



Applied Multimodal Framework for Evaluating and Forecasting Team Performance

AUTHOR(S)

Rakesh Veerabhadrapa

PUBLICATION DATE

01-01-2022

HANDLE

[10779/DRO/DU:27366996.v1](#)

Downloaded from Deakin University's Figshare repository

Deakin University CRICOS Provider Code: 00113B

Applied Multimodal Framework for Evaluating and Forecasting Team Performance

by
Rakesh Veerabhadrapa

Submitted in fulfillment of the requirements for the degree of
Doctor of Philosophy

Deakin University
January, 2022



**DEAKIN UNIVERSITY
ACCESS TO THESIS - A**

I am the author of the thesis entitled

Applied Multimodal Framework for Evaluating and Forecasting Team Performance

submitted for the degree of

Doctor of Philosophy

This thesis may be made available for consultation, loan and limited copying in accordance with the Copyright Act 1968.

'I certify that I am the student named below and that the information provided in the form is correct'

Full Name: **Rakesh Veerabhadrapa**.....
(Please Print)

Signed: Signature Redacted by Library

Date: **15/01/2022**



**DEAKIN UNIVERSITY
CANDIDATE DECLARATION**

I certify the following about the thesis entitled (10 word maximum)

Applied Multimodal Framework for Evaluating and Forecasting Team Performance

submitted for the degree of **Doctor of Philosophy**

- a. I am the creator of all or part of the whole work(s) (including content and layout) and that where reference is made to the work of others, due acknowledgment is given.
- b. The work(s) are not in any way a violation or infringement of any copyright, trademark, patent, or other rights whatsoever of any person.
- c. That if the work(s) have been commissioned, sponsored or supported by any organisation, I have fulfilled all of the obligations required by such contract or agreement.
- d. That any material in the thesis which has been accepted for a degree or diploma by any university or institution is identified in the text.
- e. All research integrity requirements have been complied with.

'I certify that I am the student named below and that the information provided in the form is correct'

Full Name: Rakesh Veerabhadrapa

(Please Print)

Signed:

Signature Redacted by Library

Date: 18/01/2022

Deakin University CRICOS Provider Code 00113B

Dedication

Special thanks to my wife Sneha, my parents, Veerabhadrappa and Parimala, and my brother Deepak, without your support this achievement would never have been possible.

Acknowledgements

First and foremost, I would like to show my sincere respects to my supervision team, Associate Professor Asim Bhatti and Dr Imali Hettiarachchi. Asim, thank you for this PhD opportunity and all your mentoring, support and enthusiasm throughout my PhD journey. Your expertise, feedbacks, management and people skills, and your calmness to confront and overcome problems were invaluable in ensuring my PhD stayed on track. Imali, without your support this PhD was never achievable. I am grateful for your insightful feedback on my reports, constructively analysing my ideas and improving them, sharing the wisdom and the patience to let me explore at my own pace. I am indebted to both of you and value the professional relationship I developed with you as a team.

I also acknowledge and thank Dr Dawei Jia, Defence Science Technology (DST) for her invaluable feedback during the design and development of the experiment process. I would like to thank Mr Mohammed Yahya Al Gumai for his promptness, support and contributions during the data collection process.

I am grateful to all the members of the Institute for Intelligent Systems Research & Innovation (IISRI), Deakin University, for their kind support, time and efforts by taking part in my experiments. Special thanks to Professor Saeid Nahavandi, Professor Chee Peng Lim and Associate Professor Samer Hanoun for encouraging and appreciating my work, helping me settle down and providing me with professional working opportunities. I acknowledge the infrastructure and support by the Institute for Intelligent Systems Research & Innovation (IISRI), Deakin University.

I would like to thank the many academics and HDR students, past and present, from IISRI. Specifically, Rehan, Rokon, Masood, Mohammed Jalali and Duy. Whether it be small chats during coffee breaks or larger discussions over lunch, the opportunity to discuss my research interests and hear alternative perspectives was immensely beneficial to shaping my PhD project.

I would like to recognise my previous colleagues and well-wishers at the Middle East College, Department of Electronics and Communication Engineering who motivated me to pursue my PhD. In particular, a special thanks to Mr Satish Masthenahally Nachappa, Dr Chandrashekar Ramaiah, Dr Arun N.S. Shankarappa, Mr Vinod Kumar Ramankutty and Mr. Yahya Yaqoob Obaid Al-Belushi.

I would like to express my sincere appreciation for the efforts and time spent by peer-reviewers of my conference and journal articles. Your critical reflections and feedback were encouraging which challenged me to put in extra effort and helped me to improve my knowledge.

In this special achievement, I would like to remember my uncles Kenchappa, Somashekar and Narayan who passed away during my PhD journey and I could not be with them during their final moments.

I express my sincere gratitude to those who made sure my relocation to Australia was not a lonely one. Specifically, the special bond I share with Tommy, Lisa, Stephan and Kanjana, I am so grateful for creating a family environment and allowing me and Sneha to stay with you. I would like to thank Saavan for sharing the house rent-free, and without your reference, our settling in Australia would never be easier. Being an International student with long PhD study periods is not easy, and COVID lock-downs were the hardest. During this period, our countless and big thanks to Thejesh and Shruthi, you both helped us not to miss home. I am so thankful for your generosity and will treasure the memories of countless dinners, monopoly, holidays and day trips, and laughter shared over glasses of scotch and wine. Whether it is cooking, helping us move houses, dinners, drinks and fun, you people were always welcoming. I feel so fortunate to have such a large group of close friends that could always be counted on for a laugh.

I am deeply appreciative to my parents, Veerabhadrappa and Parimala, brother, Deepak, sister-in-law, Rashmi, parents-in-law, Ramesh and Nirmala, friends, Raaghu, Ajay, Deepu, Kiran, Sandeep, Santhosh and Harsha and all my family members for their unconditional love and support throughout my entire PhD education journey. I cherish our close relationships and take solace in the fact that I have your support no matter the situation. To my parents, I will forever be grateful for the unconditional sacrifices and time invested in preparing me for higher education that enabled me to pursue a PhD, as well as always encouraging me to be my best.

Lastly, I would like to thank my wife, Sneha, for her understanding and unwavering support during my PhD journey. Your enduring love and constant encouragement always made the darkest times during the PhD feel brighter. Having you by my side made this whole endeavour feel much more attainable.

Publications

I Published:

1. Veerabhadrapa, R., Ul Hassan, M., Zhang, J. and Bhatti, A., 2020. Compatibility evaluation of clustering algorithms for contemporary extracellular neural spike sorting. *Frontiers in systems neuroscience*, 14, p.34.
2. Veerabhadrapa, R., Hettiarachchi, I.T. and Bhatti, A., 2021, April. Using Recurrence Quantification Analysis to Quantify the Physiological Synchrony in Dyadic ECG Data. In *2021 IEEE International Systems Conference (SysCon)* (pp. 1-8). IEEE.
3. Veerabhadrapa, R., Hettiarachchi, I.T., Algumaei, M. and Bhatti, A., 2021, October. A Deep Convolutional Neural Network Model for Classification of Emotions from Electroencephalography Data. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 1953-1958). IEEE.
4. Using Eye-tracking To Investigate The Effect of Gaze Co-occurrence and Distribution on Collaborative Performance (Accepted to be published in IEEE-Systems Conference)
5. Gaze Convergence Based Collaborative Performance Prediction in a 3-Member Joint Activity Setting (Accepted to be published in IEEE-Systems Conference)
6. Veerabhadrapa, R., Hettiarachchi, I.T., Algumaei, M. and Bhatti, A., 2021, Tank Battle Task public repository https://github.com/ghost2307/TankBattleTask_v1_github .

II Under review:

1. Fundamental Requirements for Contemporary Performance Evaluation of Emergency-Response and Safety-Critical Teams (Under-review by IEEE-Transactions on Human-machine Systems)
2. A State-of-the-art Experimental Platform for Effective Analysis of Team Processes (Under-review by IEEE-Transactions on Systems-Man and Cybernetic: Systems)
5. Cognitive Load and Gaze Recurrence based Multimodal Index as Predictor of Collaborative Team Performance (Under-review by Computers in Human Behavior) .

III In preparation:

1. Markers of Team Cognitive Process using Cardiovascular and Gaze of Activity of Three-Member Teams
2. The Causal Effects of Team Performance

3. The Impact of Communication During Three-Member Collaborative Activity
4. Evaluating Physiological Synchrony with Performance Index for Three-Member Team Processes.

AUTHORSHIP STATEMENT

Researches should review the Deakin Authorship Procedure to determine the criteria of authorship

1. Details of publication and corresponding author

Title of Publication		Publication details
Compatibility Evaluation of Clustering Algorithms for Contemporary Extracellular Neural Spike Sorting		Frontiers in Systems Neuroscience
Name of corresponding author	School/Institute/Division if based at Deakin; Organisation and address if non-Deakin	Email or phone
Rakesh Veerabhadrapa	IISRI	rveerabhadrapa@deakin.edu.au

2. Inclusion of publication in a thesis

Is it intended to include this publication in a higher degree by research (HDR) thesis?	Yes	If Yes, please complete Section 3 If No, go straight to Section 4.
---	-----	---

3. HDR thesis author's declaration

Name of HDR thesis author if different from above. (If the same, write "as above")	School/Institute/Division if based at Deakin	Thesis title
As above	IISRI	Intelligent Multimodal Experimental Framework for Analysing, Classifying and Forecasting Team Performance
If there are multiple authors, give a full description of HDR thesis author's contribution to the publication (for example, how much did the student contribute to the conception of the project, the design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)		
I declare that the above is an accurate description of my contribution to this paper, and the contributions of other authors are as described below.		Signature and date of student
		14/05/21

Signature Redacted by Library

4. Description of all author contributions

Name	Contribution(s) (for example, conception of the project, design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)
Rakesh Veerabhadrapa	Contributed to the conception of the study, Preparation of data, , Identification of clustering algorithms, validation of results, first draft of the manuscript, contributed to the manuscript revision, read and approved the submitted version.
Masood UI Hassan	Contributed to the conception of the study, Preparation of data, Identification of clustering algorithms, contributed to the manuscript revision, read and approved the submitted version.
James Zhang	Contributed to the conception of the study, Identification of clustering algorithms, contributed to the manuscript revision, read and approved the submitted version.
Asim Bhatti	Contributed to the conception of the study, Identification of clustering algorithms, Supervised manuscript revision, read and approved the submitted version

5. Author Declarations

I agree to be named as one of the authors of this work, and confirm:

- i. that I have met the authorship criteria set out in the Deakin University Research Authorship Procedure,*
- ii. that there are no other authors according to these criteria,*
- iii. that the description in Section 4 of my contribution(s) to this publication is accurate,*
- iv. that the data on which these findings are based are stored as set out in Section 7 below.*

If this work is to form part of an HDR thesis as described in Sections 2 and 3, I further

- v. consent to the incorporation of the publication into the candidate's HDR thesis submitted to Deakin University and, if the higher degree is awarded, the subsequent publication of the thesis by the university (subject to relevant Copyright provisions).*

Name of author	Signature*	Date
Rakesh Veerabhadrapa		29-06-2020
Masood Ul Hassan		30-06-2020
James Zhang		30-06-2020
Asim Bhatti		30-06-2020

Signatures Redacted
by Library

6. Other contributor declarations

I agree to be named as a non-author contributor to this work.

Name and affiliation of contributor	Contribution	Signature* and date

* If an author or contributor is unavailable or otherwise unable to sign the statement of authorship, the Head of Academic Unit may sign on their behalf, noting the reason for their unavailability, provided there is no evidence to suggest that the person would object to being named as author

7. Data storage

The original data for this project are stored in the following locations. Please ensure that this complies with the relevant procedure. .

Data format	Storage Location	Date lodged	Name of custodian if other than the corresponding author
The data is stored in *.csv and *.mat format.	Deakin RDS	30-06-2020	Asim Bhatti

This form must be retained by the corresponding author. If the corresponding author is not a Deakin staff member or student, the Deakin research team must ensure that a copy of this form is kept at Deakin.

If the publication is to be included as part of an HDR thesis, a copy of this form must be included in the thesis with the publication.

AUTHORSHIP STATEMENT

Researches should review the Deakin Authorship Procedure to determine the criteria of authorship

1. Details of publication and corresponding author

Title of Publication		Publication details
Using Recurrence Quantification Analysis to Quantify the Physiological Synchrony in Dyadic ECG Data		
Name of corresponding author	School/Institute/Division if based at Deakin; Organisation and address if non-Deakin	Email or phone
Rakesh Veerabhadrapa	IISRI	rveerabhadrapa@deakin.edu.au

2. Inclusion of publication in a thesis

Is it intended to include this publication in a higher degree by research (HDR) thesis?	Yes	If Yes, please complete Section 3 If No, go straight to Section 4.
---	-----	---

3. HDR thesis author's declaration

Name of HDR thesis author if different from above. (If the same, write "as above")	School/Institute/Division if based at Deakin	Thesis title
As above	IISRI	A Unified Intelligent Framework for Multimodal Team Performance Analysis
If there are multiple authors, give a full description of HDR thesis author's contribution to the publication (for example, how much did the student contribute to the conception of the project, the design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)		
I declare that the above is an accurate description of my contribution to this paper, and the contributions of other authors are as described below.		Signature and date of student
		13/11/20

Signature Redacted by Library

4. Description of all author contributions

Name	Contribution(s) (for example, conception of the project, design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)
Rakesh Veerabhadrapa	Contributed to the conception of the study, data analysis, performed the statistical analyses and wrote the first draft of the manuscript, contributed to the manuscript revision, read and approved the submitted version.
Imali Hettiarachchi	Contributed to conception of the study and data analysis, manuscript revision, read and approved the submitted version
Asim Bhatti	Supervised manuscript revision, read and approved the submitted version

5. Author Declarations

I agree to be named as one of the authors of this work, and confirm:

- i. that I have met the authorship criteria set out in the Deakin University Research Authorship Procedure,*
- ii. that there are no other authors according to these criteria,*
- iii. that the description in Section 4 of my contribution(s) to this publication is accurate,*
- iv. that the data on which these findings are based are stored as set out in Section 7 below.*

If this work is to form part of an HDR thesis as described in Sections 2 and 3, I further

- v. consent to the incorporation of the publication into the candidate's HDR thesis submitted to Deakin University and, if the higher degree is awarded, the subsequent publication of the thesis by the university (subject to relevant Copyright provisions).*

Name of author	Signature*	Date
Rakesh Veerabhadrapa		13/11/20
Imali Hettiarachchi		13/11/20
Asim Bhatti		

Signatures Redacted
by Library

6. Other contributor declarations

I agree to be named as a non-author contributor to this work.

Name and affiliation of contributor	Contribution	Signature* and date

* If an author or contributor is unavailable or otherwise unable to sign the statement of authorship, the Head of Academic Unit may sign on their behalf, noting the reason for their unavailability, provided there is no evidence to suggest that the person would object to being named as author

7. Data storage

The original data for this project are stored in the following locations. Please ensure that this complies with the relevant procedure. .

Data format	Storage Location	Date lodged	Name of custodian if other than the corresponding author
The data was acquired in R-Data storage (.rds) format	Plos one public repository, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0159178	14/07/2016	Plos one

This form must be retained by the corresponding author. If the corresponding author is not a Deakin staff member or student, the Deakin research team must ensure that a copy of this form is kept at Deakin.

If the publication is to be included as part of an HDR thesis, a copy of this form must be included in the thesis with the publication.

AUTHORSHIP STATEMENT

Researches should review the Deakin Authorship Procedure to determine the criteria of authorship

1. Details of publication and corresponding author

Title of Publication		Publication details
A Deep Convolutional Neural Network Model for Classification of Emotions from Electroencephalography Data		IEEE conference on Systems Man and Cybernetics
Name of corresponding author	School/Institute/Division if based at Deakin; Organisation and address if non-Deakin	Email or phone
Rakesh Veerabhadrapa	IISRI	rveerabhadrapa@deakin.edu.au

2. Inclusion of publication in a thesis

Is it intended to include this publication in a higher degree by research (HDR) thesis?	Yes	If Yes, please complete Section 3 If No, go straight to Section 4.
---	-----	---

3. HDR thesis author's declaration

Name of HDR thesis author if different from above. (If the same, write "as above")	School/Institute/Division if based at Deakin	Thesis title
As above	IISRI	Intelligent Multimodal Experimental Framework for Analysing, Classifying and Forecasting Team Performance
If there are multiple authors, give a full description of HDR thesis author's contribution to the publication (for example, how much did the student contribute to the conception of the project, the design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)		
I declare that the above is an accurate description of my contribution to this paper, and the contributions of other authors are as described below.		Signature and date of student
		14/05/21

Signature Redacted by Library

4. Description of all author contributions

Name	Contribution(s) (for example, conception of the project, design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)
Rakesh Veerabhadrapa	Contributed to the conception of the study, data analysis, modelling of neural network architecture and wrote the first draft of the manuscript, contributed to the manuscript revision, read and approved the submitted version.
Imali Hettiarachchi	Contributed to conception of the study and data analysis, manuscript revision, read and approved the submitted version
Mohammed Yahya Qaid Al-Gumaei	Contributed to conception of the study and data analysis
Asim Bhatti	Contributed to the conception of the study, Supervised manuscript revision, read and approved the submitted version

5. Author Declarations

I agree to be named as one of the authors of this work, and confirm:

- i. that I have met the authorship criteria set out in the Deakin University Research Authorship Procedure,*
- ii. that there are no other authors according to these criteria,*
- iii. that the description in Section 4 of my contribution(s) to this publication is accurate,*
- iv. that the data on which these findings are based are stored as set out in Section 7 below.*

If this work is to form part of an HDR thesis as described in Sections 2 and 3, I further

- v. consent to the incorporation of the publication into the candidate's HDR thesis submitted to Deakin University and, if the higher degree is awarded, the subsequent publication of the thesis by the university (subject to relevant Copyright provisions).*

Name of author	Signature*	Date
Rakesh Veerabhadra		14/05/21
Imali Hettiarachchi		14/05/21
Mohammed Yahya Qaid Al-Gumaei		14/05/21
Asim Bhatti		14/05/21

Signatures Redacted
by Library

6. Other contributor declarations

I agree to be named as a non-author contributor to this work.

Name and affiliation of contributor	Contribution	Signature* and date

* If an author or contributor is unavailable or otherwise unable to sign the statement of authorship, the Head of Academic Unit may sign on their behalf, noting the reason for their unavailability, provided there is no evidence to suggest that the person would object to being named as author

7. Data storage

The original data for this project are stored in the following locations. Please ensure that this complies with the relevant procedure. .

Data format	Storage Location	Date lodged	Name of custodian if other than the corresponding author
The data was acquired in mat (MATLAB) format	Public Repository http://bcmi.sjtu.edu.cn/home/seed/	2013	

This form must be retained by the corresponding author. If the corresponding author is not a Deakin staff member or student, the Deakin research team must ensure that a copy of this form is kept at Deakin.

If the publication is to be included as part of an HDR thesis, a copy of this form must be included in the thesis with the publication.

AUTHORSHIP STATEMENT

Researches should review the Deakin Authorship Procedure to determine the criteria of authorship

1. Details of publication and corresponding author

Title of Publication		Publication details
Fundamental Requirements for Contemporary Performance Evaluation of Emergency-Response and Safety-Critical Teams		IEEE Transactions on Human-machine Systems
Name of corresponding author	School/Institute/Division if based at Deakin; Organisation and address if non-Deakin	Email or phone
Rakesh Veerabhadrapa	IISRI	rveerabhadrapa@deakin.edu.au

2. Inclusion of publication in a thesis

Is it intended to include this publication in a higher degree by research (HDR) thesis?	Yes	If Yes, please complete Section 3 If No, go straight to Section 4.
---	-----	---

3. HDR thesis author's declaration

Name of HDR thesis author if different from above. (If the same, write "as above")	School/Institute/Division if based at Deakin	Thesis title
As above	IISRI	Intelligent Multimodal Experimental Framework for Analysing, Classifying and Forecasting Team Performance
If there are multiple authors, give a full description of HDR thesis author's contribution to the publication (for example, how much did the student contribute to the conception of the project, the design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)		
I declare that the above is an accurate description of my contribution to this paper, and the contributions of other authors are as described below.		Signature and date of student
		14/05/21

Signature Redacted by Library

4. Description of all author contributions

Name	Contribution(s) (for example, conception of the project, design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)
Rakesh Veerabhadrapa	Contributed to the conception of the study, first draft of the manuscript, contributed to the manuscript revision, read and approved the submitted version.
Imali Hettiarachchi	Contributed to the conception of the study, contributed to the manuscript revision, read and approved the submitted version.
Asim Bhatti	Contributed to the conception of the study, Supervised manuscript revision, read and approved the submitted version

5. Author Declarations

I agree to be named as one of the authors of this work, and confirm:

- i. that I have met the authorship criteria set out in the Deakin University Research Authorship Procedure,*
- ii. that there are no other authors according to these criteria,*
- iii. that the description in Section 4 of my contribution(s) to this publication is accurate,*
- iv. that the data on which these findings are based are stored as set out in Section 7 below.*

If this work is to form part of an HDR thesis as described in Sections 2 and 3, I further

- v. consent to the incorporation of the publication into the candidate's HDR thesis submitted to Deakin University and, if the higher degree is awarded, the subsequent publication of the thesis by the university (subject to relevant Copyright provisions).*

Name of author	Signature*	Date
Rakesh Veerabhadra		14/05/21
Imali Hettiarachchi		14/05/21
Asim Bhatti		14/05/21

Signatures Redacted
by Library

6. Other contributor declarations

I agree to be named as a non-author contributor to this work.

Name and affiliation of contributor	Contribution	Signature* and date

* If an author or contributor is unavailable or otherwise unable to sign the statement of authorship, the Head of Academic Unit may sign on their behalf, noting the reason for their unavailability, provided there is no evidence to suggest that the person would object to being named as author

7. Data storage

The original data for this project are stored in the following locations. Please ensure that this complies with the relevant procedure. .

Data format	Storage Location	Date lodged	Name of custodian if other than the corresponding author
NA			

This form must be retained by the corresponding author. If the corresponding author is not a Deakin staff member or student, the Deakin research team must ensure that a copy of this form is kept at Deakin.

If the publication is to be included as part of an HDR thesis, a copy of this form must be included in the thesis with the publication.

AUTHORSHIP STATEMENT

Researches should review the Deakin Authorship Procedure to determine the criteria of authorship

1. Details of publication and corresponding author

Title of Publication		Publication details
A State-of-the-art Experimental Platform for Effective Analysis of Team Processes		IEEE Transactions on Systems Man and Cybernetics: Systems
Name of corresponding author	School/Institute/Division if based at Deakin; Organisation and address if non-Deakin	Email or phone
Rakesh Veerabhadrapa	IISRI	rveerabhadrapa@deakin.edu.au

2. Inclusion of publication in a thesis

Is it intended to include this publication in a higher degree by research (HDR) thesis?	Yes	If Yes, please complete Section 3 If No, go straight to Section 4.
---	-----	---

3. HDR thesis author's declaration

Name of HDR thesis author if different from above. (If the same, write "as above")	School/Institute/Division if based at Deakin	Thesis title
As above	IISRI	Intelligent Multimodal Experimental Framework for Analysing, Classifying and Forecasting Team Performance
If there are multiple authors, give a full description of HDR thesis author's contribution to the publication (for example, how much did the student contribute to the conception of the project, the design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)		
I declare that the above is an accurate description of my contribution to this paper, and the contributions of other authors are as described below.		Signature and date of student
		14/05/21

Signature Redacted by Library

4. Description of all author contributions

Name	Contribution(s) (for example, conception of the project, design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)
Rakesh Veerabhadrapa	Contributed to the conception, development of the experimental platform, computer applications, testing, data analysis, draft of the manuscript, contributed to the manuscript revision, read and approved the submitted version.
Imali Hettiarachchi	Contributed to the conception of the study, revision and testing of the platform, Selection of PBR devices and computer applications, Supervised the development, read and approved the submitted version.
Mohammed Yahya Qaid Al-Gumaei	Revision and testing of the platform
Asim Bhatti	Contributed to the conception of the study, revision and testing of the platform, Selection of PBR devices, Supervised the development, read and approved the submitted version

5. Author Declarations

I agree to be named as one of the authors of this work, and confirm:

- i. that I have met the authorship criteria set out in the Deakin University Research Authorship Procedure,*
- ii. that there are no other authors according to these criteria,*
- iii. that the description in Section 4 of my contribution(s) to this publication is accurate,*
- iv. that the data on which these findings are based are stored as set out in Section 7 below.*

If this work is to form part of an HDR thesis as described in Sections 2 and 3, I further

- v. consent to the incorporation of the publication into the candidate's HDR thesis submitted to Deakin University and, if the higher degree is awarded, the subsequent publication of the thesis by the university (subject to relevant Copyright provisions).*

Name of author	Signature*	Date
Rakesh Veerabhadra		14/05/21
Imali Hettiarachchi		14/05/21
Mohammed Yahya Qaid Al-Gumaei		14/05/21
Asim Bhatti		14/05/21

6. Other contributor declarations

I agree to be named as a non-author contributor to this work.

Name and affiliation of contributor	Contribution	Signature* and date

* If an author or contributor is unavailable or otherwise unable to sign the statement of authorship, the Head of Academic Unit may sign on their behalf, noting the reason for their unavailability, provided there is no evidence to suggest that the person would object to being named as author

7. Data storage

The original data for this project are stored in the following locations. Please ensure that this complies with the relevant procedure. .

Data format	Storage Location	Date lodged	Name of custodian if other than the corresponding author
The data was acquired in *.json, *.csv and *.pkl format. Platform and applications are available in *.exe format	Deakin RDS https://github.com/ghost2307/TankBattleTask_v1_github	2021	Imali Hettiarachchi Asim Bhatti

This form must be retained by the corresponding author. If the corresponding author is not a Deakin staff member or student, the Deakin research team must ensure that a copy of this form is kept at Deakin.

Abstract

In the present day work culture, challenges in work environments are complicated, exceeding the cognitive capacity of individuals. Often, the successful achievement of such challenges is critical, demanding interdependence with more than one party. Researchers of human factors and science of teams have shown a great deal of interest especially, to assess and evaluate the performance of teams such as effectiveness, viability, and efficiency. Conventionally, subjective feedback of collaborating members was used as a basis for the appraisal of teams. During the last couple of decades, this area of the domain has witnessed a growing interest in the employment of wearable and unobtrusive technologies. In this light, physiological and behavioural recordings (PBR) have been extensively researched to unfold complex dynamics of team individuals, interactions and task characteristics. Team studies have demonstrated that these inclusions offer advantages such as identification of best practices, forecasting potential threats and prediction of future outcomes. However, the literature survey reveals that the advancements have slowly progressed such as lacking effective team task environments, naive integration of PBR-sensors and dyadic-constrained analyses. This state of research means that current approaches are primitive, limited support for multimodal study and their compliance with larger teams (three or more members in a team) are often questionable. Therefore, this work endeavoured to explore a framework that enables appraisal and predictive methods that are suitable for larger of individuals working as a team.

This thesis conducted a selective review of studies utilising PBR to explore teams, multi-party, group and collaborative problem-solving in the context of shared space and joint activities. The inputs-processes-outputs framework, popular among the PBR based team studies was used as a basis to critically review current experimental procedures, processing methods and their relationship with theories underlying processes and emergent states of teams. A paramount objective of the review was to identify whether there existed any aids to predict team performance outcomes utilising regression, neural networks and classifier algorithms. In addition, three studies were performed in the initial exploratory approach to develop an understanding of synthetic collaborative task settings, PBR and spatial-temporal data processing algorithms. These studies included a review of feature selection processes and the performance of unsupervised and supervised clustering algorithms, a study on using Heart-rate-variability (HRV) of dyads engaged

in a collaborative task, and the development of deep neural-networks (NN) for classifying emotions using electroencephalography (EEG).

The primary objective of this thesis was to explore relationships among task performance, multimodal PBR and subjective feedback in the context of processes and emergent states of teams. Hence, the knowledge gained through the review and the exploratory studies were transformed to establish an empirical team testing framework. The framework comprised of a novel 3-member cooperative joint activity integrated with four modes of PBR, i.e., EEG, electrocardiography (ECG), eye-tracking and verbal communication. The embedded system is capable of concurrently acquiring a total of 12 PBRs (3-members and 4-modalities). Seventeen 3-member teams attempted 4-trials of the experiment while their multimodal data was recorded.

This thesis separately analysed data from ECG and eye-tracking, and ultimately, proposes a method that achieves a fusion of both modalities which serves as a composite index in the evaluation and analysis of team performance. First, eye-tracking was analysed and methods to quantify gaze behaviours such as convergence, co-occurrences and distribution are introduced in this study. The relationship between gaze behaviours and task performances were explored using simple linear regression and multivariate regression. Cross-correlation between linear statistics and performance was employed to identify features and a back-propagation NN resembling regression model was achieved.

Second, R-R peaks of ECG was used to establish time-series cognitive states (CS) of individuals. Then CS of team trials were represented using 7-level categorical team cognitive states and 5-levels cognitive load compliance states. The correlation between these levels and performance revealed a significant effect. Further, in accordance with the theories of visual cognitive activity and autonomic cognitive activity, a composite index was derived combining gaze co-occurrence and CS. Due to limited data samples, the analyses were cross-validated using K-fold validation running 30 different permutations.

This novel framework enabled exploring multimodality-inspired composite indices for evaluation and prediction of team performance. The results provided a solid foundation in the context of the relationship between team processes, PBR-based compliance indices and gaze behaviour. The validity of research outcomes could certainly be improved by increasing the number of data samples.

Future research could focus on enhancing the multimodal capability incorporating EEG and verbal communication. The time-series nature of data certainly encourages the use of recurrent NN for predicting the performance of teams using the initial phase of task performance data. The in-house developed task environment encourages the incorporation of predictive methods for online prediction of team performances.

Contents

Publications	vii
Abbreviations	xxxii
1 Introduction	1
1.1 Background	1
1.1.1 Team Definitions	1
1.1.2 Conventional Methods	2
1.1.3 History, Rationale and Fundamental Components of Work-Team Research	2
1.1.4 Physiological, Behavioural and Multimodal	3
1.2 IPO Framework	4
1.2.1 Inputs	4
1.2.2 Processes or Mediators	5
1.2.3 Outputs	6
1.2.4 Emergent States	6
1.3 Ramifications for Team Performance Evaluation	7
1.4 Current gaps, Motivations and Objectives	8
1.4.1 Current Gaps	8
1.4.2 Research Motivations	11
1.4.3 Research Objectives	12
1.5 Methodology and Research Design	12
1.6 Publications	13
1.7 Thesis structure	15
2 Critical Review of Fundamental Components for Physiological and Behavioural Recording (PBR)-based Team Studies	18
2.1 Introduction	18
2.2 Measuring Team Members' Activity	19
2.2.1 Physiological Observations	19
2.2.2 Behavioural Observations	21
2.3 Team-level Measurements	21
2.3.1 Compliance Metrics	22
2.3.2 Temporal Neurodynamic Patterns of Teams	23
2.3.3 Latent Semantic Analysis (LSA)	24
2.3.4 Epistemic Network Analysis (ENA)	24
2.4 Relationships between Team Performance and Constructs of IPO	24
2.4.1 Team Performance and Team Inputs	24
2.4.2 Team Performance and Team Processes	25

2.4.3	Team Performance Outcome and Its Implications . . .	26
2.5	Multimodality in The Context of Teams	27
2.5.1	Data Fusion in Multimodal Studies	28
2.6	Simulation Environments	29
2.6.1	Current Limitations of Synthetic Environments	30
2.7	Subjective Ratings	31
2.8	Discussion and Future recommendations	31
2.8.1	Team Cognition Based Performance Evaluation	32
2.8.2	Analysis Beyond Dyads	32
2.8.3	Need for Computation Improvement	33
2.8.4	Experimental setup	33
2.9	Summary	40
3	Preliminary Study-a: Using ECG	41
3.1	Introduction	41
3.2	Materials	42
3.2.1	Participants	42
3.2.2	Experimental Setup and Procedure	43
3.3	Method	43
3.3.1	Data pre-processing	43
3.3.2	Recurrence Quantification Analysis	45
3.3.3	Physiological Synchrony	47
3.3.4	Permutation-based significance testing	48
3.4	Results and Discussion	48
3.5	Summary	54
4	Preliminary Study-b: Using EEG	55
4.1	Introduction	55
4.2	Materials and Methods	57
4.2.1	Experiment and Dataset	57
4.2.2	Data Preparation	57
4.2.3	Training and Test Data	57
4.2.4	Multi-Channel Dense Convolutional Neural Network .	59
4.3	Results and Discussion	62
4.4	Summary	66
5	A Novel Multimodal and Multiparty Collaborative Experimental Framework	67
5.1	Introduction	67
5.2	The Tank-Battle Task	69
5.2.1	The Synthetic Task	69
5.2.2	Push-pull communication	70
5.2.3	Simplified Player Object Control	70
5.2.4	Configurable Parameters	71
5.2.5	Task Intensity and Spawn Pattern	71
5.2.6	Task Summary and Performance Measures	72

5.3	Materials	72
5.3.1	Participants	72
5.3.2	Experimental Setup	73
5.3.3	Multimodal Physiological Data	73
5.3.4	Questionnaires and Subjective feedback	80
5.3.5	Experimental Procedure	83
5.4	Simulation Analytics	83
5.4.1	User-Experience	84
5.4.2	Environment Dynamics	85
5.5	Results & Discussion	86
5.5.1	Comfort-levels of using wearable devices	86
5.5.2	Task-load Index	87
5.5.3	Comparing Objective Against Subjective Performance	89
5.5.4	Task-related Questionnaire	91
5.5.5	Cooperation in Gaming Questionnaire (CGQ)	91
5.6	General Discussion	94
5.7	Summary	96
6	Gaze Co-occurrence and Distribution	98
6.1	Introduction	98
6.1.1	Pre-processing and Data Selection	99
6.1.2	Gaze data preprocessing	99
6.1.3	Paired-distance Estimation	99
6.1.4	Gaze Co-occurrences	100
6.1.5	Distributed Gaze	101
6.2	Results & Discussion	101
6.2.1	Analysis of Operating Conditions	101
6.3	Discussion	105
6.4	Summary	106
7	Gaze Convergence Using Neural Network	108
7.1	Introduction	108
7.1.1	Data Selection	109
7.1.2	Gaze Data Pre-processing	110
7.1.3	Paired-distance Estimation	110
7.1.4	Neural Network Prediction Model	111
7.2	Results	112
7.2.1	Significance Between Operating Distances and Performance	112
7.2.2	Performance Prediction	112
7.3	Discussion	113
7.4	Summary	115
8	Multimodal Index and Effects of Team Cognitive Activity	116
8.1	Introduction	116
8.1.1	Experiment	118

8.2	Methods	118
8.2.1	Processing IBI	118
8.2.2	Processing Gaze data	123
8.3	Results	126
8.3.1	Data for Analysis	126
8.3.2	Relationship between PB measures and performance .	127
8.3.3	Gaze data analysis	130
8.3.4	Team Cognitive Activity	132
8.4	Discussion	135
8.5	Summary	137
9	Conclusion, Limitations and Future Recommendation	138
9.1	Conclusion	138
9.2	Limitations	139
9.3	Future recommendations	141
A	Processing ECG to extract IBI and HRV measures	168
A.1	ECG Data	168
B	Recurrence Quantification Analysis	171
B.1	Preparing Data for RQA	171
B.2	Recurrence Plots	172
B.2.1	Recurrence Rate	173
B.3	Diagonal Profile Measures	174
B.3.1	Determinism	174
B.4	Vertical-Horizontal Profile Measures	175
C	Tank Battle Task Environment	176
C.1	Game Main Menu	176
C.2	Parameter Configuration Window	176
C.3	Survey Configuration Window	178
C.4	Survey Attempting Window	178
C.5	Task Performance Feedback Window	180

List of Figures

Figure 1.1	A generalised model of IPO framework inspired by the works of Rosen <i>et al.</i> , Kazi <i>et al.</i> , Gladstein (1984) and Tannenbaum <i>et al.</i> [1–4]	5
Figure 2.1	A summary of physiological systems (dashed-boxes) for observing internal activity from individuals. Also shown are the potential measures at individual-level that are of importance for team level analysis	20
Figure 3.1	Example showing a short extract of ECG data, R-peaks and IBI	44
Figure 3.2	Example of time-series IBI estimated from ECG	44
Figure 3.3	Histogram of observed embedding delay and dimensions, respectively, for all 38 participants considered in the study	45
Figure 3.4	A plot of Threshold vs min recurrence observed across participants data set for varying window sizes. The dashed line shows the threshold selected for the generating RP, which satisfies all recommended criteria (refer to section 3.3.2 for details) for all window sizes	47
Figure 3.5	Plot of Pearson’s correlation of all valid pairs for window size of 300 seconds to demonstrate difference between every RQA measure considered in the study. Each team consists of a p_i & n_i , indicating a programmer and a navigator, respectively. The diagonals ($i = j$) in plots represent true pairs and elsewhere ($i \neq j$) are the pseudo pairs. It should be noted that the experiment was originally conducted across two sessions. To accommodate this, pseudo pairs were formed only for the participants whose programming task, time and space synchronised. The x-axis in the plot indicate participants who started as programmers and y-axis indicate participants who started as navigators. For comparison with time-domain HRV measures, correlation observed for SDNN is also included in the subplot (f).	51

Figure 3.6	Comparison of correlation coefficient produced by the linear measure SDNN-HRV and nonlinear measures; <i>Rec</i> , <i>D</i> , <i>H</i> and <i>L</i> , respectively for real pairs using a window size of 300 seconds	52
Figure 4.1	The topological representation of electrodes orientation of the EEG cap employed in this data	58
Figure 4.2	Architecture of the multi-Channel dense convolutional neural network	60
Figure 4.3	The plot of accuracy and cross entropy loss obtained from the model employing training and testing sets generated using the SI1 method	64
Figure 5.1	Key components of the experimental framework based on the principles of IPO for team performance analysis, comprising of team task environment, PBR techniques and subjective questionnaires.	68
Figure 5.2	Main operating window of the tank battle simulation task and modalities	69
Figure 5.3	An overview of experimental setup demonstrating how every physiological and behavioural techniques are integrated and communicated with the server (Manager) and client machines (Alpha, Bravo, Charlie)	74
Figure 5.5	An overview of the custom developed windows application to simultaneously record raw IBI data from three different devices. For verifying the working of the device and to ensure transparency with participants, the raw IBI values were converted and displayed as beats per minute using eqn. 5.1.	77
Figure 5.6	An overview of gaze measurements demonstrating fixations and saccades in the task environment	78
Figure 5.7	An overview of gaze measurements demonstrating fixations and saccades in the task environment	81
Figure 5.8	Summary of team-individuals activity recorded from mouse events during a trial session. The top row is record of <i>Look</i> events showing a players visual scanning of map regions. Middle row is records of <i>Navigate</i> events showing where a player operated. And bottom row is record of <i>Engage</i> events showing where a player engaged to destroy UGVs.	84
Figure 5.9	Dynamics of participant BTs. (a). Spatial location of BTs with <i>x</i> and <i>y</i> axis representing screen coordinates in relation to 1080×1920 screen resolution. (b). Temporal location of BTs with <i>x</i> axis representing pixel value corresponding to 1080×1920 screen resolution and <i>y</i> axis representing time in milliseconds.	85

Figure 5.10	Dynamics of UGVs. (a) Spatial locations of <i>friend</i> variants and (b) <i>enemy</i> variant UGVs. The x and y axis represent screen coordinates in relation to 1080×1920 display resolution	86
Figure 5.11	Average task load indices experienced by participants in trial episodes	88
Figure 5.12	Comparison between objective (Task) and subjective performance (NASA-TLX). From left to right, plot of bonus gained vs performance, plot of health lost vs performance, and overall health vs performance. The bonus gained is a result of properly identifying more friendly variants. The health loss is a result of failing to destroy enemy variants in time. The base health (%) is consolidated version of the previous two decisive parameters i.e., the bonus gained and health lost.	88
Figure 5.13	A boxplot of task performance (BH) of the teams as they progressed through 4-trials of the experiment . .	89
Figure 5.14	A heatmap comparing correlation between task performance scores and synthetically generated from task event dynamics registered by team members. . .	90
Figure 6.1	Example showing a short excerpt of original x -coordinates of a gaze-data, threshold used to decide invalid coordinates, high resolution cleaning and re-sampling and finally, the down sampled data used for further analysis.	100
Figure 6.2	A plot of correlation coefficient's between gaze-behaviour vs distance of team trials. (A) gaze co-occurring behaviour and (B) distributed gaze behaviour, for all pixel distance between 50 and 2100 with an increment of 50-pixel distance. Note: The curve for three pairs in gaze distribution analysis (B) comprised of invalid correlation scores because the corresponding data points comprised of all zeros. This was because no team had all three pairs that outside 1350 pixel distance.	102
Figure 6.3	Example showing plot of correlation coefficient vs distance for one-pair gaze co-occurrence. For comparison the plot shows responses using all samples and 4-folds from the sample.	103
Figure 6.4	Plot demonstrating the relationship between gaze co-occurrence feature f_4 and gaze distribution feature f_8 with task performance BE.	105

Figure 6.5	Example demonstrating an impression of 300, 600, 800 and 1000 pixel diameter circles on a 1920×1080 scale LCD	106
Figure 7.1	Example showing a short excerpt of original x-coordinates of a gaze-data, threshold used to decide invalid coordinates, high resolution cleaning and re-sampling and finally, the down sampled data used for further analysis.	111
Figure 7.2	Plot showing multi-variate regression utilising average pair-distance ($\mu(d_p)$) (B,C) and standard deviation ($\sigma(d_p)$) (A,C) contributing to the normalised task performance (BH).	112
Figure 7.3	K-fold testing used for validating the neural network model. The plot shows MAE observed for 4 folds with 30 different permutations	113
Figure 8.1	Tercile to cognitive load transformation to representation of a cognitive pattern.	120
Figure 8.2	Basis for grouping team cognitive patterns to levels	121
Figure 8.3	Detailed picture of seven levels of team cognition after categorising patterns of team-member terciles. From left to right describes most patterns with lower levels of team cognition moving toward higher levels of team cognition. The labels to the left of each plot indicates the team cognitive state labels	122
Figure 8.4	Detailed picture of five levels of cognitive compliance after categorising patterns of team-member terciles. From left to right describes most patterns with lower levels of team cognition moving toward higher levels of team cognition. The labels to the left of each plot indicates the team cognitive state labels	124
Figure 8.5	A detailed picture of multiple linear regression utilising PB measures CS_{25} and f_{μ}	132
Figure 8.6	A detailed picture of multiple linear regression utilising BC measure BC_{1P} and categorical-TCS measure T_{17}	132
Figure 8.7	Boxplots of TCSs CS_{22} and CS_{24} against different stages of the task showing variations due to teams continuous involvement. Subplots A and B corresponds to counts of CS_{22} and CS_{24} , respectively, observed during the full-length of a task trial. Subplots C and D corresponds to CS_{22} counts during low intensity and CS_{24} counts during the high intensity phase of a task trial.	133

Figure 8.8	Boxplots of gaze-based s_μ (eqn. 8.3) against experiment stages and task conditions.	134
Figure A.1	Example of ECG data recorded at 500Hz sampling frequency	168
Figure A.2	Using running median filter to smoothen ECG signal	168
Figure A.3	Subtracting ECG with smoothened ECG to obtain spikes	169
Figure A.4	Square transformation to eliminate negative peaks . .	169
Figure A.5	Using threshold to detect area around R-peaks	169
Figure A.6	Binary transformation and detection of rising edges and falling edges	170
Figure A.7	R-peaks are identified by selecting the maximum ECG value between a pair of rising and falling edges	170
Figure B.1	A plot differences between raw IBI-series, after interpolating with four times the number of samples and extrapolating to exactly 1 sample/second.	171
Figure B.2	Example showing recurrence plot of IBI-series x_k . For more detail, a section the highlighted section is further zoomed in figures. B.3 and B.4	172
Figure B.3	Example showing zoomed version of the highlighted section of Figure. B.2. For more detail, the highlighted section is further zoomed in Figure. B.4	173
Figure B.4	Example showing zoomed version of the highlighted section of Figure. B.3. Also shows what constitutes a vertical, horizontal and diagonal lines from RP.	174
Figure C.1	Selection menu of tank battle task.	176
Figure C.2	Task parameters and the circular window for setting UGV spawn locations.	177
Figure C.3	Window showing all questionnaire configurable property of the task.. . . .	178
Figure C.4	Window showing editing and formatting of surveys and questions.	178
Figure C.5	Window showing participants answering a questionnaire.	179
Figure C.6	Performance feedback window showing each members task scores and the based health at the end of a trial.	180

List of Tables

Table 2.1	Summary physiological measures contributing to team performance analysis	35
Table 2.2	Summary multimodal measures contributing to team performance analysis	38
Table 3.1	Observed significance (holm-bonferroni adjusted p-values) for selected short-term window sizes estimated across the RQA measures; <i>Rec</i> , <i>D</i> , <i>H</i> , <i>L</i> and <i>L_{max}</i> obtained by comparing 95% confidence intervals of pseudo pairs with mean correlations of real pairs. The observed level of significances are * for $p < 0.05$, ** for $p < 0.01$ and *** for $p < 0.001$	53
Table 4.1	Parameters used to generate a model for training and testing	61
Table 4.2	Summary of MC-DCNN accuracies observed from different training and test data generation methods .	63
Table 4.3	Accuracies observed across all 15 participants' experiment-1 data using KNN, SVM and MC-DCNN using SD and SI2 data splitting methods. The average and standard deviation row is obtained across all 15 participants	65
Table 5.1	List of configurable objects and parameters with example settings	71
Table 5.2	Consolidated wearable device comfort response from participants	87
Table 5.3	Different stages of the experiment	91
Table 5.4	Consolidated response from participants	91
Table 5.5	Cooperation in Gaming Questionnaire	93
Table 6.1	A summary of results for the gaze behaviour measures identified in Fig. ??	104
Table 8.1	Example of interpreting categorical team cognitive load ($\vec{T}$) measures for a team-trial	120
Table 8.2	Example of interpreting BC measures for a team-trial	126
Table 8.3	Linear statistics (mean and standard deviation (SD))and Pearson correlation ($N=32$) between IBI-based measures <i>CS</i> , <i>T</i> , <i>LC</i> and task performance (<i>BH</i>)	128
Table 8.4	Multivariate regression using combinations of $CS_{16,21\&25}$ and performance <i>BH</i>	129

Table 8.5	Linear statistics (mean and standard deviation (SD))and Pearson correlation (N=28) between gaze based f_{μ} , BC and task performance (BH)	130
-----------	--	-----

Abbreviations

AETS advanced embedded training system.
ANN Artificial Neural Network.
ANS Autonomic nervous system.
BCI Brain Computer Interface.
BPOG best point-of-gaze.
BS Base Station.
BT Battle Tank.
CGQ cooperation in gaming questionnaire.
CM-TLX composite multimodal team level index.
CNS central nervous system.
CRQA Cross-recurrence quantification analysis.
CS currogator supercili.
CSCL computer supported collaborative learning.
CSCW computer supported collaborative working.
DA Directional Agreement.
DDD distributed dynamic decision-making.
DNN dense neural network.
EBR Eye Blink Rate.
ECG Electrocardiography.
EDA Electro dermal activity.
EEG Electroencephalography.
EMG Electromyography.
ENA Epistemic Network Analysis.
ER&SC emergency response and safety critical.
f-ID fixation identification number.
GID global-identification.
HR heart rate.
HRV heart rate variability.
IALD inter-areas links density.
IBI Inter-beat interval.
ICD inter-connectivity density.
ICG Impedance Cardiogram.
IDM Instantaneous derivative matching.
IMO input-mediator-output.
IMOI input-mediator-outcomes-input.
IPO inputs-processes-outputs.
KNN K-nearest neighbours.

