July 2015.

[312] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[313] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.

[314] A.-F. N. M. Perrin, H. Xu, E. Kroupi, M. Řeřábek, and T. Ebrahimi, "Multimodal dataset for assessment of quality of experience in immersive multimedia," in *Proceedings of the 23rd ACM International Conference on Multimedia*. ACM, 2015, pp. 1007–1010.

[315] A.-F. Perrin, M. Řeřábek, and T. Ebrahimi, "Towards prediction of sense of presence in immersive audiovisual communications," *Electronic Imaging*, vol. 2016, no. 16, pp. 1–8, 2016.

[316] J. Ruohonen and S. Hyrynsalmi, "Evaluating the use of internet search volumes for time series modeling of sales in the video game industry," *Electronic Markets*, vol. 27, no. 4, pp. 351–370, 2017.

[317] R. Bernhaupt and F. F. Mueller, "Game user experience evaluation," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2016, pp. 940–943.

[318] R. Bernhaupt, *Game User Experience Evaluation*. Springer Publishing Switzerland, 2015.

[319] J. M. Kivikangas, G. Chanel, B. Cowley, I. Ekman, M. Salminen, S. Järvelä, and N. Ravaja, "A review of the use of psychophysiological methods in game research," *Journal of Gaming & Virtual Worlds*, vol. 3, pp. 181–199, 09 2011.

[320] W. Ijsselsteijn, W. van den Hoogen, C. Klimmt, Y. de Kort, C. Lindley, K. Mathiak, K. Poels, N. Ravaja, M. Turpeinen, and P. Vorderer, "Measuring the experience of digital game enjoyment," *6th International Conference on Methods and Techniques in Behavioral Research, Maastricht, The Netherlands, August 26-29, 2008*, pp. 88–89, 2008.

[321] A. Drachen, L. E. Nacke, G. Yannakakis, and A. L. Pedersen, “Correlation between heart rate, electrodermal activity and player experience in first-person shooter games,” in *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games*, 2010, pp. 49–54.

[322] M. E. Ballard and J. R. Wiest, “Mortal kombat (tm): The effects of violent videogame play on males’ hostility and cardiovascular responding,” *Journal of Applied Social Psychology*, vol. 26, no. 8, pp. 717–730, 1996.

[323] A. Drachen, L. Nacke, G. Yannakakis, and A. Pedersen, “Psychophysiological correlations with gameplay experience dimensions,” *Brain, Body and Bytes Workshop, Chi 2010*, 2010.

[324] G. N. Yannakakis, J. Hallam, and H. H. Lund, “Entertainment capture through heart rate activity in physical interactive playgrounds,” *User Modeling and User-adapted Interaction*, vol. 18, no. 1, pp. 207–243, 2008.

[325] C. Pedersen, J. Togelius, and G. N. Yannakakis, “Modeling player experience in super mario bros,” in *2009 IEEE Symposium on Computational Intelligence and Games*, 2009, pp. 132–139.

[326] T. McMahan, I. Parberry, and T. D. Parsons, “Modality specific assessment of video game player’s experience using the emotiv,” *Entertainment Computing*, vol. 7, pp. 1–6, 2015.

[327] L. E. Nacke, M. N. Grimshaw, and C. A. Lindley, “More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game,” *Interacting with Computers*, vol. 22, no. 5, pp. 336–343, 2010.

[328] B. Cowley, I. Kosunen, P. Lankoski, J. M. Kivikangas, S. Järvelä, I. Ekman, J. Kemppainen, and N. Ravaja, “Experience assessment and design in the analysis of gameplay,” *Simulation & Gaming*, vol. 45, no. 1, pp. 41–69, 2014.

[329] A. Stein, Y. Yotam, R. Puzis, G. Shani, and M. Taieb-Maimon, “Eeg-triggered dynamic difficulty adjustment for multiplayer games,” *Entertainment Computing*, vol. 25, pp. 14 – 25, 2018.

[330] A. Drachen, A. Canossa, and G. N. Yannakakis, “Player modeling using self-organization in tomb raider: Underworld,” in *2009 IEEE Symposium on Computational Intelligence and Games*, 2009, pp. 1–8.

[331] J. Togelius, R. D. Nardi, and S. M. Lucas, “Making racing fun through player modeling and track evolution,” 2006. [Online]. Available: http://cogprints.org/5221/

[332] O. Missura and T. Gärtner, “Player modeling for intelligent difficulty adjustment,” in *DS ’09 Proceedings of the 12th International Conference on Discovery Science*, 2009, pp. 197–211.

[333] M. Čertický, M. Čertický, P. Sinčák, G. Magyar, J. Vaščák, and F. Cavallo, “Psychophysiological indicators for modeling user experience in interactive digital entertainment,” *Sensors*, vol. 19, no. 5, p. 989, 2019.

[334] A. Lobel, M. Gotsis, E. Reynolds, M. Annetta, R. C. Engels, and I. Granic, “Designing and utilizing biofeedback games for emotion regulation: The case of nevermind,” in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2016, pp. 1945–1951.

[335] C. L. Bethel, J. L. Burke, R. R. Murphy, and K. Salomon, “Psychophysiological experimental design for use in human-robot interaction studies,” in *2007 International Symposium on Collaborative Technologies and Systems*. IEEE, 2007, pp. 99–105.

[336] K. Itoh, H. Miwa, Y. Nukariya, M. Zecca, H. Takanobu, S. Roccella, M. C. Carrozza, P. Dario, and A. Takanishi, "Development of a bioinstrumentation system in the interaction between a human and a robot," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2006, pp. 2620–2625.

[337] E. Bekele and N. Sarkar, "Psychophysiological feedback for adaptive human–robot interaction (hri)," in *Advances in Physiological Computing*. Springer, 2014, pp. 141–167.

[338] P. Rani and N. Sarkar, "Making robots emotion-sensitive-preliminary experiments and results," in *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005*. IEEE, 2005, pp. 1–6.

[339] D. Croft, "Estimating intent for human-robot interaction," in *IEEE International Conference on Advanced Robotics*. Citeseer, 2003, pp. 810–815.

[340] F. Dehais, E. A. Sisbot, R. Alami, and M. Causse, "Physiological and subjective evaluation of a human–robot object hand-over task," *Applied Ergonomics*, vol. 42, no. 6, pp. 785–791, 2011.

[341] L. Paletta, A. Dini, C. Murko, S. Yahyanejad, M. Schwarz, G. Lodron, S. Ladstätter, G. Paar, and R. Velik, "Towards real-time probabilistic evaluation of situation awareness from human gaze in human-robot interaction," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2017, pp. 247–248.

[342] G. Podevijn, R. Oǵrady, N. Mathews, A. Gilles, C. Fantini-Hauwel, and M. Dorigo, "Investigating the effect of increasing robot group sizes on the human psychophysiological state in the context of human–swarm interaction," *Swarm Intelligence*, vol. 10, no. 3, pp. 193–210, 2016.

[343] M. Swangnetr and D. B. Kaber, "Emotional state classification in patient–robot interaction using wavelet analysis and statistics-based feature selection," *IEEE Transactions on Human-Machine Systems*, vol. 43, no. 1, pp. 63–75, 2013.

[344] L. Tiberio, A. Cesta, G. Cortellessa, L. Padua, and A. R. Pellegrino, "Assessing affective response of older users to a telepresence robot using a combination of psychophysiological measures," in *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2012, pp. 833–838.

[345] N. Goljar, M. Javh, J. Poje, J. Ocepek, D. Novak, J. Ziherl, A. Olenšek, M. Mihelj, and M. Munih, "Psychophysiological responses to robot training in different recovery phases after stroke," in *2011 IEEE International Conference on Rehabilitation Robotics*. IEEE, 2011, pp. 1–6.

[346] C.-H. Ting, M. Mahfouf, A. Nassef, D. A. Linkens, G. Panoutsos, P. Nickel, A. C. Roberts, and G. R. J. Hockey, "Real-time adaptive automation system based on identification of operator functional state in simulated process control operations," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 40, no. 2, pp. 251–262, 2010.

[347] M. Munih and M. Mihelj, "Psychophysiological responses during robot-assisted rehabilitation," in *Rehabilitation Robotics*. Elsevier, 2018, pp. 305–317.

[348] L. Zhang, J. Wade, D. Bian, J. Fan, A. Swanson, A. Weitlauf, Z. Warren, and N. Sarkar, *Multimodal Fusion for Cognitive Load Measurement in an Adaptive Virtual Reality Driving Task for Autism Intervention*. Cham: Springer International Publishing, 2015, pp. 709–720.

[349] L. Yao, Y. Liu, W. Li, L. Zhou, Y. Ge, J. Chai, and X. Sun, "Using physiological measures to evaluate user experience of mobile applications," in *International Conference on Engineering Psychology and Cognitive Ergonomics*. Springer, 2014, pp. 301–310.