KSA knowledge, skills and attitudes.
LC load compliance.
LSA latent semantic analysis.
LVC Live virtual constructive.
LVET left ventricular ejection time.
MATB Multi-Attribute Task Battery.
MC-DCNN multichannel dense convolutional neural network.
NASA National aeronautics and space administration.
NASA-TLX Task Load Index.
NIRS Near-Infrared Spectroscopy.
NS neurodynamic synchrony.
OO orbicularis oculi.
PBR Physiological and Behavioural Recording.
PCA principal component analysis.
PEP pre-ejection period.
PNS parasympathetic nervous system.
POG point-of-gaze.
RMSE root mean square error.
RMSSD Root-Mean-Squared-Successive-Difference.
ROI region of interest.
RP recurrence plot.
RQA recurrence quantification analysis.
RSA respiratory sinus arrhythmia.
SA Situational Awareness.
SAE sparse autoencoder.
SCL Skin conductance level.
SCR Skin conductance response.
SDNN standard deviation of normal-normal.
SEED SJTU Emotion EEG Database.
SM Signal Matching.
SNS sympathetic nervous system.
SPC socio-physiological compliance.
SPGQ social presence in gaming questionnaire.
STE synthetic task environment.
SVM Support Vector Machines.
TCS team cognitive states.
TEQ task experience questionnaire.
U.S. United States.
UGV unknown ground vehicle.
UID unique-identification.
WC Weighted Coherence.
ZM zygomaticus major.

1

Introduction

1.1 BACKGROUND

Domains such as health care, defence, emergency response and paramilitary pose a complicated working environment. Circumstances such as unforeseen events, tasks and objectives lacking clarity, time-constrained decision-making and errors that could endanger lives are pervasive in such environments [5–8]. Typically, these situations are ambiguous and stressful, exceeding the cognitive capacity of single individuals [9]. In addition, the recent technological revolutions have elevated the complexities precluding individuals from accomplishing tasks. As a counter-strategy, work domains seek interdependence with more than one individual to address the challenges [5, 10]. Interdependent tasks include emergency response and safety critical (ER&SC) settings such as military command and control operations (example, air traffic control, flying aircraft and operating unmanned-vehicles (UV)), coordinated health-care activities (example: carrying patients and surgery), and computer supported collaborative working (CSCW) (example, remote collaborations to solve complex programming tasks and monitoring and operation of heavy industrial equipment). Thus, a relationship exists between contributions of collaborating members, conditions of the task and the performance outcomes. It is therefore of a vital interest among human factors researchers to understand the factors that influence levels of task achievement and aid in the development of indirect methods to assess and evaluate performances of work teams [11]. Such advancements are paramount for assessing and enhancement of team training methods and goals, identifying effective practices of teams, eventually contributing to advancing team functioning [12, 13].

1.1.1 Team Definitions

This study uses “team” to represent a social context where two or more individuals are interdependent on each other’s contribution, adaptively interact, collaborate, coordinate, cooperate and pursue a common interest towards valued goals [4, 5, 14]. The term “team performance” collectively

represents the behaviours that are observed (a) during the interactivity between individual members, and (b) while accomplishing their set objectives, respectively. And the overall team performance outcomes are evaluated in terms of effectiveness, viability, and efficiency to achieve the goals [4, 15].

1.1.2 Conventional Methods

Retrospectively, team studies and team-training programs analyse subjective ratings to unfold the impact of influential factors in relation to performance outcome [16–18]. Subjective ratings are collected from team members, commonly self-reported or observer-reported and aggregated over the team [3, 19–21]. However, individuals memory (long or short term) and intent of participation can expose subjective ratings to potential biases [3, 22, 23]. On the other hand, employing human scrutineers are laborious, time-consuming and uneconomical [24], in addition to systematic biases. As an alternative, recent advancements in sensor technologies have encouraged the employment of non-invasive and unobtrusive methods to unfold the relationships between collaborating members physiological and behavioural activities and task performance [12].

1.1.3 History, Rationale and Fundamental Components of Work-Team Research

Physiological correlation between interactive dyads was reported as early as mid-1950s in a study between psychotherapists and patients [25]. One of the contributing factors for such correlations is the concomitant relationships that exist between physiological variations and psychological behaviours [26]. Indeed, Hennings summary of Smith and associates motor behaviour studies highlights that, in addition to motor activities responsible for sensory control and body locomotion, it directly affects internal physiological systems such as cardiac and respiratory functioning during the self-regulatory process of individuals [11]. Thus, forming a dynamic closed-loop behavioural organisation model where sensory feedback from internal and external sources due to motor activity collectively inhibits self-regulation. Accordingly, the context of social interaction evolves an integrated sensory-motor closed-loop system between participants which is partly responsible for necessary compromises in personalities via the joint regulation of individuals internal physiological states. Early socio-physiological investigations were facilitated by employing polygraphic recording of physiological activities such as heart rate, skin conductance and skin temperature [25–28]. Typically, emotion states such as anger, antagonism, stress and joy, and affective states such as neutral, negative and positive were captured and later used in the analyses. In the following

years, physiological correlations in social contexts where liking, disliking and mixed opinions between dyads of small groups [27], and marital satisfaction between spouses [29] were analysed. These studies contributed to the 21st-century socio-psychophysiological and behavioural methods for studying processes of work teams. Hennings (2001), demonstrated computerised adaptation of motor pursuit task to investigate socio-physiological compliance (SPC), where physiological state variations were analysed via the skin conductivity, heart rate and respiration measures.

In complement with the methods mentioned above, research on work teams witnessed the evolution of cooperative tasks to socially engage participants. Everyday tasks and activities such as two-handed saw, rowing and see-saw, to manoeuvring dual-controlled vehicles and airplanes formed the basis for task designs in early studies [28,30]. The primary motivation was to facilitate the investigation of behaviours such as cooperation, leadership, collaborative problem solving, interaction as well as knowledge transfer and learning [30–32]. In line with the socio-psychological work-team settings, one of the important goals of team research in human factors was to counter the complicated and cognitively demanding settings such as ER&SC environments [33]. In this light, researchers highlighted the need for an environment that resembles naturalistic problems, and yet offers control over the experiment and task settings [31]. It was for this purpose, United States (U.S.) military commissioned the development of a number of synthetic task environment (STE) [34]. Similar to STEs, stemming from the concepts of computer-based automated instruction system was the advanced embedded training system (AETS) [32], developed by U.S. Navy to train shipboard teams. AETS included simulated naturalistic objectives, offered real-time monitoring of performances, and structured delivery of instruction and feedback to trainees. Overall, the aforementioned advancements formed the rationale and essential platform of recent PBR based team/collaborative research, establishing non-invasive PBR techniques and STEs as fundamental components of ER&SC.

1.1.4 Physiological, Behavioural and Multimodal

PBR techniques offer advantages such as continuous monitoring, computational flexibility, long-term record keeping and minimal human interventions [35,36]. The continuous data provides better temporal specificity enabling enhanced statistical analyses. Most recently, the field of computational intelligence (CI), including concepts such as artificial neural networks, unsupervised classifiers and forecasting algorithms have been rapidly evolving. Co-jointly with CI, PBR-based team studies could contribute to modelling predictive algorithms that can forecast future performance outcomes of team operation [4,37–40]. In recent years, team research has witnessed a

steady rise in the number of studies experimenting with a combination of PBRs (also referred to as multimodal techniques) [41–43].

Aggregates of PBR employing measurement from more than one mode is referred to as multimodal measurements [44]. It is believed that the natural reactions of humans involve combinations of physiological activities and behavioural responses [4]. Multiple modalities thus can capture more than one response concurrently providing a better representation of an individual's experience. Such an approach can aid better deciphering of intricate interpersonal physiological entrainment between the team members in relation to team performance [36, 45]. Multimodality can be achieved either by a combination of physiological and behavioural techniques [45, 46], a combination of two or more physiological techniques [36, 47], or a combination of two or more behavioural techniques [48].

1.2 IPO FRAMEWORK

With simulated tasks and PBR being the fundamental components of team studies, effective investigation of research questions requires an equilibrium between the two components. In this light, eleven different models or frameworks that have contributed to team studies were selectively discussed by Salas *et al.* from a pool of 134 frameworks [6]. Inspired by the works of Gladstein (1984) and Tannenbaum *et al.* (1992) [1, 2], the IPO framework (Figure. 1.1) has been widely accepted among the teamwork research community [4], especially, in studies that employ the PBR approach [3]. Alternatively, researchers have also used terms such as input-mediator-output (IMO), since processes act as mediators, as well as, input-mediator-outcomes-input (IMOI), referring to the cyclic nature of the framework.

Typical IPO framework-based applied research experimental setup can be composed of at least one PBR technique, a task environment, post-task questionnaires, analytical methods for data pre-processing and statistical methods [4, 49–51]. Such an experimental setup is necessary to test and evaluate team related theoretical hypotheses [52]. The IPO framework can help identification of factors contributing to the overall performance of a team [3, 4, 6, 13, 22]. Understanding IPO framework helps to unravel any interrelationship between the constructs *i.e.*, the *inputs*, *processes* and *outputs* that concern team performance analysis. Attributes of teams constituted with in each construct is further elaborated as follows.

1.2.1 Inputs

The *inputs* to the teams act as a stimulus which generally set the condition for activating the relevant processes. Team structure and task structure are primary contributors of the *inputs* block [3]. The team structure concerns anatomical information such as the number of members in a team, gender

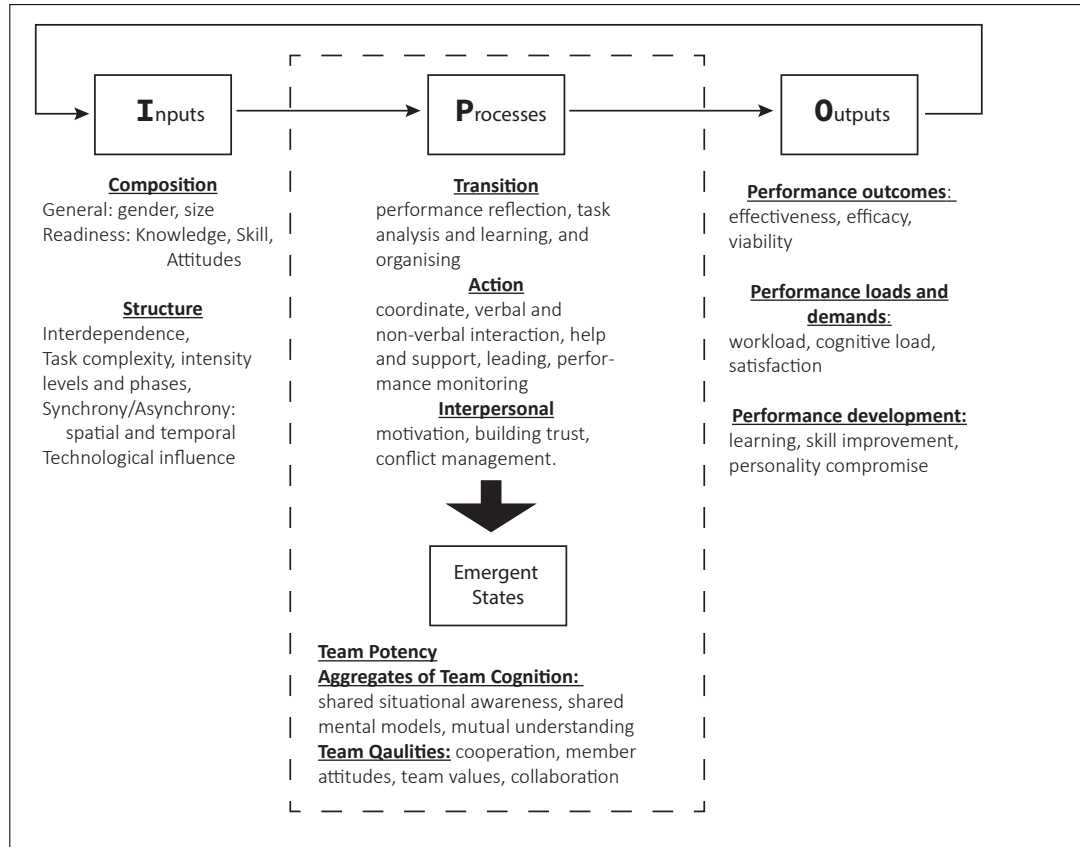


Figure 1.1: A generalised model of IPO framework inspired by the works of Rosen *et al.*, Kazi *et al.*, Gladstein (1984) and Tannenbaum *et al.* [1–4]

composition of a team [53, 54] and team members readiness (knowledge, skills and attitudes (KSA)) that defines their experience and exposure [52]. The task structure concerns the task activities such as complexity levels, phases within task episodes, and objectives of task [3].

1.2.2 Processes or Mediators

In the context of teams, often there exists an inter-connectivity between team member responses referred to as *processes*. The *processes* (also referred to as mediators in some literature) are the fundamental factors that influence the performance of a team. Marks *et al.*, conceptualises team processes based on the period of time while a team is involved in a task as well as between episodes of the task [55]. Accordingly, three different elements contribute to processes which are listed as (a) *Transition process*, the period of time occurring between trials which are responsible for teams to reflect, analyse and organise various aspects pertaining to enhancing goal achievement [56], (b) *Action process* constitutes the activities that occur while a trial is in progress which is responsible for the coordination, resource monitoring, interaction, communication, and external supporting [56] and (c) *Interpersonal*, appears at all times from the moment a team is formed and deployed. It mainly fo-

cuses on managing interpersonal relationships, responsible for maintaining harmony, boosting confidence, motivation and emotional balance [56].

1.2.3 Outputs

The *outcomes* of a team can be observed in their ability to accomplish task objectives such as effectiveness and efficiency [4]. The outcomes also represent learning in teams i.e., transformation in team members' readiness due to experience gained following completion of task routine. Furthermore, effectiveness variables such as satisfaction, viability, and load experienced by team individuals (e.g., workload, cognitive load and temporal load) [3,4] influence performance.

1.2.4 Emergent States

In addition, as a result of different processes mentioned above, extensive involvement of team members is reflected in the evolution of newer *team states* [55,57] such as team potency [58], team situational awareness [5], team stress [59], shared knowledge [10] and the trust, respect and cohesion between team members. The team states are responsible for defining qualities of a team such as member attitudes, motivations, values, cooperation, cognition and collaboration that critically define levels of team functioning, and hence, the performance.

Cognition

Collective team cognition is very critical in defining the overall performance of a team. Wildman *et al.* mentions three different approaches to perceive team cognition; generic, team knowledge and mental models and, Situational Awareness (SA) [60].

Affective states

As much as cognition is vital in a team environment, the individuals' emotions and mood directly impact productivity. In the past, affective states have been studied intra-individually as well as, a collective phenomenon [61]. Affect states significantly impact neurodynamic interconnections such as physiological activations and physiological organisations especially, when team members are enabled in a trust-building process [62].

1.3 RAMIFICATIONS FOR TEAM PERFORMANCE EVALUATION

The achievement of task objectives using objective scoring could be used as a reference for team performance evaluation [16, 52]. In this regard, the following ramifications of team performance evaluations were observed in the literature.

- *Effectiveness and Efficiency* of teams to achieve task objectives such as task completion time [63–65], account of errors accumulated by teams during a task [63, 65], differences in the quality of final product [66], teams ability to manage resources during a task [67], and efficiency in tracking [67] are some of the variables that can differentiate teams performances.
- *Team performance developmental activities and implications* such as exploring the relationship between variations in objective scores across trials and the evolution of strategies [68, 69]. The amount of practice required for transforming the level of performance to a different level (enhancement or decline in performance) [70].
- *Expert vs novice teams* describe the composition of the team based on performance levels. Similar terminologies could include experienced vs inexperienced teams [45, 70]. For example, it is of interest to understand the physiological and behavioural differences between experts handling a task complexity relative to novices. Time taken to complete a task was used as a parameter to observe the difference between elite, intermediate and poorer teams [70]. For the sake of simplicity, the term *superior performers* is adopted in this review to describe experienced [70], elite [63], high-scoring [21], low-error [71], skilled, good-quality [66] and any terminologies of similar sort.
- *Task-conditions*, differences in objective scores or any performance indicator could be used to learn the impact of various task conditions such as complexities i.e., high-intensity versus low-intensity phases of task [36, 70, 72].
- *Organisation*, differences in objective scores or any performance indicator due to differences in the organisation of the team. The influence of individual attitudes such as homogeneity, heterogeneity, the evolution of ownership and leadership, and conflict on task performance is studied in the literature [35, 66, 73].
- *Familiarity*, performance differences that could be observed between familiar team members vs non-familiar team members. Some studies employ this approach to observe performance differences between pseudo-generated teams against original teams [36].

- *Information Exchange*, performance differences observed due to collective knowledge and skills of the team members, uniqueness towards exchanging information, collective problem solving and processing and deciphering information [5,74].

In the event of multiple performance determining factors, cumulative scores representing the achievement of objectives are employed in some studies as task-specific performance scores. These objective scores offer an excellent platform to compare team performances.

Over the last two decades, the science of teams has gained significant momentum [75–77] towards understanding the factors that influence the performance of teams [13].

1.4 CURRENT GAPS, MOTIVATIONS AND OBJECTIVES

Despite modernised research methods on team performance utilising computer-simulated tasks and PBR, being practised for over 20 years, the current technology readiness level for adoption in practical and ecological settings is underdeveloped. Primarily because some of the current practices are primitive, lack structured approaches and limited complementary studies. This thesis broadly seeks to augment the research on team performance evaluation practices via the proposal of a novel multimodal applied framework. Via incorporating multiple modalities, the primary aim of this thesis endeavours to identify complementing modalities that could incorporate the influence of multiple physiological activities, behaviours and responses for effective evaluation of team performance. This is a problem of complex integration demanding a consensus between collaborative task settings, team-task objectives and IPO of teams. The process of realising this aim starts with the identification of successive gaps in the literature that will expose the limitations of current practices. The identification of gaps will help narrow the research motivations to specific areas, ultimately, aiding in the refinement of objectives of this thesis.

1.4.1 Current Gaps

As introduced in the aforementioned sections and suggested by Kazi *et al.* in a recent review [4], advancing the context of PBR teams and collaborative research demands a transdisciplinary approach. Accordingly, a meticulous literature survey revealed the following review studies that benefit the area of research: fundamental methods to improve team performance [13]; relationships between constructs of IPO framework and the importance of non-invasive and unobtrusive methods in team studies [4]; methods to utilise autonomic physiological activities to measure team performance [47]; and technologies and methods to measure aggregates of team cognition

[60]. The survey realised that there are limited reviews that catered to multimodality aspects of team studies such as data fusion models and team-level composite indices. Additionally, there is a need for a technological catch up to list advancements over the last 10 years since Salas's review [13].

This calls for a survey that critically explores previous practices employed for analysing and evaluating team performance, to identify the necessary fundamental components. Specifically, to complement the continuous nature of spatial-temporal data from technologies capturing the physiological and behavioural activity of team members, there is a need for a review that identifies the following: (a) mathematical models to establish team-level objective measures of performance, (b) simulated tasks that have been used for team performance studies to avoid any pitfalls in the previous approach and cater to multiple PBR techniques, (c) hypotheses that relate constructs of IPO with team-level PBR measures, and (d) how contemporary multimodal research could be structured. Furthermore, considering that subjective measures are currently predominant for establishing some form of ground truth (example: situation awareness global assessment technique [78]), the literature must incorporate their adaptation in previous studies, in the light of transitioning towards PBR-based measures.

Several studies are observed in the literature that incorporates multiple modalities in the investigation of team-related hypothesis. A thorough investigation reveals that although most studies collect multiple modalities, only the direct relationship between team-level measure from a single modality and a team performance aspect [11, 21, 35, 45, 79, 80] has been explored. This is primarily due to the nature of data which lacks a consensus between modalities. There is a lack of studies that focuses on merging information from multiple sources.

Theoretically, the inclusion of additional modality has the potential to enhance confidence in the relationships between task performance and PBR-based indirect measures [44]. The promotion of multi-modalities demands for identification and establishment of complementing grounds between the modalities.

Team related research is active in social, academic-learning and work settings. There is potential knowledge transferability among the settings such that findings of one could apply to the other [4]. Relatively, however, the majority of literature are observed in the context of social interactions and academic collaborative learning which predominantly investigate interpersonal and autonomic physiology. In these settings, aspects of team cognitive processes that are vital for work-team settings are lightly addressed. As a

consequence, gaze [21,81], communication [5] and body-movements behaviours [66] that are linked with team cognitive processes have received less attention. Such limitations are prevalent even in recent studies on group contexts (excluding dyads), forcing the employment of video-taping of experiment session, which is laborious to process [82,83]. Besides, the majority of studies have focused on dyadic engagements. Researchers have further argued whether dyadic settings qualify as a group [84].

In addition, researchers employ off-the-shelf components to prepare necessary team experiments. For example, commercial multi-player games are used as STEs, PBRs are often collected via separate computer applications, and paper-based techniques or web-hosted questionnaires are employed to obtain subjective feedback. However, the coherency of integration is often either loosely considered or ignored in this domain of research. See section. 2.6.1 for a detailed discussion on the limitations of team task environments used in previous studies.

Efficient addressing of hypotheses demands inter-connectivity among research questions, collaborative tasks and PBR. For this purpose, IPO framework suggests the need for a cyclic process to prepare an effective team task environment [3,4]. Feedback from *outputs* serves as a basis for tuning dynamics of a task (i.e., the *inputs*) [85] such as complexity levels, the pace of participant and simulation agents, and length and duration of tasks. Such a process critically dictates the effectiveness of team hypotheses investigation.

Albeit, when employing PBR techniques, they must not confound nor interfere participants engagement with task objectives. For example, breath-sensor could be confounding [44], and the use of finger-tip based sensors pose restricted hand movements which can limit operations that concern mouse or keyboard button-pressing. Similarly, tasks that require projection on multiple visual displays may require constant neck movement interfering with desktop-based eye-tracking devices. In recent years, the non-invasive based trends are further advancing via the employment of multiple modalities to enhance the quality of team research.

Hence, there is reliance on a selection of PBRs that does not interfere with each other, yet that can capture a broader spectrum of complementing physiological and behavioural data. This calls for a novel experimental setup that balances integration of multiple modalities, simulated tasks and still maintains the comfortableness level of participants. The task should also be capable of comprehensively engaging more than two participants enabling compatibility with larger groups.

1.4.2 Research Motivations

Since Salas's review, in addition to the number of team-based studies, much has changed in the space of analysis framework (example, IPO [3]), mathematical models, technological advancements. The literature and the cyclic nature of IPO framework further suggests that the interdependence between the blocks of the framework are contemporary. For example, the demands of a task, the objectives and participants' KSA, the interactivity between team members and the outcomes are interdependent. In this light, the identified gaps and recent technological advancements are the primary motivations of this research.

Specifically, via addressing the gaps, this thesis can contribute to the community in the following ways: The first gap aspires to contribute a library on fundamental components for analysing and evaluating team performances. Via addressing this gap, the thesis establishes a solid foundation for contemporary experiment designs aiding in identifying suitable components in relation to distinct research hypotheses. Such a contribution also serves for the establishment of effective training programs.

Hence, this research is motivated to compile a rich literature establishing necessary resources for effective exploration of the influential factors of team performance.

Technological advancement such as gamification [86] and advancements in game engines [87], motivates the development of better ways to engage members as a team. Further, advancements in behavioural recording techniques; for example, eye-tracking [88] and body movement tracking offered by modern augmented-reality/virtual-reality technologies, physiological methods that offer enhanced sensors and wireless technologies [89,90], and most importantly, the multimodal studies.

Hence, this research is motivated towards the development of an experimental platform combining multi-party cooperative tasks (engaging more than two members) and methods to acquire multimodal PBR. Additionally, this seeks to explore multimodal index/indices for an indirect measure of team performance.

In general, empirical studies based on the principles of IPO team study framework suggests that relationships between members' physiological dynamics and task outcomes could enable predictive models that could prompt teams of any adverse outcomes [4]. Additionally, recent advancements in feature engineering, unsupervised clustering [91], machine learning and Artificial Neural Network (ANN), and the critical review on team-level measurements motivates this research to address the lack of complementing grounds among data from multiple PBR.

Hence, this study aims to explore the feasibility of employing PBR for modelling a NN architecture and predictive algorithms for indirect assessment of team performance.

1.4.3 Research Objectives

- Critically review the current practices to list non-invasive, unobtrusive and continuous PBR technologies, team-level objective measures of performance, simulated task settings and their pitfalls, and subjective feedback mechanism
- Moving beyond dyadic task settings and in line with the cyclic nature of IPO, this research aims to develop a computer-based simulated task to engage a minimum of three members in a team. This will be achieved using the Unity game engine [92].
- Achieve a state-of-the-art experimental platform by combining, the simulated task and four modes of PBR techniques (i.e., EEG, ECG, eye-tracking and communication), and a mechanism to obtain feedback from participants.
- Design of experiment that could efficiently engage participants. The compliance of the task with IPO enables observation of outcomes, which in turn helps in adjusting parameters of the task such as intensity and duration to achieve efficient engagement.
- Explore relationships between the PBR measurements and task performance indices to derive team-level measurements and predictive aids.

1.5 METHODOLOGY AND RESEARCH DESIGN

As will be established in Chapter. 5, ensuring compliance with the principles of IPO framework, an empirically viable transdisciplinary approach is adopted to oversee the objectives of this thesis. This allowed the development of a novel experimental setup involving a fresh concept of the task environment that can engage a minimum of three members in a team, PBR techniques that allow for dynamically acquiring measures from individual team members [4] and subjective ratings for establishing abstract ground truth were carefully chosen. Methodologically, irrespective of a top-down or bottom-up approach, these recordings are considered as lower-level components providing inputs only from an individuals' perspective [4]. As such, this thesis embarked on an exploration of quantitative relationships associating individuals physiological measures with influential factors [4]. Ultimately, the data contribute to the exploration of multimodal indices

that derive from cardiovascular, visual cognitive behaviour, central nervous system and voice, speech and linguistic features.

1.6 PUBLICATIONS

The findings of this research work have been published in several journals and conferences. The publications include three journal articles (one published and two under review) and four conference papers (two published and two under review).

Publication 1, journal article (Published). Veerabhadrappe, R., Ul Hassan, M., Zhang, J. and Bhatti, A., 2020. Compatibility evaluation of clustering algorithms for contemporary extracellular neural spike sorting. *Frontiers in systems neuroscience*, 14, p.34. <https://doi.org/10.3389/fnsys.2020.00034>

The journal *Frontiers in systems neuroscience* specialises in cognitive and computational neurosciences. It is one of the highest-ranked journals and maintains a quartile-1 status in this specialisation. The primary contribution of this study was to explore unsupervised clustering algorithms for processing time-series data. The contributions of this article can be observed in the preliminary exploration Chapter.

Publication 2, journal article (Under review). Veerabhadrappe, Rakesh, Imali T. Hettiarachchi, and Asim Bhatti. "Fundamental Requirements for Contemporary Performance Evaluation of Emergency-Response and Safety-Critical Teams" (2021). *IEEE Transactions on Human-Machine Systems*

Tech-archive online availability: https://www.techrxiv.org/articles/preprint/Fundamental_Requirements_for_Contemporary_Performance_Evaluation_of_Emergency-Response_and_Safety-Critical_Teams/15133923

The journal *IEEE Transactions on Human-Machine Systems* specialises in human factors and cognitive ergonomics, signal processing, computer science applications, data processing in human-computer interaction systems and organisation. It is one of the highest-ranked journals and maintains a quartile-1 status in this specialisation. The primary contribution of this study was to selectively review the fundamental concepts necessary for the contemporary experimental study of team performances. The contributions of this article can be observed in chapter-2 exploration Chapter.

Publication 3, journal article (Under Review). Veerabhadrappe, Rakesh, Imali T. Hettiarachchi, and Asim Bhatti. "A State-of-the-art Experimental Platform for Effective Analysis of Team

Processes" (2021). *IEEE Transactions on Systems, Man, and Cybernetics: Systems*

The journal *IEEE Transactions on Human-Machine Systems* specialises in computer science applications, data processing in human-computer interaction and electrical and electronic engineering. It is one of the highest-ranked journals and maintains a quartile-1 status in this specialisation. The primary contribution of this study was to demonstrate the capability of the experimental platform in accordance with IPO framework, participant engagement, participant satisfaction, human-human group experiment design and multimodal data acquisition. The contributions of this article can be observed in the chapter-3 experimental platform and team experiment design.

Publication 4, conference paper (Published)

Veerabhadrappe, R., Hettiarachchi, I.T. and Bhatti, A., 2021, April. Using Recurrence Quantification Analysis to Quantify the Physiological Synchrony in Dyadic ECG Data. In *2021 IEEE International Systems Conference (SysCon)* (pp. 1-8). IEEE.

Publication 5, conference paper (Under Review)

Veerabhadrappe, R., Hettiarachchi, I.T. and Bhatti, A., 2021, April. Gaze Convergence Based Collaborative Performance Prediction in a 3-Member Joint Activity Setting. In *2021 IEEE International Systems Conference (SysCon)* (pp. 1-8). IEEE.

Publication 6, conference paper (Under Review)

Veerabhadrappe, R., Hettiarachchi, I.T. and Bhatti, A., 2021, April. Using Eye-tracking To Investigate The Effect of Gaze Co-occurrence and Distribution on Collaborative Performance In *2021 IEEE International Systems Conference (SysCon)* (pp. 1-8). IEEE.

The *IEEE International Systems Conference (SysCon)* comprises of a range of systems integration, engineering and application concepts. These articles targeted the sensor integration and applications for team performance analysis. In particular, the contribution of publication 4 was a preliminary analysis to understand the effects of HRV to understand the concepts of physiological synchrony. Publication 5 and 6 describes empirical methods to observe effects of eye-tracking such as co-occurrence and distribution of gaze data in team performances. Further, outlines a method to employ deep neural networks for the prediction of team performance. Details of the findings and contributions are located in Chapter 6 of this thesis.

Publication 7, conference paper (Published)

Veerabhadrapa, R., Hettiarachchi, I.T., Algumaei, M. and Bhatti, A., 2021, October. A Deep Convolutional Neural Network Model for Classification of Emotions from Electroencephalography Data. In 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 1953-1958). IEEE.

The *IEEE International Conference on Systems, Man, and Cybernetics* addresses concepts in system science and engineering, human-machine systems, and cybernetics. This article targeted the Human Factors area of Human-Machine Systems. The primary goal of this article was a preliminary analysis is the use of a deep neural network algorithm to recognise human emotions using EEG. Details of the findings and contributions are located in Chapter 6 of this thesis.

1.7 THESIS STRUCTURE

This thesis comprises seven chapters, including this chapter. The remaining six chapters are organised as follows.

Chapter 2: Critical Review of Fundamental Components for PBR-based Team Studies is the first contribution of this thesis in the form of a theoretical study performing a detailed review of related literature. The review aims to collate information from studies that employ single or multiple PBRs in a team context. Previous attempts are critically examined to identify mathematical models, the methodological approach adopted and hypotheses investigated linking with the type of team performance and team-level aggregation of PBR data. The study establishes the necessary theory in line with principles of the IPO enabling applied frameworks for future multimodal research on team performance and team-based training paradigms.

Chapter 3: Preliminary Study-a represents the first of two preliminary studies conducted in this thesis via the investigation of SPC in academic computer supported collaborative learning (CSCL) using ECG data. The study explores RQA with various levels of HRV data as a non-linear approach to observe physiological compliance between collaborating participants. The study was used to gain exposure to team research and ECG data, processing and team-level representation of physiological measurements and experimental processes.

Chapter 4: Preliminary Study-b represents the second of two preliminary studies conducted in this thesis via the employment of artificial neural networks with EEG data for online classification of human emotions. Human emotions are primary contributors of interpersonal relationships that significantly impacts team performance. Employing EEG for detection of emotions using any classical classifiers requires pre-processing, filtering

and complex feature extraction processes, which are not suitable for online and real-time applications. This study explores the feasibility of utilising pre-processed and filtered EEG data with the customised architecture of a deep neural network. This study provided exposure to a different type of physiological data, and processing concepts based on artificial neural networks. This indirectly builds the knowledge to explore early prediction of team performances using human emotions.

Chapter 5: A Novel Multiparty Collaborative Experimental Framework is one of the major contributions of this thesis. Having established the necessary knowledge through the theoretical review and practical applications, this chapter demonstrates their integration to achieve a novel empirical framework. This chapter establishes the necessary methodology and methods for the exploration of the research questions. Inspired by previous team task environments and to augment beyond dyadic settings, a 3-member collaborative task is introduced in this chapter. The chapter introduces computer applications to integrate tasks with four modes of PBR techniques enabling multimodal based team studies and a mechanism to collect subjective feedback. The chapter outlines the parameters to design team task objectives that will be employed during the experiment. Experimental procedures and details regarding ethics approval are included in this chapter. One of the paramount focuses of this chapter is a detailed evaluation of the framework in line with the principles of IPO, and its capability to explore collaborative and team-based hypotheses.

Chapter 6: Analysis-1: The Impact of Gaze Co-occurrence and Distribution is the first of the three analyses performed with the novel data. This chapter presents an analysis of how co-occurring and distributed events contribute to performance. The Chapter outlines some important steps for pre-processing the gaze data to augment its usage in a team context. Novel methods to quantify these events are introduced in the chapter. The analysis presented here explores the possibility of identifying a suitable operating radius between team members and its relationship with performance.

Chapter 7: Analysis-2: Employing Gaze Convergence in Neural Networks for Team Performance Prediction is the second analysis of the novel data. Teams engaged in desktop-based collaborative problem solving has been increasing in recent years. The reliance on gaze-based measures on performance was established in the previous chapter. Translating such observations via meaningful representation using appropriate features can aid in modelling predictive aids. This chapter endeavoured to use gaze convergence indices to model a simple neural network algorithm to predict performance. Linear statistics of Euclidean distance between pairs of team members gaze is explored to quantify gaze convergence. Relationships between gaze convergence and performance were analysed and suitable features were selected for modelling the neural network.

Chapter 8: Analysis-3: Multimodal Index and Effects of Team Cognitive Activity is the third analysis and another novel contribution of this thesis. This chapter explored markers of cognitive activity and collaborative activities in the context of three-member task which is highly demanded in education, ER&SC domains. To this end, previous chapters explored relationships between single-mode PBR based measures and team performance. Continuous data from cardiovascular activity HRV, neural brain activity (EEG) and eye-movements (gaze fixations, co-occurrence, distribution and convergence) were investigated. The multimodal analysis presented in this chapter attempts to merge measures from cardiovascular and gaze activity to investigate the influence of cognitive load states from Autonomic nervous system (ANS) and visual-cognitive activity, respectively, on performance. The analysis introduces methods to derive team level indices based on compliance in cognitive load and gaze-based measures. First, the indices derived for each modality are individually analysed with relation to their task performance. Then, a multivariate regression model was implemented that combined the team level indices from both modalities resulting in a composite multimodal team level index (CM-TLX). The index achieves a significant correlation with task performances. Ultimately, demonstrating the possibility of employing CM-TLX as an objective measure of performance. In addition, team cognitive states and gaze-based measures that shows the effects of members continuous involvement is analysed.

Chapter 9: Conclusion, Limitations and Future Recommendation The chapter revisits the set objective and reflects some of the shortcomings and limitations of the current thesis. And, the recommendation provides details on how to overcome the limitations and what avenues are available for future exploration.

An *Appendix*, three appendices are provided explaining the background processing of ECG data, RQA and some pictorial evidence of configuring and hosting the Tank Battle Task. Appendices are placed after the bibliography section.

2

Critical Review of Fundamental Components for PBR-based Team Studies

2.1 INTRODUCTION

This chapter presents the first contribution of the thesis on synthesising a library of fundamental components in line with the constructs of IPO framework, suitable for contemporary and empirical research on team performance. A pool of studies primarily focusing on work-team settings and multimodal approaches in group context were targeted for the review. There were limited studies in this area, and hence, related works from academic collaborative working and social group settings were selected for reviewing. Overall, the goal was to identify and collate previous methods and technologies with examples of how they were employed in team settings, as well as the state of multimodality in the context of teams. This comprised of a detailed compilation of current trends in PBR technologies, qualitative and quantitative methods for analysing and evaluating team and collaborative performance, tasks and experiments, and subjective performances. Following this is a brief explanation that reflects team performance relationship with the underlying principles of inputs, processes and outputs of empirical team studies. Accordingly, this review briefly addresses the following:

- (a) analyses and procedures employed to process PBR data to represent at team-level,
- (b) how PBRs can foster a better deciphering of the relationship between team performances, constructs of IPO framework, statistical aggregations and subjective ratings,
- (c) finally, how future team research experimental setup can be shaped.

Insights and suggestions presented in this review can promote newer ideas, strengthen a better understanding of underlying team attributes, and foster robust multimodal approaches for effective evaluation of team performance.

Finally, the discussion highlights some of the gaps in this research domain and provides recommendation on potential future prospects that could better address the gaps.

2.2 MEASURING TEAM MEMBERS' ACTIVITY

The performance of teams is primarily dependent on individuals behaviours (e.g., skills that dictate the capacity of an individual to achieve desired task objectives) and team members' interaction dynamics (e.g., the degree of effective collaboration, coordination, and collective knowledge) [3, 4, 13, 20]. Carefully examining the interactions among team individuals could provide insight into the influential attributes such as affect, behavioural and cognitive states [55]. Such an approach can help understanding intra-team relationships as well as factors influencing at individual level that led to different levels of performance outcome [13, 93].

For understanding how PBR have contributed to team studies, this section briefly introduces different data acquisition methods observed in the literature. Cowley [44] mentions over ten possible sources to record the activity from individuals. However, techniques that are continuous, unobtrusive and non-confounding have been an integral part of team research studies. Focusing on the primary motive of this review, only methods that could have contributed to team performance related studies, in any closely related form are selected for the following discussion. The Table.2.1 briefly summarises all the techniques and their related measures discussed in this section.

2.2.1 Physiological Observations

The physiological methods are primarily concerned with the internal activity from the human nervous system. Figure. 2.1 summarises the organisation of the human nervous system, key technologies for recording the activity from individuals, and various levels of measurements that are potentially useful for performance analysis of teams.

The human nervous system is composed of the central nervous system (CNS) and the parasympathetic nervous system (PNS). The activity of CNS can be captured using EEG and Near-Infrared Spectroscopy (NIRS). Typically, wavelet or Fourier transformation could be used to decompose the temporal voltage information from EEG channels into six frequency components referred to as *delta* (1-3Hz), *theta* (4-7Hz), *alpha* (8-13Hz), *beta* (14-30Hz), *gamma* (31-50Hz) and *mu*. In addition, these frequency components are used to compute different measures of cognitive states such as engagement EEG_E [94] and workload EEG_{WL} [72, 94, 95] of individuals. Furthermore, the EEG data can also be processed to account for eyes' blink referred to Eye Blink Rate (EBR) which is an estimate of blinks per

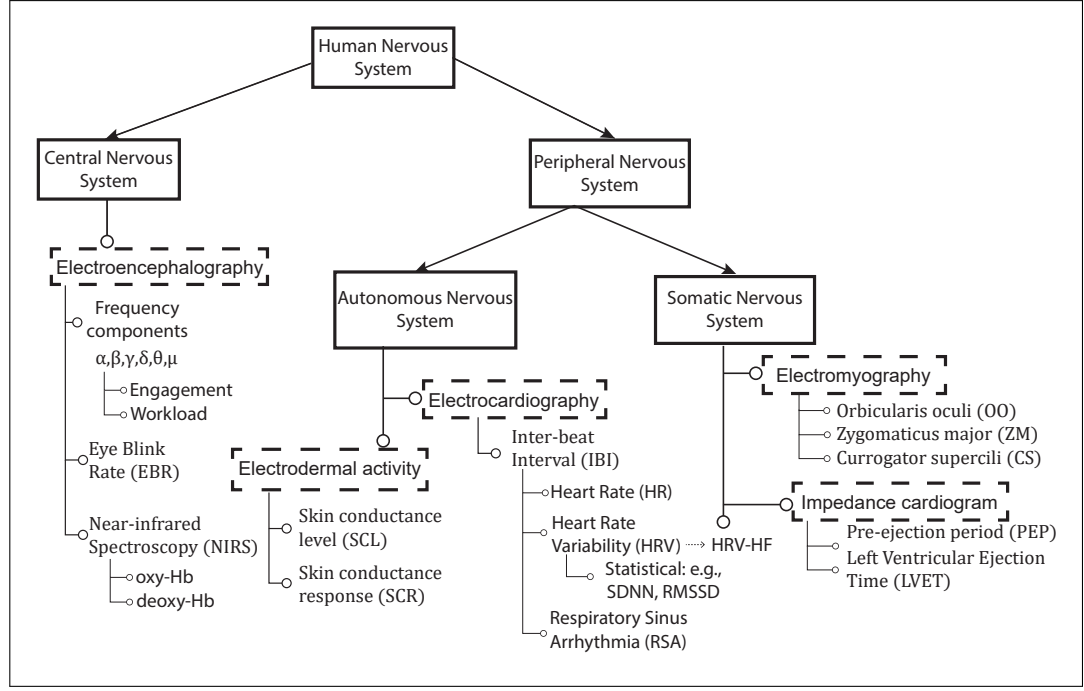


Figure 2.1: A summary of physiological systems (dashed-boxes) for observing internal activity from individuals. Also shown are the potential measures at individual-level that are of importance for team level analysis

minute [36]. From NIRS, levels of oxygen in haemoglobin in the form of oxy-Hb and deoxy-Hb can be recorded from individuals [53,65].

The activity of PNS can be further sub-categorised into ANS and sympathetic nervous system (SNS) [4]. ANS has attracted special attention in team context because it reflects the activities due to interpersonal relationship [89]. The activity in ANS can be captured using Electro dermal activity (EDA), ECG and respiration [4,35,67,80,96]. From EDA, measurements such as Skin conductance level (SCL) and Skin conductance response (SCR) (also referred to as galvanic skin response) are of interest in team studies [35,67,97]. From ECG, the intervals between R-peaks referred to as IBI are used to measure heart rate (HR), HRV, and respiratory sinus arrhythmia (RSA) of individuals [12]. In addition, HRV-based measures such as (a) the high frequency component of HRV between 0.15Hz-0.40Hz (HRV-HF) [67], and (b) Statistical measures such as standard deviation of normal-normal (SDNN) [80,96] and Root-Mean-Squared-Successive-Difference (RMSSD) [80], are of interest in team studies.

The activity of SNS can be observed in HRV-HF, facial Electromyography (EMG) and Impedance Cardiogram (ICG) [47]. From EMG, measurements such as orbicularis oculi (OO), zygomaticus major (ZM) and currogator supercili (CS) are of interest in team studies [4,54,69]. From ICG, measures such as pre-ejection period (PEP) and left ventricular ejection time (LVET) have been used in team context [47].

2.2.2 Behavioural Observations

The behaviours of individuals can be observed in techniques such as communication recordings [5,70,98], eye-gaze tracking [44,81] and body movements [63,66] which convey valuable information in analysing and evaluating team performance.

Communication records can convey greater insights about shared cognition, team situational awareness, shared mental models and team-level interaction processes [13]. The team communication can be processed using either verbal or non-verbal approach [16]. Verbal contents include processing of audio transcripts to extract speech and linguistic features. This has been achieved using LSA [45], IBM speech to text services [43,99] and Linguistic Inquiry and Word Count text analysis program [9]. Non-verbal content based measurements can include information about number of communication attempts such as button-clicks and acknowledgements to attempted communication [100], by extracting vocal tone frequencies [39,46] and team interaction instances of attempted problem-solutions, general meetings, debriefing and brainstorming [42]. Further, coordination was also measured from communication such that for every information shared by a team member towards a specific member of the team who returned an acknowledging negotiation. This process was summarised using a cooperation index which is a ratio of number of communication attempts to cooperated acknowledgements. Complete details of non-content based measures that could be extracted from communication is discussed by Cooke *et al.* [101].

Eye movements are low-level manifestations of visual-cognitive interactions. Gaze data enables observing visual interests of individuals providing valuable inferences underlying cognitive processes. Eye-gaze tracking offers a range of metrics such as fixations and saccades [102] to quantify the gaze behaviour of an individual. The measures can be used to observe the relationship about team members gaze behaviour and performance. Previously, gaze based measures such fixations around certain regions of interests, are employed to understand the impact of joint-activity on performance [63,81].

Finally, team members actions have been analysed using accelerometers [66] or video recordings [63,98,103]. Physical bodily movements such as postural sway, hand coordination and face and head movements are also used to evaluate join-activity in relation to performance [42,43,63,99].

2.3 TEAM-LEVEL MEASUREMENTS

In a team context, the PBR at individual-level must be collectively analysed [12]. Such analyses are important for deriving quantitative and qualitative measures that could be used to evaluate, analyse and rationally comprehend the influence of constructs of IPO in relation to performance. The following

metrics have been observed in the literature that could be used as indices of collective team response.

2.3.1 Compliance Metrics

Previously, compliance was quantified via neurodynamic synchrony (NS) which is a measure of similarity exhibited by team members' physiological measurements [4]. Note that, only physiological signals recorded from an identical source are deemed valid for deriving compliance metrics. Some of the compliance based measures are as follows:

Directional Agreement (DA)

represents a high-level estimate of compliance [12,67]. When comparing physiological signals from two members of a team, DA represents the number of points that directionally matched either in the positive cycle or on the negative cycle in relation to their previous points. DA is expressed as percentage of points directionally matching from an entire data set.

Signal Matching (SM)

It is considered as a good measure of compliance [12,67]. SM measures the area between two team members' physiological signals. Larger area indicates less similarity whereas smaller area indicated more similarity.

Weighted Coherence (WC)

It is a measure of similarity between two team members' physiological signals observed at specific frequency bands of interest [11,12,67].

Instantaneous derivative matching (IDM)

It represents a similarity measure between the curves of two team members' physiological signals [67]. The measure accounts for average number of differences between the derivatives observed at common reference points from the corresponding physiological signals [12].

Regression

Guastello demonstrated the employment of regression based models to measure the overall linear and nonlinear synchronisations between team members' autonomic arousals [97]. In this light, measures such as coefficient of determination, auto-regression weight and synchronisation weight were derived from linear auto-regression, and lyapunov exponent and dimension from auto-regression weight of the exponential series. Furthermore, method

such as cross-correlation was also used to measure the similarity between team members physiological signals [21,65,67,96]. Besides, in the presence of multiple channel data (e.g., EEG data with N-channels), regression based methods have been employed to account for neurodynamic linkages i.e., active linkages in the brain regions of cooperating individuals [36,72]. In this light, multivariate auto regression and partial directed coherence were used to derive the measurements, inter-connectivity density (ICD) and inter-areas links density (IALD) [36] also referred to as *strength density* and *clustering coefficient* [72], respectively.

Cross-recurrence quantification analysis (CRQA)

The CRQA is an alternative method to quantify nonlinear synchronisations between team members physiological or behavioural signals. It combines the concepts of RQA and recurrence plot (RP) [104] to be extended for team based studies. The diagonal, horizontal and vertical profiles, and the density of points in the recurrence plot, respectively, are used to observe nonlinear properties such as, determinism, laminarity, entropy and Rate of Recurrences. CRQA has been employed to observe the relationships such as joint activity and coordination between the team members [21,66,68,81,96]. Other versions of CRQA are multidimensional-RQA (Md-RQA) [43,66] also referred to as multivariate-RQA (Mv-RQA) [62] offer similar measurements but with an advantage of comparing multiple signals originating from two team individuals.

2.3.2 Temporal Neurodynamic Patterns of Teams

Physiological signals of team members have also been used to construct neurodynamic patterns. For instance, temporal measures of engagement EEG_E , workload EEG_{WL} and IBI from individual members of a team could be collectively represented to form team-level neurodynamic patterns [45,70,71,105]. Such patterns represent instantaneous states of the team's engagement-level, workload-level and cognitive-load. Quantification of such patterns could be achieved using the concept of Shannon's Entropy where sliding windows of appropriate length are used to account for evolution of teams neurodynamic states [70]. For example, if a team experienced a variety of states during a window then the resulting Entropy would be high and *vice versa*. Such approaches enable observation of temporal variations in Entropy levels. Furthermore, assuming team individuals represent a complex dynamic systems the temporal Entropy series also reveals attractor basins. In addition, the neurodynamic patterns were also used to construct a 1-second lag transition matrix [70,71]. The matrix revealed overall concentrations of different levels of teams engagement and workload.

2.3.3 Latent Semantic Analysis (LSA)

LSA can extract discriminating parameters from communication transcript such as semantic vector length and cosine-parameter that have been employed to observe differences between teams performance [45]. Semantic vector length is a measure of number of words either during a piece of discourse, phase of task, or during an entire task period. On the other hand, the LSA cosine-parameter measures the similarity between two transcripts or two pieces of discourse [106].

2.3.4 Epistemic Network Analysis (ENA)

ENA employs singular value decomposition at its kernel which is a type of unsupervised clustering. It constructs network such that nodes of the network represent key variables such as regions of interest and discussion topics, and the connection between nodes represent strength of interaction between the key variables. Such networks are used to analyse speech data, interaction between individuals in a group [107,108] and also to understand gaze coupling between members of a team. Epistemic modelling is also advantageous in processing multimodal signals [42].

2.4 RELATIONSHIPS BETWEEN TEAM PERFORMANCE AND CONSTRUCTS OF IPO

This section looks at how the analyses and techniques mentioned in the previous section transforms raw physiological and behavioural signals into team related dynamics. Also, the section briefly unfolds the relationships among team dynamics, performance and the constructs of IPO framework.

2.4.1 Team Performance and Team Inputs

The inputs to teams i.e., the composition of teams and the structuring of task conditions significantly impacts team performance. Evaluating the relationship between team composition and task aids in unravelling the physiological operating dynamics of superior performers [52]. For instance, the transition matrix constructed using temporal EEG_E and EEG_{WL} states of the team demonstrated that teams composed of experienced members had more concentrations in the region of higher levels of team engagement and workload [70,71]. Further, using hierarchical clustering algorithm on the LSA of cosine measured between all pairs of teams clearly clustered experienced teams and novice teams [70]. The sliding-window based entropy levels constructed from temporal EEG_E and EEG_{WL} demonstrated that experienced teams had relatively higher entropy levels indicating a

greater neurodynamic flexible characteristics [45,109]. A non-verbal team communication analysis by Cooke *et al.* [100] revealed that experienced teams exhibited a lower coordination to communication ratio.

The relationship between task structure and team performance could be used to identify conditions of task that impact performances. Indeed, LSA of team communication transcripts was employed to learn correlations between team performance scores and task conditions [45]. For example, vector lengths derived from LSA were used to identify the active phase of a task [45] such that, the phase of a task represented by smaller vector lengths was prominent during high active phase relative to other phases of the task.

In addition, the ICD indices could also be used to identify harder phases of a task [36,72]. Any elevations in the regions of the brain's activity due to higher task complexity could be observed via the IALD analysis [36,72]. Besides, the ICD also demonstrated a significantly higher neurodynamic linkage in the real teams during an elevated cooperation phase of the task relative to pseudo teams (synthetically generated from real teams' data) [36].

2.4.2 Team Performance and Team Processes

Understanding the relationships between performance and processes of teams can help explain the impact of qualities and emerging states on team performance.

In the context of cooperation and coordination, Wallot *et al.*'s work [66] enabled to observe the relationship between information sharing, behavioural coordination and performance. Accordingly, during an industrial production line simulation task, the homogeneity (egalitarian) condition, yielded better design of products relative to other conditions. Further, the quality of the product design inversely correlated with the determinism and laminarity (from MdrQA) between team individuals' hand movements (quantified using accelerometer readings). Similarly, FNIRS measure oxy-hb demonstrated that wavelet coherence between mixed gender pairs has significantly higher correlations under cooperating phases of a task [53].

The relationship between coordination and task performance could also be analysed using compliance metrics. As such, in a Multi-Attribute Task Battery (MATB) environment, a lower root mean square error (RMSE) level during tracking was used as indicator of superior performance [67]. Montague *et al.* reports a significant positive correlation between the DA of team members SCR and performance [67]. Further, larger mean absolute deviation indicated poor resource management which correlated with lower SM indices (indicator of higher NS).

Compliance metrics are also employed to analyse the relationship between coordination and task conditions as a function of performance. The WC of HRV-HF captured the drop in performance under increased task

complexity conditions [67] such that the WC resulting in higher averages under hard and low reliability condition relative to normal condition. In another custom developed tracking task, team performance was measured based on completion time, collision damage, RMS error (deviation from the centre path) as a direct objective of team coordination [110]. Henning *et al.* observed that WC of EDA and HR was related with drop in completion time. The WC of EDA, HR and respiration was also related with drop in RMS error. The CC of HR also demonstrated significant relationship with drop in completion time and RMS error. In their following study, CC of HR was able to predict performance changes as a function RMS error with significant negative correlations [111].

The neurodynamic linkages measured using ICD and IALD [36,72] also corresponded to cooperation, and the corresponding analysis enabled observation of the relationship between levels of cooperation and performance.

In the context of joint activity, Tolston *et al.* [81] observed that the recurrence between the team members' gaze activity negatively correlated with performance scores. Video analysis has also been used to observe coordination as a function of team composition [63]. A video analysis of laparoscopic surgery task revealed that teams composed of elite surgeons exhibited low de-synchronisations relative to the other composition of teams.

Communication has been employed to understand the relationship between aggregates of team cognition and performance [5,101]. Relationship between team performance and experience was explored by Cooke *et al.* [5]. Accordingly, it was observed that team performance improved as teams gained experience, which was resultant of improvements in team interaction (process behaviours such as coordination and communication). This indicates that the development of shared mental models improved performance [5]. Further, it was observed that the teams which resumed task after shorter breaks performed better than teams that took longer breaks between tasks. This decline performance was observed due to loss of skill by individuals rather than team interaction (interaction-based measures) [101]. This shows that degradation of skill was significant compared to shared mental models.

2.4.3 Team Performance Outcome and Its Implications

Understanding the implications of team performance outcomes is essential for effective evaluation of underlying team inputs and processes. Previously, this approach has been used to understand shared decision making due to performance feedback. In a production line task involving crafting of origami sail boats [68,69], correlation between NS and performance developments were reported. During a negative performance development, teams took a call for changing the strategy. When teams were presented with new strategy and post adoption of the strategy, the NS between ZM

and SCR of team members were observed to be low. On the other hand, NS of SCR increased in teams which opted to continue without adopting the presented strategy.

This context of analysis can also be applied in the prediction of critical performance outcomes. As such, in a programming scenario, where performance of teams was decided via pass or fail outcome of the task, SCR was able to distinguish the effect of outcome in advance. During a passed outcome, the dyad demonstrated significantly higher SCR levels compared to failed outcome [35].

Indeed, this approach was employed to observe the effect of feedback on cooperation. Using NIRS measures of oxy-Hb level, Balconi *et al.* observed a drop in inter-brain synchronicity following an induced negative feedback [65]. Also, Cheng *et al.* observed a positive correlation between task performance and synchronicity [53]. Such activities in specific regions of brain can be used as markers of cooperation [53,65,112].

The performance outcome can also be used in objective identification of skill composition or characteristics of teams. It was observed that as novice teams gained experience via subsequent repeating runs through the task, there was a gradual rise in the mean entropy [70] which could be used as indicators of a teams performance evolution.

2.5 MULTIMODALITY IN THE CONTEXT OF TEAMS

The team inputs, processes, states, qualities and outcomes are co-dependent entities. In an ecological context, combination of these entities interplay defining the natural behaviour of humans [4]. In this light, multiple-PBRs have been previously employed to study human emotions to understand collaboration, cooperation and coordination [35,36,46,79]. Table. 2.2 provides a summary of methods contributing to multimodality in team context.

Individual emotions can be associated with multiple interaction processes which can be effectively evaluated using multimodality. For example, the 'threat' is an emotional state in humans that can elevate vasoconstriction and reduce in cardiac functioning which is directly responsible for inducing stress [46]. The heightened stress levels can diminish cognitive functioning (the ability to think and response) causing ineffective communication. This phenomenon was observed and validated by Neubauer *et al.* using speech and cardiac data [46]. Vocal-tone variations from speech data and total peripheral resistance from ECG measurements were used in their definition of a multimodal index [46]. The index was used to understand the impact of stress leading to communication response.

Analogously, the investigation of *cross-level effects* by Gorman *et al.* observed simultaneous developmental changes in cognitive and behavioural functioning due to team interactions [45]. EEG and communication data were analysed to understand entrainment between team members as they

progressed through a task. Such analysis enables better task related performance evaluation that result from team-level entrainment, for example, in understanding leader driven theories [73, 113, 114] where speech from the leader induces motivation that affects team-processes eventually reflecting in enhanced performance [45].

Multimodal approaches have also been favoured in enhancing team training outcomes and assessments [41, 45]. For example, concurrent data from eye-tracking and EEG were collectively analysed to observe the impact of cohesiveness and expertise in a cyber-security setting [41]. Collective analysis using additional modality also provide advantage by potentially substituting for any gaps that might be suffered through a single mode [46, 115]. Toppi *et al.* observed that HR and EBR, which are regulated by CNS activity demonstrated strong correlations with different task conditions [36]. The correlations also significantly varied depending on operator roles and responsibilities. Such analyses can be employed to understand relationship between team inputs i.e., team composition and task conditions.

Multimodal research studies enable identification of substitute sources that offer cost-effective and better alternative solutions. For example, by comparing concurrent measures of EDA and ECG, Ahonen *et al.* was able to observe that SCR (from EDA) offered better and faster method to convey information originating from ANS activation when compared with HRV features (from ECG) [35].

Multimodality aids to explore the impact of certain stimuli on team outcomes. For example, (a) Murray *et al.* explored feature vectors comprising of features from both linguistic and acoustic speech modulations to observe relationship with task performance [48], and (b) Neubauer *et al.* explored the relationship between acoustic voice features and physiological states [46]. In the context of studies mentioned in (a) and (b), multimodal analysis can be used to link both the physiological and behavioural responses to better explain task outcomes.

Overall, it can be established that multimodal approaches have the capability to promote significant and distinct observations, where a single modality would not have comprehensively explained the scenario [42, 43].

2.5.1 Data Fusion in Multimodal Studies

In addition to the employment of multiple modalities, a collective analytical index which mathematically and rationally unifies data from multiple modalities is very important. Only a few studies have attempted data fusion from multiple modalities or collective analysis in some form [42, 43, 45, 47, 68, 79].

The method demonstrated by Eloy *et al.* employs measurements from three different modalities [43]. In this method, for any member the three

PBR signals were first averaged. Then, the averaged signals from all team members were analysed using MdrQA to estimate unified percentage recurrence. Echeverria *et al.* employs a multimodal matrix mechanism to generate an interactivity network to observe teams' performances at various scenarios [42]. Monster *et al.*'s method employs joint activity measurements derived from ZM and SCR of all the teams [68], where principal component analysis was later employed to observe distinct groups. Finally, the procedure employed by Gorman *et al.* analyses each modality individually, however, discussions link observations from EEG and speech data analysis, explaining cross-level effects [45].

2.6 SIMULATION ENVIRONMENTS

Simulation environment also referred to as STE are integral components for contextualising team research studies [13, 85, 116]. They offer control, cost-effective and safer platforms for team individuals to interact and share a visual space. Such platforms enable easy modifications of task conditions, regenerate and adjust task elements according to required activity [117] thereby, enabling enhanced training outcomes and economical team training methods [116]. Ideally, a platform that can emulate natural settings is a preferred choice.

In this light, team studies have employed environments ranging from a simple board game [118] to multiplayer computer games (e.g., Bomberman, Ibb & Obb, Tetris and Hedgewars) [12, 21, 50, 54, 119–121] and role play [122] which offer choices to observe coordination and cooperation. To observe compliance in behavioural responses against team performance, physical tasks such as origami construction [68, 69], lego building [66], or even head movements [43] have been observed in the literature.

In recent years, simulated environments have attracted attention in Defence and civil sectors as an efficient and effective mechanism for training [123]. STEs that emulate real world scenarios offer the perfect testing grounds for teams. This includes ecological settings such as flight simulator [36, 38] and hands on surgical setup [42, 63, 124]. Tactical environments that offer command and control operations are typical choices for team cognition study. Example of tactical environments include submarine piloting and navigation [45, 70], unmanned aerial vehicle-STE [100, 101], research environment for supervisory control of heterogeneous unmanned vehicles [81], distributed dynamic decision-making (DDD) [125], NeoCITIES [126], and many similar forms can be observed in [116, 127, 128].

In addition, Live virtual constructive (LVC) environments are fully integrated simulators specifically employed by Defence Forces in many countries [117, 129–131]. It integrates the concepts of augmented reality and virtual reality to provide near realistic air-combat and land-combat training experience to defence personnel.

Non-tactical approaches offering controlled environments to study cognition included choices such as MATB [67, 72], process control synthetic environments [47], customised cooperative task [53, 112], and CSCL [35, 79, 80, 82, 96].

2.6.1 Current Limitations of Synthetic Environments

Recently, Lawson *et al.* reviewed 57 commercially available and licenced computerised task environments for assessment of military team [116]. The review, however, did not focus on the convenience of employing such environments with PBR devices. It is very important to ensure that the PBR do not interfere participants' commitment with the task. For example, finger-tip connectable devices may interfere when using mouse clicking or keyboard operation. Similarly, tasks projected on multiple screens result in constant neck movement interfering with desktop based eye-tracking devices.

Additional limitations could also be observed in team task objectives offered by current process. For example, the multi-player cooperative games were primarily intended for entertainment purposes, hence, may not effectively meet research intentions. In team research employing physical tasks, the participation could be impacted from physical competency. Besides, there are limited study suggesting the effectiveness of these task environments in the light of team-based participant engagement.

The modification and tuning process offered by existing task environments are arduous and limited. And, not all STEs support logging user-input data such as mouse locations and button-press events [17]. Such data are important to derive other measures of performance such as time management, resource management, cognitive application and task-engagement. The lack of which impedes comprehending any correlations between task dynamics and PBRs. Where possible, requires the knowledge of complex programming skills limiting its employment [116]. Furthermore, some STEs such as counter-strike are so complex to handle that they require participants to master gadget usage such as joystick and keyboard where several keys must be simultaneously comprehended [132]. In conclusion, existing STEs have their own limitations, hindering the participation from diverse demographic groups due to factors including lack of tuning flexibility, dependency of domain specific knowledge, gadget mastering, and limits on physical demands. These limitations make them incompatible with the principles of IPO framework.

Further, socio-dyadic tasks such as doctor-patient, mother-child and coupled tasks [67, 133] inherently promote complementary behaviours [4], hence are limited to certain social aspects of engagement and beyond the scope of collaborative problem solving domains.

2.7 SUBJECTIVE RATINGS

Subjective rating are questionnaires which are used in studies to obtain feedback from participants [54,111]. Such an approach has been traditionally used where individuals rate their experience concerning task difficulty, emotion, and personality-traits [69,120,134]. NASA Task Load Index (NASA-TLX) is one of the commonly employed subjective rating techniques to understand the task workload experienced by members of a team [72,124,127]. Later, a team-level index can be obtained by globally averaging the scores from individual's subjective ratings.

Further, in the light of "human emotions", the Dundee Stress State Questionnaire [9], the Trait Anxiety-Taylor Manifest Anxiety Scale [97], the State-Trait Anger Expression Inventory-2 [16,134], the Profile of Mood States [135], the Positive and Negative Affect Scale [54], the Self-Assessment Manikin (SAM) [54,136]. In addition, questionnaires such as Emotional Intelligence [97], Coping flexibility [97] and Empathy Assessment Index [97] were found in the literature that specifically reflect human emotions.

To assess the competitiveness and cooperativeness in computer game environments or social gaming platforms, questionnaires such as the Competitive and Cooperative Presence in Gaming Questionnaire [119], the Social Presence in Gaming Questionnaire [54], and the checklist for trust and team affect questionnaire [128] have been employed.

In the light of team cognition, task specific questionnaires, situational Awareness Rating Technique [127,137], Crew Awareness Rating Technique [137,138] and Situation Awareness Global Assessment Technique [137] were observed in the literature.

2.8 DISCUSSION AND FUTURE RECOMMENDATIONS

Salas emphasised the need for modern technologies to promote extensive and good quality team research [13]. Current advancements in wearable sensor technologies could boost multimodal-based team research [85]. Several low-cost, low-powered and smart devices are currently flooding the consumer market enabling better integration of off-shelf wearable and portable PB sensors and technologies into experiments [139]. Observing this trend, it can be realised that PBR technologies have been very well advancing, relatively however, the processing and analysing of the data has slowly progressed, more so, methods for collective team related analysis. Some of the constraints that have been constantly highlighted by researchers are (a) computationally demand (b) expensive setup [48] (c) identification of right combination of recording systems [35] and (d) effective fusion of data from multiple modalities are sparsely observed in the literature. Further, following are some of the key areas that studies concerning team performance analysis could focus in future research.

2.8.1 Team Cognition Based Performance Evaluation

Team cognition has been highly valued among work teams [13, 140], and interactive dynamics strongly reflects these aspects. This review reveals that team cognition has been mostly studied in terms of interactivity using communication and eye-tracking [5, 140]. Literature mentions physiological measures such as cardiac [80, 85, 120] as markers for analysing team cognitive activity, however, markers informing team cognition aggregates such as shared mental models needs empirical reinforcements. For example, comparing studies [46] and [5], where both employ some form of measurements derived from communication. Towards generalisation and in natural scenarios, the ITC study [5] would also benefit with the inclusion of additional modes such that, any interpersonal and emotional factors that may have affected or promoted team performance [65] could be accounted via an additional modality.

Literature mentions three primary means to observe cognition in teams: interactions, team knowledge and mental models, and team situational awareness (SA) [60, 85]. In this light, real-time concurrent measures from EEG, eye-tracking and communication transcripts were emphasised in the literature [60]. Further, Cooke *et al.* states that studies must link performance outcomes with emergent states such as cognition rather than linking to physiological measures at individual level [5]. Thus, multimodality could be key for efficient deciphering of attributes underlying team performance.

2.8.2 Analysis Beyond Dyads

Majority of the team studies observed in the literature are mostly concerned with dyads [67, 133, 140] where complementary emergent states are inherent. In contrast, limited studies have explored teams comprising of a minimum three persons or larger. Amongst these, physiological studies investigating neurodynamic organisation in six-member teams [45, 70], NS analysis in four-member teams [12] and joint activity computed for three-member teams were observed in the team performance analysis literature [68]. Further, in some studies though teams larger than dyads were considered, the PC and joint activity was computed between any two members (hereafter referred to as sub-dyads) of a team [12, 68]. Later, the physiological interactivity that was measured across all possible sub-dyads of a team was globally averaged to compute a team-level interactivity score.

Future team experiments could engage three or more members in team to move team-based analysis beyond dyads. Ideally, studies must focus on concurrently analysing multiple team members data that could be represented using a unified index such as the neurodynamic organisation [45, 70], or interactivity index using multimodal features by Eloy *et al.* [43]. Further, methods employed in sub-dyadic approach [12, 68], could have better im-

pact via more empirical evidence. Such advancements are necessary for standardising and generalising any innovative methods across team study community. Addressing such challenges fosters newer analytical methods for efficient deciphering of attributes underlying team performance.

2.8.3 Need for Computation Improvement

Transferability of computation methods across the broad ramifications of team performance can enable newer data fusion methods. Further, empirical evidence concerning data processing methods are relatively less observed in the literature. For example, nonlinear dynamics in team analyses were observed to be less compared to the use of linear statistics. In addition, the employment of artificial neural networks is still basic. Only one study was observed to use artificial neural network based future outcome prediction [38]. Our review also shows that temporal prediction capabilities of recurrent neural networks have not been observed in team settings. There is a need for maximising the sample count using bootstrapping [80] and pseudo team data generation [36] from small collection. Such techniques can encourage the employment and development of deep neural networks based team research studies.

2.8.4 Experimental setup

The experimental setup should support continuous acquisition of PBRs, and concurrently from all the members of team. There is a need for task environment that can constantly engage members of team as opposed to the nature of distributed task environment [97]. Such environments can induce interactions enabling evaluation of team performance in relation to adaptability, emergence of leadership and coordination [141].

Task environments that can engage general public improve sample count as well enable exploring a diverse pool of skills. The environment must offer engagement of more than two participants to advance team research beyond coupled dynamics. The stimulus or testing tasks built into the environment must be strong enough to test the targeted individual behaviours and performances [66]. The environment must offer compactness, portability, easy configurability and cross-domain compliant [123]. It must be capable of testing cooperation and coordination, and provide qualitative or quantitative feedback of team performance. Such feedback mechanism reflects behaviour related to cognitive activity [65]. In addition, task-event data that can temporally unfold dynamics of the task such as mouse clicks, movements, key-element usage and key-logs, must be accessible from the environment [17].

This review revealed that majority of team studies have experimented with physiological systems compared to behavioural systems. In this light, future research could focus on a combination of physiological and behavioural system targeting greater flexibility in team studies. Specifically, the employment of eye-tracking and communication is encouraged to better explain team cognitive aggregates such as share mental models and team situational awareness [60]. Finally, well targeted subjective ratings must be included as a means to collect direct feedback from participants.

Table 2.1: Summary physiological measures contributing to team performance analysis

Individual level	Processed	Team-level	team-size	Performance
EEG				
<i>alpha, beta, gamma, theta and delta</i> frequency spectrum	integer levels indicating EEG_E and EEG_{WL} all team members were combined to form patterns	Sliding window based Entropy was estimated from temporal EEG_E patterns	six	Discriminant function analysis revealed more experienced teams had higher mean entropy compared to the less experienced teams [45]
		1-second lag transition matrix to detect attractor basins from EEG_E	six	Attractor basins of experienced teams were different from inexperienced teams [70]
	EEG_E	From temporal EEG_E , sliding window based Entropy was constructed	six	Inexperienced teams were characterised by more irregularities in the Entropy compared to experienced teams [70]
	EEG_{WL}	From temporal EEG_{WL} , sliding window based Entropy was constructed	dyadic	mean of the Entropy correlated with mistakes [71]
		averaged inter-connectivity map, strength density (ICD), clustering coefficient (IALD)	dyadic	Inter connectivity map showed strong connections in α and θ bands during harder phase of task which also correlated with reduced performance, and contrasting results were observed in easy conditions. This phenomenon was also complemented by higher values of ICD and IALD in specific region of brain [72]
		ICD, IALD	dyadic	ICD of α and θ bands was significantly higher in real teams during high workload phase of the task compared to pseudo generated teams and IALD showed the region of brain activity [36]

Individual level	Processed	Team-level	Team-size	Performance
FNIRS	oxy-hb, deoxy-hb	Pearson CC used to establish inter-brain synchronicity (IBS) between participants brain regions	dyadic	IBS decreased post negative feedback [65]
	oxy-hb	wavelet coherence used to calculate average IBS	dyadic	change in coherence correlated with cooperation phase of the task [53], and was significant in mixed gender groups
ECG				
IBI		WC (High frequency and low frequency components) [67]	dyadic	
		cross-correlation, cross-fuzzy entropy (CFE)	dyadic	High score correlated with low cross-correlation between team members IBI, and CFE negatively correlated with coupling [21]
	mean-IBI and from HRV MSD, RSA and log-RSA	SM, DA, IDM and CC	4	correlation was significantly high for high performing teams compared to low performing teams [12]. Trend-level significant difference was observed with DA
ECG and ICG combined analysis	log-RSA association with PEP and LVET, respectively to obtain a composite score	cross-correlation to derive team level index	dyadic	From ordinary least squares, composite scores demonstrated significant associations to team error as well as distinctly associated with individual operators [47]
Actions				
accelerometer voltage recordings		MdRQA %Determinism and %Laminarity	dyadic	During the egalitarian condition, lower values corresponded with better looking designs [66]
video		de-synchronisation	dyadic	de-synchronisations were less among elite teams who also took shorter duration to complete the task [63]

Individual level	Processed	Team-level	Team-size	Performance
EDA				
EDA		R^2 , lyapunov exponents, synchronisation weight	dyadic	Pearson-r showed correlation between synchronisation weight and false alarm rates [97]
	SCR	DA	dyadic	DA was related with higher performance levels [67]
	SCR	SM	dyadic	SM was related performance in resource management [67]
	SCR	WC (High frequency and low frequency components) [67]	dyadic	
	SCR	MWE	dyadic	MWE significantly distinguished fail and pass conditions [35]
				higher SCR levels shared relation with failed condition visually distinguishing from a passed condition [35]
	SCL and SCR			SCL and NSCR levels tended to be higher in positive outcome groups [135]
EMG				
OO, ZM			dyadic	Have been employed to observe participants responses post task outcome [54]
SM,CS			dyadic	To observe emotional changes due to chance to employ performance enhancing routine [69]
Eye-tracking				
Eye-gaze	fixations	CRQA, %RR	dyadic	%RR decreased as team performance increased [81]

Individual level	Processed	Team-level		Team-size	Performance
		Communication			
speech transcript	LSA	semantic length	vector	six	
	LSA cosine	hierarchical clustering	clus-	six	More experienced teams clustered from less experience teams [45]
	Non-content based	CCR index		three	Experienced teams showed significantly higher CCR index compared to inexperienced teams [100]

Table 2.2: Summary multimodal measures contributing to team performance analysis

Individual level	Processed	Team-level	team-size	Performance
	Interaction: idle, construction and execution; Speech: silence, controller speaking and contributor speaking; body movement: low and moderate	Single mode: Only interaction measures, dual mode: All combinations of Interaction and Speech measurements, multimode: Combination of measurements from all three individual modes	3	Only specific combination of measurements demonstrated significant correlation with task scores [99]
Speech features and ECG	Speech: fundamental frequency and peak slope, ECG: cardiac output and total peripheral resistance		2	Only relation to stress, and no relation to team performance [46]
EDA, ECG, Respiration, EMG			2	Cooperation versus competition [50]
Facial expression, ECG, Linguistic Inquiry	Facial expression: joy, anger, fear, positivity and confusion; ECG: HRV, Speech: word count, sadness, general, insights and cognitive processing	Facial expression and Speech: mean counts and HRV: RMSSD	2	Impact of emotions [9]

Individual level	Processed	Team-level	team-size	Performance
EDA, ECG, Respiration, EMG, EEG			2	Cooperation and collaboration [120]
audio, video, ECG, EDA			2	Annotated multimodal data for analysing collaboration versus affective states [136]
Speech, body movement and EDA	Signals from individual mode was z-score normalized averaged across participants.	MdRQA was used to estimate the %recurrence	2	No significance between recurrence and task score [43]
EEG, Eye-tracking			2	No collective analysis observed. Expert teams were able to quickly identify areas of interest compared to novice teams by approximately 3 times longer [41]
ECG, EEG, Respiration and GSR	features were extracted for each modality using statistical and signal processing approach		2	Classifiers trained with bench-mark data were used to classify each participants modality features. Multimodal and individual classification was performed to observe the cognitive state of participant [38]. Team of 2 members participated together in flight simulation task, however, there was no team level analysis.
gaze and speech	gaze measures were used to compute cognitive load and windowed-entropy, and speech audio: four different prosody tones	proportional similarity between gaze-entropy and cognitive load of the participants	2	LSTM to predict learning gain and post-test score. Input data comprised of gaze and audio provided best prediction of post-test score [39]
physical position, speech and interaction, EDA, actions	For every participant, 5 location data, binary coded presence or absence of speech and listening, respectively and EDA peaks was acquired	Each modality form every participant, was compiled into multimodal matrix	2-3	task specific analysis which majority depended on the patient's needs [42]

2.9 SUMMARY

Literature concerning performance analysis of ER&SC teams using physiological and behavioural approach are not exclusive, however, are critical for many team-based research. Furthermore, the literature is significant for employment of multimodality in ER&SC team research. The literature pool suggests that the advancement of multimodality in general demands for a strong multidisciplinary approach. As such, this review briefly summarised single modality ER&SC team performance studies and methods employed for data processing and how contemporary experiments could be shaped. The need for dedicated STE and sensor integration realises that team research can benefit from the involvement of broader research communities. Such developments can promote advanced nonlinear neurodynamic analysis of teams and newer mathematical models such as data fusion techniques. Overall, contributing to enriching the literature pool, empirical evidence, and subjective performance quantification and analysis of ER&SC teams.

The next two chapters builds on the concepts presented in this chapter to establish a groundwork in the form of preliminary studies. It explores PBR devices, data formats and pre-processing, and evaluates the suitability of incorporating in the current design. Additionally, it provides a brief outlook of identifying participants pool to establish teams, experiment setting, procedures to collect team data and some insights of potential neural network applications. Specifically, chapter 3 explores the paramount and critical team related concept of socio-physiological compliance using cardio-vascular raw ECG data. As mentioned in the above sections, HRV is one of important measures in team settings, and hence, the preliminary study-1.

3

Preliminary Study-a: Using ECG

3.1 INTRODUCTION

SPC is highly regarded in team settings and has attracted significant interest from science of teams researchers [11]. Over a considerable period of involvement in an activity, the members of a team develop certain qualities such as cooperation, cognition and shared knowledge [5]. Often, controlled collaborating activities are employed in team research to gauge such interactive qualities of teams. Synthetic environments hosting collaborative problem solving and group working on computers [79,80] are emphasised in professional courses. Empirical team research employs such environments to test, analyse and validate team related hypotheses [35,80].

Previously, Ahonen *et al.* conducted a study to analyse patterns of SPC [80] in a dyadic team context using ECG recordings. The study aimed to address that, under the context of a shared cognitive task, the degree of synchrony will serve to establish metrics that can gauge collaboration. In this light, linear measures of HRV was adopted to observe and analyse SPC by evaluating physiological interconnection between members of collaborating teams. Ahonen *et al.* observed that SDNN of IBI was effective in demonstrating the significant difference between real pairs and pseudo pairs (carefully generated from real pairs data).

Further, it has been established that ECG is mainly influenced by the autonomic nervous system which in turn is regulated by complex interconnection of sympathetic and parasympathetic nervous systems [142,143]. This regulation is observed to be dynamic with activations being either complementary or co-related. Both mental [46,143,144] and physical stress [142,145], respectively can cause variations in the vasoconstriction and vasodilation processes, which can be captured using ECG. Analysis of HRV is a quick, sensitive and replicable procedure for assessing cardiovascular autonomic activity.

In this light, literature shows that linear analysis techniques such as time and frequency domain parameters have been employed to assess the physiological compliance in teams [11,36,80,110]. However, research has been encouraged to explore nonlinear options as an alternative solution

to process ECG data [142, 146, 147]. For example, Gonzalez *et al.* explores the concept of RQA [142] to extract HRV features, to observe the effects of physical demands on heart rate. Accordingly, recent literature suggests that nonlinear measurements such as RQA has gained a significant momentum in the measurement of physical stress experienced by individuals using the ECG recordings.

Thus, the study presented in this chapter focuses on non-linear analysis of HRV to assess team collaboration. Due to the non-stationarities or changes of state of the HRV signal, RQA is a well suited non-linear analysis technique for HRV. Additionally, the parameter *recurrence* is a basic attribute observed in many dynamical systems, including regulatory cardiac response. The corresponding quantitative analysis can measure repeating events of a given state of the system. Although, nonlinear options to process HRV is growing in popularity, however, are primarily motivated towards physical health of the patients, and less concerned with the mental processing demands of individuals [148]. Previous employment of concepts of RQA on team physiological or behavioural responses is observed to be sparse, with some works attempting via the Cross-RQA analysis [21, 68, 81] or the Multivariate-RQA analysis [62].