[350] N. Kumar and J. Kumar, "Measurement of cognitive load in hci systems using EEG power spectrum: an experimental study," *Procedia Computer Science*, vol. 84, pp. 70–78, 2016.

[351] S. Puma, N. Matton, P.-V. Paubel, ÃĽric Raufaste, and R. El-Yagoubi, "Using theta and alpha band power to assess cognitive workload in multitasking environments," *International Journal of Psychophysiology*, vol. 123, pp. 111 – 120, 2018.

[352] G. M. Wilson, "Psychophysiological indicators of the impact of media quality on users," in *CHI'01 extended abstracts on Human Factors in Computing Systems*. ACM, 2001, pp. 95–96.

[353] R. D. Ward and P. H. Marsden, "Physiological responses to different web page designs," *International Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 199–212, 2003.

[354] A. Bruun, "It's not complicated: A study of non-specialists analyzing gsr sensor data to detect ux related events," in *NordiCHI 2018Nordic Conference on Human-Computer Interaction*, 2018, pp. 170–183.

[355] T. Lin, M. Omata, W. Hu, and A. Imamiya, "Do physiological data relate to traditional usability indexes?" in *Proceedings of the 17th Australia conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future*. Computer-Human Interaction Special Interest Group (CHISIG) of Australia, 2005, pp. 1–10.

[356] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks Jr, "Physiological measures of presence in stressful virtual environments," in *ACM Transactions on Graphics (TOG)*, vol. 21, no. 3. ACM, 2002, pp. 645–652.

[357] I. Gerě and N. Jaušcvec, "Multimedia: Differences in cognitive processes observed with eeg," *Educational Technology Research and Development*, vol. 47, no. 3, pp. 5–14, 1999.

[358] N. S. Al Madi and J. I. Khan, "Measuring learning performance and cognitive activity during multimodal comprehension," in *Information and Communication Systems (ICICS), 2016 7th International Conference on*. IEEE, 2016, pp. 50–55.

[359] D. Novak, M. Mihelj, and M. Munih, "Dual-task performance in multimodal human-computer interaction: a psychophysiological perspective," *Multimedia Tools and Applications*, vol. 56, no. 3, pp. 553–567, 2012.

[360] M. Grassmann, E. Vlemincx, A. von Leupoldt, J. M. Mittelstädt, and O. Van den Bergh, "Respiratory changes in response to cognitive load: A systematic review," *Neural Plasticity*, vol. 2016, 2016.

[361] S. Lee, A. Matteson, D. Hooshyar, S. Kim, J. Jung, G. Nam, and H. Lim, "Comparing programming language comprehension between novice and expert programmers using EEG analysis," in *Bioinformatics and Bioengineering (BIBE), 2016 IEEE 16th International Conference on*. IEEE, 2016, pp. 350–355.

[362] E. Ferreira, D. Ferreira, S. Kim, P. Siirtola, J. Röning, J. F. Forlizzi, and A. K. Dey, "Assessing real-time cognitive load based on psycho-physiological measures for younger and older adults," in *2014 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*. IEEE, 2014, pp. 39–48.

[363] Y. Liu, J. M. Ritchie, T. Lim, Z. Kosmadoudi, A. Sivanathan, and R. C. Sung, "A fuzzy psycho-physiological approach to enable the understanding of an engineer's affect status during cad activities," *Computer-Aided Design*, vol. 54, pp. 19–38, 2014.

[364] T. A. Nguyen and Y. Zeng, "A physiological study of relationship between designer's mental effort and mental stress during conceptual design," *Computer-Aided Design*, vol. 54, pp. 3–18, 2014.

[365] T. A. Nguyen and Z. Yong, “Effects of stress and effort on self-rated reports in experimental study of design activities,” *Journal of Intelligent Manufacturing*, vol. 28, no. 7, pp. 1609–1622, 2017.

[366] P. Sedgwick, “Pearson’s correlation coefficient,” *British Medical Journal*, vol. 345, p. e4483, 2012.

[367] E. Pereda, R. Q. Quiroga, and J. Bhattacharya, “Nonlinear multivariate analysis of neurophysiological signals,” *Progress in Neurobiology*, vol. 77, no. 1-2, pp. 1–37, 2005.

[368] C. Lazar, J. Taminau, S. Meganck, D. Steenhoff, A. Coletta, C. Molter, V. de Schaetzen, R. Duque, H. Bersini, and A. Nowe, “A survey on filter techniques for feature selection in gene expression microarray analysis,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, vol. 9, no. 4, pp. 1106–1119, 2012.

[369] K. Torkkola, “Feature extraction by non-parametric mutual information maximization,” *Journal of Machine Learning Research*, vol. 3, no. Mar, pp. 1415–1438, 2003.

[370] F. Fleuret, “Fast binary feature selection with conditional mutual information,” *Journal of Machine Learning Research*, vol. 5, no. Nov, pp. 1531–1555, 2004.

[371] A. Kraskov, H. Stögbauer, and P. Grassberger, “Estimating mutual information,” *Physical Review E*, vol. 69, no. 6, p. 066138, 2004.

[372] G. Niso, R. Bruña, E. Pereda, R. Gutiérrez, R. Bajo, F. Maestú, and F. del Pozo, “Hermes: towards an integrated toolbox to characterize functional and effective brain connectivity,” *Neuroinformatics*, vol. 11, no. 4, pp. 405–434, 2013.

[373] R. Vijayalakshmi, D. Nandagopal, N. Dasari, B. Cocks, N. Dahal, and M. Thilaga, “Minimum connected component–a novel approach to detection of cognitive load induced changes in functional brain networks,” *Neurocomputing*, vol. 170, pp. 15–31, 2015.

[374] A. Roebroeck, E. Formisano, and R. Goebel, “Mapping directed influence over the brain using granger causality and fmri,” *Neuroimage*, vol. 25, no. 1, pp. 230–242, 2005.

[375] F. De Vico Fallani, L. Astolfi, F. Cincotti, D. Mattia, M. G. Marciani, A. Tocci, S. Salinari, H. Witte, W. Hesse, S. Gao *et al.*, “Cortical network dynamics during foot movements,” *Neuroinformatics*, vol. 6, no. 1, pp. 23–34, 2008.

[376] W. Hesse, E. Möller, M. Arnold, and B. Schack, “The use of time-variant EEG granger causality for inspecting directed interdependencies of neural assemblies,” *Journal of Neuroscience Methods*, vol. 124, no. 1, pp. 27–44, 2003.

[377] D. Marinazzo, M. Pellicoro, and S. Stramaglia, “Kernel-granger causality and the analysis of dynamical networks,” *Physical Review E*, vol. 77, no. 5, p. 056215, 2008.

[378] C. Diks and V. Panchenko, “A new statistic and practical guidelines for nonparametric granger causality testing,” *Journal of Economic Dynamics and Control*, vol. 30, no. 9-10, pp. 1647–1669, 2006.

[379] N. Ancona, D. Marinazzo, and S. Stramaglia, “Radial basis function approach to nonlinear granger causality of time series,” *Physical Review E*, vol. 70, no. 5, p. 056221, 2004.

[380] L. A. Baccalá and K. Sameshima, “Partial directed coherence: a new concept in neural structure determination,” *Biological Cybernetics*, vol. 84, no. 6, pp. 463–474, 2001.

[381] F. He, S. A. Billings, H.-L. Wei, and P. G. Sarrigiannis, “A nonlinear causality measure in the frequency domain: Nonlinear partial directed coherence with applications to eeg,” *Journal of Neuroscience Methods*, vol. 225, pp. 71–80, 2014.

[382] G. G. Supp, A. Schlögl, N. Trujillo-Barreto, M. M. Müller, and T. Gruber, "Directed cortical information flow during human object recognition: Analyzing induced EEG gamma-band responses in brain's source space," *PloS One*, vol. 2, no. 8, pp. 1–11, 08 2007.

[383] J. R. Sato, D. Y. Takahashi, S. M. Arcuri, K. Sameshima, P. A. Morettin, and L. A. Baccalá, "Frequency domain connectivity identification: an application of partial directed coherence in fmri," *Human Brain Mapping*, vol. 30, no. 2, pp. 452–461, 2009.

[384] C. E. Shannon, "Prediction and entropy of printed english," *Bell System Technical Journal*, vol. 30, no. 1, pp. 50–64, 1951.

[385] B. GourÃľvitch and J. J. Eggermont, "Evaluating information transfer between auditory cortical neurons," *Journal of Neurophysiology*, vol. 97, no. 3, pp. 2533–2543, 2007.

[386] S. A. Neymotin, K. M. Jacobs, A. A. Fenton, and W. W. Lytton, "Synaptic information transfer in computer models of neocortical columns," *Journal of Computational Neuroscience*, vol. 30, no. 1, pp. 69–84, Feb 2011.