The current study presents an analysis employing RQA in a team setting using the HRV data derived from ECG recordings. From individuals, RQA is employed to extract nonlinear features of HRV, and the physiological synchrony between cognitively engaging participants is quantified via Pearson's correlation coefficient. Most importantly, the study explores the effectiveness of these nonlinear HRV features as alternate options to assess SPC. In terms of capabilities to distinguish collaborating pairs', the RQA measures; recurrence rate, laminarity, determinism and entropy, were observed to be robust in short-term window lengths and equally sensitive compared to linear features of HRV.

3.2 MATERIALS

To facilitate our investigations, the data previously presented in studies by Ahonen *et al.* [80, 149] is employed. The data is available to download from <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0159178>. The corresponding study was approved by the ethical review of Hospital District of Helsinki and Uusimaa, Finland, adopting the guidelines from the Declaration of Helsinki for human experiments.

3.2.1 Participants

Originally, forty-six novice programming students had contributed to the study. The ECG data from eight participants were discarded due to artefacts.

The data from 38 participants comprised of sixteen females and twenty-two males with three having dominant left hand participated in this study. There was one participant aged over thirty-four, nine participants were aged between twenty-four and twenty-nine, and the remaining twenty-eight were aged between eighteen and twenty-three.

3.2.2 Experimental Setup and Procedure

Initially, experimenters briefed all the participants and obtained a written consent from them. Next, every participant was attached with ECG acquisition sensors such that, the electrodes were positioned on the lower left rib-cage and on right coracoid process. Unobtrusive sensors, eMotion Faros 180° by Mega Electronics Ltd were employed to record ECG data. The clocks of the recording device were synchronised with the experimenter's laptop providing master timing for measurements. The acquisition of physiological measurements started immediately after sensors were in contact with the participants, and ended only when the sensors were removed.

All pairs initially watched a 7-minute baseline video, and then proceeded to attempt the task. A classroom based programming task from their curriculum was adopted for the study. Cooperation between the participants was necessary to achieve the objectives of the task. One workstation was allotted to each pair, where individual members switched between 'programmer' role, the participant who performed the main programming task and 'navigator' role, the participant who assisted the programmer. Duration of the experiment lasted 90-minutes with roles being switched at every seventh minute.

3.3 METHOD

3.3.1 Data pre-processing

The ECG data was originally recorded at 500 Hz sample frequency as shown in Figure 3.1. The data was pre-processed using colibri R-package [150] and R-peaks were extracted, which in turn was used to construct time series R-R interval representing IBI. Detailed steps to process R-peaks from ECG data is provided in Appendix-A. As such, the R-peaks were first detected from the ECG data using the process described by Daskalov and Christov in [151] as shown in Figure 3.1. The difference between two successive R-peaks represents an IBI value. Successive differences between the R-peaks represent the time-series IBI as shown in Figure 3.2.

The procedure was cross-verified using the pan-tompkins process for extracting IBI series and observed no significant difference [144, 145, 152].

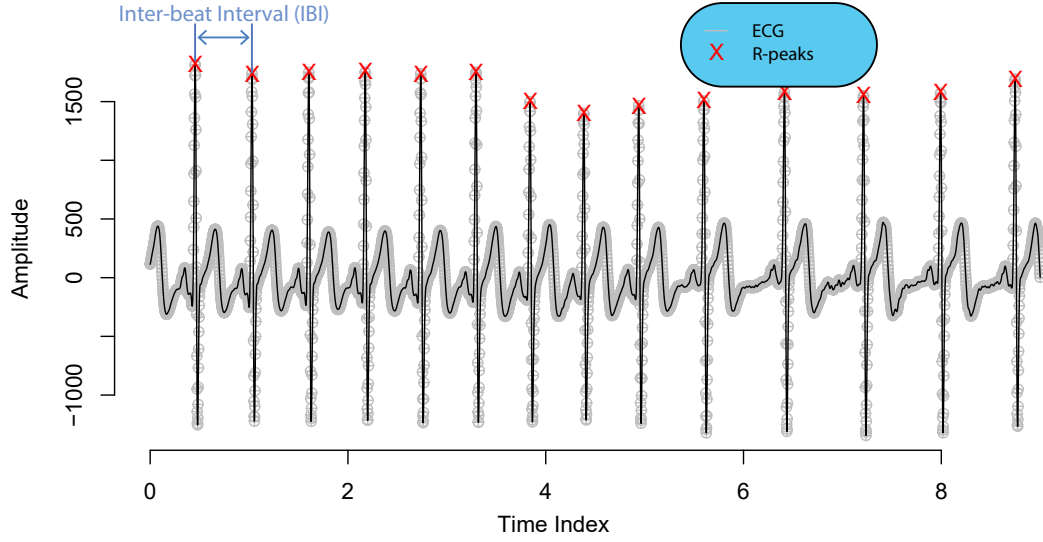


Figure 3.1: Example showing a short extract of ECG data, R-peaks and IBI

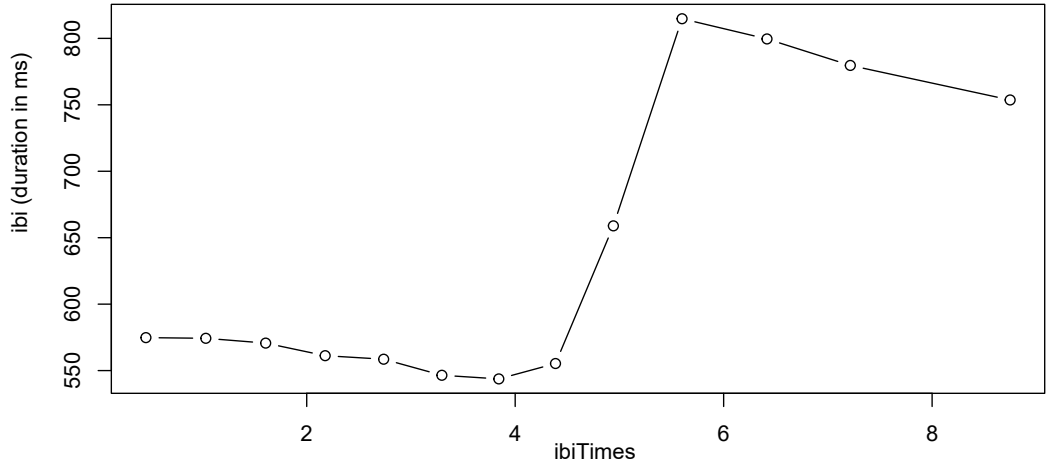


Figure 3.2: Example of time-series IBI estimated from ECG

For uniformity in comparison, IBI intervals derived using colibri R-package was adopted for the rest of the analysis.

The remaining of the analysis was performed using Matlab. The extracted IBI time-series is inherently unequally time-stamped (Figure 3.2). Since, RQA requires data timestamps to be evenly spaced, the IBI time series of each participant was first linearly interpolated [153] to four times the amount of original samples. Then, IBI was linearly extrapolated with time stamps evenly spaced at 1-second interval, which will be referred to from here on as $\bar{IBI}$. This is performed to avoid any length dispute that may arise when estimating correlation between two participants windowed RQA measures.

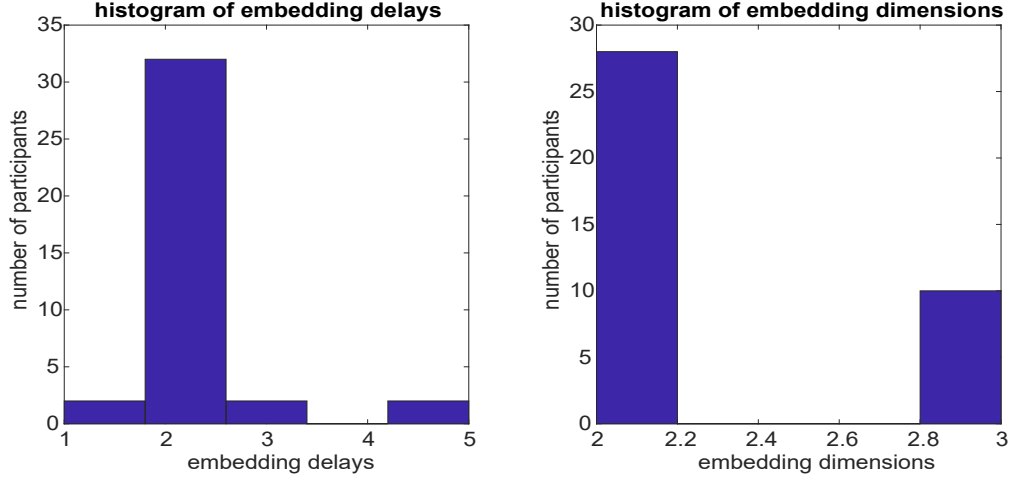


Figure 3.3: Histogram of observed embedding delay and dimensions, respectively, for all 38 participants considered in the study

3.3.2 Recurrence Quantification Analysis

The RQA processing is twofold; first, RP is derived from the IBI time-series data, and second, the quantification metrics are extracted. From the RP some of the commonly used horizontal, vertical and diagonal profile-based measures are adopted in our study:

- recurrence rate (Rec), estimate of points revisiting a phase space
- determinism (D), estimate of points forming continuous diagonal lines
- entropy (H), Shannon's entropy from the diagonal lines
- laminarity (L), estimate of points forming continuous vertical and horizontal lines, respectively
- maximum diagonal line (L_{max}), the longest diagonal line observed in the RP

For each participant, five RQA measures Rec , D , H , L_{max} and L , were extracted. Because the RP is symmetric meaning $X_{m,n} \equiv X_{n,m}$, the measures are estimated only from the upper triangle of the RP. Complete details of estimating RP and examples of quantification process is provided in Appendix-B.

Recurrence Plot

The first step towards RQA is to generate recurrence plot (RP) from the one-second interval spaced IBI data, $\tilde{IBI}$. The RP enhances the chances of observing any recurring and deterministic properties in the data, which otherwise would not be visible [104]. The RP procedure mentioned in

[104] was adopted in our analysis. Accordingly, a symmetric matrix X was constructed as shown below

$$X_{m,n}(\varepsilon) = \Theta(\varepsilon - \|\vec{I\tilde{B}I}_m - \vec{I\tilde{B}I}_n\|), \quad m, n = 1, 2, \dots, N \quad (3.1)$$

where N represents the number of samples in time series $I\tilde{B}I$, ε is the minimum radius between any two consecutive points, $\Theta(\cdot)$ is a Heaviside function such that, when $a < 0$, $\Theta(a) = 0$ and $a \geq 0$, $\Theta(a) = 1$ and $\|\cdot\|$ is the Euclidean distance. The CRQA toolbox for Matlab from Marwan *et al.* was employed to generate RPs [104]. To generate RP, the cross recurrence plot function from the toolbox requires appropriate selection of three parameters; embedding delay (minimum delay), embedding dimension and the threshold radius (ε).

Estimating Embedding Parameters

The procedure mentioned in [154] is adopted to estimate minimum embedding delay and embedding dimension. First, the delay parameter was calculated using the concept of average mutual information, and then, the dimension parameter was calculated as explained in [142, 154]. For comparison, histograms of embedding delays and dimensions, respectively, observed across all participants can be seen in Figure 3.3. It is clear that majority of the participants had the embedding delay and embedding dimension estimated as 2, which was adopted for further analysis. Although, some articles suggested the employment of embedding dimension as 10 [155], however, other articles recommend employing embedding dimension greater than the dimension of the data under consideration [156]. Embedding dimension of 10 were tried however, since short-windowed analyses were adopted, there were higher chances of encountering sparse spaces in RP where no result was produced by the tool.

Estimating Threshold Radius (ε)

The final parameter to construct an appropriate RP is to select the threshold radius (ε). The method proposed by Zbilut *et al.* [157] was adopted. Accordingly, Rec was calculated for a range of radius between 0.01, 0.02, ..., 0.1, 0.2, ..., 1, 2, ..., 10, and for all short-term window sizes adopted in the study. A plot of thresholds vs recurrence rates was constructed using a log scale as shown in Figure 3.4. From the plot, the condition *foot-after-noise* was located by observing the point after which there was gradual and linear rise in the Rec . All window sizes satisfied the criteria, however, for the window size of 60, valid Rec only starts at the threshold 0.7 where a steady rise was observed (indicated by dashed line in the Figure 3.4). Thus, to accommodate the smallest window size of 60, a threshold

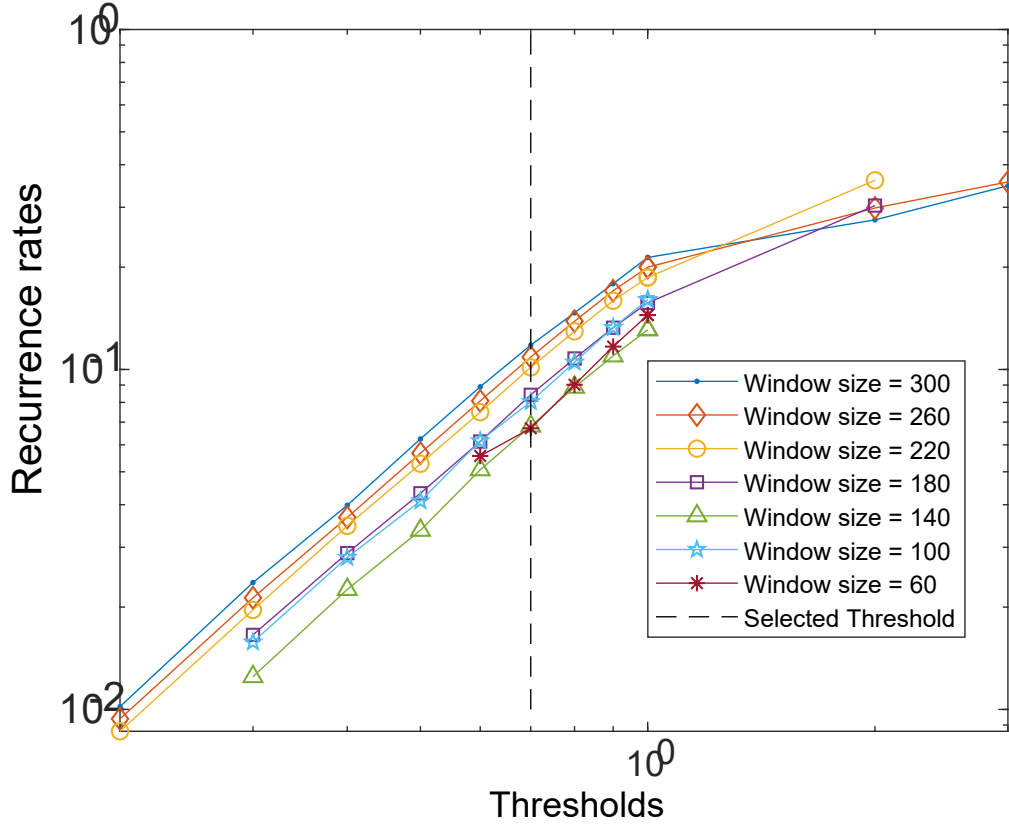


Figure 3.4: A plot of Threshold vs min recurrence observed across participants data set for varying window sizes. The dashed line shows the threshold selected for the generating RP, which satisfies all recommended criteria (refer to section 3.3.2 for details) for all window sizes

ε of 0.7 was considered for subsequent analysis. Further, the corresponding Rec of 0.067 satisfied the suggested nominal recurrence rate of 0.01 or 1%.

3.3.3 Physiological Synchrony

To be compliant for comparison with the results of originals study [80], the PS is estimated using the concept of Pearson's correlation coefficient. Further, it is also a well established measure representing SPC [12, 21]. From the real pairs, the data was randomly mixed, and pseudo pairs were synthetically generated. It was ensured that every pseudo pair was made up of complementing programmer-navigator pair and synchronised in time. Again, to ensure consistency in the comparison, same participant ID's that formed pseudo pairs in the original study was adopted. The correlation was evaluated for all real and pseudo pairs, respectively.

3.3.4 Permutation-based significance testing

To investigate the hypothesis that the mean correlations of real pairs estimated using nonlinear measures will be significantly different from the mean correlation of pseudo pairs, the permutation-based significance testing framework mentioned in [158] was adopted, similar to test concept mentioned in the original study. Correlations estimated using the RQA measures was considered as the statistics under evaluation. The correlations computed for pseudo pairs was used to generate 10000 pseudo sets, and the mean correlations were estimated for all 10000 pseudo sets. The mean correlations of real pairs and pseudo pairs, respectively was used to compute the significance probability p-value and their corresponding confidence intervals.

The process is repeated for all window sizes and across all five test statistics (RQA measures). Holm-bonferroni adjustments to the p-value was used to eliminate any Type-I and Type-II errors [159].

3.4 RESULTS AND DISCUSSION

The present research aimed to investigate whether nonlinear dynamic measures of HRV can offer alternate options to linear HRV measures in assessing the SPC under a dyadic collaborating task. One of the commonly employed and previously established nonlinear analysis method, RQA and its measures, Rec , D , H , L and L_{max} were adopted. To facilitate the investigation, SPC measures derived from the RQA of HRV were compared with the measures mentioned in the original study.

In the original study by Ahonen *et al.* [80], ECG data was collected from participants engaged in a classroom environment task based on computer supported collaborative work (CSCW). Three linear measures; mean (μ -HR), root mean square of standard deviation (RMSSD) and SDNN were used to establish measurements of an individuals' physiological responses. The synchronicity between the physiological response of members of a dyad was measured using the concept of correlation coefficient. The study aimed to address that, under the context of a shared cognitive task, the degree of synchrony will serve to establish metrics that can gauge collaboration. To support the research, pseudo pairs were carefully generated. The study reported that only SDNN was effective in demonstrating the significant difference between real pairs and pseudo pairs as an indicator of collaboration.

In accordance with the original study, correlation coefficient was estimated using the RQA measures for window sizes, 60, 100, 140, 180, 220, 260 and 300. For visual comparison, example of correlation coefficients for the window size 300 across the RQA measures is shown in Figure 3.5. The diagonals in the figure indicated correlation observed for real pairs, and

elsewhere indicated the correlation for pseudo pairs. Any blank spaces in the figure indicate that no pairing was formed as participants belonged to different sessions. This ensured that the study was conducted only on legitimate and valid pairs who synchronised in time and location.

The results of the permutation test for selected window sizes across the RQA measures is summarised in the Table 3.1. From the table it can be observed that, except for L_{max} , all of the RQA measures considered in the context significantly passed the hypothesis. These measures were observed to be significantly greater in the real pairs under 95% confidence intervals. Further, it shows that the RQA metric Rec was relatively more sensitive, and the RQA metric H was observed to be relatively less sensitive.

From the results presented, it was observed that the RQA measures; Rec , D , H and L offered alternate features which were equally sensitive to pairwise physiological changes observed in HRV relative to its corresponding linear measures. The measures were used to assess the variability and complexity of the IBI time-series. The physiological synchrony measured using Pearson's correlation coefficient of collaborating pairs was observed to be significantly stronger compared to synchrony observed between non-collaborating pairs. Further, a one-way analysis of variance (ANOVA) test was performed to observe if there was any statistical difference between the performance of time-domain HRV measure SDNN and RQA measures. Figure. 3.6 shows the result of ANOVA, where it is evident that there is no statistical difference between SDNN and the RQA measures; Rec , D , H and L . This also confirms the presence of sensitive nonlinear features in HRV dynamics under the short-term window sizes between 60-300. These dynamics were sensitive to cognitive challenges faced by the participants working in a team of two. Overall, it can be established that nonlinear dynamic measures of HRV offer parameters that are sensitive to distinguish collaboration via the concept of PS.

In a previous study by Sharif *et al.* [148] RQA was employed in the analyses of team ECG data to distinguish relaxed vs stressed teams, respectively. The study however, reported contrasting results where L_{max} demonstrated significant results relative to other RQA measures. Further critically analysing the results, SDNN in the original study had suffered to produce comprehensive significance when estimated for window size of 220. Contrastingly, nonlinear measures, Rec and L did not suffer from any marginal or borderline significance. However, the parameter H , failed to produce significance for window size 60. Although, not in the border, D and H often were less significant under window sizes 180 and 260. Coincidentally, with L_{max} not satisfying the significance, and D and H demonstrating less significance on couple of instances, it can be established that the diagonal profiles of RP could be relatively less sensitive. On the contrary, measures Rec and L which depended on density and vertical-horizontal profiles of RP, respectively offer relatively higher sensitive measures. In accordance

with the reliability, comparing linear features vs nonlinear features of HRV, the later offered 4 sensitive measures compared to only one by the former approach. Because, the RQA measures provided significant correlation in all window sizes, they can be considered more reliable and short-term robust parameters for measuring SPC.

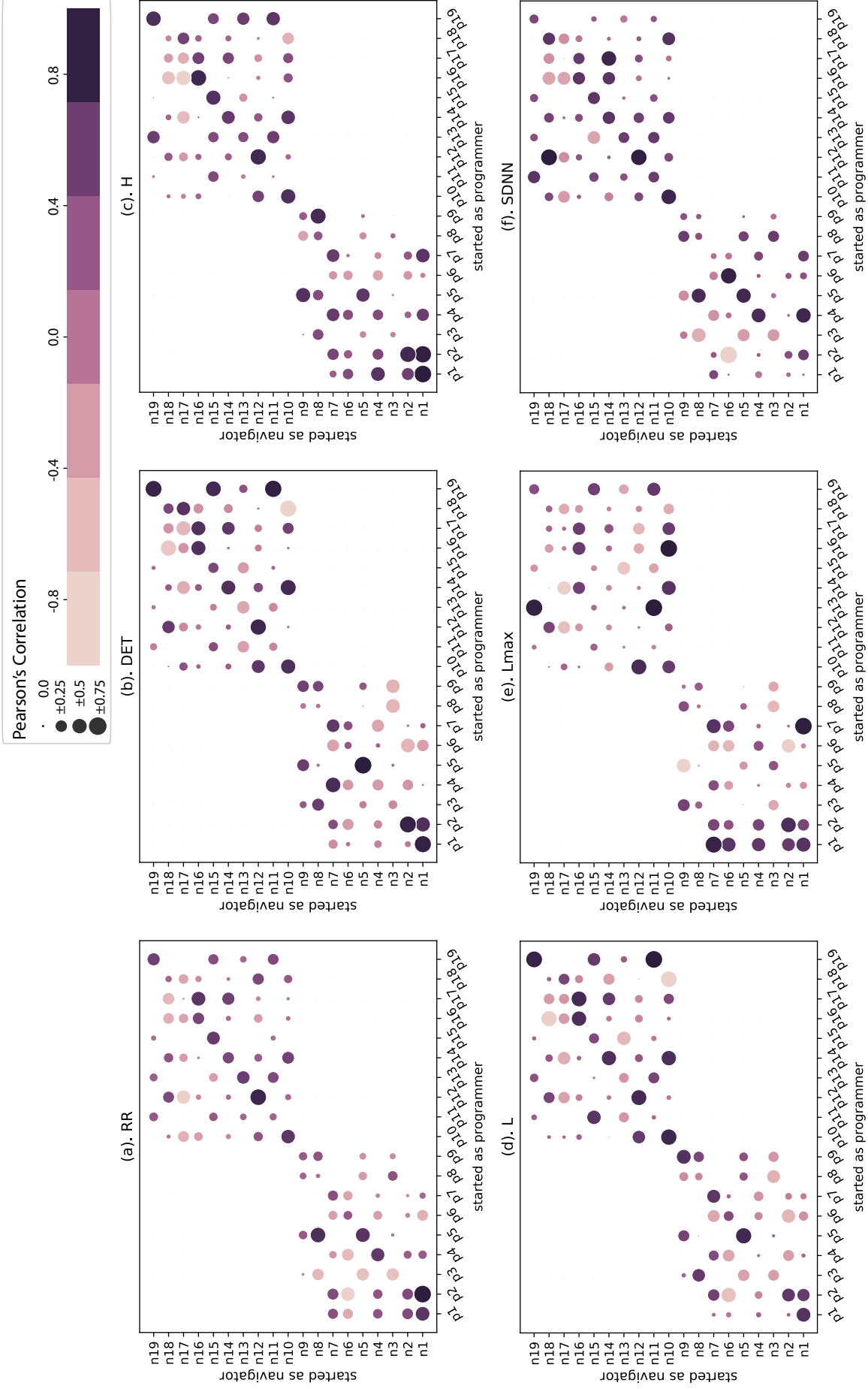


Figure 3.5: Plot of Pearson's correlation of all valid pairs for window size of 300 seconds to demonstrate difference between every RQA measure considered in the study. Each team consists of a p_i & n_i , indicating a programmer and a navigator, respectively. The diagonals ($i = j$) in plots represent true pairs and elsewhere ($i \neq j$) are the pseudo pairs. It should be noted that the experiment was originally conducted across two sessions. To accommodate this, pseudo pairs were formed only for the participants whose programming task, time and space synchronised. The x-axis in the plot indicate participants who started as programmers and y-axis indicate participants who started as navigators. For comparison with time-domain HRV measures, correlation observed for SDNN is also included in the subplot (f).

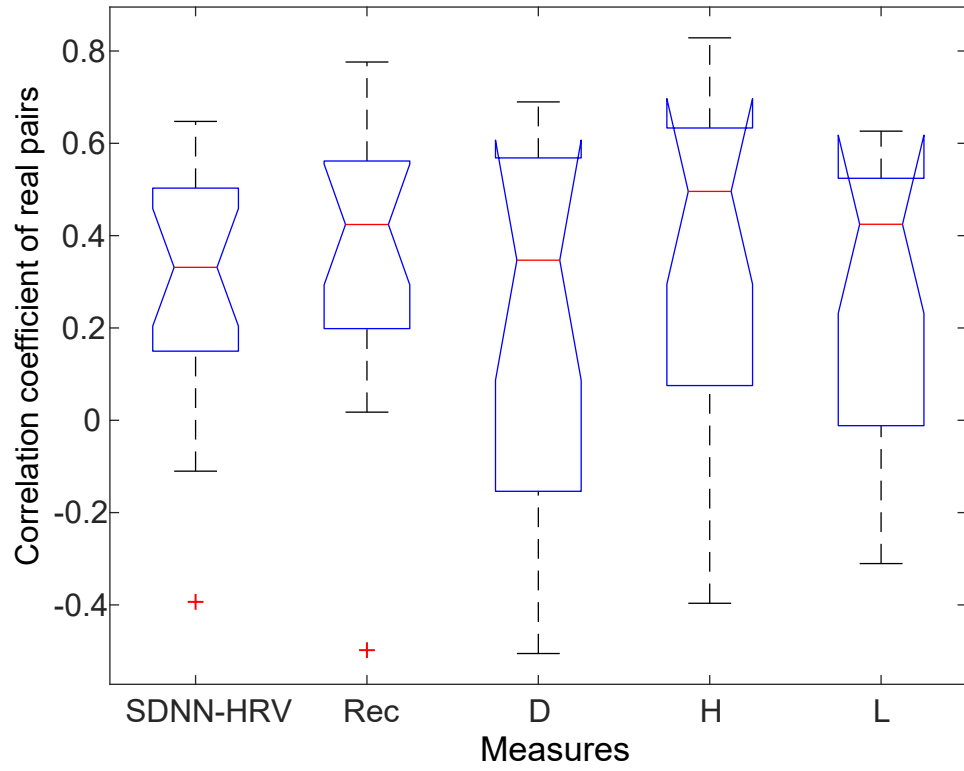


Figure 3.6: Comparison of correlation coefficient produced by the linear measure SDNN-HRV and nonlinear measures; *Rec*, *D*, *H* and *L*, respectively for real pairs using a window size of 300 seconds

Table 3.1: Observed significance (holm-bonferroni adjusted p-values) for selected short-term window sizes estimated across the RQA measures; *Rec*, *D*, *H*, *L* and *L_{max}* obtained by comparing 95% confidence intervals of pseudo pairs with mean correlations of real pairs. The observed level of significances are * for $p < 0.05$, ** for $p < 0.01$ and *** for $p < 0.001$

Window sizes (in seconds)	RQA measure	Pseudo pairs confidence intervals		Real pairs Mean correlation	Adjusted p-value
		2.50%	97.50%		
60	<i>Rec</i>	0.041	0.17	0.192	0.018*
	<i>D</i>	-0.05	0.079	0.108	0.018*
	<i>H</i>	-0.015	0.11	0.106	0.065
	<i>L</i>	-0.027	0.092	0.11	0.018*
	<i>L_{max}</i>	-0.026	0.092	0.047	0.292
100	<i>Rec</i>	0.026	0.195	0.238	0.008**
	<i>D</i>	-0.064	0.102	0.135	0.013*
	<i>H</i>	0.003	0.16	0.177	0.016*
	<i>L</i>	-0.049	0.114	0.149	0.012*
	<i>L_{max}</i>	-0.025	0.125	0.058	0.391
140	<i>Rec</i>	0.025	0.24	0.292	0.01*
	<i>D</i>	-0.072	0.137	0.162	0.027*
	<i>H</i>	0.027	0.227	0.244	0.027*
	<i>L</i>	-0.061	0.155	0.194	0.017*
	<i>L_{max}</i>	-0.032	0.142	0.051	0.517
180	<i>Rec</i>	0.018	0.251	0.304	0.009**
	<i>D</i>	-0.072	0.165	0.177	0.044*
	<i>H</i>	0.037	0.265	0.272	0.044*
	<i>L</i>	-0.056	0.176	0.229	0.009*
	<i>L_{max}</i>	-0.05	0.144	0.034	0.58
220	<i>Rec</i>	0.005	0.262	0.32	0.005**
	<i>D</i>	-0.083	0.177	0.204	0.028*
	<i>H</i>	0.036	0.274	0.286	0.029*
	<i>L</i>	-0.078	0.172	0.249	0.004**
	<i>L_{max}</i>	-0.046	0.164	0.083	0.302
260	<i>Rec</i>	-0.022	0.254	0.317	0.008**
	<i>D</i>	-0.077	0.197	0.223	0.029*
	<i>H</i>	0.04	0.299	0.307	0.041*
	<i>L</i>	-0.064	0.205	0.263	0.008**
	<i>L_{max}</i>	-0.025	0.199	0.062	0.646
300	<i>Rec</i>	-0.053	0.254	0.357	0***
	<i>D</i>	-0.102	0.198	0.236	0.019*
	<i>H</i>	0.031	0.325	0.35	0.025*
	<i>L</i>	-0.091	0.19	0.255	0.008**
	<i>L_{max}</i>	-0.047	0.212	0.096	0.419

3.5 SUMMARY

The main objective of the study was to investigate whether RQA measures provide significant indications of collaborative activity during a cognitive challenging task. The results and analysis presented in the article comprehensively support our investigation. It was clear that physiological synchrony between real pairs were observed to be significantly different from that of pseudo pairs. Compared to time-domain HRV measures, the dynamics of the HRV captured via RQA measures offer an attractive alternative to SPC analysis during team studies.

Having established the concept of socio-physiological compliance from team physiological data. The following Chapter explores EEG data in the paradigm of an import affect aspect of individuals i.e., emotions. As mentioned in the Chapter. 2, EEG is one of the popular physiological data in team settings. Most importantly, one of the objectives of this thesis to achieve neural network based predictive algorithm. Thus, the following chapter collectively addresses the concepts of EEG, human emotions and ANN.

4

Preliminary Study-b: Using EEG

4.1 INTRODUCTION

Human emotion is a collective reflection of psychophysiological and behavioural processes [160]. Investigating the influence of emotions has been a focal point of various research domain such as individual performance, performances in work-team settings [97], couple dynamics [160], monitoring cognitive states [161–163], human-machine teaming and Brain Computer Interface (BCI) studies [164–166]. Identifying markers or patterns pertaining to emotions can certainly benefit such studies.

The ability of EEG to capture accurate responses concurrently from multiple regions of the brain has been a primary encouragement for emotions-based human performance and BCI research [44, 167]. However, the low signal to noise ratio and higher dimensions of EEG signals, respectively, objective identification of emotions is computationally intensive. Thus, affective computing research community has been very active in this domain with a purpose to better identify patterns and improving classification accuracies [168]. Such advancements critically define the widespread usage and broader applications surrounding human emotions. However, majority of the studies are limited to laboratory environment. With the increased discussion about autonomous systems, and human machine teaming [161–163], trust and emotions are becoming widely discussed topics. In order to aim for real-world applications in these fast evolving domains, there is an increased need for real-time monitoring of emotions in practical applications. While wearable devices are providing a means of real-time access to the data, advanced computational models are required to handle the real-time processing for emotion recognition [169, 170].

A typical study employing EEG signals involves decomposing time domain channel data into frequency bands [171, 172] or wavelet components [173], followed by application of a suitable feature extraction methods and training a classifier model. Conventionally, EEG signals are decomposed into 5 frequency bands α (8-13Hz), β (14-30Hz), δ (1-3Hz), θ (4-7Hz) and γ (31-50Hz) [44]. Feature extraction methods such as Pearson Correlation Coefficient [174, 175], differential entropy [176], differential asymmetry,

rational asymmetry, power spectral density [177] and differential causality [178] have been proposed to train a classifier model [167, 179]. Literature also mentions the use of dimension reduction methods such as principal component analysis (PCA) [177], kolmogrov-smirnov tests (ks-test) [91, 176], fisher projections [178] and minimal-redundancy-maximal-relevance [177] where applicable. In emotion classification domain, classifiers such as SVM, KNN) have been employed for classifying the features [91, 97, 166, 176]. The advancements in artificial intelligence computation has additionally fostered neural network algorithms such as dense belief networks [178], dynamic graph convolutional neural networks [167] and dense neural network (DNN) with sparse autoencoder (SAE) [174]. From the literature, it is observed that artificial intelligence based algorithms relatively have achieved superior results [164, 167, 168, 174, 180, 181].

The artificial intelligence based models are very popular among multimedia data or structured data where explicit labelling is available and favouring real-time applications [178]. Unlike video or image classification, the low signal to noise ratio of EEG signals require several transformations before a formal application [178]. Such processes may not be favouring a real-time classification problem. Further, it can be observed that, the conventional dimension reduction techniques such as PCA is not applicable because during online processing the test data is unknown [91]. The dimension reduction techniques, for example, PCA generally operates by taking the entire data set, learns correlations within the dataset and then provides best features that can improve classification accuracy [91, 182]. Such procedure performs better when entire data pool is available.

To overcome the critical needs of real-time emotion classification from multi-channel EEG signals, the article proposes a MC-DCNN. The article investigates whether such a neural network can extract appropriate features with minimal data processing, and still achieve good classification accuracies and in real-time. Publicly accessible dataset SJTU Emotion EEG Database (SEED) [183] is employed in the study. From literature [167, 168, 174, 177, 178] employing SEED dataset, two methods are proposed to generate training and test data such as subject independent and subject dependent. The objective is to utilise labelled samples from previous users to model a classifier for predicting emotions of new users [184, 185]. From the literature [168, 180], it is observed that the subject independent method has been challenging resulting in relatively lower accuracies or very high standard deviations. MC-DCNN is evaluated using all three types of data splitting methods mentioned above, and the results are compared with SVM and KNN.

4.2 MATERIALS AND METHODS

4.2.1 Experiment and Dataset

The SEED dataset comprises of data collected from a cohort of 15 participants (7 Males and 8 Females). Each participant performed three experiments on three different days. During each experiment, a participant completed 15 trials with every trial lasting approximately 4-minutes. A total of 15 Chinese movie clips with 5 clips in each of neutral, positive, or negative emotion category were randomly played during the experiment. EEG data was collected at a sampling frequency of 1000Hz using a 62-channel cap with electrode placements following 10-20 international system as shown in the Figure. 4.1. The public repository [183] contains both the pre-processed EEG data and features extracted using five different methods. The procedure for processing the data including removal of unwanted artefacts and feature extraction is explained in detail by Zhang and Lu [178]. The repository comprised of two feature sets for each feature extraction methods, one set was smoothed using moving average filter, and the other using linear dynamical system.

4.2.2 Data Preparation

The analysis presented in the article uses only data from experiment-1 of all 15 participants. This is to ensure a fair comparison with the DNN-SAE emotion recognition model by Luo *et al.* [174]. Accordingly, there were a total of 225 (*trials*) data from $1(\text{experiment}) \times 15(\text{participants}) \times 15(\text{trials})$. The lengths of EEG data varied between 30000 (minimum) and 54000 (maximum), which would result in unequal number of samples. The data was cleaned using the method mentioned by Luo *et al.*. The process removes any spurious disturbances or noise at the beginning and end of EEG signals and ensured each class had equal same of samples. Hence, from each 4-minute trial data, 80-second segment (16000 samples) from the centre was extracted. Further, Luo *et al.* evaluated their model using two different excerpts sizes; 12-second and 8-second, and demonstrated that the latter produced better accuracies. In this light, 80-second segmented data, 8-second excerpts (1600 samples) were extracted with an overlap of 4-seconds. Thus, post cleaning each trial contributed $19(\text{excerpts}) \times 62(\text{channels}) \times 1600(\text{samples})$.

4.2.3 Training and Test Data

In the context of SEED data, three different methods were found in the literature to generate training and test data as discussed below. Irrespective

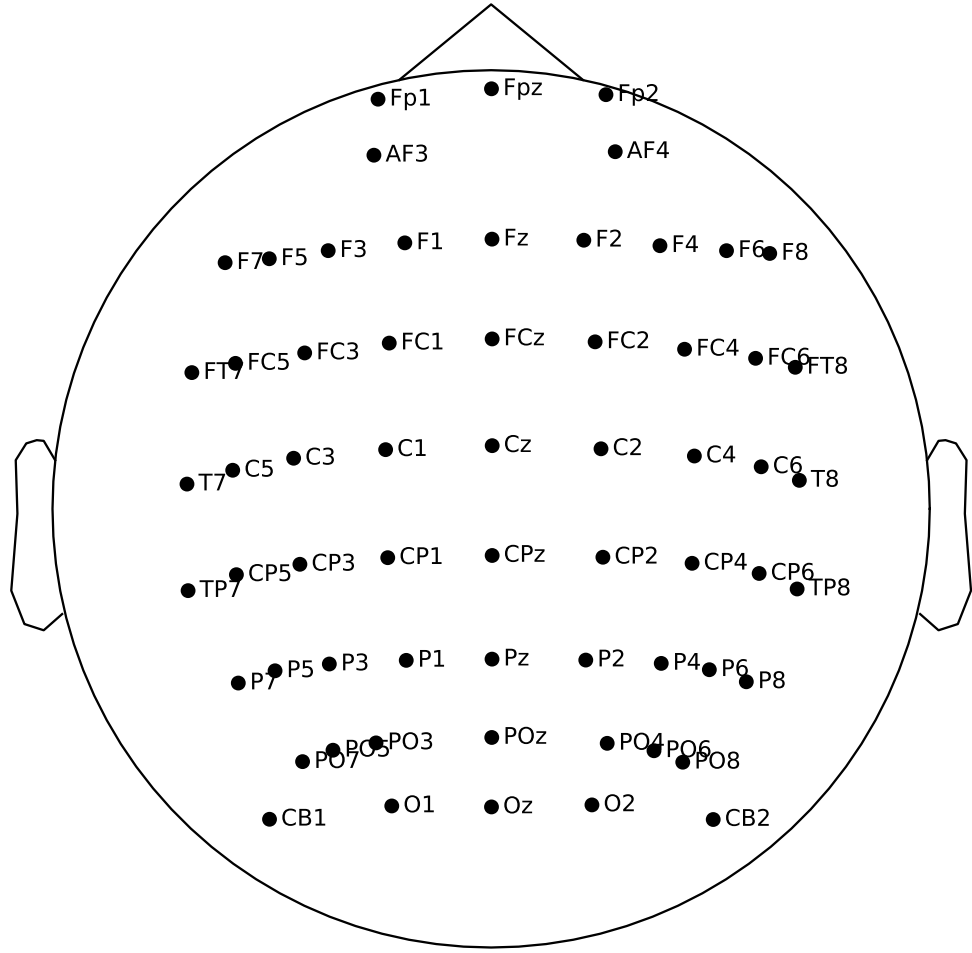


Figure 4.1: The topological representation of electrodes orientation of the EEG cap employed in this data

of the approach, both the training and the test sets were randomly shuffled before employing in a classifier model.

Subject Independent

s11 The method was adopted by Luo *et al.* [174] which is also considered a standard norm to evaluate a neural network based classifier. Here, data of all 15 participants from experiment-1 were first pooled providing a total of 4275 (from 225(*trials*)) excerpts. The pool was then split into 80% training (3420 excerpts) and 20% for testing (855 excerpts).

s12 The method was formally introduced by Wang *et al.* [179] as "leave one subject out", and was also adopted by Li *et al.* in [180]. Accordingly, data from 14 participants; 3990 ($14(\text{participants}) \times 15(\text{trials}) \times (19\text{excerpts})$) excerpts are used as training set and data from one participant; 285 ($15(\text{trials}) \times (19\text{excerpts})$) excerpts were used for testing. This procedure was repeated

15 times such that every participant was tested against a MC-DCNN model that was trained using data from remaining 14-participants.

Previously under SI2 procedure, SVM and KNN produced the best performance using linear dynamical system-smoothed differential entropy features (hereafter, referred to as DE_{LDS}) [176, 178]. Energy of all five frequency bands from each of 62-channels (total of 310 features) achieved the best performance. For detailed analysis about performance results from other tests, readers may refer to original studies [176, 178]. In this light, present study compares the performance of MC-DCNN with SVM and KNN as shown in Table. 4.3. To ensure a fair comparison, when regenerating results of SVM and KNN, centre 80-second segment was selected from the feature-set and no features were discarded.

Subject Dependent (SD)

The method was first mentioned by Duan *et al.* [177] when differential entropy feature extraction method was proposed for emotion classification. First 8-trials of an experiment were used as training and next 4-trials from the same experiment were used for testing. The second mention of a similar method was found in the article [178] by Zheng *et al.* where data from 9-trials of an experiment were used for training and next 6-trials from the same experiment were used for testing. Zheng *et al.* used the method to compare performance of dense belief networks with SVM and KNN methods, respectively [178]. The method mentioned second was also repeated by Song *et al.* in [167], Wang *et al.* in [179], Wang *et al.* in [168] and Li *et al.* in [180].