[387] M. H. I. Shovon, D. N. Nandagopal, B. Cocks, and R. Vijayalakshmi, "Capturing cognition via eeg-based functional brain networks," in *Emerging Trends in Neuro Engineering and Neural Computation*. Springer, 2017, pp. 147–172.

[388] R. Vicente, M. Wibral, M. Lindner, and G. Pipa, "Transfer entropy—a model-free measure of effective connectivity for the neurosciences," *Journal of Computational Neuroscience*, vol. 30, no. 1, pp. 45–67, Feb 2011.

[389] L. Barnett, A. B. Barrett, and A. K. Seth, "Granger causality and transfer entropy are equivalent for gaussian variables," *Physical Review Letters*, vol. 103, no. 23, p. 238701, 2009.

[390] B. Gourévitch and J. J. Eggermont, "Evaluating information transfer between auditory cortical neurons," *Journal of Neurophysiology*, vol. 97, no. 3, pp. 2533–2543, 2007.

[391] M. Chávez, J. Martinerie, and M. Le Van Quyen, "Statistical assessment of nonlinear causality: application to epileptic EEG signals," *Journal of Neuroscience Methods*, vol. 124, no. 2, pp. 113–128, 2003.

[392] S. Sabesan, K. Narayanan, A. Prasad, L. Iasemidis, A. Spanias, and K. Tsakalis, "Information flow in coupled nonlinear systems: Application to the epileptic human brain," in *Data Mining in Biomedicine*. Springer, 2007, pp. 483–503.

[393] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," *Physics Reports*, vol. 424, no. 4-5, pp. 175–308, 2006.

[394] O. Sporns, D. R. Chialvo, M. Kaiser, and C. C. Hilgetag, "Organization, development and function of complex brain networks," *Trends in Cognitive Sciences*, vol. 8, no. 9, pp. 418–425, 2004.

[395] M. H. I. Shovon, D. N. Nandagopal, R. Vijayalakshmi, J. T. Du, and B. Cocks, "Transfer entropy and information flow patterns in functional brain networks during cognitive activity," in *International Conference on Neural Information Processing*. Springer, 2014, pp. 1–10.

[396] C. J. Stam and E. A. De Bruin, "Scale-free dynamics of global functional connectivity in the human brain," *Human Brain Mapping*, vol. 22, no. 2, pp. 97–109, 2004.

[397] B. C. Van Wijk, C. J. Stam, and A. Daffertshofer, "Comparing brain networks of different size and connectivity density using graph theory," *PloS One*, vol. 5, no. 10, p. e13701, 2010.

[398] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: simple building blocks of complex networks," *Science*, vol. 298, no. 5594, pp. 824–827, 2002.

[399] M. D. McDonnell, Ö. N. Yaveroğlu, B. A. Schmerl, N. Iannella, and L. M. Ward, "Motif-role-fingerprints: the building-blocks of motifs, clustering-coefficients and transitivities in directed networks," *PloS One*, vol. 9, no. 12, p. e114503, 2014.

[400] E. A. Leicht and M. E. Newman, "Community structure in directed networks," *Physical Review Letters*, vol. 100, no. 11, p. 118703, 2008.

[401] C. Yan and Y. He, "Driving and driven architectures of directed small-world human brain functional networks," *PloS One*, vol. 6, no. 8, p. e23460, 2011.

[402] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world'networks," *Nature*, vol. 393, no. 6684, p. 440, 1998.

[403] V. Latora and M. Marchiori, "Economic small-world behavior in weighted networks," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 32, no. 2, pp. 249–263, 2003.

[404] W. Liao, J. Ding, D. Marinazzo, Q. Xu, Z. Wang, C. Yuan, Z. Zhang, G. Lu, and H. Chen, "Small-world directed networks in the human brain: multivariate granger causality analysis of resting-state fmri," *Neuroimage*, vol. 54, no. 4, pp. 2683–2694, 2011.

[405] M. D. Humphries and K. Gurney, "Network 'small-world-ness': a quantitative method for determining canonical network equivalence," *PloS One*, vol. 3, no. 4, p. e0002051, 2008.

[406] M. H. I. Shovon, D. N. Nandagopal, J. T. Du, R. Vijayalakshmi, and B. Cocks, "Cognitive activity during web search," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2015, pp. 967–970.

[407] R. Salvador, J. Suckling, M. R. Coleman, J. D. Pickard, D. Menon, and E. Bullmore, "Neurophysiological architecture of functional magnetic resonance images of human brain," *Cerebral Cortex*, vol. 15, no. 9, pp. 1332–1342, 2005.

[408] S. Micheloyannis, E. Pachou, C. J. Stam, M. Vourkas, S. Erimaki, and V. Tsirka, "Using graph theoretical analysis of multi channel EEG to evaluate the neural efficiency hypothesis," *Neuroscience Letters*, vol. 402, no. 3, pp. 273–277, 2006.

[409] N. Langer, A. Pedroni, L. R. Gianotti, J. Hänggi, D. Knoch, and L. Jäncke, "Functional brain network efficiency predicts intelligence," *Human Brain Mapping*, vol. 33, no. 6, pp. 1393–1406, 2012.

[410] M. Rubinov, S. A. Knock, C. J. Stam, S. Micheloyannis, A. W. Harris, L. M. Williams, and M. Breakspear, "Small-world properties of nonlinear brain activity in schizophrenia," *Human Brain Mapping*, vol. 30, no. 2, pp. 403–416, 2009.

[411] M. Jalili and M. G. Knyazeva, "Eeg-based functional networks in schizophrenia," *Computers in Biology and Medicine*, vol. 41, no. 12, pp. 1178–1186, 2011.

[412] W. de Haan, Y. A. Pijnenburg, R. L. Strijers, Y. van der Made, W. M. van der Flier, P. Scheltens, and C. J. Stam, "Functional neural network analysis in frontotemporal dementia and alzheimer's disease using EEG and graph theory," *BMC Neuroscience*, vol. 10, no. 1, p. 101, 2009.

[413] M. Ahmadlou, H. Adeli, and A. Adeli, "Graph theoretical analysis of organization of functional brain networks in adhd," *Clinical EEG and Neuroscience*, vol. 43, no. 1, pp. 5–13, 2012.

[414] S. J. Leistedt, N. Coumans, M. Dumont, J.-P. Lanquart, C. J. Stam, and P. Linkowski, "Altered sleep brain functional connectivity in acutely depressed patients," *Human Brain Mapping*, vol. 30, no. 7, pp. 2207–2219, 2009.

[415] J. Wu, J. Zhang, C. Liu, D. Liu, X. Ding, and C. Zhou, "Graph theoretical analysis of EEG functional connectivity during music perception," *Brain Research*, vol. 1483, pp. 71–81, 2012.

[416] F. Weichert, D. Bachmann, B. Rudak, and D. Fisseler, "Analysis of the accuracy and robustness of the leap motion controller," *Sensors*, vol. 13, no. 5, pp. 6380–6393, 2013.

[417] G. Marin, F. Dominio, and P. Zanuttigh, "Hand gesture recognition with leap motion and kinect devices," in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 1565–1569.

[418] AutoCAD, "ObjectARX developer's guide," *AutoDesk Inc*, 2002.

[419] C. Dussault, J.-C. Jouanin, M. Philippe, and C.-Y. Guezennec, "EEG and ecg changes during simulator operation reflect mental workload and vigilance," *Aviation, Space, and Environmental Medicine*, vol. 76, no. 4, pp. 344–351, 2005.

[420] M. H. I. Shovon, N. Nandagopal, R. Vijayalakshmi, J. T. Du, and B. Cocks, "Directed connectivity analysis of functional brain networks during cognitive activity using transfer entropy," *Neural Processing Letters*, vol. 45, no. 3, pp. 807–824, 2017.

[421] S. F. Witelson, "The brain connection: the corpus callosum is larger in left-handers," *Science*, vol. 229, no. 4714, pp. 665–668, 1985.

[422] I. Crk and T. Kluthe, "Toward using alpha and theta brain waves to quantify programmer expertise," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. IEEE, 2014, pp. 5373–5376.

[423] U. Herwig, P. Satrapi, and C. Schönfeldt-Lecuona, "Using the international 10-20 EEG system for positioning of transcranial magnetic stimulation," *Brain Topography*, vol. 16, no. 2, pp. 95–99, 2003.

[424] T. Manual, "Testbench user manual – crossroads academy," https://www.crossroadsacademy.org/crossroads/wp-content/uploads/2016/05/Test-Bench-Manual-.pdf/, [Online; accessed 06-Jun-2019].

[425] MathWorks, "Matlab," https://au.mathworks.com/products/matlab.html/, [Online; accessed 06-Jun-2019].

[426] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, 2004.