In line with methods mentioned in literature [167, 168, 179, 180], MC-DCNN was evaluated individually for each participant. First 9 trials of a participant's data (171-excerpts) were used for training and the remaining 114 excerpts were used for testing. Similar to SI2, the performance of MC-DCNN were compared with SVM and KNN as shown in Table. 4.3. The centre 80-second segment of DE_{LDS} were utilised for generating results of SVM and KNN.

4.2.4 Multi-Channel Dense Convolutional Neural Network

A generic version of the proposed MC-DCNN classifier model is shown in Figure. 4.2. The model was constructed using *Tensorflow* and *Keras* libraries [186]. In abstract, MC-DCNN is a fully connected model comprised of three blocks of neural networks. The extracted EEG signals excerpts are passed to convolutional layers to extract features, which is then encoded and decoded by SAE and finally, a classifier based on dense layers of perceptron's generates the output.

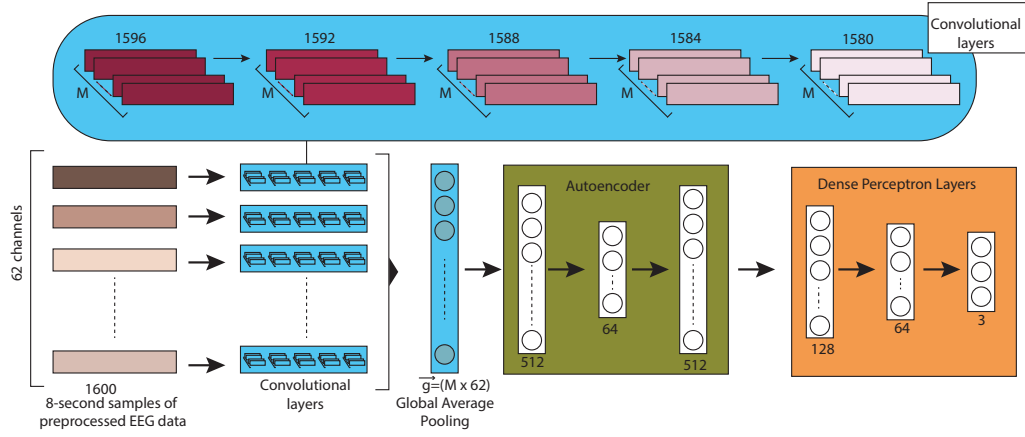


Figure 4.2: Architecture of the multi-Channel dense convolutional neural network

Convolution layers

The convolutional layers are intended to extract features from the training excerpts. Convolution operations were achieved using 1-dimension convolution function such that during training the impulse response of filters developed to resemble finite impulse response filter or wavelet type filter. Hence, the size of filters k was varied between 5, 7, 9 and 11. No significant variations were observed in the performance for difference values of k and an optimal value of $k = 9$ was chosen for generating results. The number of training epochs was set to 100, with a trigger mechanism to stop training when over-fitting was detected. The training was stopped if any of the following three different conditions was detected.

- If the average accuracy in previous 3 epochs exceeded 99%.
- If the moving average test loss for previous 20 epochs continued to increase.
- If the moving average of the difference between training loss and testing loss for previous 10 epochs continued to increase.

During a training epoch, n excerpts of size $62(\text{Channels}) \times 1600$ were fed i times through the model. Where, n and i represent batch size and step size such that $n \times i \approx \text{number of training excerpts}$. And, as shown in Fig. 4.2 there were M filters in each layer which were activated using the rectified linear (*reLU*) function. As shown in eqn. 4.1 the *reLU* function ensures only positive values progressed through the network.

$$\text{reLU}(t) = \max(0, t) \quad (4.1)$$

The parameters used to configure the MC-DCNN model corresponding to each of the data preparation methods is detailed in Table. 4.1. Since the number of training excerpts in subject dependent method were only 171, to avoid potential over-fitting the number of filters M is set to 9.

Table 4.1: Parameters used to generate a model for training and testing

Train and test set generation method	Batch size (n)	Step size (i)	Filters (M)
SI1	32	107	32
SI2	32	125	32
SD	19	9	9

In the global average pooling layer, the output from each of the filters were first averaged and then flattened which results in averaged feature vector of size $62(Channels) \times M(filters)$.

Autoencoder

The output vector of global averaged pool layer was then fed to the SAE-block which learns low-level features. The encoding and decoding operation also serve as smoothing operation to remove noise, which is similar to feature smoothing operation mentioned by Duan *et al.* in [177] achieved using moving average filter or linear dynamical system methods. Although here the layer-weights of SAE-block are adjusted automatically during the training process. After training, the weights of SAE layers will be tuned to recognise patterns in the incoming data enabling the same encoding and decoding operation during the compression and decompression phases.

As shown in Fig.4.2, the SAE operation was achieved using three layers with first and third layers, respectively had 512 perceptron's, and the middle layer had 64 perceptron's. The output of a perceptron is as represented by $ReLU(z)$ where z is as defined in the eqn. 4.2,

$$z = w \times x + b \quad (4.2)$$

w and b represent the weight and bias, respectively, leading to a perceptron, and x represents input to the perceptron from its previous layer. Since, the model is trained with a certain batch size n , eqn 4.2 can be represented as shown below

$$Z = W \times X + \vec{b} \quad (4.3)$$

where, the element x in eqn. 4.2 is represented as a matrix X in the eqn. 4.3. The size of X corresponds to $l \times n$, where, l represents number of outputs from the previous layer. Consequently, the weights and the bias' are also represented by matrix W and $\vec{b}$, with their corresponding sizes which depend on the number of elements between the layers.

Dense Perceptron Layers

The dense layers emulate multivariate logistic regression operation. As shown in Fig.4.2, the block consists of three layers with 128 perceptron's in layer-1, 64 perceptron's in layer-2 and 3 perceptron's in layer-3 corresponding to each class of data. The output of perceptron's in layer-1 and 2 is as represented in eqn. 4.2. The classifier output $\vec{y}$ from layer-3 follows *softmax* activation function as shown in eqn. 4.4,

$$\vec{y} = \frac{e^z}{\sum_{c=0}^2 e^{z_c}} \quad (4.4)$$

where, $c \in 0, 1, 2$ represent emotion classes in the dataset.

Loss Calculation and Training Optimisation

Since there were three different classes, loss was calculated using *sparse categorical crossentropy* function as shown in eqn. 4.5 in the process to achieve convergence.

$$\mathcal{L}(\vec{y}, \hat{y}) = \sum_{c=0}^2 y_c \log \hat{y}_c \quad (4.5)$$

where, $\vec{y}$ and $\hat{y}$ represents the expected output and estimated output, respectively.

Adaptive momentum *Adam* optimisation function was employed to optimise the learning process, avoid potential over-fitting, and achieve better convergence. The hyper-parameters of *Adam* optimiser were retained in its default settings such as, mean of the moment, $\beta_1 = 0.9$, exponential weighted average, $\beta_2 = 0.999$, division by zero correction parameter, $\epsilon = 10^{-8}$ and the learning rate, $\alpha = 0.001$ [187]. On the back propagation, the hyper-parameters tune the weights W and the bias $\vec{b}$ corresponding to a layer.

4.3 RESULTS AND DISCUSSION

The primary purpose of MC-DCNN model is to achieve good results for a real-time emotion recognition, where training can be limited to an existing pool of data. The performance is quite unknown when the classifier model encounters new features. To cater for such a scenario, the proposed model was tested on three different types of training data, targeting subject independent and subject dependent methods. The two cases of subject independent training results of MC-DCNN presented in Table. 4.2, aims to check the proposed techniques robustness to completely unknown data.

The training and testing accuracy using the SI1 method is shown in Figure. 4.3 demonstrating that MC-DCNN achieves a best accuracy of

Table 4.2: Summary of MC-DCNN accuracies observed from different training and test data generation methods

Train and test set generation method	Accuracy	Average accuracy from 15 participants	Accuracy standard deviation from 15 participants
SI1	97.42	-	-
SI2	-	67.27	7.02
SD	-	75.47	10.8

97.42%. It can be observed from the graph that the training was stopped automatically when the average training accuracy in the previous three epochs exceeded 99%. Correspondingly, the average testing accuracies in the epochs neither improved nor declined. However, the steady rise in the loss towards the end of training indicates that the testing accuracy would have dropped. Such trends indicate over-fitting, and the training was terminated automatically. In comparison, using the same data split method, the DNN-SAE model by Luo *et al.* [174] achieved a maximum accuracy of 96.77% using PCC features. However, the procedure of obtaining such an accuracy required the convolutional and SAE layers of the DNN-SAE model to be trained with all of the data. Contrastingly, MC-DCNN presented in the article achieves a better accuracy on a completely unseen 20% test data.

The results of SI2 and SD for all 15-participants is summarised in Table. 4.3. In the context of SD, the accuracies of MC-DCNN varied between a minimum 60.52% (P-5) and a maximum 96.49% (P-9) at an average of 75.47%, and standard deviation of 10.8%. In comparison SVM and KNN achieved average accuracies of 80.34% and 71.72%, respectively. However, the standard deviations of SVM and KNN accuracies were 16.55% and 15.18%, respectively. It can be inferred that although SVM produced high average accuracy, the classification errors are relatively higher. In summary, the results clearly show that MC-DCNN is capable of achieving better accuracies compared to KNN and SVM, even when dimension of data is large and volume of data is less.

In the context of SI2, accuracies of MC-DCNN varied between the minimum 55.08% (P-13) and maximum 76.84% (P-8) at an average of 67.27% and standard deviation of 7.02% as can be seen in Table. 4.3. In comparison, SVM and KNN produced relatively lower accuracies of 42.78% and 55.00%. Overall, it can be established that MC-DCNN can produce better performance irrespective of the data preparation methods.

Using SEED dataset, Li *et al.* [180] reports a maximum accuracy of 93.12% and 85.40% for SD and SI2, respectively. Comparing overall accuracies, Luo *et al.* [174] reports the best accuracy of 96.77%. In this regard, MC-DCNN

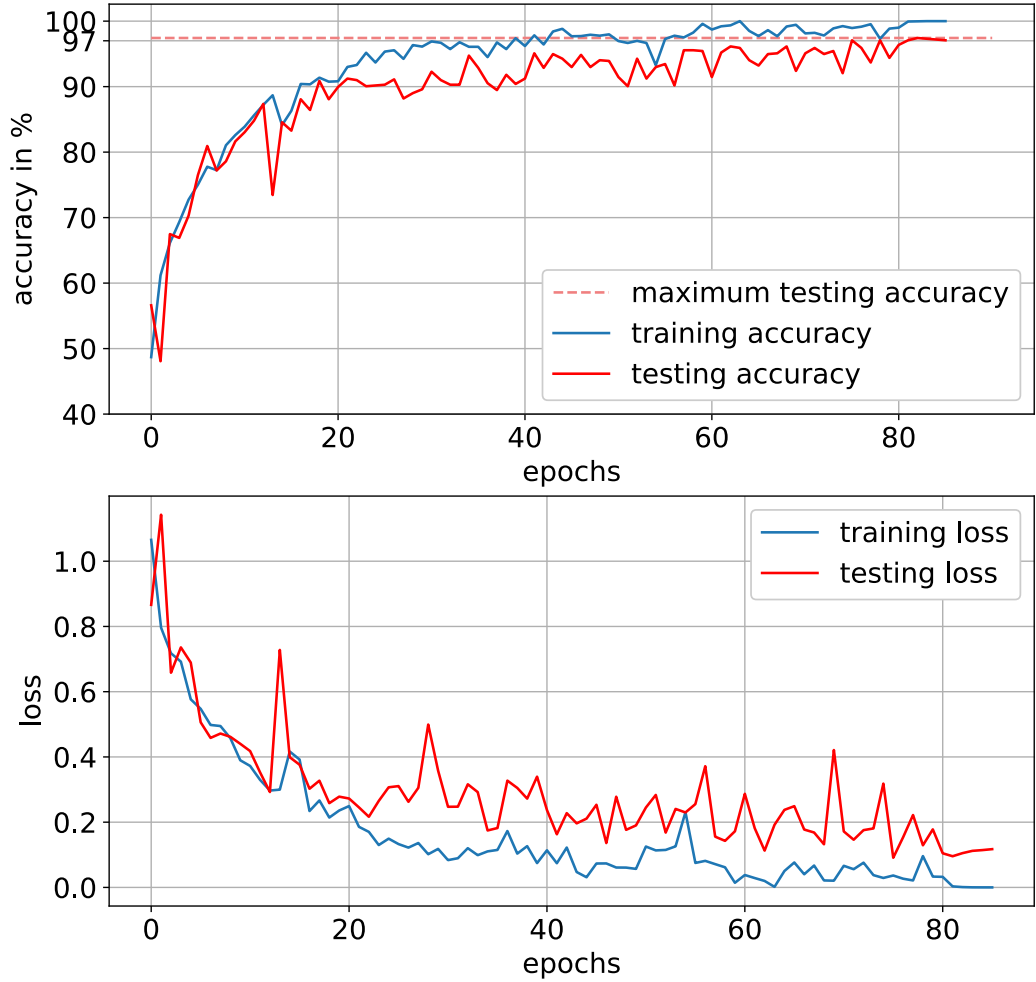


Figure 4.3: The plot of accuracy and cross entropy loss obtained from the model employing training and testing sets generated using the SI1 method

achieves the best accuracy of 97.42%. However, on closer inspection, we observed that the data selection processes employed in literature [174, 177, 180] are not same. Such comparisons may not provide a fair estimation of performance, hence, we restrict our recommendations solely based on results presented in Table 4.3 and 4.2.

Another advantage of the proposed method is the use of only pre-processed data as input to the model of emotion recognition. Computational cost encountered due to dimension reduction and feature derivation may not offer a practical real-time solution. Thus, the proposed MC-DCNN enables better classification of emotion classes via the modelling of 62-channels pre-processed EEG data.

Since, MC-DCNN achieves robust performance for unseen data (i.e., SI1 and SI2), the need for repeated training can be eliminated. Thus, although the convergence requires 80 epochs which could be time consuming, however, is a one-time process which is an advantage in real-time application.

Table 4.3: Accuracies observed across all 15 participants' experiment-1 data using KNN, SVM and MC-DCNN using SD and SI2 data splitting methods. The average and standard deviation row is obtained across all 15 participants

Participant	SD			SI2		
	KNN	SVM	MC-DCNN	KNN	SVM	MC-DCNN
P-1	83.33	96.04	90.35	60.83	61.50	65.26
P-2	65	93.33	63.15	42.50	44.83	62.79
P-3	71.45	93.54	83.33	53.16	58.60	59.29
P-4	100	96.25	78.07	35.33	64.58	74.03
P-5	45	77.08	60.52	40.75	70.91	69.12
P-6	97.77	100	85.08	61.00	46.66	63.85
P-7	50	78.95	73.68	42.50	47.08	60.35
P-8	77.91	70.41	84.21	54.91	65.33	76.84
P-9	86.04	90.2	96.49	33.33	60.00	57.89
P-10	64.5	66.66	62.28	39.41	59.16	67.71
P-11	67.08	77.77	72.8	53.08	73.16	76.48
P-12	72.29	67.08	81.57	27.91	44.14	74.69
P-13	73.5	75.62	63.15	33.33	38.41	55.08
P-14	55.2	88.95	70.71	30.41	91.75	69.47
P-15	66.66	33.33	66.66	33.33	53.33	75.72
Average/ Standard deviation	71.72/ 15.18	80.34/ 16.55	75.47/ 10.8	42.78/ 10.74	58.63/ 13.79	67.27/ 7.02

The classification of emotion states could often be an abstract problem as emotions can vary from subject to subject and strongly depends on stimulation type. Irrespective of data preparation method, the variation in accuracies explains that the participant achieving lowest accuracy demonstrated that there were high variances between the training features and test features leading to an under-performing model. It must be also noted that the volume of data considered in the current context is less than half the size considered in previous studies.

Future research could focus on observing the performance of MC-DCNN when trained on large volumes of data. Or some form of feature extraction where channel data could be selectively used in training instead of using data from all the channels. We strongly believe that, using a full set of data, and appropriate channel selection method [178], the model can perform better classification. However, this is beyond the current intentions of the model which is primarily focusing on online classification with real-time application.

4.4 SUMMARY

The article demonstrates a suitability of using MC-DCNN for online classification of emotions from EEG data. The performance comparison with the offline techniques (SVM and KNN) shows that MC-DCNN achieves a competitive performance when testing features are unknown. Pre-processed EEG data could be directly fed to MC-DCNN enabling a streamlined detection of emotion classes. The article demonstrates that MC-DCNN can handle large dimension of data achieving reliable accuracies of 75.47% even when 40% of the data is completely unknown to the trained model. The model is capable of achieving the best testing accuracy of 97.42% when trained on larger volume of training samples.

The following chapter sees the employment of ideas, concepts and the knowledge established through the review and preliminary studies. It starts by addressing one of short comings of current approach, i.e., via establishing novel empirical framework built on the fundamental concepts of team studies and principles of IPO framework. As such, it introduces a novel team task environment and carefully chosen PBR devices, participants recruitment and experimental procedure. It introduces capabilities of task dynamics in the exploration of team related hypothesis and some of the preliminary task performance results.

5

A Novel Multimodal and Multiparty Collaborative Experimental Framework

5.1 INTRODUCTION

Experimental platforms are an integral requirement of studies researching team performance. According to IPO framework [3], conditions of task and properties of PBR techniques [44] play a vital role in the investigation of team based hypotheses. Thus, an experimental platform comprising of team task environment, PBR techniques and subjective questionnaires as shown in Figure. 5.1 form essential requirements for indirect objective evaluation of team related attributes. Moreover, in the current era where automation and real-time data are becoming vital, there is increasing reliance on such platforms in the context of collaborative tasks [188].

Previously, team research has employed various task environments, however certain drawbacks and limitations are inherent as mentioned in sections. 2.6 and 2.6.1. In summary, team research integrate off-the-shelf PBR with STE to establish an experimental platform, and internet-based survey platforms or paper-based methods are employed for collecting subjective feedback [132]. The integration requires compatibility between software applications to ensure valid reference markers are available in the data. Often, such unification lacks in coherency and compatibility issues posing additional drawbacks, and hampering effective and efficient investigations.

This chapter introduces a state-of-the art experimental platform specifically designed to address team based hypotheses. The platform integrates a novel team task with four types of PBR techniques; EEG, ECG, eye-tracking and communication, and team based questionnaires. Data collected from seventeen 3-member teams were used for evaluating the platform. Evaluation parameters included integration compatibility, participation experience using task load indices and implications of task dynamics data for performance evaluation of teams. Discussion reveals, how this platform overcomes some of the current limitations, as-well-as, the capability of the data to promote a broad spectrum of possible team related analyses with minimal unwanted demands on the participants. With the current trend focusing

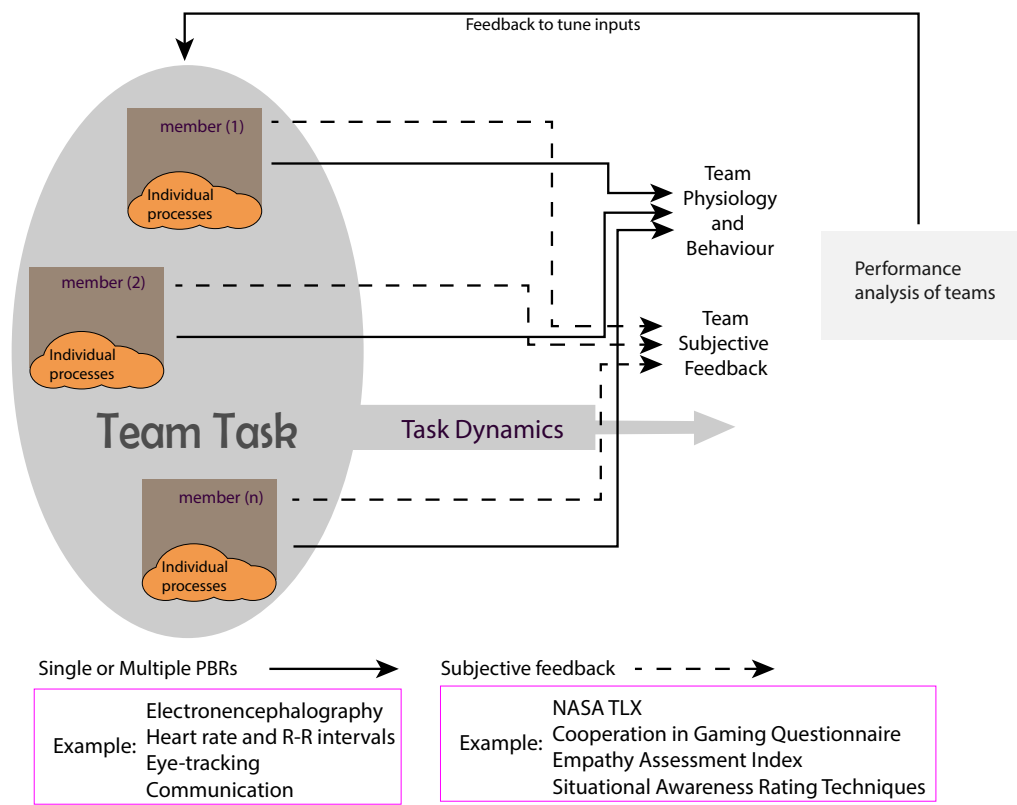


Figure 5.1: Key components of the experimental framework based on the principles of IPO for team performance analysis, comprising of team task environment, PBR techniques and subjective questionnaires.

on PBR based objective evaluation of team-related attributes, this platform could serve as a model for preparing better experiments.

5.2 THE TANK-BATTLE TASK

5.2.1 The Synthetic Task

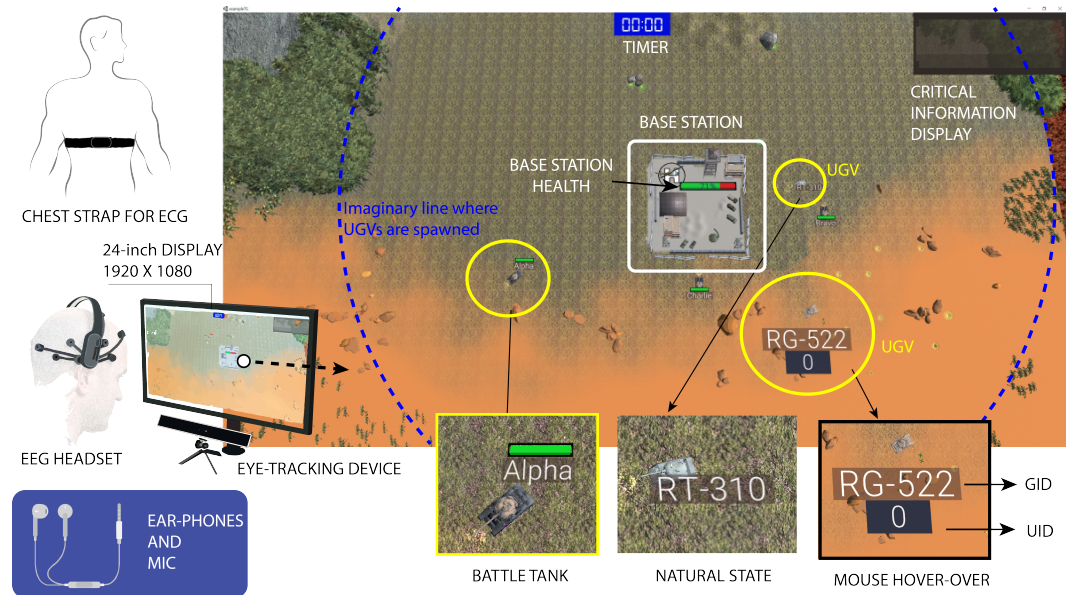


Figure 5.2: Main operating window of the tank battle simulation task and modalities

A novel tactical monitoring and protecting task namely “The Tank-Battle Simulation” was developed for this study. The Task is a 3-member multi-party collaborative and cooperative task environment. Unity® game engine [92] and mirror networking for Unity [189] was used to develop this task. The main simulation window is as shown in the Figure. 5.2. There are three components in the main window: (a) a 3-dimensional fixed-view of the task area, (b) a panel conveying critical information to the members, and (c) a timer window.

In the current configuration, the simulation regularly spawns UGVs at the circular edge of the map (marked with blue broken-lines in the Figure. 5.2) in a predefined pattern and approach a central Base Station (BS). The main objective of the task is to maintain the health of the BS as high as possible. From the spawn position, some UGVs continuously travel around the BS, whereas others move in a straight line. Two variants of UGVs, i.e., the *enemy* and the *friend* approach the BS. The *enemy* variants attack the BS by firing projectiles, which reduces the *health*. And the *friend* variants bring aid to the BS such that if the aid is successfully received, the health of BS increases. Teams composed of three members; Alpha, Bravo and Charlie, must work together cooperatively and strategically to maintain a good health of the BS. Each player will be assigned a semi-autonomous BT that can destroy UGVs. Teams must protect the BS from incoming enemy attacks, and avoid destroying friendly UGV to successfully receive the aid. This task incorporates an effective push-pull communication mechanism for enhanced

collaborative activity (section. 5.2.2), a method that implements physically effortless operation to attract wider demographics (section. 5.2.3), a pre-defined complex pattern enabling inter team comparison (section. 5.2.5), and a wide range of configurable parameters for tuning task dynamics (section. 5.2.4).

5.2.2 Push-pull communication

To encourage effective collaboration through communication, information push and pull mechanism is embedded into the task environment. This is achieved by assigning each UGV with a global-identification (GID) and unique-identification (UID)s. The GID serves as a global-reference (g) and the three unique-references u_n , one for each member i.e., $n \in \{1, 2, 3\}$ resemble an identifiable parameter such as vehicle speed or radar detection frequency. The reference g is alpha-numeric, visible for all members, and will not be the same for any two UGVs. And the reference $u_n \in \mathbb{Z}, \{5 \leq u_n \leq 30, \text{ and } u \bmod 5 = 0\}$, is unique for each member and hidden. A UGV will be enemy if $\sum_n(u_n) = 30$, otherwise the UGV will be a friend. For any UGV, a team-member can only view their u , and are obliged to verbally exchange the information with others to determine whether the UGV is an enemy or an ally. To view the reference u_n of a UGV, a member must bring a mouse-pointer on the UGV. By inducing the information push-pull mechanism, the environment generates a team-level information void that can be filled with successful communication. This enables players to verbally exchange their UIDs with each other to judge the variant of a UGV before attacking or passing. Incorrectly judging the variant of a UGV or delaying to determine the identity of the UGVs, can cost BS health either by missing aid or attack from enemies. Intelligible communication is key for a successful exchange of UGV information enabling efficient identification of enemy and friend variants. Hence, successful task performance is possible only through effective communication. Further, to ensure that there was less demand on individuals' mental load through summing, only multiples of 5 and < 25 were used as UID's.

5.2.3 Simplified Player Object Control

Team members must control their assigned a semi-autonomous BT, which can be used to engage in attack (fire projectiles at enemies), and navigate to a location on the map. This could be achieved with just a push of left-click mouse button. Simply clicking anywhere on the map will command the BT to navigate to the clicked location. And simply clicking on any incoming UGV will command their BT to fire projectiles at the UGV.

5.2.4 Configurable Parameters

Table 5.1: List of configurable objects and parameters with example settings

Object/Parameter	reference	current setting
number of BT projectiles	P	20
BT energy recovery	t_E	20 seconds
BT projectile fire rate	BT_{PFR}	0.5 seconds
BT acceleration	BT_{acc}	2000 units
UGV spawn interval	t_{UGV}	7 seconds
UGV projectile fire rate	UGV_{PFR}	0.5 seconds
UGV acceleration	UGV_{acc}	1500 units
number of waves	W	6
number of enemies/wave	UGV_F	38
number of friends/wave	UGV_E	39
configurable spawn locations		
BS starting health	H	20000 units

For efficient adaptability of the system for different testing scenarios, the STE allows configuration of many objects and parameters. All configurable parameters of the STE are listed in the Table. 5.1 along with its current settings.

The BTs can be configured to carry P number of projectiles. When a player exhausts all their projectiles, their BT enters energy recovery mode for time t_E seconds. UGVs in the simulation are spawned regularly with an interval t_{UGV} and the spawn locations are at equal distances from the centre of the BS such that they take equal time to enter the scene.

At the start of a trial, the health of BS is allotted H units which corresponds to 100%. As the BS receives aid from friend variants, the health improves, and when attacked by enemy variants, the health decreases. Any variation in health of BS is reflected in the base health (%), Figure. 5.2).

Overall, early mitigation of enemies minimises loss of BS health, and successful retention of friend variants till their end of life, rewards health to BS. The BTs carry limited ammunition, and the operating space is large, hence, members must effectively collaborate to maximise their productivity.

5.2.5 Task Intensity and Spawn Pattern

This is an example scenario which is designed to last approximately 11-minutes. The scenario comprises of six-waves; the first three are

low-intensity (360-seconds) and remaining three are high-intensity (318-seconds). A wave in low-intensity phase spawns 8-friends and 5-enemies at t_{UGV} interval and gap between waves was 30 seconds. A wave in high-intensity phase spawns 4-friends and 9-enemies at t_{UGV} interval and gap between waves was 15-seconds. In any wave, the spawning was such that first 9-UGVs were spawned sequentially from three different corners of the map, and last 4-UGVs were spawned from one section. Spawn locations and variant type followed a defined pattern, however, patterns were different for all 6 waves and all four trials of the experiment.

5.2.6 Task Summary and Performance Measures

The primary objectives of the task can be summarised as:

- communicate with other members of the team to recognise the threat by identifying *enemy* and *friend* variants
- destroy *enemy* UGVs to protect the health of BS.

Accordingly, points are awarded for the maintaining overall health of base station as *BH*, for correctly recognising friends to receive bonus *gain*, and points are deducted for getting attacked by enemies *loss*.

5.3 MATERIALS

5.3.1 Participants

Recruitment

A list of eligibility criteria was defined to ensure integrity and effectiveness of results, and to safeguard potential ill-effects on the experiment due to susceptible participants. Accordingly, the first safety guideline instructed participants with any psycho-pathological disorders (e.g., post-traumatic stress disorder), neurological conditions including pregnancy must refrain from the experiment. This was to avoid any influence on physiological measurements [190, 191]. The second guideline was in line with the suggestions from Laborde *et al.*, which restricted any participation from people suffering any physical health conditions (e.g., medical implants, sitting for a 1.5 hours) [192].

Age restriction was followed to ensure only participants between 18-60 years were included in the experiment. The broadened age cap promoted a diverse group of adults to participate in the study. However, it has been reported that, participants beyond this age demonstrated decline in cognitive abilities such as abstract thinking, learning and problem solving [193–195]. Only dominant and naturally right-handed mouse users were selected for

participation [191]. This minimised any influence of system error where decision-making process was not affected by hand-usage orientation.

All the aforementioned criteria and guidelines were listed on the advertisements and conveyed to interested participant via the email. The invitation email to all participants requested that they refrain from consuming any caffeinated beverages 2-hours before arriving and refrain from consuming any alcohol the previous day. In addition, participants were informed via the plain language statement before starting the experiment. Participants were recruited through advertisement flyers, snowball sampling and word-of-mouth methods. A soft copy of the advertisement was posted on the social media platforms (e.g., LinkedIn, Facebook) and Deakin University related posts and pages were targeted.

Demography

To test the capabilities of the task environment, seventeen 3-member teams were invited to participate in the study. The novelty of the task meant that all members had no prior experience with the task, and members of a team were not familiar with each other. The participants pool comprised of 36-males and 3-females (age: mean (μ) = 31, standard deviation (σ) = 7) with normal or corrected-to-normal vision and naturally right-handed computer mouse using capabilities. Informed consent was obtained from all the participants. The authors acted as facilitators to help the participants wear the PBR devices.

5.3.2 Experimental Setup

The experiment was setup using four desktop computers. One computer hosted the STE (Server/Manager) and three computers for team members. All computers were running Microsoft Windows 10 operating system with core i-7 processor and NVIDIA GeForce GTX 1080Ti Graphic Processing Unit. The task was projected on 24 inch Monitor at 1080×1920 resolution.

5.3.3 Multimodal Physiological Data

The modalities were chosen such that they offered continuous, unobtrusive, non-invasive and preferably wireless means to acquire physiological and behavioural data. The following four selected methods serves this purpose.

Electroencephalography

As introduced in Chapter. 4 and remarks in previous literature [44] suggests that human cognitive and emotion studies benefit from electrical brain activity recordings. EEG is one of the techniques that meets such a demand

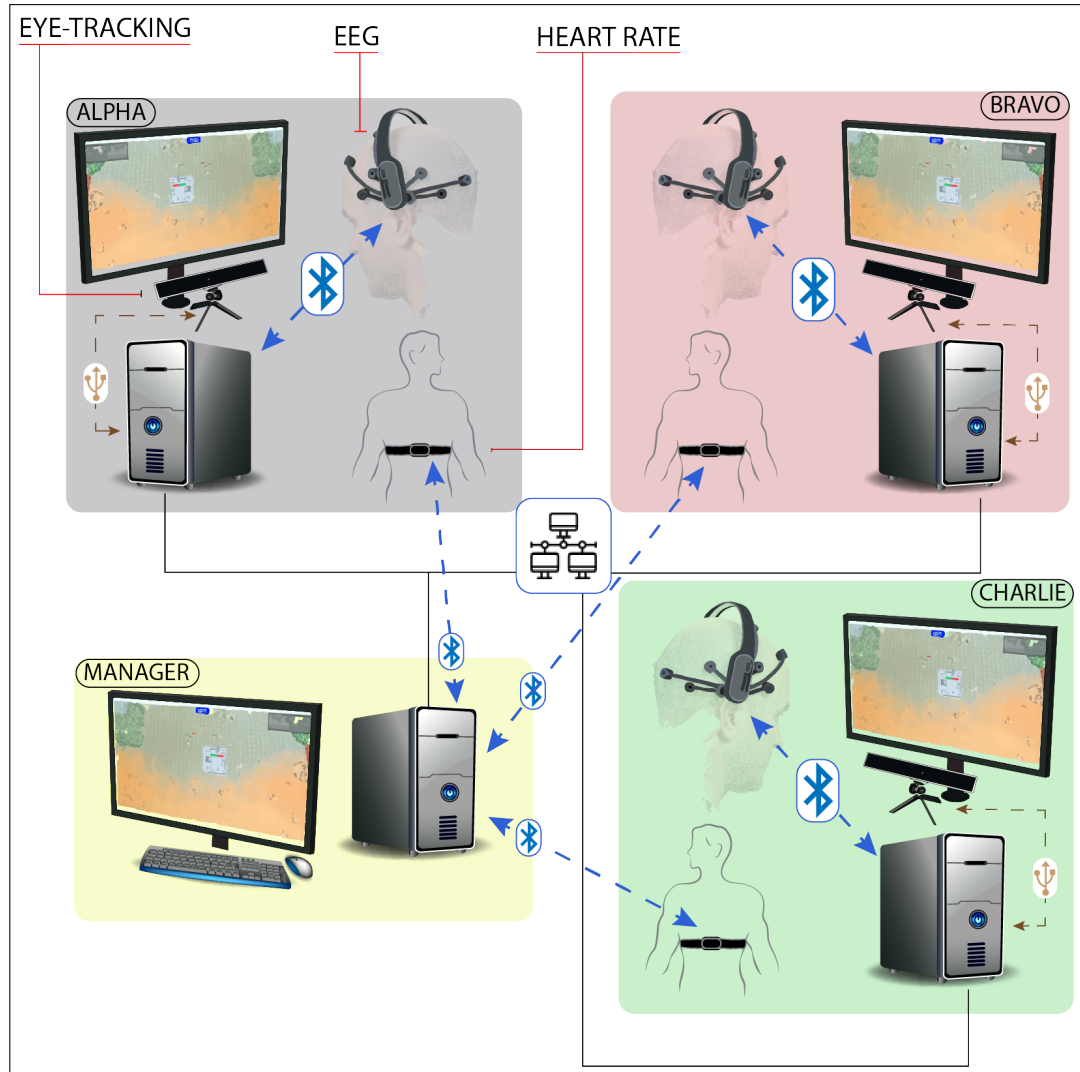


Figure 5.3: An overview of experimental setup demonstrating how every physiological and behavioural techniques are integrated and communicated with the server (Manager) and client machines (Alpha, Bravo, Charlie)

and fits the team study requirements mentioned in sections. 2.2.1 and 2.3.2. Additionally, compared to other techniques such as FNIRS, the bluetooth connectivity and software flexibility allows outdoor mobile and operational convenience. EEG data was collected using wireless Emotiv® EPOC+ with 14-channels using 10-20 international system electrode placement [90] as shown in Figure. 5.4a and 5.4b. Three devices (one for each participant) were connected to participants' computers using bluetooth dongles as shown in Figure. 5.3. The software stores all the data on a cloud server, and offers support for formats European data format (*.edf) and comma separated value (*.csv)

EEG represents compound responses from different parts of brain i.e., Thalamus, Hippocampus and Parietal lobe [44]. The method employs array of non-intrusive electrodes distributed over scalp, a technique to record electrical activity (voltage differences) between electrodes [44] as shown

in Figure. 5.4a and 5.4b. The electrical activity is a result of rhythmic oscillations due to neuronal activity (action potentials) originating from corresponding area of brain [44]. A channel of EEG-hardware records a collective activity from a population of neurons in the vicinity of the electrode. Rhythmic oscillations elicited from this synchronised activity can be observed via distinct frequencies. Subsequently, these oscillations could be detected by sensitive electrodes at the scalp. Lopes (2013) [196] defines these neural oscillations as a significant increase in peak power of the EEG signal observed at a particular frequency spectrum. Spectrum's that are of interest in this thesis were introduced in sections. 2.2.1 and 4.1, and is brief explained below.

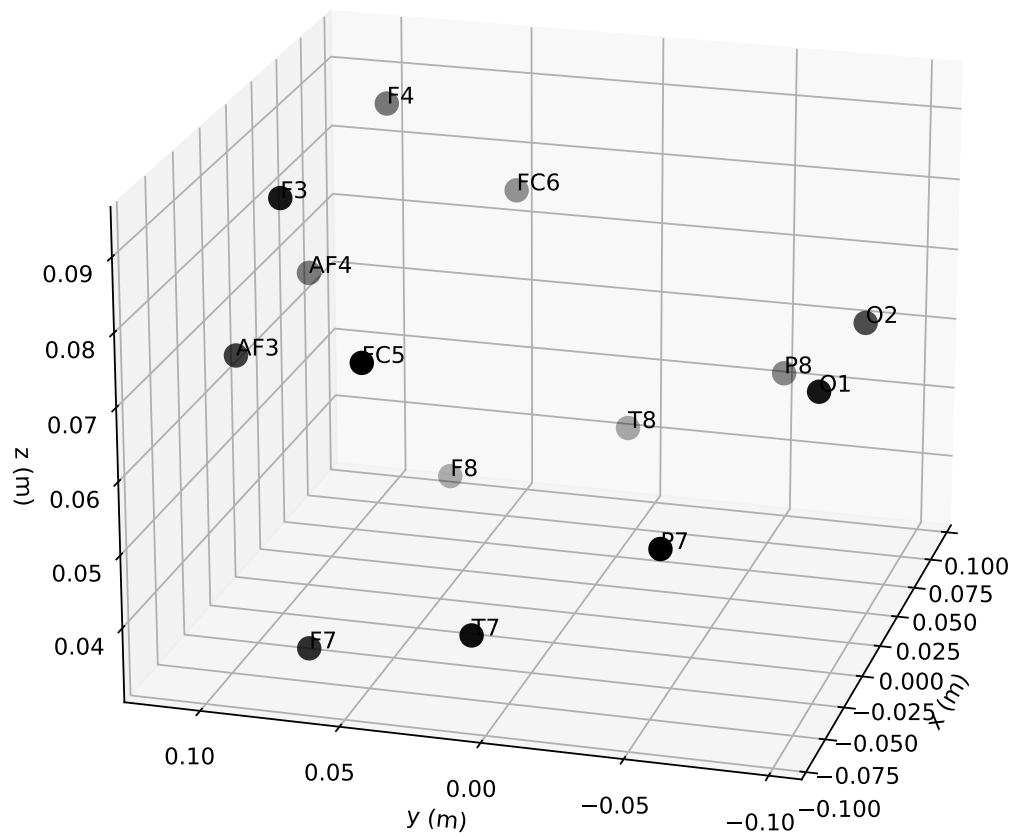
- δ (Thalamus) is mostly observed during deep sleep [44, 197].
- θ (Hippocampus) are observed during motion (limbic activity) and are also signs of familiarity indications [44, 198]
- α (Thalamus) waves indicate alertness, focus, and working memory. Presence of α characterises alert wakefulness yet relaxed state of a person [44, 199].
- μ (Parietal lobe) is combination of mostly α with small influences of β , a characteristic of maintaining physical stillness in awake [44, 200].
- β (Parietal lobe) waves with lower frequency are indication of concentration and busy thinking, also representation of involvement in emotion induced tasks [44, 201].
- γ (Hippocampus) is related with problem solving, sense of fear, mental processing and consciousness [44, 202].