[427] A. Delorme, T. Mullen, C. Kothe, Z. A. Acar, N. Bigdely-Shamlo, A. Vankov, and S. Makeig, "Eeglab, sift, nft, bcilab, and erica: new tools for advanced EEG processing," *Computational Intelligence and Neuroscience*, vol. 2011, p. 10, 2011.

[428] Y. Li, Z. Ma, W. Lu, and Y. Li, "Automatic removal of the eye blink artifact from EEG using an ica-based template matching approach," *Physiological Measurement*, vol. 27, no. 4, p. 425, 2006.

[429] A. Kandaswamy, V. Krishnaveni, S. Jayaraman, N. Malmurugan, and K. Ramadoss, "Removal of ocular artifacts from eeg-a survey," *IETE Journal of Research*, vol. 51, no. 2, pp. 121–130, 2005.

[430] B. Zhou, X. Wu, J. Ruan, L. Zhao, and L. Zhang, "How many channels are suitable for independent component analysis in motor imagery brain-computer interface," *Biomedical Signal Processing and Control*, vol. 50, pp. 103–120, 2019.

[431] U. Gal, K. Lyytinen, and Y. Yoo, “The dynamics of it boundary objects, information infrastructures, and organisational identities: the introduction of 3D modelling technologies into the architecture, engineering, and construction industry,” *European Journal of Information Systems*, vol. 17, no. 3, pp. 290–304, 2008.

[432] T. A. Nguyen and Y. Zeng, “Analysis of design activities using EEG signals,” in *ASME 2010 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers, 2010, pp. 277–286.

[433] D. Kahneman, *Attention and effort*. Prentice-Hall Englewood Cliffs, NJ, 1973, vol. 1063.

[434] Y. Tang and Y. Zeng, “Quantifying designer’s mental stress in the conceptual design process using kinesics study,” in *DS 58-9: Proceedings of ICED 09, the 17th International Conference on Engineering Design, Vol. 9, Human Behavior in Design, Palo Alto, CA, USA, 24.-27.08. 2009*, 2009.

[435] I. Chiu and L. Shu, “Potential limitations of verbal protocols in design experiments,” in *ASME 2010 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers, 2010, pp. 287–296.

[436] H. Kuusela and P. Pallab, “A comparison of concurrent and retrospective verbal protocol analysis,” *The American Journal of Psychology*, vol. 113, no. 3, p. 387, 2000.

[437] M. Kavakli, M. Suwa, J. Gero, and T. Purcell, “Sketching interpretation in novice and expert designers,” in *Visual and Spatial Reasoning in Design*, vol. 99. Citeseer, 1999, pp. 209–219.

[438] M. Kavakli and A. Boyali, “De-sign: robust gesture recognition in conceptual design, sensor analysis and synthesis,” in *Advances in Robotics and Virtual Reality*. Springer, 2012, pp. 201–225.

[439] D. G. Ullman, T. G. Dietterich, and L. A. Stauffer, “A model of the mechanical design process based on empirical data,” *Ai Edam*, vol. 2, no. 1, pp. 33–52, 1988.

[440] M. B. Waldron and K. J. Waldron, “A time sequence study of a complex mechanical system design,” *Design Studies*, vol. 9, no. 2, pp. 95–106, 1988.

[441] M. Suwa, T. Purcell, and J. Gero, “Macroscopic analysis of design processes based on a scheme for coding designers’ cognitive actions,” *Design Studies*, vol. 19, no. 4, pp. 455–483, 1998.

[442] M. Kavakli and J. S. Gero, “Sketching as mental imagery processing,” *Design Studies*, vol. 22, no. 4, pp. 347–364, 2001.

[443] M. Kavakli and J. S. Gero, “The structure of concurrent cognitive actions: a case study on novice and expert designers,” *Design Studies*, vol. 23, no. 1, pp. 25–40, 2002.

[444] C. J. Atman, M. E. Cardella, J. Turns, and R. Adams, “Comparing freshman and senior engineering design processes: an in-depth follow-up study,” *Design Studies*, vol. 26, no. 4, pp. 325–357, 2005.

[445] M. Suwa and B. Tversky, “What do architects and students perceive in their design sketches? a protocol analysis,” *Design Studies*, vol. 18, no. 4, pp. 385–403, 1997.

[446] A. Gevins, M. E. Smith, H. Leong, L. McEvoy, S. Whitfield, R. Du, and G. Rush, “Monitoring working memory load during computer-based tasks with EEG pattern recognition methods,” *Human Factors*, vol. 40, no. 1, pp. 79–91, 1998.

[447] A. Al-Qahtani, A. Nasir, M. Z. Shakir, and K. A. Qaraqe, "Cognitive impairments in human brain due to wireless signals and systems: An experimental study using EEG signal analysis," in *e-Health Networking, Applications & Services (Healthcom), 2013 IEEE 15th International Conference on*. IEEE, 2013, pp. 1–3.

[448] M. Z. Baig and M. Kavakli, "Expertise classification using functional brain networks and normalized transfer entropy of EEG in design applications," in *Proceedings of the 2019 11th International Conference on Computer and Automation Engineering*. ACM, 2019, pp. 41–46.

[449] C. H. Seng, R. Demirli, L. Khuon, and D. Bolger, "Seizure detection in EEG signals using support vector machines," in *Bioengineering Conference (NEBEC), 2012 38th Annual Northeast*. IEEE, 2012, pp. 231–232.

[450] A. Bashashati, M. Fatourechi, R. K. Ward, and G. E. Birch, "A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals," *Journal of Neural Engineering*, vol. 4, no. 2, p. R32, 2007.

[451] P. Shenoy and R. P. Rao, "Dynamic bayesian networks for brain-computer interfaces," in *Advances in Neural Information Processing Systems*, 2005, pp. 1265–1272.

[452] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for eeg-based brain–computer interfaces," *Journal of Neural Engineering*, vol. 4, no. 2, p. R1, 2007.

[453] K. Polat and S. Güneş, "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast fourier transform," *Applied Mathematics and Computation*, vol. 187, no. 2, pp. 1017–1026, 2007.

[454] B. Schack, N. Vath, H. Petsche, H.-G. Geissler, and E. Möller, "Phase-coupling of theta–gamma EEG rhythms during short-term memory processing," *International Journal of Psychophysiology*, vol. 44, no. 2, pp. 143–163, 2002.

[455] J. Decety and J. Grèzes, "Neural mechanisms subserving the perception of human actions," *Trends in Cognitive Sciences*, vol. 3, no. 5, pp. 172–178, 1999.

[456] J. Duncan and A. M. Owen, "Common regions of the human frontal lobe recruited by diverse cognitive demands," *Trends in Neurosciences*, vol. 23, no. 10, pp. 475–483, 2000.

[457] S. K. Card, A. Newell, and T. P. Moran, *The Psychology of Human-Computer Interaction*. Hillsdale, NJ, USA: L. Erlbaum Associates Inc., 1983.

[458] J. N. Saby and P. J. Marshall, "The utility of EEG band power analysis in the study of infancy and early childhood," *Developmental Neuropsychology*, vol. 37, no. 3, pp. 253–273, 2012.

[459] Y. Lin, Z. Liu, and X. Gao, "Alpha-band oscillation during speech recognition under different sensory conditions," in *Biomedical Engineering and Informatics (BMEI), 2015 8th International Conference on*. IEEE, 2015, pp. 153–157.

[460] M. Gola, M. Magnuski, I. Szumska, and A. Wróbel, "EEG beta band activity is related to attention and attentional deficits in the visual performance of elderly subjects," *International Journal of Psychophysiology*, vol. 89, no. 3, pp. 334–341, 2013.

[461] W. H. Miltner, C. Braun, M. Arnold, H. Witte, and E. Taub, "Coherence of gamma-band EEG activity as a basis for associative learning," *Nature*, vol. 397, no. 6718, p. 434, 1999.

[462] S. H. Fairclough, L. Venables, and A. Tattersall, "The influence of task demand and learning on the psychophysiological response," *International Journal of Psychophysiology*, vol. 56, no. 2, pp. 171–184, 2005.

[463] O. Solomon Jr, "Psd computations using welch's method," *NASA STI/Recon Technical Report N*, vol. 92, 1991.

[464] P. C. Yuen, Y. Y. Tang, and P. S. P. Wang, Eds., *Multimodal Interface for Human-machine Communication*. River Edge, NJ, USA: World Scientific Publishing Co., Inc., 2002.

[465] H. P. A. Boshuizen, R. Bromme, and H. Gruber, Eds., *Professional learning: Gaps and transitions on the way from novice to expert*. Dordrecht: Kluwer, 2004.

[466] Looxids, "Looxid labs," http://looxidlabs.com/, [Online; accessed 06-Jun-2019].

[467] M. S. Cain, A. N. Landau, and A. P. Shimamura, "Action video game experience reduces the cost of switching tasks," *Attention, Perception, & Psychophysics*, vol. 74, no. 4, pp. 641–647, 2012.