Electrocardiography

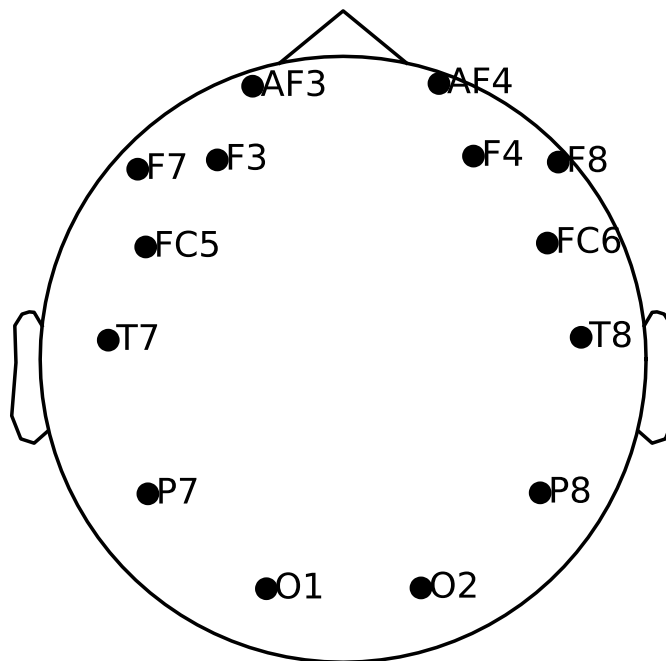
The preliminary study in Chapter.3 demonstrated the use of ECG signals in the study of SPC. Primarily, ECG represent cardiovascular activity which is a collective response from the sympathetic and parasympathetic divisions of the ANS. Various measures of ECG that are of interest in team settings is addressed in section. 2.2.1. ECG offers advantage such as quick reflections, short and long-term variability to psychosocial stress and workload demands [44]. In addition to physiological synchrony, measures of ECG data are used to represent temporal variations in cognitive states [203].

$$HR = 60(s) / ibi(ms) \quad (5.1)$$

Polar H10 devices were used to collect heart-rate (HR) and raw R-R intervals data [204]. The quality of data from this particular device has been



(a) 3-dimensional topological representation



(b) 2-dimensional topological representation

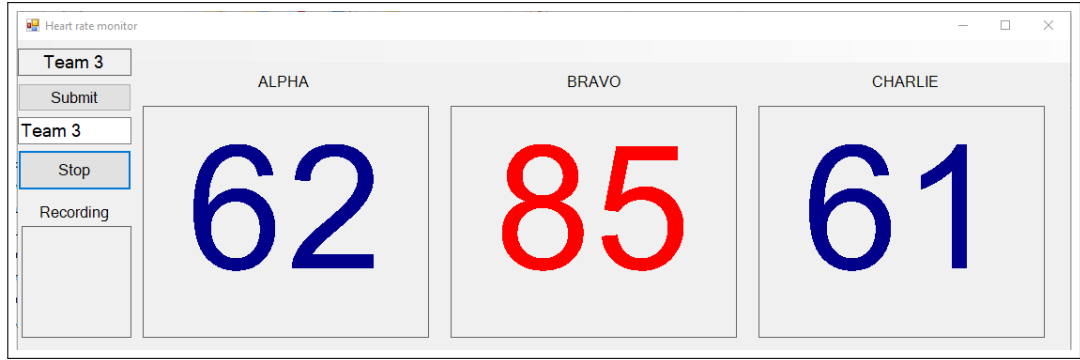


Figure 5.5: An overview of the custom developed windows application to simultaneously record raw IBI data from three different devices. For verifying the working of the device and to ensure transparency with participants, the raw IBI values were converted and displayed as beats per minute using eqn. 5.1.

previously verified by Gilgen-Ammann *et al.* [204]. The device resembles a belt that could be wrapped around the chest region and uses surface electrodes enabling non-invasive method to acquire the data. A custom application was designed to meet the software requirements (Figure. 5.5) such that three devices were connected to the server via bluetooth as shown in Figure. 5.3. The application continuously records raw IBI with universal time stamps data into three different (*.csv) formats; one for each participant. Rows in the raw data consisted of universal time stamp and IBI in milliseconds. In the light of this thesis, the hardware and software perfectly fit the requirement. More details on raw IBI series is provided in Appendix-A.

Eye-tracking

The activity of human eye is dependent on task and operating contexts [205]. These activities can be categorised as eye-movements tracking and pupillometry [44]. Fixations and saccades are common measures derived from eye-movements. *Fixation* accounts for the ability of humans to regulate the retina image stabilisation with an effort to extract information, and *saccadic* measurement accounts for movements between point of gazes to acquire correlated information within the visual space as shown in Figure. 5.6. On the other hand, pupillometry accounts for pupil dilation where the size of pupil is regulated by muscles, the *sphincter pupillae* and the *dilator pupillae*. Contraction and dilation processes are sensitive to cognitive and mental demands, and during affective and attention processing [206]. However, the process is more sensitive to changes in lighting to regulate the amount of light that enters the cornea [207].

Retrospectively, pupillometry and eye-movements measurements required different approaches. Early eye-movements measurements were achieved using multiple cameras such that one camera was located on a

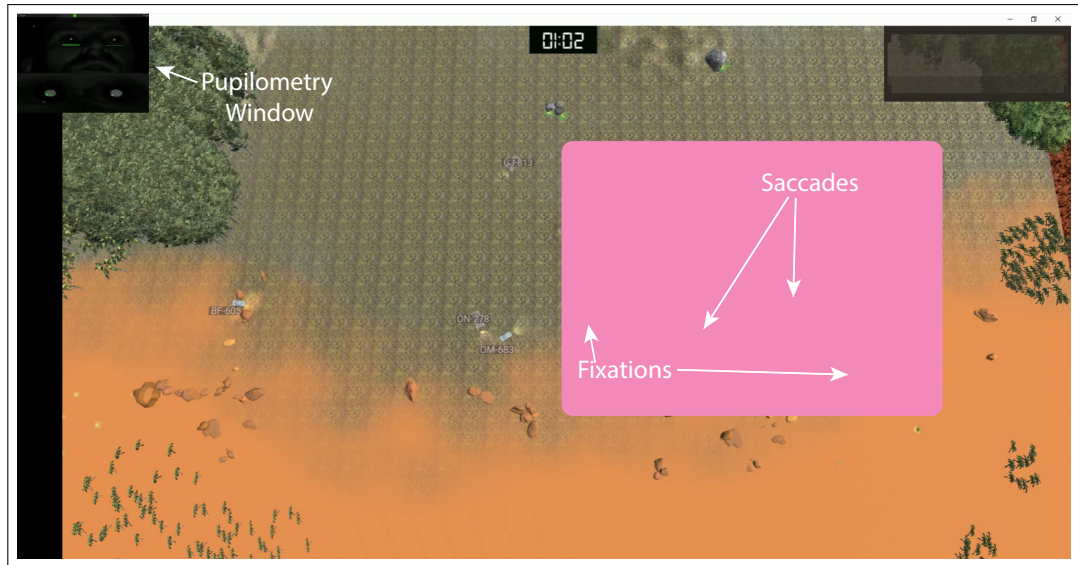


Figure 5.6: An overview of gaze measurements demonstrating fixations and saccades in the task environment

desktop focusing the subject and the other was on the head of the subject [208]. And pupillometry was achieved using video oculography, where infrared cameras recorded the pupils and image processing was used to account for variation in sizes. As explained in section. 2.2.2, the above mentioned gaze measures are of greater significance in team-level investigations. Recently, technological advancements such as Gazepoint GP3 enable real-time acquisition of both measurements, and hence was adopted in this research. Gazepoint GP3 is a desktop-based eye tracking device, capturing participants eye gaze data at 60Hz sample rate [88]. The devices were placed under the visual display and the data was recorded using Gazepoint Analysis software. The gazepoint data was locally recorded on participants' computers.

Gaze and Eye-related measurements

The software provides universal time-stamped measures of best point-of-gaze (BPOG) (x and y coordinates in relation to the size of the visual display), fixation point-of-gaze, fixation identification number (f-ID), saccade magnitudes, left and right pupil diameters, blink rate, blink duration, blink identification number and binary validation indices for all measures. All the data can be extracted in *.csv format.

Importance of Gaze in Team Settings

As introduced in section. 2.2.2, the investigation of eye movement behaviours to unfold collaborative processes has garnered significant interest among researchers in recent years [79, 209, 210]. During collaborative activities, members of a team coordinate and cooperate over a shared space to

achieve common task objectives [211]. As a result, interacting members sub-consciously and consciously exhibit eye-movement behaviours to learn each other's intentions, and to demonstrate one's intentions for others [211, 212]. Such behaviours are conceptualised as joint attention [133, 213–215] and distributed gaze [21, 81, 216]. Analysing such behaviours could help unravel the intelligibility of interaction which in turn provides insights into verbal (communication) and non-verbal information exchange (gestures), attitudes, and preparedness [133]. It also fosters newer assessment and objective evaluation methods of performances [209]. Additionally, there is a growing interest in the development of predictive models utilising gaze-based measures [209, 217]. Such advancements hold a significant importance in evaluation and enhancement of performance, learning, training and adaptation, specifically in academic [218], defence [32], aviation [208, 219], health-care [220] and sports [211] sectors.

Previously, Cherubini *et al.* had hypothesised that in a remote collaborative working condition [221], gaze convergence between pairs should positively correlate with performance. Indeed, joint attention studies provide empirical evidence for relationships between gaze-recurrence on collaborating members and their performance. For example, in a speaker-listener task setting, the gaze of good-listeners was observed to follow the interests of the speaker [214]. Coco and Dale [214] employed RQA to quantify gaze recurrence between dyads and observed a positive correlation with good-listeners. Similarly, in a surgical task setting, it was observed that the doctor and nurses while mostly focused on different region of interest (ROI), however, occasionally sharing each other's gaze fixations was equally critical for better interaction [216]. Similarly, Fasold *et al.* observed that referees in a basketball game evidently distributed their gaze to improve coverage of the shared space [211]. Fasold *et al.* also reports a contrasting effect in a handball game setting, where the majority of referees gaze attention was around the ball [222]. Visual correlations are also employed as indicators of team cognitive activity such as team situational awareness [60]. Traditionally, fixation revisits have been analysed in relation to performance [208]. Eye tracking offers as an substitute to video analysis providing an array of gaze-dynamics that are beneficial for team related analysis. Thus, reliance on gaze co-occurrence and distribution of gaze within the interaction space is task-specific, and are important in ecological settings.

Further, studies demonstrate application of gaze co-occurrences as a categorical classification (experts vs novices) using hidden-markov models [223], assessment of collaboration using unsupervised clustering [209], and collaborative performance [210].

Verbal Communication

Human voice can be defined as a complex pattern of sounds composed of multiple harmonics with fundamental frequency. In addition to the linguistic feature, the speech conveys valuable information in the form of utterances, quality, prosody and intonations to the listeners. Cognitive demands, pathological conditions and emotions states are known to modulate the voice production, and as explained in section. 2.2.2 forms of communication are greatly valued in collaborative investigations.

An internet-of-things based communication application was developed using zero-mq application peripheral interface (API) protocol to exchange low-latency voice message over local area network. The application enables participants to communicate with each other using stereo in-ear earphones with microphone built into the device. Since, the experiment was designed such that the participants were seated in the same room, only voice information was recorded. Microsoft® voice recorder application was used to locally record voice data in MPEG layer3 (*.mp3) format. Further, the recorded voice data was processed using a pitch filter to hide the identities of the participants.

5.3.4 Questionnaires and Subjective feedback

This comprised of a demographic questionnaire, National aeronautics and space administration (NASA)-Task Load Index (NASA-TLX), cooperation in gaming questionnaire (CGQ), and task experience questionnaire (TEQ). All the questionnaires were built into the STE and the responses could be exported to the server-machine in *.csv format.

NASA-TLX

One of the primary targets of this thesis was to minimise external influence on performance in the form of physical demands. In addition, the configuration parameters aimed to maintain an average temporal and mental demands, so it offers an even field for engaging a diverse group of participants. In order to examine the influence of these properties, the NASA-TLX subjective workload rating technique was employed in this study. The format introduced by Hart and Staveland has been widely adopted in human factors research, across various team contexts including, training and simulations [127, 224, 225]. Examining these properties offers a better controlled environment for testing and evaluation of team related hypothesis. In addition, Rubio *et al.*'s study reports validation where the NASA-TLX and other workload scales demonstrated a significantly strong correlation with performance feedback ratings [226]. To this end, team literature supports the employment of NASA-TLX as a suitable tool to

evaluate subjective workload demands [35,227]. In the context of teams and groups, global index derived by averaging across team members has also been observed in the literature [19,228].

NASA-TLX built into the task environment is shown in Figure. 5.7. Complete details of questions and descriptions that were collected during the experiment is given below. The results of subjective feedback from participants are discussed in section. 5.5.2

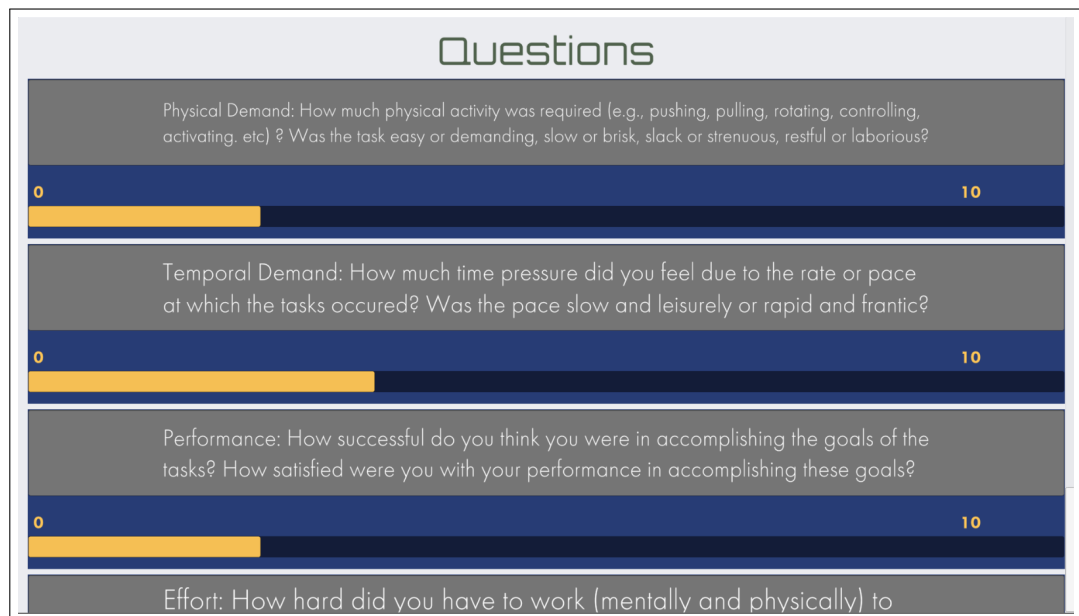


Figure 5.7: An overview of gaze measurements demonstrating fixations and saccades in the task environment

- Physical Demand: How much physical activity was required (e.g., pushing, pulling, rotating, controlling, activating, etc) ? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
- Temporal Demand: How much time pressure did you feel due to the rate or pace at which the tasks occurred? Was the pace slow and leisurely or rapid and frantic?
- Performance: How successful do you think you were in accomplishing the goals of the tasks? How satisfied were you with your performance in accomplishing these goals?
- Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?
- Frustration: How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Cooperation in Gaming Questionnaire

Literature suggests that online team games promote communication and cooperation, thereby, improving the overall team experience [229, 230]. In virtual collaborative environments, as opposed to the general presence, the term *social presence* represents “being together with one another” [231]. It is important in virtual environments to establish a social connection between each other. Thus, the degree of social presence decides the strength of bonding, contributing to team cognitive processes such as mutual attention, awareness and understanding in work settings [5, 232, 233].

Various levels of social presence have been observed in the literature, varying between a stronger perception of behavioural engagement and mutual presence, a mid-level psychological interaction to a lower level sensing of miscellaneous social entities [234, 235]. To this end, literature evolved an appropriate scale for measuring social presence referred to as social presence in gaming questionnaire (SPGQ). One of the early employments of the questionnaire was demonstrated by De Kort *et al.* (2007) [236].

Team literature reports that SPGQ cannot be readily adopted as its original intention was for competitive games [237], and hence includes items such as ‘revenge’ and ‘schadenfreude’, which do not fit cooperative and collaborative contexts. Hence, Hudson and Cairns (2014) devised a specific tool that accounted for competitive and cooperative presence in gaming questionnaire (CCPIG version-1.2) [232]. This included two competitive aspects i.e., awareness and engagement, and two cooperative aspects (total 14 questions) i.e., team cohesion and team involvement (total 25 questions). In the interest of this research only cooperative aspects of the questionnaire was adopted, and will be referred to as CGQ. Similar to NASA-NASA-TLX, CGQ was built into the task environment and the results are discussed in section. 5.5.5.

Task Experience Questionnaire

The questionnaire was designed to examine participants understanding of engagement rules and objectives of the task. This included a series of basic task related questions that is important for successful participation in the task. Further, it also aimed a learning during which trial the cohort became comfortable, whether the instruction video clearly conveyed the objectives and whether the practice session served its purpose. Similar approaches have been previously observed in the literature. For example, Chanel *et al.* employed game experience questionnaire [50] to assess players experiences during participation. Similar to the above questionnaires, TEQ was built into the task environment and the results are discussed in section. 5.5.5.

5.3.5 Experimental Procedure

The study was approved by the Human Ethics Advisory Group (HEAG) of the Faculty of Science Engineering and Built Environment, Deakin University, Australia.

All the teams participated between 10:00 am and 12:30 pm. However, the overall duration of participation was a maximum 90-minutes (approx). The participation duration can be broken down as 20-minutes (approx) for setting-up and calibrating the PBR devices and 60-minutes (approx) for experiment-participation and questionnaire completion.

On arrival, participants signed the consent form and completed the demographic questionnaire. All participants were provided new identities as Alpha, Bravo and Charlie, and were instructed to use these identities when addressing each other during the participation. Next, the devices were setup using the following four steps. Step-1, participants were instructed to wear the Polar H10 chest strap on the upper rib-cage such that the electrodes touched the skin. Step-2, facilitators helped participants to wear EEG headset, and also performed the necessary adjustments to ensure good EEG quality. Step-3, participants were requested to sit in a comfortable position to operate the computer mouse and then their eye-trackers were calibrated. Step-4, participants were instructed to put on the in-ear earphones.

After the initial setup, all participants watched an instruction video which explained rules and objectives of the task. In addition, participants were encouraged to actively communicate with each other at any point during the participation and in-between the task-episodes, and instructed that they could develop any strategy by communicating with each other. Then, participants were allowed to clarify any doubts related to the task. Finally, participants engaged as a team in a practice task session which lasted 1-minute (approx) to familiarise with the task environment before starting actual trials.

All teams completed four-trials with each trial lasting 11-minutes (approx). At the end of each trial, a performance feedback score appeared on all participants' screen. After reviewing the feedback, participants completed NASA-TLX. At the end of fourth-trial, in addition to NASA-TLX, participants also completed cooperation in gaming questionnaire and task-understanding questionnaire.

5.4 SIMULATION ANALYTICS

One of the main features of the simulation is the access to continuous spatial and temporal dynamics of participants' activities and the environment. The server records UGV and BT events and the trial start and end times. In this section we use some example data to demonstrate the analytic capabilities provided by the proposed environment. The client machines

continuously record screen coordinates of mouse and BTs. The screen coordinates are recorded in $\{x,y\}$ format and relative to the size of the display, and simulation times use universal timestamps milliseconds format. All recordings are continuous at 30 samples/second rate and available for analysis in *.json files.

5.4.1 User-Experience

Participants use mouse to perform activities that can be summarised into events as *look*, *navigate* and *engage*. The *look* event is when a participant brings the mouse-pointer over a UGV, to view the UID. The *navigate* event is a record of participants commanding their BT to navigate to a region of the map. And the *engage* event is record of participants commanding their BT to attack a UGV. Figure 5.8 shows the spatial representation of participants mouse event activity. The *look* event shows participants engagement in regions of the map demonstrating their contribution in the task. The plots of *navigate* and *engage* events shows region of the map where a participant would mostly operate. In addition, non-evented data such as mouse-positions is logged at 6Hz sampling frequency.

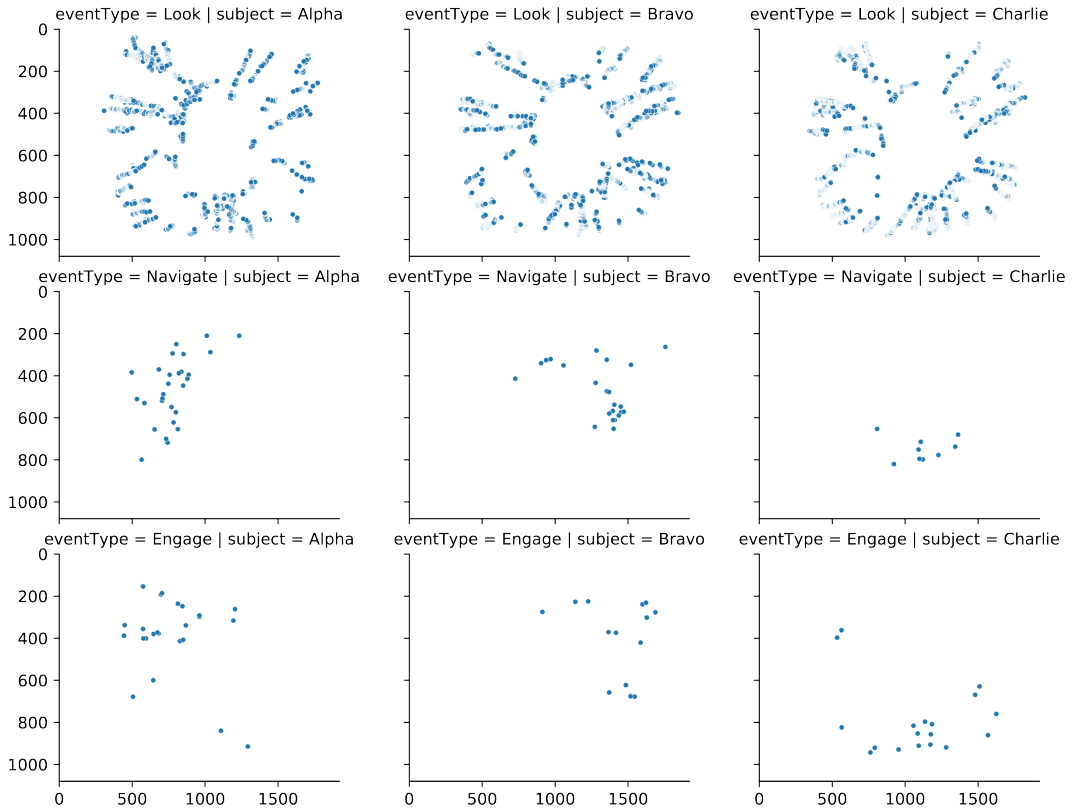


Figure 5.8: Summary of team-individuals activity recorded from mouse events during a trial session. The top row is record of *Look* events showing a players visual scanning of map regions. Middle row is records of *Navigate* events showing where a player operated. And bottom row is record of *Engage* events showing where a player engaged to destroy UGVs.

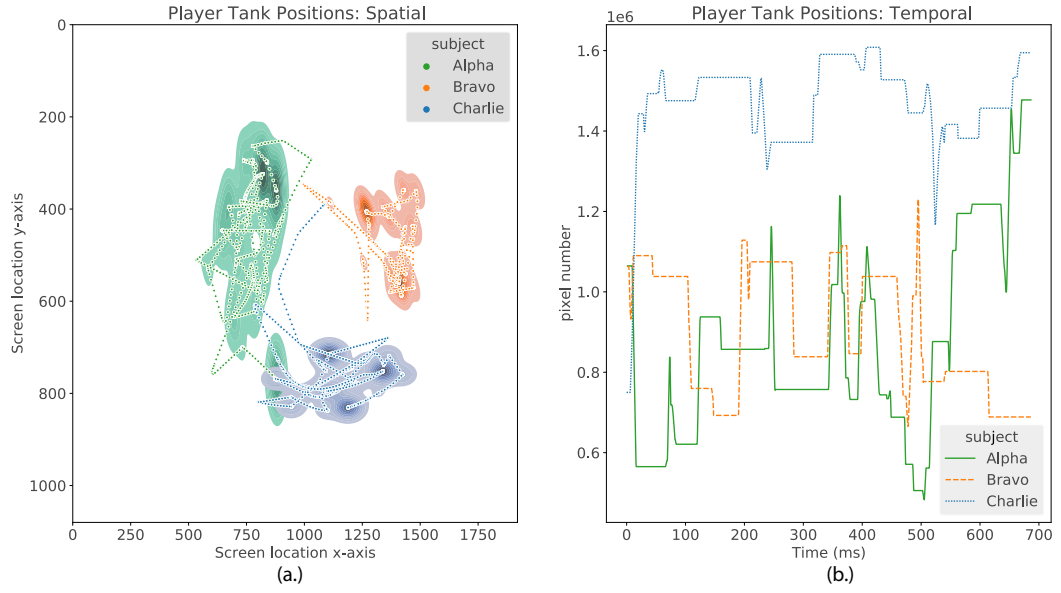


Figure 5.9: Dynamics of participant BTs. (a). Spatial location of BTs with x and y axis representing screen coordinates in relation to 1080×1920 screen resolution. (b). Temporal location of BTs with x axis representing pixel value corresponding to 1080×1920 screen resolution and y axis representing time in milliseconds.

Differences between the events *look* and *engage* can be used to understand average reaction time, taken by teams to judge if a UGV is *friend* or an *enemy*.

5.4.2 Environment Dynamics

Participant BT Events

BT objects represent participants in the simulation which share relationship with mouse events i.e., *navigate* and *engage*. The location of BTs on the map shows the region where the participants operated during a trial instance as shown in Figure. 5.9(a). The X-Y location information of BTs was used to derive pixel-value corresponding to the projected screen. Using time information, it is possible to temporally unfold activities of participants as shown in Figure. 5.9(b).

Dynamics of the Challenge

The two variants of UGVs, *enemy* and *friend* are further split into two types of vehicles such as *Semi* and *T90*. The *Semi* type UGV are simple in challenge as they only move in straight line, halts when arrived in the proximity of BS, and starts to fire projectiles. The *T90* type UGV poses a complex challenge to participants by moving in straight line till it arrives in the proximity of BS, and continuously goes around the BS in circle while also firing projectiles. This combination elevates the challenge to the team, demanding strategies

to improve achievement of task objectives. The *.json file contain UGV movements on the screen (x,y coordinates), and instantaneously records of UGV-events such as aid, fire, spawn and destroy,

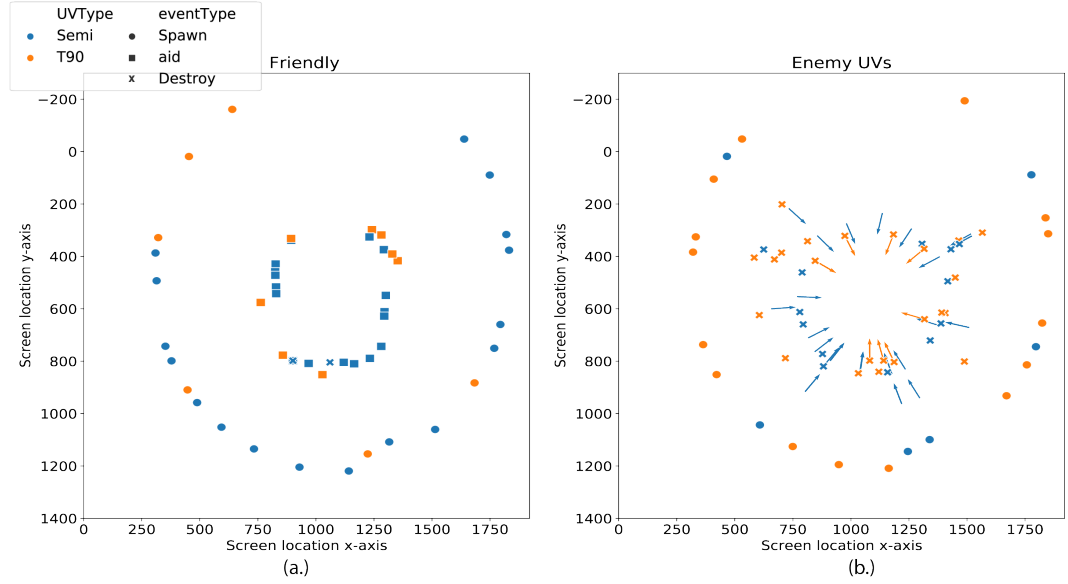


Figure 5.10: Dynamics of UGVs. (a) Spatial locations of *friend* variants and (b) *enemy* variant UGVs. The x and y axis represent screen coordinates in relation to 1080×1920 display resolution

The UGVs are tagged corresponding to their type and variant. Such tagging helps to unravel spatial locations of UGVs under various events during a trial. The events of UGVs can be summarised as *spawn*, *aid*, *fire* and *destroy*. UGVs are spawned by the STE and disappear when attacked by participants or after delivering aid (only friend variants). Figure. 5.10 demonstrates spatial locations events corresponding to each variant of UGVs. Accordingly, spawn points associated with *enemy* and *friend* variants are shown in Figure. 5.10. Figure. 5.10(a) shows locations of *friend* variant delivering aid or when it was incorrectly destroyed. While Figure. 5.10(b) shows location at the time when *enemy* variants were destroyed. The figure also shows projectile fire direction indicated by arrows that points in the direction of BS.

5.5 RESULTS & DISCUSSION

5.5.1 Comfort-levels of using wearable devices

The questionnaire collected participants feedback about the level of comfort while wearing Emotiv EEG and Polar H10 chest strap. Participants rated their experience on a scale of 1 to 5 with 1 being the least comfortable and 5 being the most comfortable, after approximately 1-hour of usage. Partici-

pants overall experience in Table. 5.2 shows they were fairly comfortable, with the chest strap being more comfortable compared to EEG headset.

Table 5.2: Consolidated wearable device comfort response from participants

wearable device	mean (μ)	standard deviation (σ)
Overall	3.11	1.32
Emotiv EEG	2.77	1.51
Polar H10 Chest strap	3.52	1.30

5.5.2 Task-load Index

NASA Task-load Index (NASA TLX) was collected from participants to unfold load dynamics experienced by them as they progressed through task episodes. Participants responded to task-load index questions on a scale of 1-10 such that scores of 1 and 10 represented the minimum and the maximum load index, respectively.

Figure. 5.11 shows the average load indices experienced by participants in each trial. From the results, it could be understood that the overall physical demand required to perform the task was relatively less. Also, the temporal, mental and efforts required to perform the task was at mid-level.

Task Performance

Task performance of teams was analysed and observed that BH scores had gradually increased as teams progressed through trials. This observation was in line with theories of learning, adaptation and experience development [238,239]. As shown in the Figure. 5.13, a one-way analysis of variance (ANOVA) was conducted to compare the effect of performance on trial progression and observed a significant effect, $F(3,36) = 8.25$, $p < 0.001$. A detailed analysis was conducted with Bonferroni correction and observed that the scores in trial-1 (-760.2 ± 431.73) were significantly lower from trial-3 (-179.2 ± 222.37) and 4 (-120.2 ± 174.27). There was no other statistical significance observed. Thus, suggesting that early trials could include learning and experience development effects as observed in other similar studies.

Synthesised Task Performance Measures

In addition to the three performance measures mentioned above, additional measures could be synthesised solely from the dynamics of the task. As an example, three measures are derived as follows

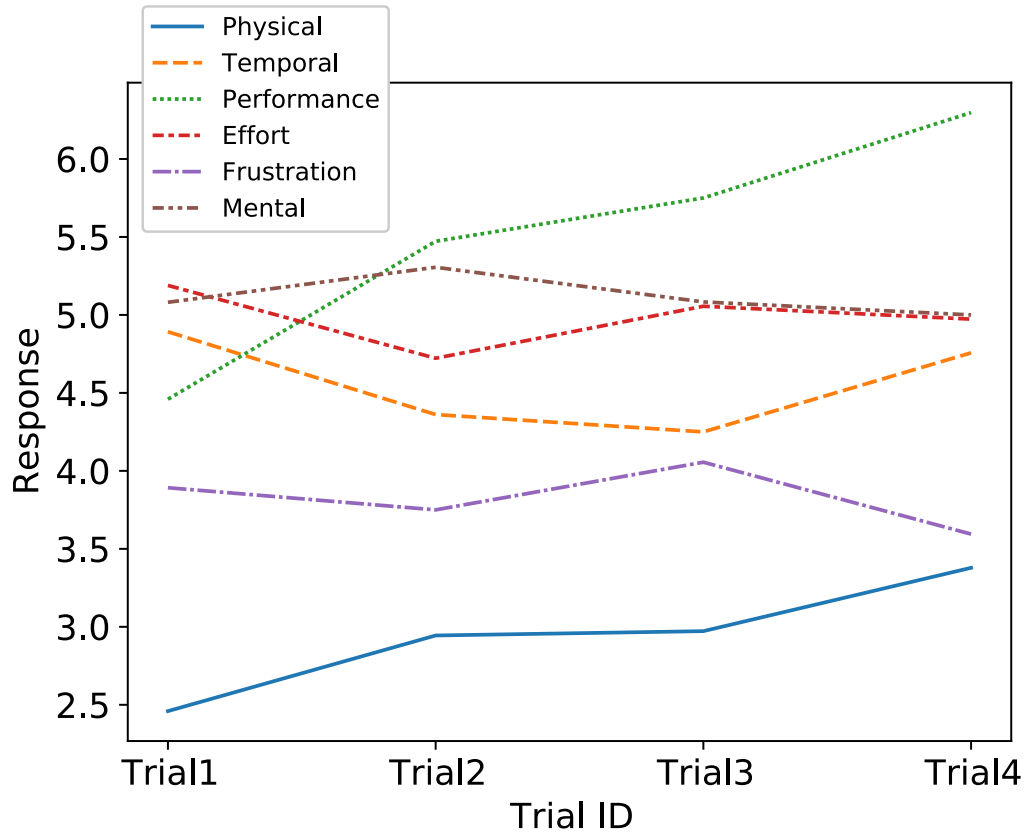


Figure 5.11: Average task load indices experienced by participants in trial episodes

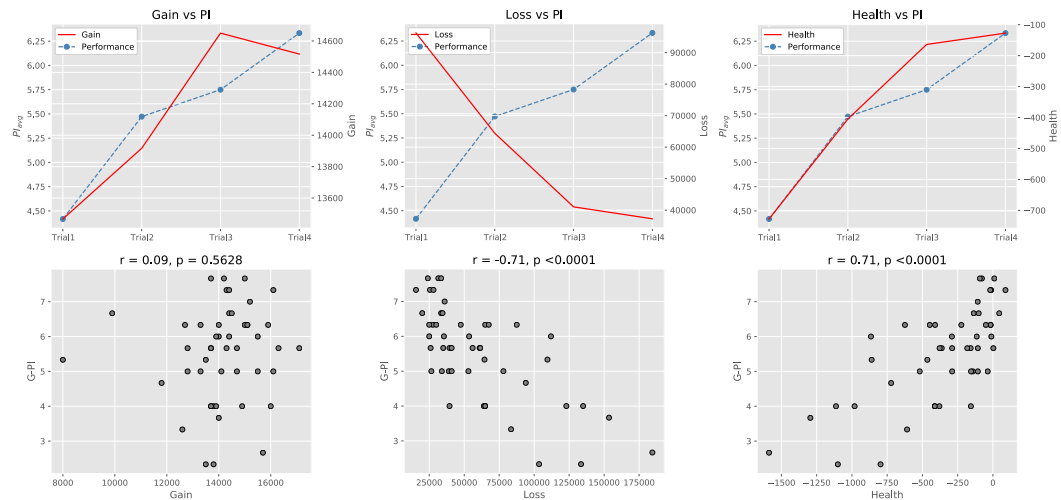


Figure 5.12: Comparison between objective (Task) and subjective performance (NASA-TLX). From left to right, plot of bonus gained vs performance, plot of health lost vs performance, and overall health vs performance. The bonus gained is a result of properly identifying more friendly variants. The health loss is a result of failing to destroy enemy variants in time. The base health (%) is consolidated version of the previous two decisive parameters i.e., the bonus gained and health lost.

- kill duration, time taken by teams to destroy an enemy tank measured from the time it was first spawned on the map.

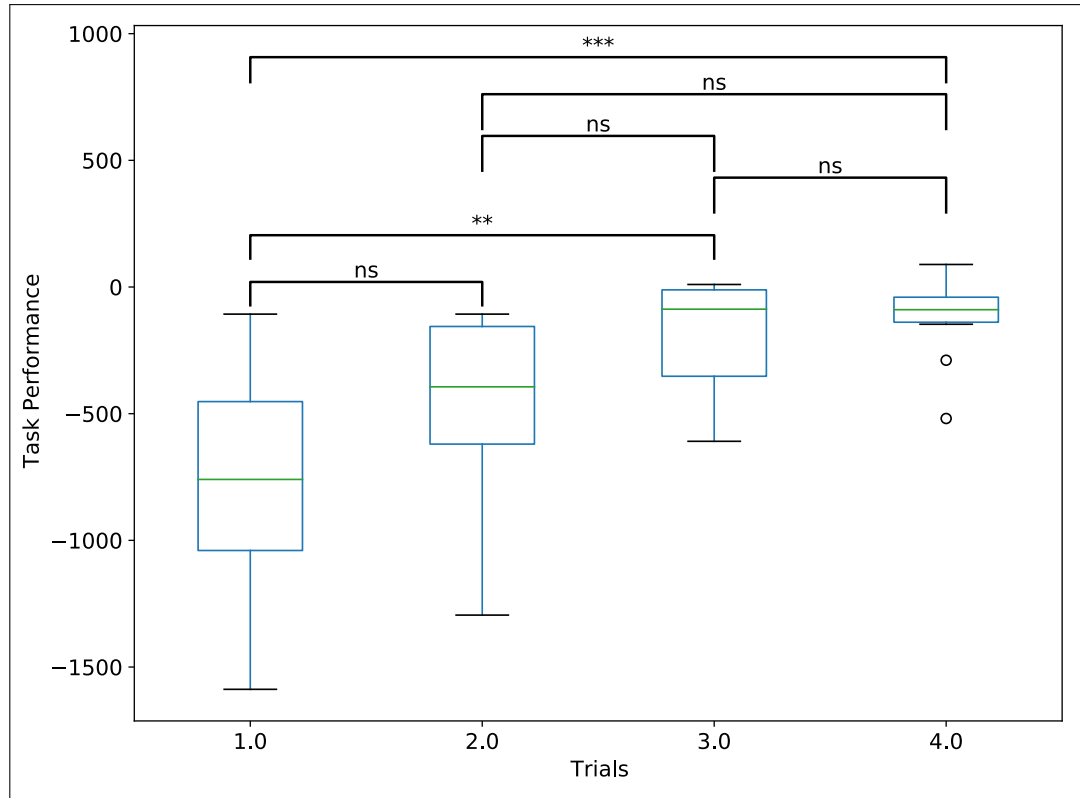


Figure 5.13: A boxplot of task performance (BH) of the teams as they progressed through 4-trials of the experiment

- First Response, time taken by all three members to successfully register look events on any UGV. It is measured as time between first look event of the first registering member and first look event of the registering last member.
- First Response to kill duration, time between the completion of first response and destroying the enemy.

The average and standard deviation of all the synthetically generated performance measures were estimated and compared against the original measures using correlation coefficient. The result demonstrated in Figure. 5.14 shows that except for bonus gained score, most of the measures strongly correlates among each other. Thus, establishing additional reference measures for further endeavours of this thesis.

5.5.3 Comparing Objective Against Subjective Performance

From NASA-TLX, the performance index (*PI*) reported by participants were compared against achievement of task measured via the scores gain, loss and overall health. This enables understanding the relationship between performance perceived by participants as a team and overall execution of task objectives. From Figure. 5.12 (top row), results shows that the averaged

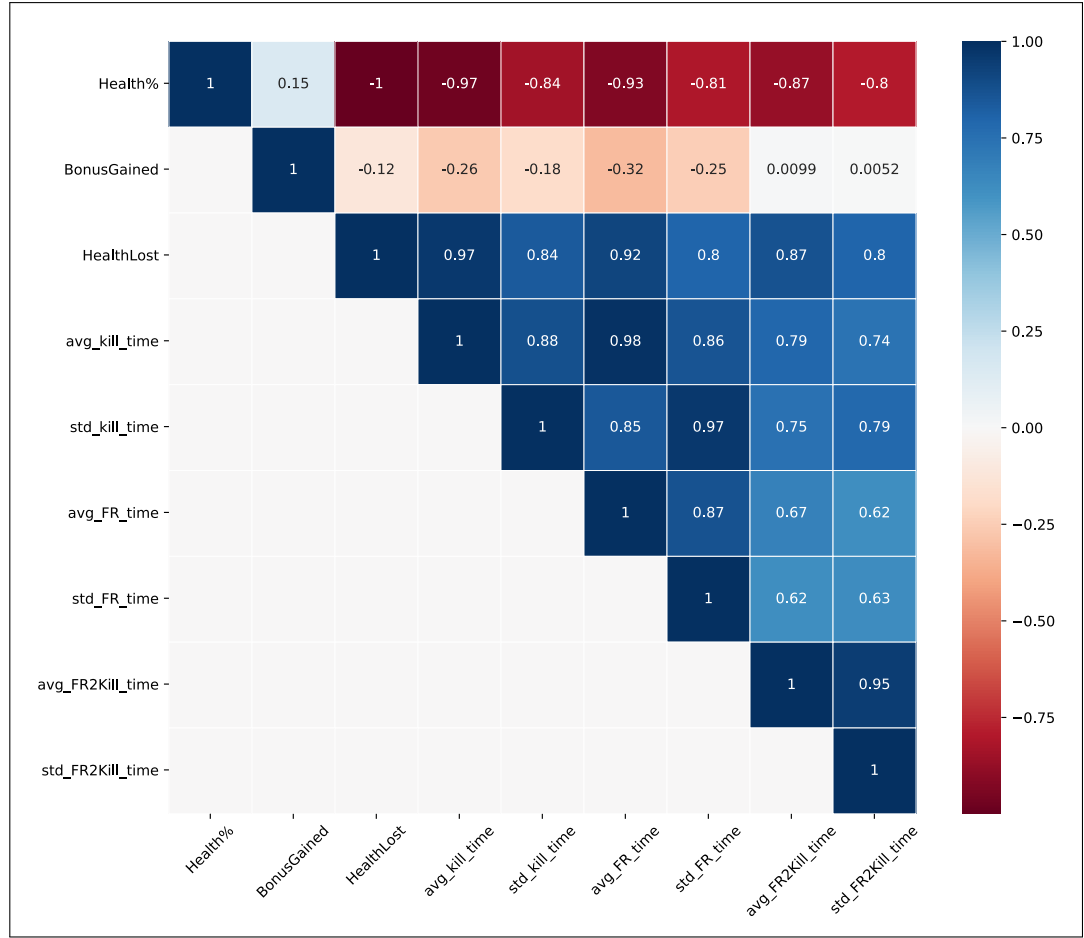


Figure 5.14: A heatmap comparing correlation between task performance scores and synthetically generated from task event dynamics registered by team members.

performance index (PI_{avg}) perceived by participants during trials demonstrated a relationship with task scores averaged across trials. In comparison with the subjective reporting of PI, the gain and the overall health steadily increased and the loss gradually reduced as teams progressed through the trials.

Further, a detailed analysis was performed to observed if there was any significant relationship between PI perceived by team members during a trial and their task scores observed during the trial. For this purpose, PI of team members during a trial were averaged to obtain a team-level global performance index (G-PI). Then, Pearson correlation was estimated between G-PI and task scores as shown in Figure. 5.12 (bottom row). Results in the figure shows that G-PI demonstrated a significant relationship with the health lost parameter *loss* ($r = -0.71, p < 0.0001$) and the *overall health (%)* ($r = 0.71, p < 0.0001$), however, no significant relationship was observed with respect to the bonus gained parameter (*gain*).

5.5.4 Task-related Questionnaire

Participants understanding of task objectives and related dynamics is key to validate their prominent participation. Accordingly, this questionnaire aimed to learn the stage when participants became comfortable to competitively perform in the task. Participants responded to eight basic task-related questions on a scale of 1-7 as described in Table. 5.3:

Table 5.3: Different stages of the experiment

Stage	Indicator
Instruction video	1
Practice session	2
Trial-1	3
Trial-2	4
Trial-3	5
Trial-4	6
Never	7

From the Table. 5.4, it can be inferred that, participants understood most of the task-related information from the instruction video and the practice session. However, teams learnt the whereabouts of the other members only during the first trial session.

Table 5.4: Consolidated response from participants

Question	Experiment Stage mean $\pm$ standard deviation ($\mu \pm \sigma$)
Whereabouts of team-members on the map	2.43 ± 1.40
Learn to navigate the BTs	1.62 ± 0.97
Learn to attack UGV	1.57 ± 0.87
Sharing UIDs was key	2.06 ± 1.17
Recognise friend variants	1.18 ± 0.51
Recognise enemy variants	1.18 ± 0.52
Only UGVs firing were enemies	2.00 ± 1.02
Destroying enemies is strategy	2.12 ± 1.13

5.5.5 Cooperation in Gaming Questionnaire (CGQ)

Cooperation is a key attribute of team that this task setting aims to achieve. This questionnaire aids in determining the level of cooperation, communication and collaboration between the team members. Participants responded to 25-questions on a scale of 1-5 as described in Table. 5.5 such that a scores

of 1 and 5 represented *not at all* and *very much so* responses, respectively. The table suggests that most participants agreed to all the questions inclining to the fact that aggregates of cooperation was visible in the task. From the table it could be understood that the task has potential to test various attributes and qualities of collaborative and joint activity processes.

Table 5.5: Cooperation in Gaming Questionnaire

Question	Experiment Stage mean $\pm$ standard deviation ($\mu \pm \sigma$)
I acted with my team-mates in mind	3.74 $\pm$ 1.04
I considered my team-mates' possible plans/thoughts	3.66 $\pm$ 1.04
It was as much about the team as about my own game	3.83 $\pm$ 1.11
I felt my team shared a common overall aim	4.09 $\pm$ 0.94
I did not want my team to think I had let them down	3.76 $\pm$ 1.26
I felt my team-mates were looking out for me	3.60 $\pm$ 1.34
I felt I contributed to the team	4.17 $\pm$ 0.94
I felt my actions made a difference to my team-mates	3.86 $\pm$ 1.07
My actions were determined by the objectives of the team	3.69 $\pm$ 1.17
I felt my team was committed to working together	3.97 $\pm$ 1.18
The actions of my team-mates affected my thoughts and actions	3.83 $\pm$ 1.16
Being part of a team motivated me	4.14 $\pm$ 0.96
My team communicated well	3.91 $\pm$ 1.00
The team had a mutual understanding	3.86 $\pm$ 1.02
I put the performance of the team over my personal performance	4.06 $\pm$ 1.04
My team-mates played a significant role in my experience of the game	4.00 $\pm$ 1.01
I wanted my team to value me	3.66 $\pm$ 1.12
I felt responsible for achieving the objectives of the team	3.94 $\pm$ 0.98
I felt a social connection to my team-mates (camaraderie)	3.89 $\pm$ 0.89
I made an effort to work with my team-mates	4.29 $\pm$ 0.88
I felt like I was part of a team	4.23 $\pm$ 0.90
I felt my team shared common short term goals	4.03 $\pm$ 0.97
I was aware of my team	3.86 $\pm$ 1.51
My team-mates were useful	4.40 $\pm$ 0.73
I felt the team helped me	4.43 $\pm$ 0.73

5.6 GENERAL DISCUSSION

The primary purpose of this work was to address the need for an integrated experimental platform for testing team related hypotheses. Constructs of IPO theory are used as basis for development of the platform to ensure compatibility with PBR based team research. Table. 5.1 shows a range of configurable parameters that allows modification and tuning of the task dynamics such as intensity-levels and duration. Current configuration is used to demonstrate capabilities of the platform and the feasibility of employing multiple modalities in the context of team research. Data of thirteen teams composed of three members with no prior experience and not familiar with each other were used to validate the platform. The platform comprehensively overcomes current drawbacks which can be briefly explained as follows:

- Where consideration of dyads as group was previously questioned [84,240], this study presents results using three members in a team, overcoming dyadic constraints
- Configurable parameters overcome the need for complex programming
- The accessibility of user-input data acquisition is advantageous
- Minimal physical demand and gadget mastering of the study helps overcome unwanted stress
- The diverse pool of participants invited via social media platforms and their responses to task-related questionnaire suggests the possibility of engaging mixed demographic participation
- The STE is a simple Microsoft®Windows application which could be started with click of a button allowing the adoption by wider research groups investigating team processes.

Further, the results of CGQ suggests that the task in its current configuration comprehensively addresses collaborative processes. The participants feedback in Table. 5.2 clearly shows that participants were comfortable wearing PBR devices for a period of 90 minutes. It should be noted that there were 12 modalities (3 participants $\times$ 4 modalities) concurrently recorded, demonstrating the feasibility of employing multimodality in team context.

The configurable parameters allow easy portability of the STE to test various attributes of teams [116]. For example, varying the intensities of the task promotes exploring physiological and behavioural patterns related to team stress and emotions [59]. The duration parameter flexibility enables the tasks to be completed in one session, multiple trials of short intervals (current setting) as well as, multiple sessions [5] promoting an effective

team research concerning working memory and aggregates of cognitive processes. Further, it also facilitates preparation of tasks to suit either a broader or a selective demographic participants-pool.

The availability of continuous and spatio-temporal nature of user-input data (i.e., mouse locations and key-press logs) conveys human-computer interaction information. Such data jointly with PBR promotes empirical measure of performance which is a primary requirement of an integrated observational platform, highly demanded in the space of cyber-security operations [17]. In this light, comparing mouse-events and task engagement events in Figure. 5.8 and 5.10, respectively, it can be noticed that there exists a relationship between UGVs locations and participants' *look* events, indicative of cooperative engagement. In addition, the availability of user performance data enables effective exploration of empirical relationship between the task performance and emergent states of team such as cooperation and team situational awareness and PBR-based objective team performance [5].

Concurrent acquisition of PBR data fosters efficient observation of physiological aggregations between team members such as physiological and behavioural synchrony/compliance and correlated activations [4]. Specifically, EEG, ECG and eye-tracking are employed in such investigations [35, 81, 96, 241]. Statistical aggregations are used to understand processes underlying emergent states such as cooperation and joint activity between participants [35]. The broad spectrum of multimodality also promotes team studies to model predictive algorithms [39, 40, 79]. The choice of multimodal PBR achieves more informative data conveying the activity of autonomic nervous system, peripheral nervous system, visual-cognitive behaviour and raw form of communication. Most recent approach in a three-member collaborative context have used primitive approaches using EDA and video analysis [82, 242, 243]. While EDA is of significant importance, however, video analysis requires laborious interpretation.

The availability of PBR along and subjective feedback complies with the international standards organisation's (ISO 17488, 2016) behavioural demands [87], psychophysiological measurements [244] and the reliable retrospective questionnaires [245]. The autonomous nature of BT's, the configurability offered by the STE and simplicity of participation using simple click of mouse button, offer support to perform team experiments in context of detection-response task (DRT) [87]. DRT is essential of exploring cognitive-workload experienced by team-members, and thus, this platform aids in the analysis of effects of team cognitive-workload on performance outcomes. Similar to the human-in-the-loop models [188], the attributes of this STE such as temporal specificity coupled with continuous nature of PBR can enable modelling of machine learning algorithms [246] to achieve team-in-the-loop models.

Since all members of team share equal responsibilities and roles, communication data can be used to observe conflict and emergence of leadership

roles. Previously, communication utterances have also been used as marker of team expertise and performance, shared mental models, shared knowledge, collective decision-making and information processing [45, 246, 247]. The availability of concurrent PBR fosters indexing such behaviours. The current configuration of the simulation enables analyses that concern the influence of emergent states on task performance [52].

The predefined variety of wave patterns in the current configuration offered several advantageous. It ensured teams did not get a chance to learn the spawn pattern or guess friends and enemies without exchanging UIDs. On analysing communication data, it was further confirmed that attempts to guess variants were incorrect. However, since the pattern was predefined, it encourages between team analyses. Additionally, it enables generation of pseudo team data for observing collaborative indicators in the PBR [35, 96].

The theme of this task using unmanned vehicles is primarily driven by future technological revolution involving teams for supervisory control of air, ground and under-water autonomous vehicles [188]. The overall functioning of this platform is in line with the architectures of advanced embedded training system by Zachary *et al.* [32]. Such applications are not just encouraged in defence but also in non-military activities such as wildlife surveillance, controlling forest fires and under-sea explorations [248]. The human-in-the-loop nature of the STE means that by modifying the agents of the platform [249], it is possible to tasks in health-care delivery environment.

The entire setup could be privately hosted on a server-machine thereby, allowing an additional layer secure collection of data concealing identities of participants yet enabling efficient research. Except eye-tracking, all the devices can be remotely controlled at the server-machine. The platform facilitates observing task performance of teams in the form of BS health percentage, BS health lost and bonus gained reflecting details of teams performance. A comparison between objective and subjective performance is shown in Figure. 5.12. From the figure and corresponding results shows a clear correlation between participants responses and their task performance. The current setting enables observation of levels of progression faster with a significant trial-trial performance variation. The testing and evaluation method presented here could be used as a framework to validate contemporary experimental platform.

5.7 SUMMARY

This chapter introduces an integrated experimental platform aiming towards advancing the exploration of team related questions. The PBR data collection presented in this chapter are investigated in the chapters-6, 7 and 8 to analyse any team related relationships. Specifically, the next chapter explores pre-processing of gaze data to investigate the effects of gaze convergence and distribution. Methods to pre-process gaze data and representation as a

team-level index is the main contributions of this chapter. Considering the three member nature of the data, the methods and findings presented in this chapter is one of novel contributions of the thesis.

Primarily, focusing on the area of research investigating physiological and behavioural relationships concerning the processes, qualities and emergent states in multiparty joint activity context. Such platforms are in demand where collaboration, cooperation and coordination critically decide the level of task achievement. Results and discussion suggest that the novel task introduced in this article qualifies as better alternative to entertainment based STEs as well as serves as a model for enhancing current practices. Coupled with the choice of PBRs, user-inputs and subjective feedback offers as a state-of-the-art model to explore relationships that concern team functioning.

6

Gaze Co-occurrence and Distribution

6.1 INTRODUCTION

How important is it to look together relative to looking apart? In collaborative problem-solving tasks, there is a mixed opinion about the relationship between gaze behaviours and task performance. As mentioned in sections 2.2.2 and 5.3.3, studies exploring gaze behaviours in socio-dyadic contexts report a positive effect on performance. Contrastingly, work teams and tactical scenarios with complex problems have reported a negative effect on performance. Solely, focusing on group problem-solving in natural settings, this chapter aims to investigate the effect of gaze co-occurring and distribution of gaze instances on performance. This analysis explains how measures of gaze behaviours i.e., co-occurrence and distribution could play a vital role in the successful achievement of task objectives. The study investigates whether such measures provide appropriate features for a predictive model.

The information push-pull mechanism (section 5.2.2) stimulates gaze co-occurring events. On the other hand, dynamics of the task such as speed of BT, the span of the map, sequential spawning of UGVs, and their spawn and travel patterns demand vigilance from teams. Thus, stimulating the distribution of gaze to achieve better coverage. Members must intelligibly exchange spatial references g and u , and distribute work to maximise their performance. In this light, methods to quantify gaze co-occurrence and distribution for 3-member teams is introduced in the chapter. The effectiveness of the method is evaluated by analysing the relationship between task performance and visual coordination measures. Pearson's correlation coefficient is used to quantify the relationships and also serves as important predictive features. A multivariate regression is attempted to achieve a composite index combining measures from both gaze co-occurring and distribution.

6.1.1 Pre-processing and Data Selection

Following analysis was conducted to observe if there were any biases in the gaze data due to learning and training development effects.

Selection of Data

Gazepoint GP3 was used to record the eye movements of members at 60 samples/second after performing a 9-point calibration. From the gaze point analysis software time-series point of gaze (BPOG) in $\{x, y\}$ coordinates were used for the remainder of the analysis.

After analysing the progression of team scores in trials (Figure. 5.13) and a summary of the feedback in Table. 5.4, the effect of learning could be inferred. Teams required trials 1 and 2 to gain an understanding of task objectives. A similar effect was previously observed by Argote et al. in a 3-member task setting with trials of 12-minutes duration [239]. Hence, the remainder of the analysis only uses data from trials 3 and 4. Further, on inspecting the data pool, 3-teams were removed due to technical error, where the gaze of at least 1-team member was not available.

6.1.2 Gaze data preprocessing

First, using the trial start and end time as a reference, member-trials raw time-series BPOG data were extracted for all teams. The data was next cleaned to remove any invalid data such that x-coordinates exceeding 1920, y-coordinates exceeding 1080, and any negative $\{x, y\}$ coordinates were removed and refilled using linear interpolation. The cleaned data was then up-sampled to 4-times the sampling using linear interpolation, before down-sampling to 10 samples per second using linear extrapolation. Thus, retaining enough gaze samples to analyse members eye-movement behaviours as shown in Figure. 6.1.

6.1.3 Paired-distance Estimation

The cleaned BPOG $\{x, y\}$ coordinates of member trials were used to calculate time-series gaze distance d between all possible pairs using the equation below.

$$\vec{d}_p(n) = ||M(x_n, y_n) - N(x_n, y_n)|| \quad (6.1)$$

where, n is the discretised timestamp, $p \in \{(Alpha, Bravo), (Bravo, Charlie), (Alpha, Charlie)\}$ representing all possible pairs in a team, and M and N represent corresponding parties of any pair p . For example, time-series gaze distance between a

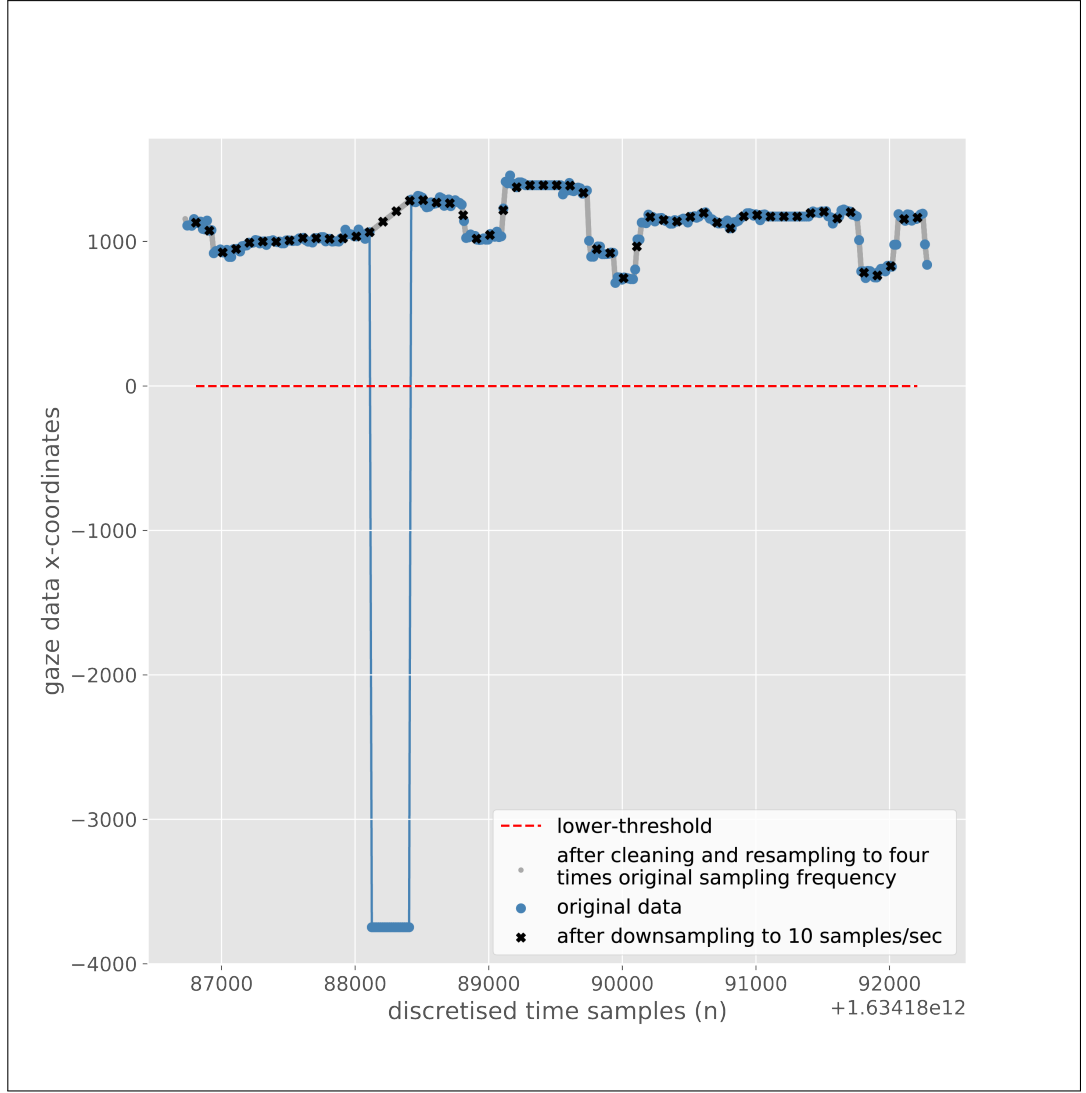


Figure 6.1: Example showing a short excerpt of original x-coordinates of a gaze-data, threshold used to decide invalid coordinates, high resolution cleaning and resampling and finally, the down sampled data used for further analysis.

pair, $p = (Bravo, Charlie)$ at a discrete-time n can be estimated as, $\vec{d}(n)_{p=(Bravo, Charlie)}$, and M and N in the equation represent instantaneous BPOG coordinates of Bravo and Charlie, respectively. Ultimately, the trial of a team was represented using three vectors, one for each possible pair in p .

6.1.4 Gaze Co-occurrences

Gaze co-occurrences were accounted for by estimating the number of pair-classes i.e., any pair, any two-pairs and all three-pairs co-occurring within a certain distance r . Thus, co-occurring instances were estimated for all distances ranging between 5 and 2100, with an increment of 50. Here 5 and 2100 correspond to the shortest and furthest possible distance, respectively, between the BPOG of a pair. Note that the display size used for this study

was 1920×1080 , and hence, the maximum possible distance between a pair was 2202.

Thus, for any distance d , the gaze co-occurrence g_c of a team trial was represented using four values. For example, $g_c = [627, 2686, 831, 2706]$ indicated that within a certain distance d , there were 2686, 831, 2706 number of instances that any pair, any two pairs and all three pairs operated, respectively and on 627 instances gaze of no pairs co-occurred.

6.1.5 Distributed Gaze

Similarly, distributed gaze was accounted for by estimating the number of pair-classes operating outside of a certain distance d . The distributed gaze was also estimated for all distances ranging between 5 and 2200 with an increment of 50. For a certain distance d , the distributed gaze g_d of a team-trial was also represented using a vector of four variables. For example, $g_d = [1075, 365, 4078, 133]$ indicated that a team trial had 365, 4078, 133 instances that any pair, any two pairs and all three pairs had BPOG outside distance d , respectively, and there were 1075 instances where the gaze of no pairs was observed outside the distance.

6.2 RESULTS & DISCUSSION

6.2.1 Analysis of Operating Conditions

Pearson's correlation was performed to identify the best operating distance. Hence, correlation coefficients were calculated between pair-classes and their task performance (BE), for all distances mentioned above. Plots corresponding to gaze co-occurrence and gaze distribution are shown in Fig. 6.2, respectively. Although, here a range of pixel distances was considered for the analysis, however, only fewer distances produced a significant effect on performance. In addition, a skewness test was adopted to discard any data that exceeded the limits $-0.5 \leq x \leq 0.5$, where x indicates the skewness test score. Thus, ensuring that only normally distributed and significantly effective data were considered, as indicated by bubble markers in the plot. The selected data, corresponding radius and the decisive parameters are summarised in Table. 6.1.

Following results were observed with relation to the current 3-member challenge:

Significance of Gaze Distribution

The highest positive significant effect of the performance was observed if any one-pair operated outside 600-pixel distance. Contrastingly, if any one-pair, any two-pairs and all three-pairs operated outside 1350 and 100-

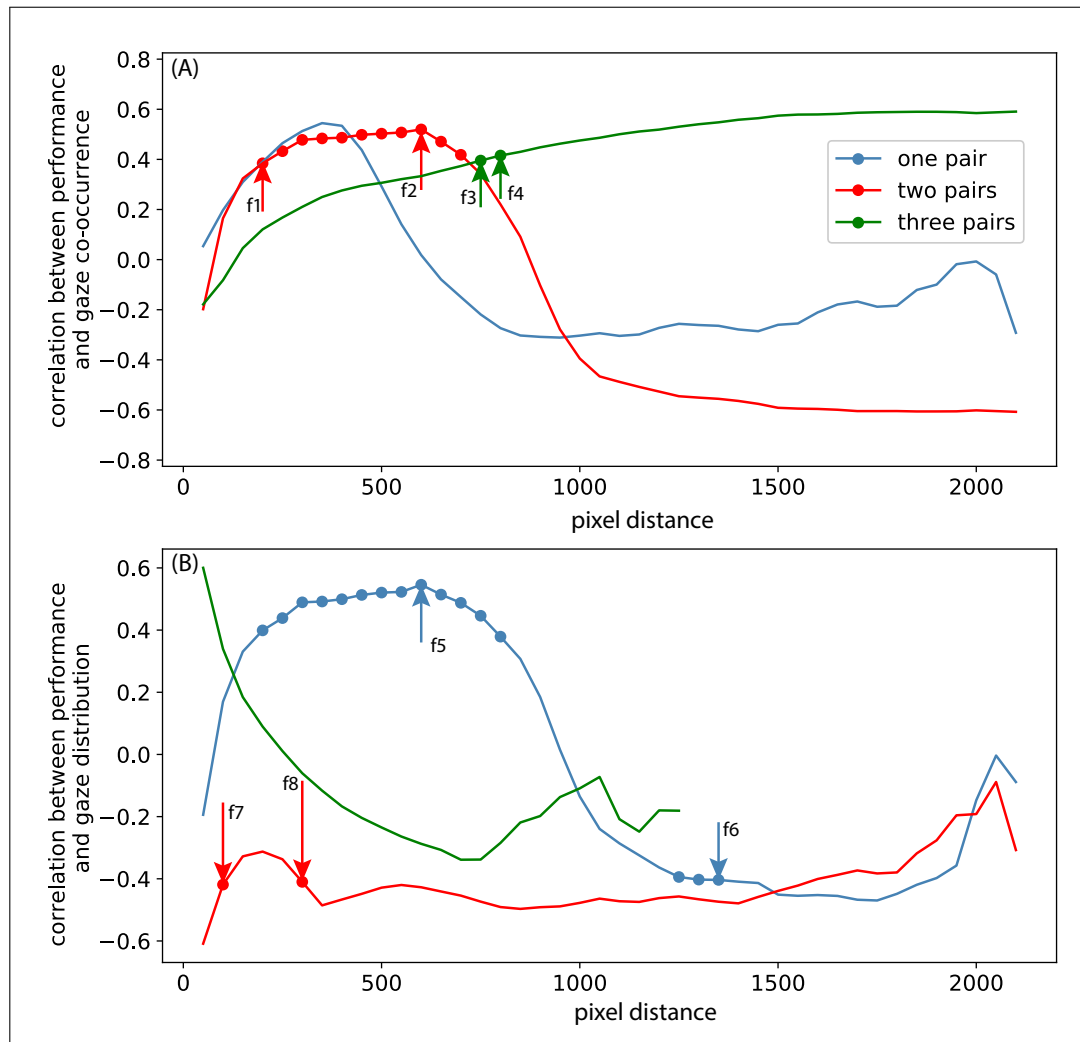


Figure 6.2: A plot of correlation coefficient's between gaze-behaviour vs distance of team trials. (A) gaze co-occurring behaviour and (B) distributed gaze behaviour, for all pixel distance between 50 and 2100 with an increment of 50-pixel distance. **Note:** The curve for three pairs in gaze distribution analysis (B) comprised of invalid correlation scores because the corresponding data points comprised of all zeros. This was because no team had all three pairs that outside 1350 pixel distance.

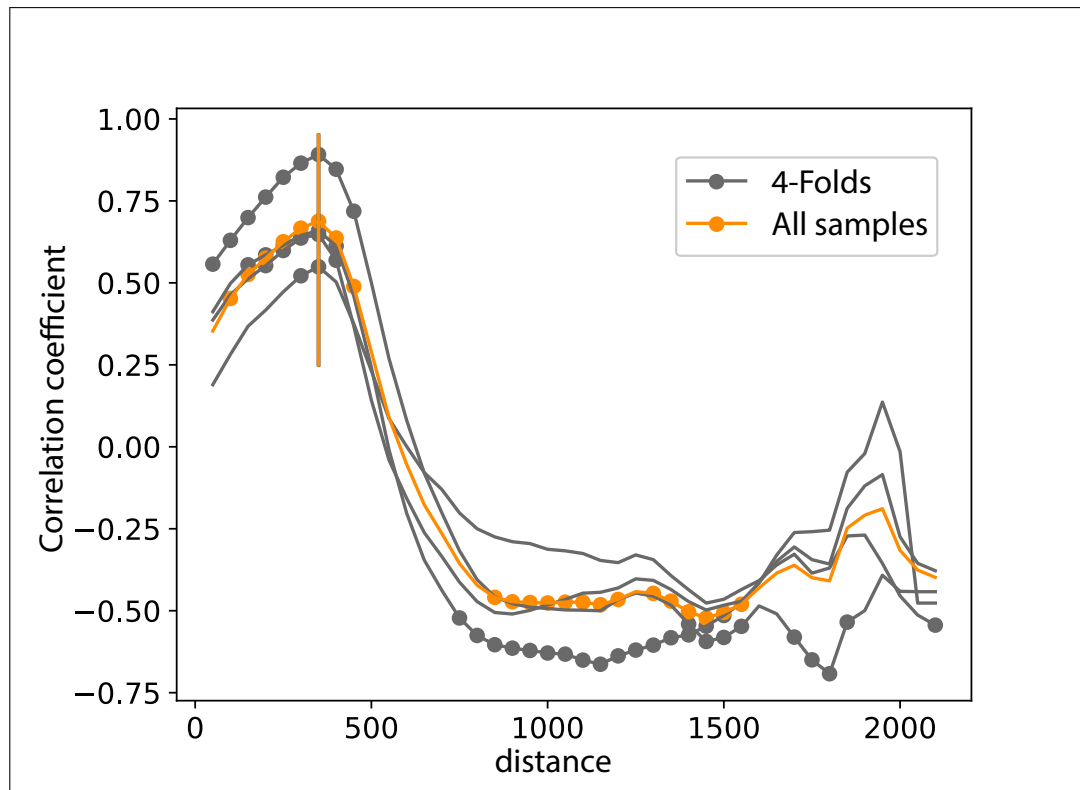


Figure 6.3: Example showing plot of correlation coefficient vs distance for one-pair gaze co-occurrence. For comparison the plot shows responses using all samples and 4-folds from the sample.

pixel range’, respectively resulted in a significant negative effect of the performance.

This effect reflected the visual span of teams, operating at a distance that ensured better coverage of the map. Often, teams arrived at such a decision between the trial episodes when they could regroup and form a strategy. However, only fewer teams managed to maintain their positions resulting in a good performance. Whereas, poor performing teams were distracted due to task pressure (more incoming tanks and no organisation), never arrived at a strategy, or demonstrated no interaction between trial episodes.

Significance of Gaze Distribution

The highest positive significant effect of the performance was observed if any one-pair operated outside 500-pixel distance. Contrastingly, if any two-pairs and all three-pairs operated outside 400 and 1000-pixel range’, respectively resulted in a significant negative effect of the performance.

Multi-variate Regression

The observed significant features reported in Table. 6.1 were used to estimate a multivariate regression equation. The rationale was to observe any improvement in performance due to multiple features. Specifically, this study

investigated whether combining measures from gaze co-occurrence g_c and distribution of gaze g_d improve predictive capability. In this light, all possible 247 combinations i.e., $\{(f1, f2), (f1, f3), \dots, (f1, f2, f3, f4, f5, f6, f7, f8)\}$ were tried. Although, most combinations delivered a significant correlation, however, no combination outperformed the highest observed *adjusted* r^2 of 0.58 ($f3$ in Table. 6.1).

Table 6.1: A summary of results for the gaze behaviour measures identified in Fig. ??

Graph reference	pixel distance	Gaze behaviour	pearson-r, p-value	adjusted r^2	skewness
f1	200	g_c	0.38, p=0.04	0.14	-0.07
f2	600	g_c	0.52, p=0.004	0.27	0.02
f3	750	g_c	0.39, p=0.03	0.15	-0.38
f4	800	g_c	0.41, p=0.02	0.17	-0.44
f5	600	g_d	0.54, p=0.002	0.29	0.02
f6	1350	g_d	-0.40, p=0.03	0.16	0.49
f7	100	g_d	-0.41, p=0.02	0.17	-0.05
f8	300	g_d	-0.40, p=0.03	0.16	0.23

Multi-variate Regression

The observed significant and normally distributed features reported in Table. 6.1 were used to estimate a multivariate regression equation. The rationale was to observe any improvement in performance due to multiple features. Specifically, this study investigated whether combining measures from gaze co-occurrence g_c and distribution of gaze g_d would improve predictive capability. In this light, all possible 248 combinations i.e., $\{(f1, f2), (f1, f3), \dots, (f1, f2, f3, f4, f5, f6, f7, f8)\}$ were tried. The results demonstrated that a best effect of performance ($F(2, 25) = 7.580, p = 0.002$) with an r^2 of 0.37 and adjusted- r^2 of 0.33 was achieved using the features ($f4, f8$). The relationship between ($f4, f8$) with performance could be visualised in Fig. 6.4. Accordingly, the estimated performance can be formulated as $BE = -0.135 f4 + 0.062 f8$.

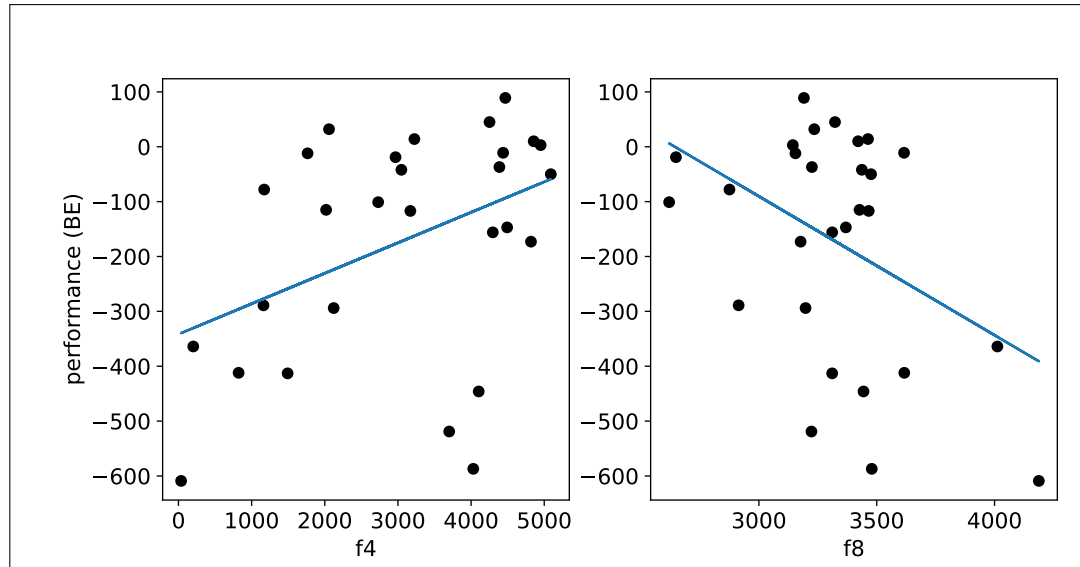


Figure 6.4: Plot demonstrating the relationship between gaze co-occurrence feature f_4 and gaze distribution feature f_8 with task performance BE.

The above analysis indicates that there were instances during the task when all three pairs were required to operate under 800-pixel distance and any two pairs to operate above 300-pixel distance. The corresponding values in the plot (Fig. 6.4), further confirm that overall allocation of gaze would result in higher gaze co-occurrence and lower gaze distribution, collectively impacting task performance.

6.3 DISCUSSION

This work aimed to investigate the importance of gaze co-occurring and distribution instances in relation to performance. To facilitate our investigation the gaze data collected from 3-member multi-party joint collaborative activity (Chapter. 5) was chosen. Objectives of the task resembled natural work settings demanding cooperation, coordination and intelligible communication for successful performance. Eye movements of participants were collected and analysed off-line.

A thorough analysis of task performance scores and subjective feedback enabled the identification of trials with learning and experience development effects. This observation was in line with the findings previously reported by Argote et al. [239]. The employment of Euclidean distance to establish a pair-gaze relationship was in line with previous studies [214, 221, 250]. The time series pair-distances were later transformed to compute gaze behaviour measures, co-occurrence g_c and distribution g_d . Via co-occurrence, the method accounted for instances demanding joint attention. The distribution, on the other hand, accounted for better visual coverage demonstrating team awareness and vigilance. The correlation results suggest that both gaze co-occurrence and distribution impacted during

the task. This finding was in line with Andersen *et al.*'s observation [216]. Further, our attempts to arrive at a composite score was valid, since combining g_c and g_d measures using multivariate regression significantly improved performance prediction capability.

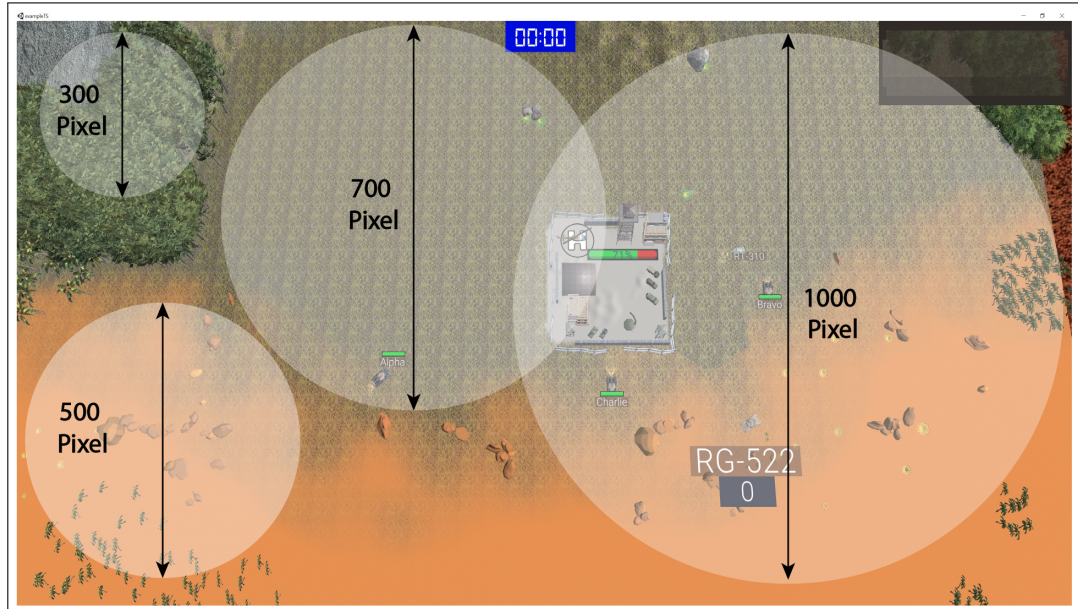


Figure 6.5: Example demonstrating an impression of 300, 600, 800 and 1000 pixel diameter circles on a 1920×1080 scale LCD

In addition, the study by Tolston *et al.* suggested a radius of 250 pixels between a dyad achieved a significant relationship with performance. However, no rationale was explained in the study for choosing the radius. In this light, this study observes that when the distance between gaze coordinates of a dyad was less than 600 pixels, resulted in a significant positive correlation with performance. Observe Fig. 6.5 to visualise various pixel distances on a 1920×1080 LCD. Moreover, the best effect of the performance was observed when the gaze of all three members was within 800 pixels.

Further, according to the input-mediators-output team study framework, task attributes contributes to the inputs [3]. The framework is a cyclic process [4] where outcomes could be used to adjust inputs. In this light, this study presents a platform enabling the understanding of task attributes. Also, the Tank Battle Task presented here allows modification of task settings thus, this analysis method and task environment complement the IPO framework.

6.4 SUMMARY

According to task settings defined in this study, gaze co-occurrence and gaze distribution were evident, however, independently contributed to performance. Although such a gaze behaviour was previously declared

as task-specific, the evidence presented here highlights its importance in the evaluation of team performance. Future research could focus on the exploration of other gaze behaviours and attempt deriving composite index, which could aid better team performance analyses and predictive aids.

Having established the concepts of co-occurrence and distribution in this chapter, and learning the best operating radius, the following chapter explores the impact of gaze convergence on performance. The next chapter takes a step closer by implementing a neural network model in the form of back propagating algorithm to resemble a logistic regression problem. Lastly, statistical analysis is employed to compare the performance of neural network model and its equivalent multivariate regression model using K-fold cross validations.

7

Gaze Convergence Using Neural Network

7.1 INTRODUCTION

Recent studies in the space of collaborative and multi-party joint activity context have demonstrated a growing interest in the development of predictive models utilising eye-movement based measures [209, 210, 217]. Although, literature demonstrate a relationship between members' gaze and performance, however, the practicality of application in larger groups and intelligent predictive modelling has slowly progressed. One common constraint to this is that much of the studies were in dyadic team setting, and when larger groups are considered studies have resorted to a laborious examination of videos. Further, in the context of prediction, regression and binary classification are observed in the literature. Thus, indicating that the selection of appropriate team members gaze features can aid in the development of predictive models. This chapter aims to advance gaze-based collaborative research by overcoming some of the drawbacks. As mentioned in section. 5.3.3, the hypothesis described by Cherubini *et al.* [221] is revisited here, however, with a different controlled task setting. The promising results observed in previous Chapter, and digging deeper into some of notable studies [209, 210, 214, 223] mentioned in section. 5.3.3 formed a strong rationale for the current investigation.

To this end, literature has established that multi-party joint activities such as CSCW demands intelligible communication for efficient coordination and cooperation. As a result of coordinated interactions, the gaze of collaborating members aligns to form patterns [133]. Typically, paired social contexts such as parent-child, teacher-student and doctor-patient exhibit joint attention, where members share similar ROI [67]. For example, in a speaker-listener task setting, the gaze of good-listeners was observed to follow the interests of the speaker [214]. Coco and Dale [214] employed RQA to quantify gaze recurrence between dyads and observed a positive correlation with good-listeners. Contrastingly, in a visually dynamic tactical scenario, such as a reconnaissance task, where one person was tasked to fly unmanned

aerial vehicle, while a partner was tasked to take pictures, Tolston *et al.* observed that gaze recurrence between dyads demonstrated a negative effect on performance [21, 81]. Employing RQA-based approach required x and y screen coordinates to be separately analysed. Such limitations discourage their expansion in natural settings such as radar monitoring or crowd monitoring [211] where ROIs are arbitrary. In addition, methods employing epistemic network analysis rely on predefined ROI [133] which are used as spatial-references for analysing co-occurrences. Such context demanded non-recurring gaze events for better coverage of the visual space. Overall, such advancements hold a significant importance in defence [32], aviation [208, 219], health-care [220] and sports [211].

Advancing beyond objective evaluation models, this study seeks ANN methods to predict performance using gaze behaviours of team members. Euclidean distance is used to establish gaze convergence. Initially, linear statistics are explored to analyse relationship between gaze convergence and performance using Pearson's correlation. The results are used to identify suitable features for modelling multivariate regression equation. Further, a ANN in the context of structured data [251] was modelled. The performance of the neural network was verified using K-fold cross-validation running 30 different permutations. The mean absolute errors both approaches are compared revealing promising results and encouraging the employment of team members' gaze behaviour for objective assessment and prediction of collaborative performance.

Gaze Data

The following analysis uses best-point-of-gaze (BPOG) measure $\{x, y\}$ co-ordinates of participants' gaze. Irrespective of screen dimension, BPOG is available in the normalised format ranging between 0 and 1.

7.1.1 Data Selection

From the feedback summary in Table. 5.4, it could be understood that most information was conveyed in the introductory video and practice session, however, participants required trial-1 and 2 to perform comfortably. Thus, observing task performance and feedback, it can be inferred that trials 1 and 2 demonstrated adaptation and experience development effect. In addition, trials 1 and 2 of some teams also suffered technical issues such as server freezing, resulting in uneven data samples corresponding to trials. Thus, to overcome technical issues and to avoid any biases due to learning, only data of trials 3 and 4 are considered for the remainder of the analysis. Further, on inspection of gaze data, three teams were discarded due to technical issues that resulted in gaze missing from at least one member of a team.

Thus, at the time of this analysis there were a total of 20 team-trials data; 10-teams with 2 trials in each.

7.1.2 Gaze Data Pre-processing

A similar process as mentioned in section. 6.1.2 was followed, however, the raw BPOG information was utilised. First, BPOG in (x,y) coordinates were extracted for all member-trials by matching timestamp information of gaze-data with the start and the end timestamps of task trials. Next, the data was cleaned to remove any invalid coordinates, due to blinks and eye-tracking missing pupils. This was achieved by removing invalid coordinates from the raw data followed by interpolation using original timestamps. Fig. 7.1 shows an example where x-coordinates of raw gaze-data outside the normalised width of 0 and 1 is removed. Then for further analysis, the cleaned and resampled data were first interpolated to with 4-times the original sampling frequency and down-sampled to 10 samples per second. The same process was also adopted to clean and generate y-coordinates of member-trials gaze data, by removing any height measures outside the normalised range of 0 and 1.