[468] C.-H. Wu, Y.-M. Huang, and J.-P. Hwang, "Review of affective computing in education/learning: Trends and challenges," *British Journal of Educational Technology*, vol. 47, no. 6, pp. 1304–1323, 2016.

[469] Z. Kosmadoudi, T. Lim, J. Ritchie, S. Louchart, Y. Liu, and R. Sung, "Engineering design using game-enhanced cad: The potential to augment the user experience with game elements," *Computer-Aided Design*, vol. 45, no. 3, pp. 777–795, 2013.

[470] M. Hajinoroozi, Z. Mao, T.-P. Jung, C.-T. Lin, and Y. Huang, "EEG-based prediction of driver's cognitive performance by deep convolutional neural network," *Signal Processing: Image Communication*, vol. 47, pp. 549–555, 2016.

[471] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[472] J. Hu, "Comparison of different features and classifiers for driver fatigue detection based on a single EEG channel," *Computational and Mathematical Methods in Medicine*, vol. 2017, 2017.

[473] O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih, and U. R. Acharya, "Deep learning for healthcare applications based on physiological signals: A review," *Computer Methods and Programs in Biomedicine*, vol. 161, pp. 1–13, 2018.

[474] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for eeg-based brain–computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, p. 056013, 2018.

[475] L. Ma, J. W. Minett, T. Blu, and W. S. Wang, "Resting state eeg-based biometrics for individual identification using convolutional neural networks," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 2848–2851.

[476] Z. Tang, C. Li, and S. Sun, "Single-trial EEG classification of motor imagery using deep convolutional neural networks," *Optik-International Journal for Light and Electron Optics*, vol. 130, pp. 11–18, 2017.

[477] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Computers in Biology and Medicine*, vol. 100, pp. 270–278, 2018.

[478] M. Kallenberg, K. Petersen, M. Nielsen, A. Y. Ng, P. Diao, C. Igel, C. M. Vachon, K. Holland, R. R. Winkel, N. Karssemeijer *et al.*, "Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1322–1331, 2016.

[479] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in mri images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.

[480] N. Hatipoglu and G. Bilgin, "Cell segmentation in histopathological images with deep learning algorithms by utilizing spatial relationships," *Medical & Biological Engineering & Computing*, vol. 55, no. 10, pp. 1829–1848, 2017.

[481] M. Liang and X. Hu, "Recurrent convolutional neural network for object recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3367–3375.

[482] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, "A convolutional neural network cascade for face detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5325–5334.

[483] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2014, pp. 655–665.

[484] U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ecg segments with convolutional neural network," *Information Sciences*, vol. 405, pp. 81–90, 2017.

[485] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, "Application of deep convolutional neural network for automated detection of myocardial infarction using ecg signals," *Information Sciences*, vol. 415, pp. 190–198, 2017.

[486] M. Z. Baig, E. Javed, Y. Ayaz, W. Afzal, S. O. Gillani, M. Naveed, and M. Jamil, "Classification of left/right hand movement from EEG signal by intelligent algorithms," in *Computer Applications and Industrial Electronics (ISCAIE), 2014 IEEE Symposium on*. IEEE, 2014, pp. 163–168.

[487] Z. J. Koles, M. S. Lazar, and S. Z. Zhou, "Spatial patterns underlying population differences in the background eeg," *Brain Topography*, vol. 2, no. 4, pp. 275–284, 1990.

[488] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve bci designs: unified theory and new algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2011.

[489] Y. Bengio *et al.*, "Learning deep architectures for ai," *Foundations and Trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[490] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[491] H. Jahani and M. Kavakli, "Exploring a user-defined gesture vocabulary for descriptive mid-air interactions," *Cognition, Technology & Work*, vol. 20, no. 1, pp. 11–22, 2018.

[492] J. A. Blalock and T. E. Joiner, "Interaction of cognitive avoidance coping and stress in predicting depression/anxiety," *Cognitive Therapy and Research*, vol. 24, no. 1, pp. 47–65, Feb 2000.

[493] S. J. Dollinger and F. T. L. Leong, "Volunteer bias and the five-factor model," *The Journal of Psychology*, vol. 127, no. 1, pp. 29–36, 1993.

[494] C. A. Boneau, "The effects of violations of assumptions underlying the t test," *Psychological Bulletin*, vol. 57, no. 1, pp. 49–64, 1960.

[495] H. O. Posten, H. Cheng Yeh, and D. B. Owen, "Robustness of the two-sample t-test umdeh violations of the homogeneity of farjance assumptioi," *Communications in Statistics-Theory and Methods*, vol. 11, no. 2, pp. 109–126, 1982.

Appendix A

Ethics Approval

The result of Ethics approval application is attached in this section. The approval was received before starting the experiments. The information about the project, including the aims of the project, the research plan, and the methods were given to the Macquarie University Ethics Committee for approval. We also included a description of the projected number, sex, and age range of participants and provided a detailed description of what will be required of participants. All the experiments were conducted in accordance with the accepted ethical principles governing research involving humans at the VR Lab, Simulation Hub at Macquarie University.

Gmail

Muhammad Zeeshan Baig <baig.mzeeshan@gmail.com>

Fwd: Ethics application 5201700784 Final Approval

1 message

Manolya Kavakli-Thorne <manolya.kavakli@mq.edu.au> Thu, Aug 24, 2017 at 11:18 AM
To: Muhammad Baig <baig.mzeeshan@gmail.com>

Congratulations Zeeshan!
Cheers,
Manolya

Associate Professor Manolya Kavakli

Director of Postgraduate Coursework Program, E6A 372,
Department of Computing, Faculty of Science & Engineering,
Director of Virtual Reality Lab at theSimulation Hub, Y3A,
Macquarie University, Sydney NSW 2109, Australia
tel: (61 2) 9850 9572 fax: (61 2) 9850 9551 email: manolya.kavakli@mq.edu.auhttp://www.comp.mq.edu.au/~manolya/

Simulation Hub: http://research.mq.edu.au/research_facilities/simulation_hub/_nocache
VISOR (Virtual and Interactive Simulations of Reality) Research Group:http://web.science.mq.edu.au/groups/visor/

From: Faculty of Science Research Office <sci.ethics@mq.edu.au>
Sent: Thursday, August 24, 2017 11:03 am
Subject: Ethics application 5201700784 Final Approval
To: Manolya Kavakli-Thorne <manolya.kavakli@mq.edu.au>
Cc: fse.ethics <fse.ethics@mq.edu.au>, Katherine Shevelev <katherine.shevelev@mq.edu.au>, Cathi Humphrey-Hood <cathi.humphrey-hood@mq.edu.au>

Dear A/Prof Kavakli-Thorne

RE: Ethics project entitled: "Biofeedback in Design Expertise and Adaptive Multi-Modal System Design using EEG Signal Analysis and Motor Cognition"

Ref number: 5201700784

The Faculty of Science and Engineering Human Research Ethics Sub-Committee has reviewed your application and granted final approval, effective 24/08/2017. You may now commence your research.

This research meets the requirements of the National Statement on Ethical Conduct in Human Research (2007). The National Statement is available at the following web site:

http://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/e72.pdf.

The following personnel are authorised to conduct this research:

A/Prof Kavakli-Thorne
Mr Muhammad Zeeshan Baig
Mr John Porte

NB. STUDENTS: IT IS YOUR RESPONSIBILITY TO KEEP A COPY OF THIS APPROVAL EMAIL TO SUBMIT WITH YOUR THESIS.

Please note the following standard requirements of approval:

1. The approval of this project is conditional upon your continuing compliance with the National Statement on Ethical Conduct in Human Research

(2007).

2. Approval will be for a period of five (5) years subject to the provision of annual reports.

Progress Report 1 Due: 24/08/2018
Progress Report 2 Due: 24/08/2019
Progress Report 3 Due: 24/08/2020
Progress Report 4 Due: 24/08/2021
Final Report Due: 24/08/2022

NB. If you complete the work earlier than you had planned you must submit a Final Report as soon as the work is completed. If the project has been discontinued or not commenced for any reason, you are also required to submit a Final Report for the project.

Progress reports and Final Reports are available at the following website:

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/forms

3. If the project has run for more than five (5) years you cannot renew approval for the project. You will need to complete and submit a Final Report and submit a new application for the project. (The five year limit on renewal of approvals allows the Committee to fully re-review research in an environment where legislation, guidelines and requirements are continually changing, for example, new child protection and privacy laws).

4. All amendments to the project must be reviewed and approved by the Committee before implementation. Please complete and submit a Request for Amendment Form available at the following website:

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/forms

5. Please notify the Committee immediately in the event of any adverse effects on participants or of any unforeseen events that affect the continued ethical acceptability of the project.

6. At all times you are responsible for the ethical conduct of your research in accordance with the guidelines established by the University. This information is available at the following websites:
http://www.mq.edu.au/policy/

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/policy

If you will be applying for or have applied for internal or external funding for the above project it is your responsibility to provide the Macquarie University's Research Grants Management Assistant with a copy of this email as soon as possible. Internal and External funding agencies will not be informed that you have final approval for your project and funds will not be released until the Research Grants Management Assistant has received a copy of this email.

If you need to provide a hard copy letter of Final Approval to an external organisation as evidence that you have Final Approval, please do not hesitate to contact the Ethics Secretariat at the address below.

Please retain a copy of this email as this is your official notification of final ethics approval.

Yours sincerely,
Human Research Ethics Sub-Committee
Faculty of Science and Engineering

Macquarie University
NSW 2109

Appendix B

Information Consent Form

The information consent form that was used for recruiting the participants was given in this appendix. Participants were provided with the information sheet in the VR lab that contained the information about the experiment. If they were interested, then we asked them to sign the consent form. The considerate language was used, and there was no attempt to convince individuals to participate if they indicate an unwillingness to participate. The participants received no financial or other benefits as a result of participation.

Department of Computing
Faculty of Science and Engineering
MACQUARIE UNIVERSITY NSW 2109
Phone: +61 (0)2 9850-9105
Fax: +61 (0)2 9850 9551

MACQUARIE University SYDNEY·AUSTRALIA

Chief Investigator's / Supervisor's Name & Title: A/Prof Dr. Manolya Kavakli

Participant Information and Consent Form

Name of Project: **Analysis of EEG signals and Cognitive activity in 3D modeling for a Multimodal interaction system**

This research project will study gender differences in

- speech and hand gestures,
- cognitive processing, and
- brain activities.

This experiment will be recorded, either by a digital camera and/or by a microphone embedded in the camera.

The tools that are going to be used for this experiment are a digital camera, microphone, computer, Emotiv EEG Headset, Empatica wrist band; head mounted displays (HMDs) and semi-immersive cylindrical projection system.

These tools are safe and publicly available. The non-standard saline solution may cause allergy to sensitive skin. The likelihood is minimal in this study. However, you CANNOT participate if you are sensitive to the saline. A verbal warning will be given prior to commencing the study. The HMDs and VR screen may cause simulator sickness.

Please read the following points carefully:

- Should you decide to participate, you may quit anytime during the study but please remain in the VR lab and wait until the researcher has removed the Emotiv EEG Headset and Empatica wristband.
- Whenever you feel uncomfortable during the experiment, please immediately let the conductor know. Conductor will be in the VR lab during the entire session.
- Should you decide to stay during the study and experience severe discomfort, we will refer you to on-campus medical service.

The location of on-campus medical service is included in this information statement and consent form.

At the beginning of the experiment, you will be given a fifteen-minute tutorial on the purpose of the experiment and how to use the necessary applications, during which time you will be introduced to the system. Feel free to ask any questions you may have about the experiment or about the system. The total time commitment involved is estimated to be 30 minutes.

All material, including video recordings, will be kept strictly confidential and will not be made available to any persons outside this project. The researchers have no material interest in the outcome of this experiment. The results will be presented at departmental research seminars, peer-reviewed Australian and International conferences, and via peer-reviewed journal articles. **We will only use the images and speech in the video clips after the participants' identity is obscured in presentations and publications. The de-identified data would be retained for inclusion in related research by the investigators in the future**

Participation in this study is entirely voluntary. You are under no obligation to participate and may withdraw your consent to participate at any time without consequence to you. If you are interested in this study, A/Prof. Manolya Kavakli and Muhammad Zeeshan Baig will be happy to discuss it further with you and answer any queries you may have. Please feel free to contact on (02) 98509572.

Participants can obtain feedback regarding the results of the project from the Interactive Systems and Virtual Reality Research Group website located at http://web.science.mq.edu.au/groups/visor/

Thank You.

www.research.mq.edu.au/researchers/ethics/human_ethics/forms/

Medical Service on campus:

Suite 305, Level 3

Macquarie University Clinic Building (F10A)

2 Technology Place

Macquarie University NSW 2109 Tel: (02) 9812 3944 or (02) 9812 3096

For further queries about this study, please contact:

Dr. Manolya Kavakli (Chief Inv.)	02 9850 9572	manolya.kavakli@mq.edu.au
Muhammad Zeeshan Baig (PhD student)	02 9850 9530	muhammad.baig@students.mq.edu.au

I, _____________________________ have read (or, where appropriate, have had read to me) and understood the information given and any questions I have asked, have been answered to my satisfaction. I agree to participate in this research study, entitled **Biofeedback in Design Expertise and Adaptive Multi-Modal System Design using EEG Signal Analysis and Motor Cognition**, which is conducted by

Dr. Manolya Kavakli (A/Prof., Dept. of Computing, Macquarie University),

Muhammad Zeeshan Baig (Ph.D. student, Dept. of Computing, Macquarie University),

knowing that participation is entirely voluntary and I can withdraw from further participation in the research at any time without consequence.

I allow ☐ / do not allow ☐ the de-identified data to be retained for inclusion in related research by the investigators in the future.

I have been given a copy of this signed form to keep.

Participant's Name: ___________________________ (block letters)

Participant's Signature: ________________________ Date: ___________

Investigator's Name:_______________________ (block letters)

Investigator's Signature: ________________________ Date: ___________

The ethical aspects of this study have been approved by the Macquarie University Human Research Ethics Committee. If you have any complaints or reservations about any ethical aspect of your participation in this research, you may contact the Committee through the Director, Research Ethics (telephone (02) 9850 7854; email ethics@mq.edu.au). Any complaint you make will be treated in confidence and investigated, and you will be informed of the outcome.

INVESTIGATOR'S COPY

I, ___________________________ have read (or, where appropriate, have had read to me) and understood the information given and any questions I have asked, have been answered to my satisfaction. I agree to participate in this research study, entitled **Biofeedback in Design Expertise and Adaptive Multi-Modal System Design using EEG Signal Analysis and Motor Cognition**, which is conducted by

Dr. Manolya Kavakli (A/Pro., Dept. of Computing, Macquarie University),

Muhammad Zeeshan Baig (Ph.D. student, Dept. of Computing, Macquarie University),

knowing that participation is entirely voluntary and I can withdraw from further participation in the research at any time without consequence.

I allow ☐ / do not allow ☐ the de-identified data to be retained for inclusion in related research by the investigators in the future.

I have been given a copy of this signed form to keep.

Participant's Name:__________________________ (block letters)

Participant's Signature: _______________________ Date:__________

Investigator's Name:__________________________ (block letters)

Investigator's Signature: _______________________ Date: __________

The ethical aspects of this study have been approved by the Macquarie University Human Research Ethics Committee. If you have any complaints or reservations about any ethical aspect of your participation in this research, you may contact the Committee through the Director, Research Ethics (telephone (02) 9850 7854; email ethics@mq.edu.au). Any complaint you make will be treated in confidence and investigated, and you will be informed of the outcome.

PARTICIPANT'S COPY

Appendix C

Questionnaire

Name:

Age:

Gender:

Handedness:

Previous experience with AutoCAD:

The scale for score ranges from 1 (Very bad) to 7 (Excellent)

Question	Keyboard and Mouse	Gesture and Speech
Performance of the commands		
How easy was it to draw in AutoCAD?		
How well could you examine objects from multiple viewpoints?		
How well could you move or manipulate objects in the virtual environment?		
How much are you satisfied with the visual aspects of the object?		
How much delay did you experience between your instructions and expected outcomes?		
How responsive was the computer to the actions you initiated?		
Fatigue felt		
Effort: How hard was it to accomplish your level of performance?		
Did you feel any fatigue while drawing?		

User perception in interaction with computer		
How much did you feel in control when you were instructing the computer?		
How natural was the interaction with the computer?		
How natural was the interaction with the designed object?		
How aware were you of your display and control devices?		
How involved were you in the production of the 3D model?		
How distracting was your interaction with the computer?		
How proficient in interacting with the computer did you feel at the end of the experience?		
How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform the design?		
Were you involved in the experimental task to the extent that you lost track of time?		
Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you?		

Appendix D

Instruction for Drawing in AutoCAD

Drawing in AutoCAD

A- First experimentation: draw with keyboard and mouse

Global information

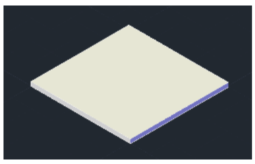
To move the camera:

- Wheel Click to pan the camera
- Maj + wheel click to move the camera
- Or choose the view with the orbit

Type undo or control Z to undo the last action.

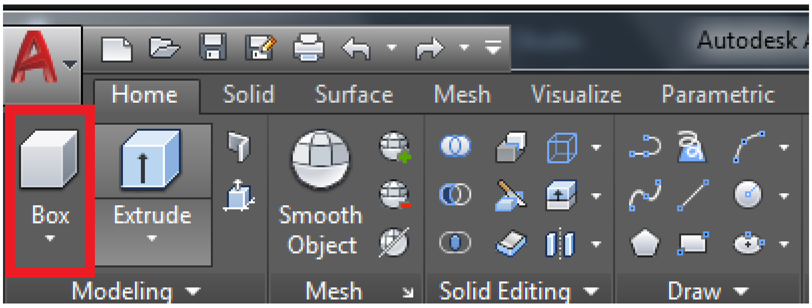
With mouse and keyboard

1- The base:



The Rectangular Base:

a. Choose 3DTool, click on box



b. Click for the first position

c. Click for the size (width and length and height) or you can type the number separated by ',' ex: type 3,3,42

2- The Pillar:

The vertical cylinder:

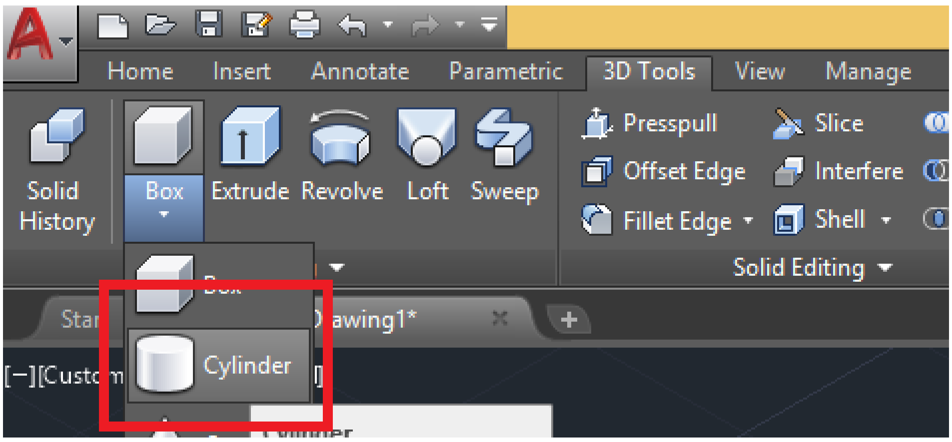
a. Choose the cylinder option from the box dropdown menu.

b. Click for the first position (middle of the base)

c. Click for the radius (approximately 1 inch) or enter it by keyboard

d. Click for the height (approximately 7 inch) or enter it by keyboard

3- The top:



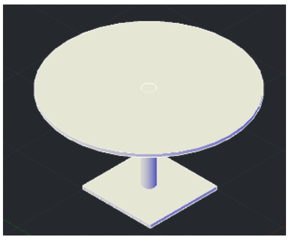
a. Choose 3DTool, click on Cylinder

b. Click for the first position (middle of the pillar)

c. Click for the radius (approximately 6 inch) or enter it by keyboard

d. Click for the height (approximately 1 inch) or enter it by keyboard

4- Change the colour:

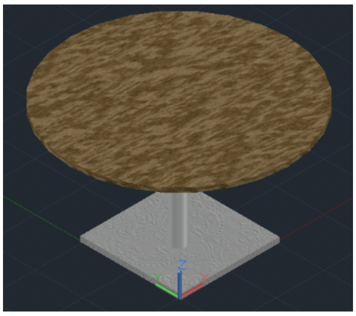


a. Type properties

b. Select the object

c. On the colour; you can change the colour

5- Change the materials:



a. Change the mode view: left click on Conceptual, then choose realistic

b. Type: Materials

c. Select the object

d. Click the material to apply

Task to perform:

Draw the 3D shape

Camera Manipulation

B- Second experiment: drawing with gesture and speech

You should use the same scenario used with mouse and keyboard. To use the gesture and speech, please read the following instruction.

For Gestures:

1- Left hand: manage the camera view

a. To activate the movement, close the hand and open it

b. To move you hand to control the camera

c. To rotate the camera, don't move your hand, just turn your wrist

d. To zoom in, perform a pinching gesture. For zoom out, perform pinch gesture with middle finger

2- Right hand: Control the cursor (mouse)

To left click: pointing gesture with index finger.

To right click: pointing gesture with thumb.

For Speech:

A. To choose the shape:

Say: ***draw a box or the shape is cylinder*:**

- For a cylinder: CIRCLE, CYLINDER, ROUND, TUBE

- For a Box: BOX, SQUARE, BARS, BAR, LAYER, RECTANGLE

- For a sphere: SPHERE

- For a cone: CONE

- For a wedge: WEDGE

- For a torus: TORUS, DONUTS,

- For a free design: POLYSOLID, FREE

B. To specify the position, say:

- The position is number, number, number

C. To specify the size (width and length), say:

- The size is number to number

- Number could be the 0 to 20, 39, 42

D. To specify the height, say:

- The height is number

E. To specify the radius, say:

- The radius is number

F. To specify the tube radius, say:

- The tube radius is number

G. To manage the camera, say:

- Camera

- Then choose between the following words the action:

- Move, displacement, orientation or if you want to change the orbit, 'Home'

- Then choose the direction between the follow words

- UP, DOWN, LEFT, RIGHT, FORWARD, BACK, BOTTOM, TOP, FRONT, SOUTH EAST, SOUTH WEST, NORTH EAST, NORTH WEST.

Script for speech :

1. Draw a box

a. The position is 0,0,0

b. The size is 6 by 6

c. The height is 1

2. Draw a cylinder

a. The position is 3,3,1

b. The radius is 1

c. The height is 5

3. Draw a cylinder

a. The position is 3,3,6

b. The radius is 6

c. The height is 1

Task to perform:

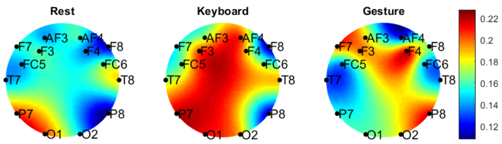
Draw the 3D shape using speech and gestures

Camera Manipulation

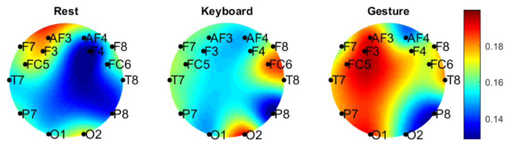
Appendix E

Additional Information

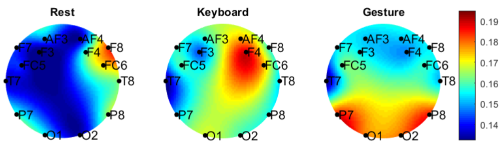
E.1 Chapter 8: Comparative Analysis of Cognitive Activity: using Power Spectral Density



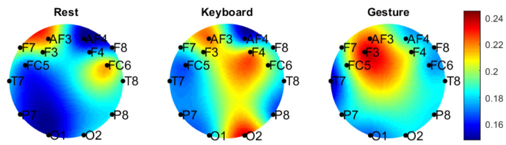
(a) Competent-User 1



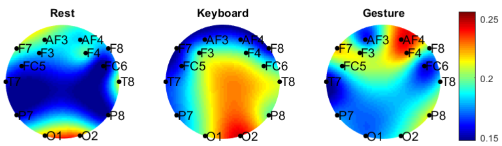
(b) Novice-User 1



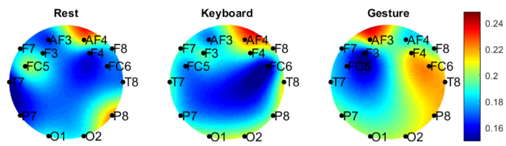
(c) Competent-User 2



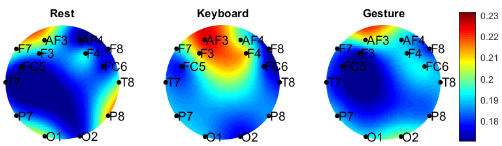
(d) Novice-User 2



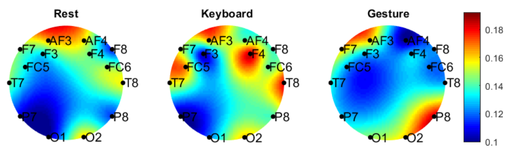
(e) Competent-User 3



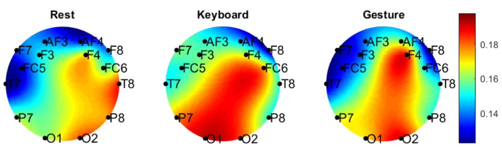
(f) Novice-User 3



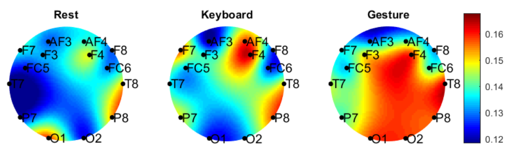
(g) Novice-User 4



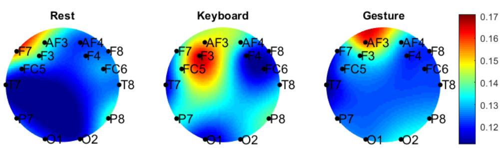
(h) Novice-User 5



(i) Novice-User 6

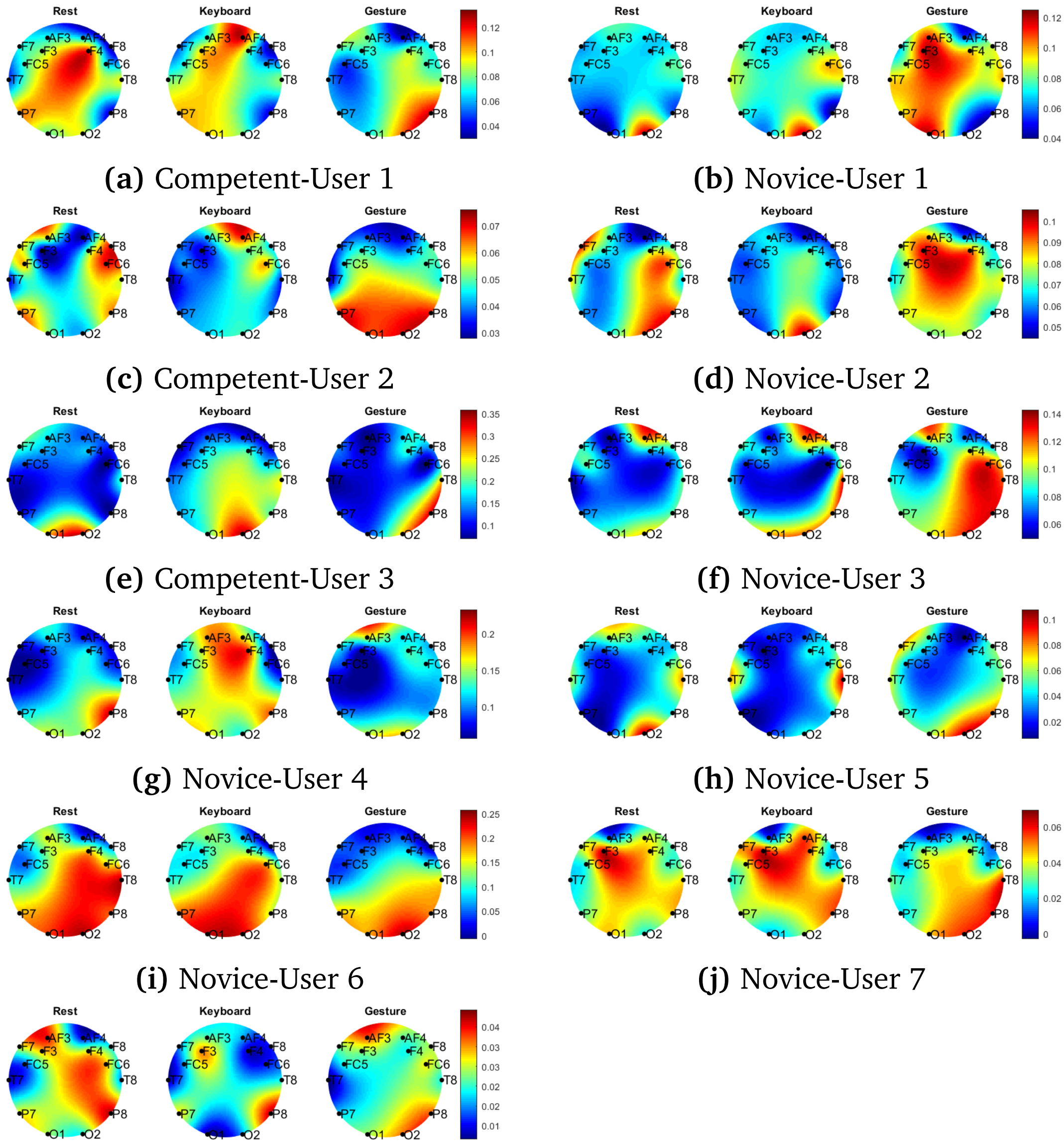


(j) Novice-User 7



(k) Novice-User 8

Figure E.1: Averaged theta activity at rest, keyboard and gesture states for three competent and three novice users



(a) Competent-User 1

(b) Novice-User 1

(c) Competent-User 2

(d) Novice-User 2

(e) Competent-User 3

(f) Novice-User 3

(g) Novice-User 4

(h) Novice-User 5

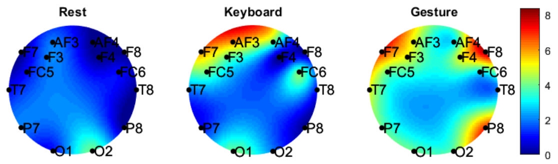
(i) Novice-User 6

(j) Novice-User 7

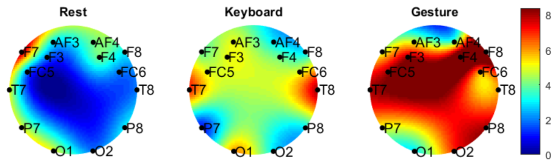
(k) Novice-User 8

Figure E.2: Averaged alpha activity at rest, keyboard and gesture states for three competent and three novice users

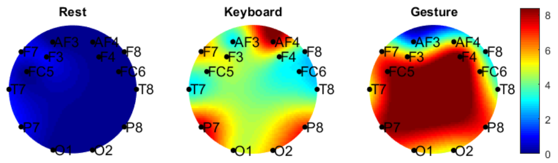
E.2 Chapter 9: Connectivity Analysis: using Transfer Entropy and Functional Brain Networks



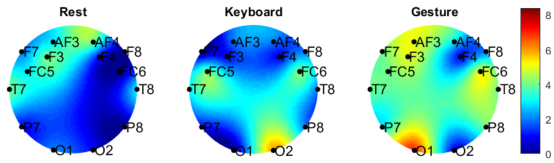
(a) Competent-User 1



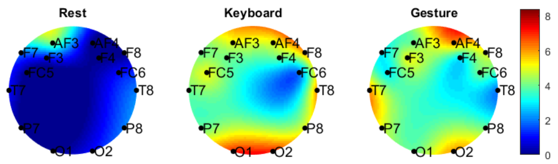
(b) Competent-User 2



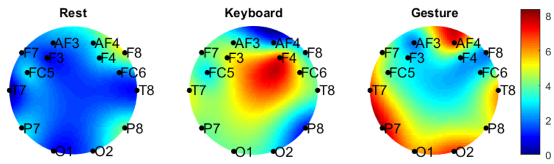
(c) Novice-User 1



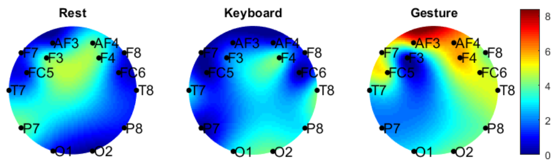
(d) Competent-User 3



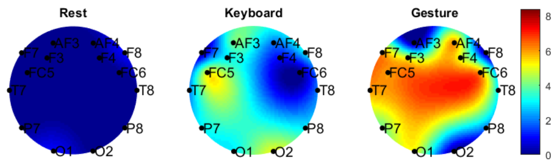
(e) Novice-User 2



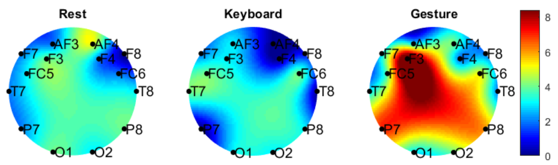
(f) Novice-User 3



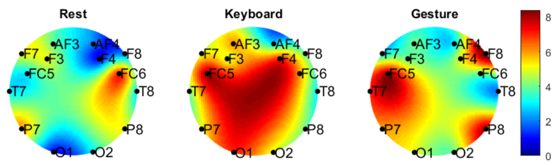
(g) Novice-User 4



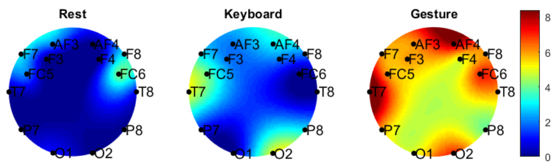
(h) Novice-User 5



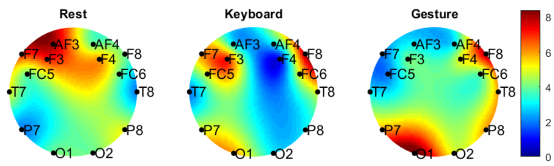
(i) Novice-User 6



(j) Competent-User 4



(k) Novice-User 7



(l) Novice-User 8

Figure E.3: Degree centrality topographical plot of all users during rest, keyboard and gesture drawing states