7.1.3 Paired-distance Estimation

In the light of joint attention, studies have relied on Euclidean distance for quantifying concepts such as gaze co-visitation [214], gaze recurrence [250] and gaze convergence [221]. Specifically, in desktop-based settings such as reconnaissance task [81], CSCL and CSCW [250]. Hence, this analysis demonstrates the employment of Euclidean distance to quantify group gaze behaviours. Next, time-series Euclidean distance d between gaze coordinates was estimated for all possible pairs within teams as shown in eqn. 7.1.

$$\vec{d}(n)_p = ||M(x_n, y_n) - N(x_n, y_n)|| \quad (7.1)$$

Thus, every team-trial was represented using three vectors $\vec{d}(n)_p$, where, n is the discretised timestamp, $p \in \{(Alpha, Bravo), (Bravo, Charlie), (Alpha, Charlie)\}$ representing all possible pairs, and M and N represent corresponding parties of any pair p . Example, the time-series distance vector between gaze coordinates of a pair *Alpha* and *Bravo* can be represented in eqn.7.1 as $\vec{d}(n)_{p=(Alpha, Bravo)}$, and M and N in the equation corresponds to time-series BPOG coordinates of *Alpha* and *Bravo*, respectively.

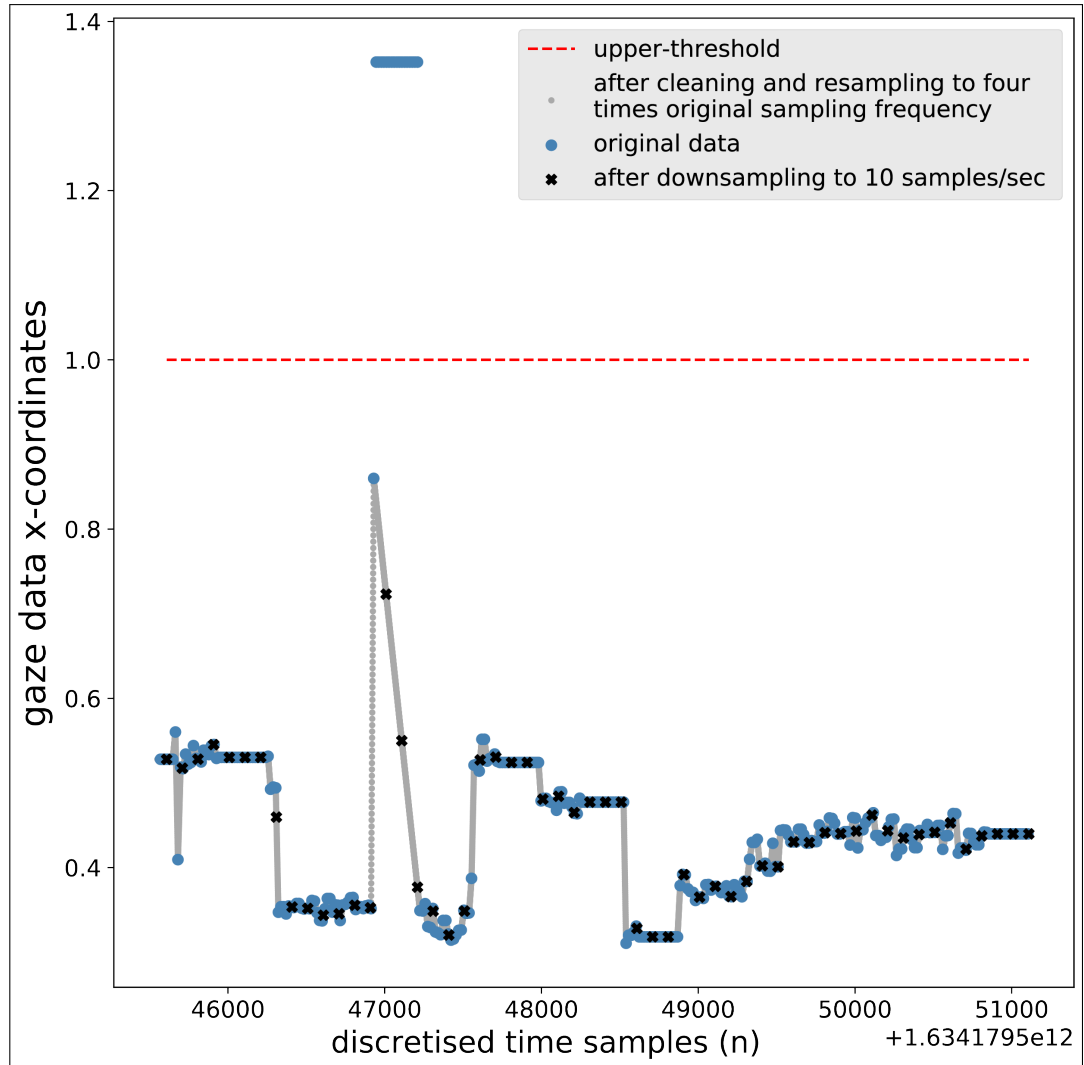


Figure 7.1: Example showing a short excerpt of original x-coordinates of a gaze-data, threshold used to decide invalid coordinates, high resolution cleaning and resampling and finally, the down sampled data used for further analysis.

7.1.4 Neural Network Prediction Model

Since, the gaze measures and the task performance form structured data, in this exploratory approach a standard multilayer back-propagating neural network was chosen to generate a predictive model. A similar architecture was previously evaluated for structure data, i.e., prediction of housing prices [251]. Tensorflow keras libraries were used to develop the neural network.

A prediction problem follows nonlinear regression function, hence, rectified linear (*relu*) and hyperbolic tangent (*tan h*) activation functions were tried for input and hidden layers. The output was activated using a linear function to ensure better numerical range compatibility. The concept of *mean squared error* (MSE) and *mean absolute error* (MAE) functions were tried to achieve loss function. To ensure better performance, both adaptive

momentum (adam) and rms prop were tried. Since normalised coordinates were employed, the resulting d_p of pairs were also normalised. Task performance scores were also normalised such that the maximum value in the pool represented 1 and the minimum value represented a 0.

7.2 RESULTS

7.2.1 Significance Between Operating Distances and Performance

Pearson's correlation was performed to investigate whether there was any significant relationship between operating distances and performance (BH). The analysis demonstrated a significant negative correlation between average of gaze distances (μ) and BH ($r = -0.55$, $p = 0.01$) and between standard deviation (σ) of gaze distances and BH ($r = -0.69$, $p = 0.0005$).

Following the significant relationship, a multivariate regression was performed and found ($F(2, 17) = 8.278$, $p = 0.003$) with an r^2 of 0.49 and mean absolute error (mae) of 0.13 as shown in Fig. 7.2. That is, teams predicted trial BH = $0.9149 - 0.35 \mu(d_p) - 5.335 \sigma(d_p)$, measured in normalised pixel distances.

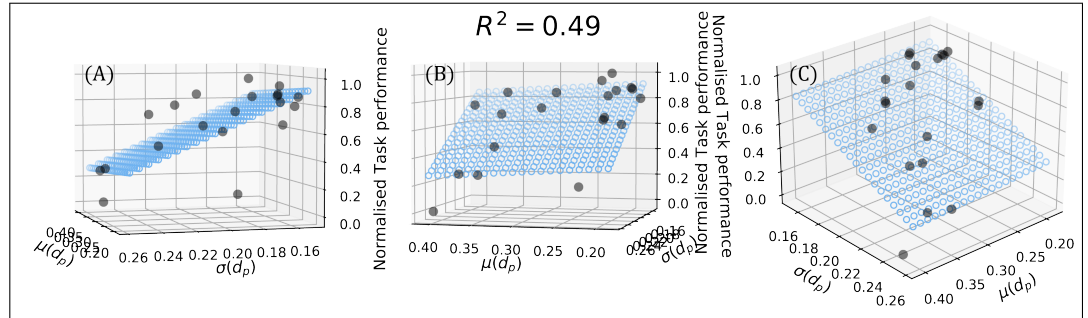


Figure 7.2: Plot showing multi-variate regression utilising average pair-distance ($\mu(d_p)$) (B,C) and standard deviation ($\sigma(d_p)$) (A,C) contributing to the normalised task performance (BH).

7.2.2 Performance Prediction

The measures μ and σ of d_p used in multivariate regression were chosen as features to train the neural network. The training was validated using K-fold validation with 4 splits, meaning for every 15 training samples there were 5 testing samples. The validation process was repeated for 30 different permutations to compensate for the low sample size and observed an average MAE of 0.06 with a standard deviation of 0.02 as shown in Fig. 7.3.

The neural network produced the least MAE when using two hidden layers with 1-neuron in each hidden layer, loss function = MAE and *RMSprop*

learning optimisation. There was no significant difference observed in the MAE when using *tanh* or *relu* activation functions.

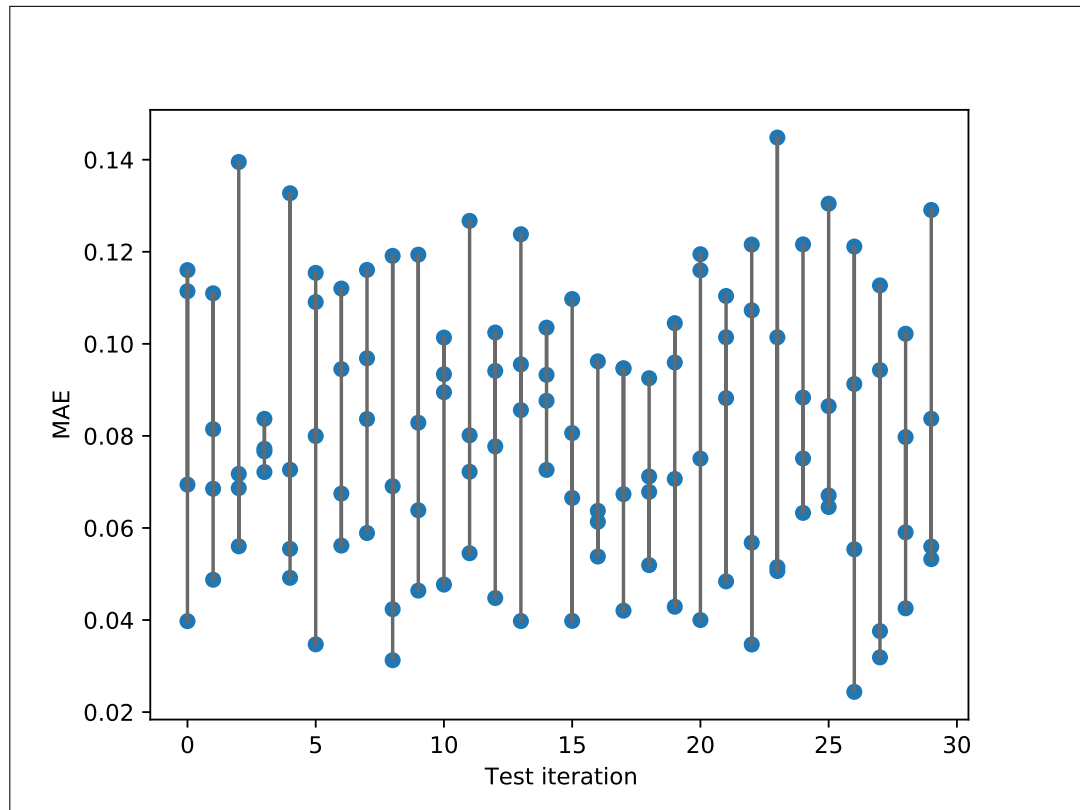


Figure 7.3: K-fold testing used for validating the neural network model. The plot shows MAE observed for 4 folds with 30 different permutations

7.3 DISCUSSION

This work aimed to promote the integration of eye-tracking for the objective prediction of group collaborative performance. A 3-member team task environment (Fig. 5.3) was used to facilitate the evaluation process. Intelligible communication via exchanging spatial references was key for better performance. Features based on linear statistics were explored to model a neural network as a predictive aid.

Analysing task performance scores in Fig. 5.12 and participants feedback in Table. 5.4 aided in elimination of data that could be contaminated from learning and experience development biases. Similar observations were made by Argote et al., where trials 1 and 2 in a 3-member task with 12-minutes trial duration had demonstrated a gradual rise in performance, and following trials demonstrated steady performance [239]. In this light, the duration of trials adopted in this study was 11:43-minutes and trials 1 and 2 demonstrated a similar effect.

This study demonstrated a gaze processing method to clean raw-gaze data (Fig. 7.1) and team-level representation using euclidean distance. By

considering instantaneous gaze locations, the method accounts for dynamic changes without relying on fixed ROIs. Thus, enabling a robust application of eye-tracking technologies relative to previous methods. As opposed to the need for identification and defining suitable ROIs mentioned in previous approaches [214, 240], the method presented here accounts for dynamic ROIs. Further, in line with the previous methods [221], Euclidean distance between participants gaze was used to objectively establish gaze-based operating dynamics. The method also simplifies computations, where previous applications [81, 250] had analysed x and y coordinates separately.

The significant correlation between the average pair-distance $\mu(d_p)$ and task performance scores indicate that during any trial if the average distance between the gaze of team members was smaller the better was the performance (BH). From the dynamics of the task, this observation relates to instances when the gaze of members converged to determine whether a UGV was an enemy or friend. Since there was a new UGV spawned every 7-seconds, the task demanded frequent gaze-convergence to achieve better performance. The correlation between standard deviation $\sigma(d_p)$ and BH further indicate that better performance was a result of organised gaze behaviour where participants responded better to each other requests. In the context of the task, this observation indicates a better gaze convergence due to participants responding quickly, intelligibly exchanging spatial reference and causing minimal confusion. There was no change in the significance and correlation effect using raw BH scores. However, to minimise the effects of biases and variances resulting from limited data samples, normalised distances are recommended for training neural networks

Coupling the average pair-distance $\mu(d_p)$ and the standard deviation $\sigma(d_p)$ in multivariate regression further strengthened the hypothesis that gaze convergence between pairs has a significant effect on performance. The results were in line with previous observations by Tolston *et al.* [81], Strang *et al.* [21] and Coco and Dale [214]. As opposed to previous nonlinear methods such as RQA, here linear statistics in the form of *mean* and *standard deviation* proved effective.

In an attempt to advance gaze-based application beyond classification [209, 223] and regression problem [210], a simple backpropagation neural network is modelled in the study. Comparing the results i.e., MAE of multivariate regression and with that of neural network clearly shows that the latter produced better results and the observed standard deviation demonstrates the consistency of the model to produce good results. The nature of the neural network resembles logistic regression, however, the training and learning enhance the prediction capability.

Further, eye-tracking in general offers excellent alternatives to traditional human intervened assessor rating approach. Such approaches are still practised in the context of CSCL and CSCW to rate proceedings either by examining video [82, 83, 240] and scrutinising in person [216, 252] as ob-

served in the assessment of team situation awareness. The work presented here could advance the research domain via employing gaze based measurements in a multi-party setting. The need for such an approach has been previously emphasised in team context literature [216].

Some studies had previously raised concerns that dyad research could not transformed to group and multi-party settings [84, 240]. In this light, studies have opted for 3-member task setting [82, 240], however, had to resort for laborious video examination. Via adopting a 3-member task environment, this study advances gaze-based collaborative research beyond conventional dyadic task settings observed in previous studies [79, 81, 133, 214, 250], yet the methods achieve objective prediction capability.

7.4 SUMMARY

The primary motivation of the study was to advance gaze-based collaborative performance prediction research beyond regression models. The significant relationship between gaze behaviours and performance enabled the identification of suitable features to train a neural network. The results provide a solid groundwork for future studies exploring ANN-based performance prediction in this research domain. Utilising gaze-based dynamics, this work can help minimise human intervention based assessments, alternatively, can also act as an assessment-supporting tool.

The next chapter explores the main contributions of this thesis via utilising the cardiovascular and gaze data. The chapter explores the effects of cardiovascular and gaze-based measures for analysing relationship with PBR on performance, and their effects on the development of team cognitive processes. Methods to quantify indices of team cognitive states, members-compliance in cognitive load and gaze based team-level measures are explored. Ultimately, relationship between the indices and performance is analysed using multivariate regression.

8

Multimodal Index and Effects of Team Cognitive Activity

8.1 INTRODUCTION

The literature review in section. 2.5 suggests that the employment of multiple modalities offer advantages over single modality based analyses. In the context of teams, research has employed a range of PB measurement techniques for multimodal analyses [35, 50, 79, 80, 120, 121]. Gorman employed ECG and voice communication; Neubauer *et al.* analysed [9] visual activity, linguistic measures and HRV; Cardiac activity and gaze behaviour has been previously analysed in dyadic settings to observe emotion and convergence of collaborating members [79]; and the activity of ANS and visual behaviour has been analysed using EDA and video analysis to observe self-regulation in collaborative learning [82, 253]. Such analyses enable better insights of team members affect, cognition and behavioural processes underlying different levels of performance. However, at team level the employment of multi-modality based analyses has lacked strong empirical evidence. Despite the advantages offered via the concurrent measures from multiple reactions of humans, however, data fusion attempts to unify information from multiple modalities has received less attention in team studies.

To this end, methods and techniques to assess group activities such as collaborative problem solving is highly emphasised in this modern day education and work settings [5–8, 82, 96, 242, 253]. Advancements in smart-devices and embedded technologies has prompted team research to employ wearable and unobtrusive technologies to unfold the physiological and behavioural (PB) dynamics underlying performance of teams [4, 254]. As a result, team research in recent years has witnessed a gradual rise in the number of studies employing multiple modalities to evaluate processes of teams [36, 43, 45, 46]. Multi-modal approach are known to offer benefits through additional supporting information, act as better alternatives and enable metadata PB analyses from multiple human responses [46, 79, 115]. There is an increasing demand for studies to explore unified analyses

combining information from more than one physiological and behavioural modality.

Hence, this Chapter aims to present data processing and analysis procedure enabling better fusion of more than one modality for objective evaluation of team performance. The analysis presented here explores the feasibility of employing team members ANS activity and gaze behaviours to unfold physiological and behavioural dynamics concerning task performance.

During cooperative and collaborative activities, members of a team develop team-processes such as mutual understanding and shared mental models [5]. And, in a cognitively challenging activity, the demand experienced by members of a team correlates spatially and temporally [81, 105]. In education settings, the knowledge and the experience of students are relatively similar. As such, academic tasks based problem solving presents a homogeneous context and physiological correlations can be observed in the form of temporal synchrony [35, 96, 253]. However, in a heterogeneous context, the cognitive demand experienced by members of a team are different. For example, during surgical procedures demands experienced by a team of surgeon, perfusionist and anaesthesiologist varies from one another [105]. Hence, physiological activations are highly-likely to form distinctive patterns. Albeit, team literature suggests that, such developments are reflected in the activities of human nervous system [89, 105] and behaviours of humans, that could be employed as an objective index of overall task performance [66, 81].

Homogeneous conditions are not just restricted to academic group tasks, but also could be observed during training the teams of defence personnel, paramilitary and health care professionals, respectively. Recent literature concerning collaboration in homogeneous conditions have mostly analysed EDA and video [80, 82, 105, 242, 253]. Relatively less is observed to employ heart rate variability for analysing team-related hypotheses [35, 96]. In a heterogeneous context, Dias demonstrates the employment of heart-rate variability based measures to observe distinct patterns in team members cognitive load [105]. Such observations could also benefit homogeneous conditions.

Spatial and temporal correlation could be observed in members behaviours such as visual activity [63, 81, 82, 255]. The importance of gaze-measures has been established in section. 5.3.3 and previous two chapters. Further, literature review Chapter. 2 suggests a range of team-level measurements including Recurrence between gaze of collaborating members measured using cross-recurrence quantification analysis (C-RQA) [81] and epistemic network analysis [255]. In this light, analysing the relationship between team gaze compliance and team task performance in a complex training and learning context encourages the development of diagnostic methods for evaluating teams [256].

Collectively, eye-tracking and chest-strap based cardiovascular monitoring devices provide wireless, unobtrusive and sensitive methods for evaluating team cognition [81, 203]. Hence, this analysis attempts the employment of gaze-based dynamics and IBI to evaluate performance of three-member team based collaborative activity in a homogeneous context. Further, in the light of multimodal analysis, this study aims to develop a unified model utilising both IBI and gaze measures. Ultimately, exploring the effectiveness of such a model as an objective measure of performance.

This study investigates whether:

- IBI could be used to observe the relationship between team cognitive load and task performance
- IBI could be used to observe the relationship between team physiological compliance and task performance
- IBI provides any valid markers of team cognitive processes
- gaze of team members could be used to observe relationship between co-occurrence of members gaze activity and task performance.
- gaze activity such as fixation and saccades of team members demonstrate relationship with task performance.
- gaze-based measures offer any valid markers of team cognitive processes
- Finally, the article evaluates whether a unified model combining the analyses of cardiovascular activity and gaze could improve correlation with task performance.

8.1.1 Experiment

Successful achievement of the objectives of the tank battle task demands collaboration, cooperation and coordination among three members in team. Physiological and behavioural activity of the participants recorded in the form of cardiac activity and eye-gaze are analysed.

8.2 METHODS

8.2.1 Processing IBI

The timestamps of IBI were matched with the starting and ending timestamps of a task trial. Thus, IBI series was extracted for every team-member-trial. This also ensured that the start and the end time references of all team members synchronised across trials.

As explained in Chapter. 3 and Appendix. A, heart beats do not follow a fixed rate. And since raw ECG was unavailable with the device employed in this study, the IBI data was cleaned and processed as follows. First, the raw IBI series of each team-member-trial was interpolated to four times the number of available samples using cubic spline interpolation. Then, the interpolated series was extrapolated to exactly 2-second interval [35,96]. Thus, the resampled IBI series was averaged to 2-second discrete and equally spaced samples [105]. This also eliminates any unwanted data that could arise due to IBI record exceeding 1000 ms duration. Following the resampling and cleaning process, there were equal number of samples across all members of a team-trial.

Team Cognitive States

The resampled IBI-series were then used to estimate team-level cognitive load using a similar method as mentioned by Dias *et al.* [105]. First, the IBI series of team-member-trials were transformed to cognitive load series using tercile transformation. Accordingly, if IBI_n represented IBI-series of a team-member-trial then its cognitive load transformation C_n was obtained using eqn. 8.1.

$$C(n) = \begin{cases} 3 & IBI(n) > q3 \\ 1 & q1 \leq IBI(n) \leq q3 \\ -1 & IBI(n) < q1 \end{cases} \quad (8.1)$$

where, $q1$ and $q3$ represented 25th and 75th percentile, respectively, of an IBI time series.

Next, cognitive load series of members are collectively analysed to obtain team-level cognitive-state series. At any discrete point in the series, the cognitive states of members can be collectively represented to form pattern as shown in Figure. 8.1.

This resulted in a possible of 27 patterns representing TCS as shown in Figure. 8.3. As such, every two second instance of a team-trial could have one of 27 possible patterns, and the trial of team could be represented using the series of TCS. The transformed series were used to derive three different measures as mentioned below.

Categorical Team Cognitive States

As opposed to distributed responsibility presented in the work by Dias *et al.*, the homogeneous nature of task means that the responsibility distribution in the current experiment is equal for all members of the team. Thus, at any instance a cognitive state pattern will be weighted irrespective of member position. For example, patterns in from the figure. 8.2 demonstrates two example scenarios, where scenario-one and two have same weight i.e

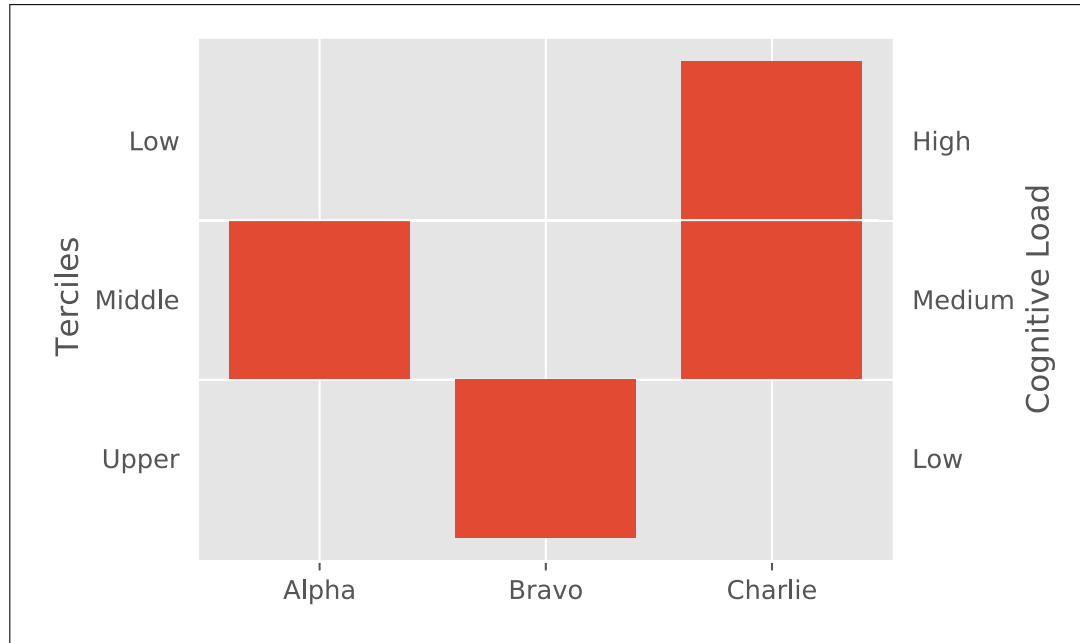


Figure 8.1: Tercile to cognitive load transformation to representation of a cognitive pattern.

3. Contrastingly, the original study would have weighted each scenario differently depending on responsibility on team members [105] where, relatively more importance was given to the surgeon. Or, in other use cases artificial neural networks were trained to assign weights to patterns depending on the frequency of their appearance during the task [45].

Accordingly, in the 3-member team-task, the pool of 27 TCS were categorised into one of the 7-level team cognitive load as show in the figure. 8.3. Then, the categorical-TCS (T) for a team-trial was estimated using a vector of 7-values such that the vector comprised of number of observations in each category of team load. Thus, the categorical-TCS of a team-trial was represented with a vector of 7-values. For example, if $\vec{T}$ of a team-trial was 12, 82, 146, 202, 147, 86, 10, it could be interpreted as:

Table 8.1: Example of interpreting categorical team cognitive load ($\vec{T}$) measures for a team-trial

Cognitive Load Level	-3	-1	1	3	5	7	9
Count	12	82	146	202	147	86	10

Cognitive Load Compliance

Similar to the previous approach, the members compliance between cognitive load were used as a categorical factor. Accordingly, the pool of 27 TCS were categorised into one of the 5-level team cognitive-load compliance

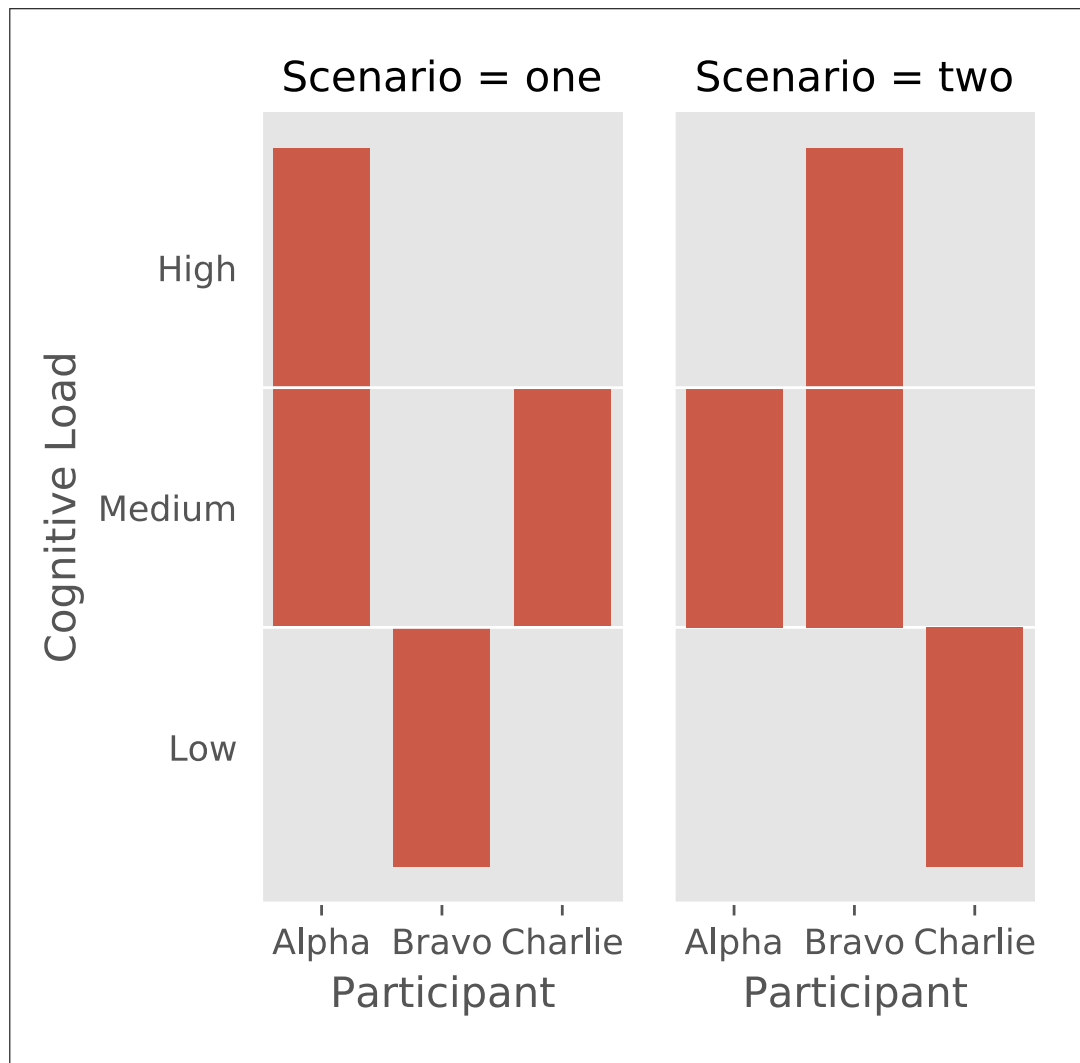


Figure 8.2: Basis for grouping team cognitive patterns to levels

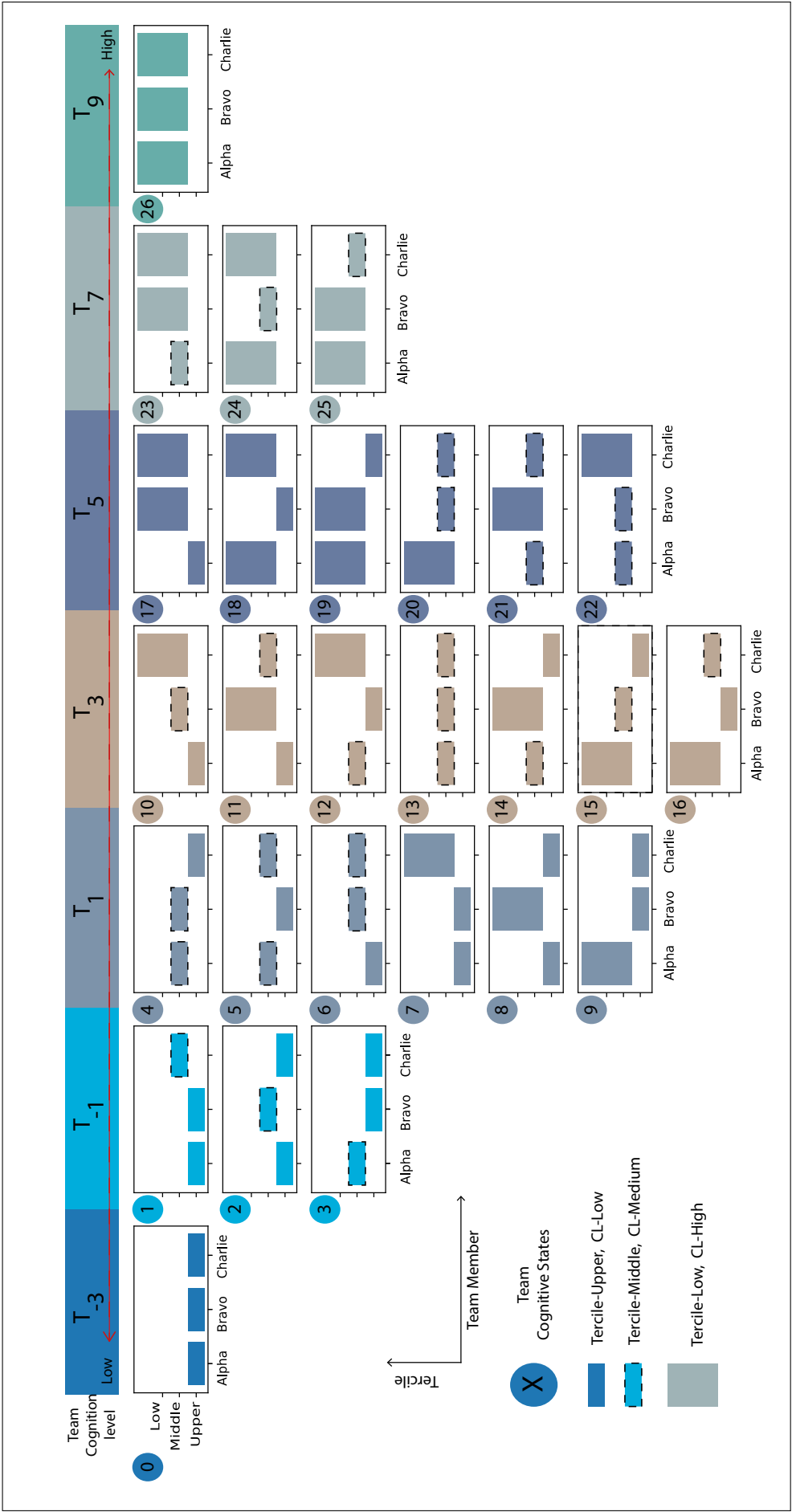


Figure 8.3: Detailed picture of seven levels of team cognition after categorising patterns of team-member terciles. From left to right describes most patterns with lower levels of team cognition moving toward higher levels of team cognition. The labels to the left of each plot indicates the team cognitive state labels

(LC) as shown in the figure. 8.4. Such that if all three members shared the same cognitive load then it indicated the state of high compliance, and when all three members shared different cognitive load then it indicated the low compliance.

8.2.2 Processing Gaze data

From the gaze point analysis software, the parameters gaze dynamic point-of-gaze (POG) and fixations identification number were used in the current analysis (see section. 5.3.3 for details). Similar to IBI data cleaning process, gaze of participants was also cleaned and pre-processed as shown in section. 6.1.2.

Average of Overall Fixations

As described in the section. 5.3.3, fixations accounted for instances where participants spent considerable time to stabilise on an image and retrieved information. In the current scenario, a part of events contributing to fixations was during communicated events, increased situational awareness such as trying to find nearby BTs and learning critical information and shared understanding to cover each other when player BTs are immobile due to low-energy as described in figures. 5.2 and 5.6 and section. 5.2.4. Other events such as continuous tracking of UGVs also contribute to fixation count, however, such effects could counter-balance the former events. In general, the average of overall fixations is expected to show signs of performance improvement.

For simplicity, symbols a , b and c are used in reference to participant IDs Alpha, Bravo and Charlie, respectively. During any trial, the overall number of unique f-IDs (f) for any team-member was estimated and averaged across the members as

$$f_{\mu} = \frac{\sum_K f_K}{3}, \text{ where } K \in \{a, b, c\} \quad (8.2)$$

Team-average of Median of Saccade Magnitudes

Similar to the previous measure, as described in the section. 5.3.3 the parameter saccadic magnitudes accounted for distance moved in pixels between point-of-gaze (POG)s. One common strategies observed across all teams is that they tended scatter their BTs across the region to achieve better coverage, hence is expected to reflect in their region of focus. During, exchange of information these events cause an elevated saccadic magnitudes due to distance covered across the map. This only happens when there are UGVs appearing from one end of the map. These events could happen during the spawn of first 9-UGVs and hence serve as good marker of

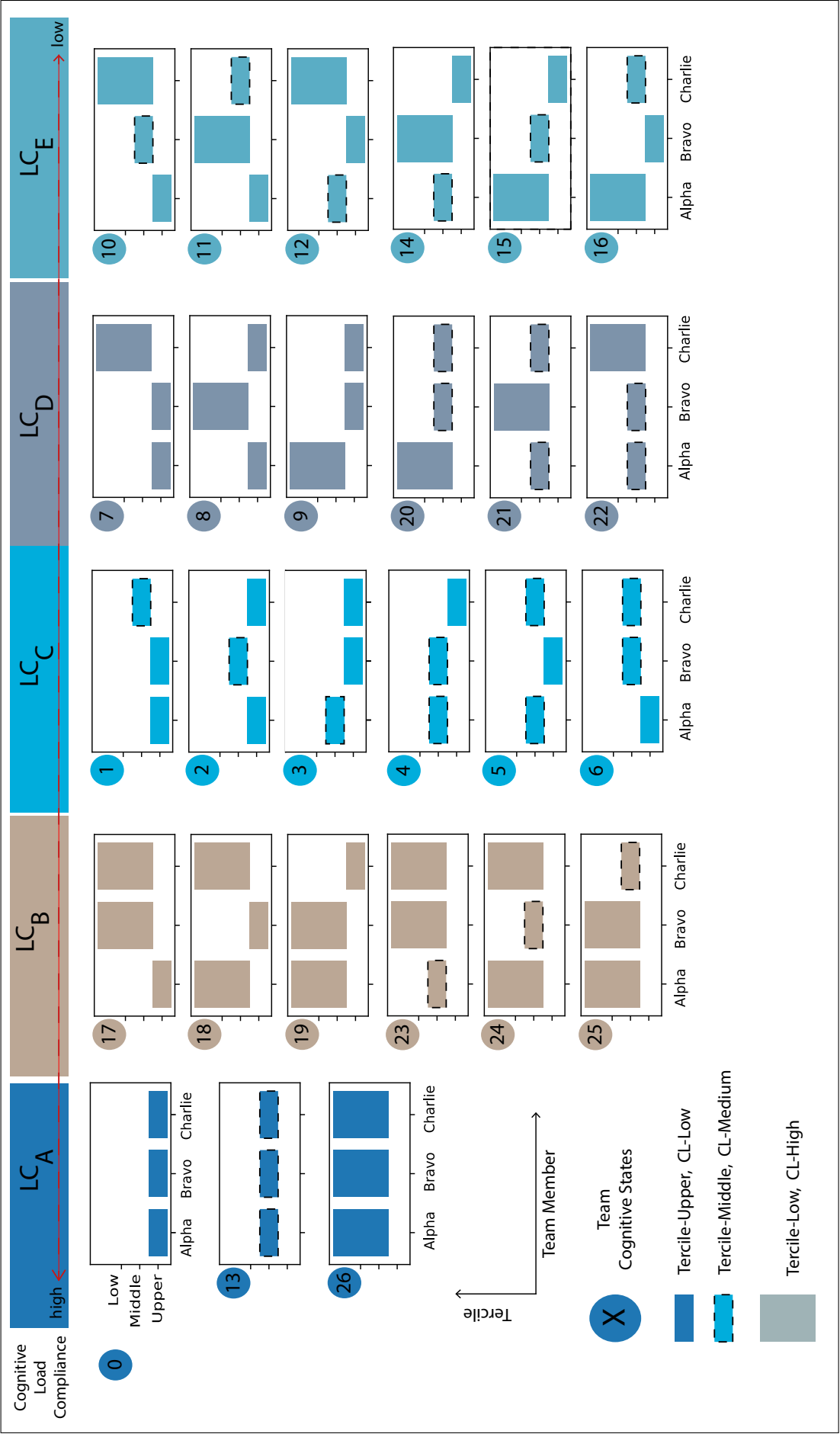


Figure 8.4: Detailed picture of five levels of cognitive compliance after categorising patterns of team-member terciles. From left to right describes most patterns with lower levels of team cognition moving toward higher levels of team cognition. The labels to the left of each plot indicates the team cognitive state labels

team cognitive activity. One downside of this observation could arise from the effects of continuous tracking events which result in more unwanted short saccades. Hence, to observe these effects and limit shorter saccadic magnitude, the median is used. Accordingly, during a trial, the median of saccades magnitude (s) for any team-member was first estimated and averaged across the members as

$$s_{\mu} = \frac{\sum_K f_K}{3}, \text{ where } K \in \{a, b, c\} \quad (8.3)$$

Team Behavioural Compliance

The processed POG series of team members was used to compute the gaze co-visitations between all pair of team members. At any discrete time (n) during a trial, the euclidean distance between all possible pairs were calculated as shown below

$$d_p(n) = ||M(x_n, y_n) - N(x_n, y_n)|| \quad (8.4)$$

where, $d(n)$ is the distance computed between team member pairs M and N at a discrete time n and $p \in \{(a, b), (b, c), (a, c)\}$ representing possible pairs. Here, M and $N \in \{a, b, c\}$ such that $M \neq N$ representing any two participant.

Then GC series was constructed for each trial based on the number of pairs accounting for gaze co-occurrence with 50 pixel radius as shown in eqn. 8.5.

$$(GC_{0P}^{\rightarrow}(n), GC_{1P}^{\rightarrow}(n), GC_{2P}^{\rightarrow}(n), GC_{3P}^{\rightarrow}(n)) = \begin{cases} (1, 0, 0, 0) & \text{if } c1 \\ (0, 1, 0, 0) & \text{if } c2 \\ (0, 0, 1, 0) & \text{if } c3 \\ (0, 0, 0, 1) & \text{if } c4 \end{cases} \quad (8.5)$$

$$\begin{aligned} c1 &= d_p(n) > 50, \\ \forall p &= \{(a, b), (b, c), (a, c)\} \end{aligned} \quad (8.6)$$

$$\begin{aligned} c2 &= d_p(n) \leq 50 \text{ and } d_q(n) > 50 \text{ and } d_r(n) > 50, \\ &\text{where, } p, q, r \in \{(a, b), (b, c), (a, c)\}, \\ &\text{and } p \neq q \neq r \end{aligned} \quad (8.7)$$

$$\begin{aligned}
c3 &= d_p(n) \leq 50 \text{ and } d_q(n) \leq 50 \text{ and } d_r(n) > 50, \\
&\text{where, } p, q, r \in \{(a, b), (b, c), (a, c)\}, \\
&\text{and } p \neq q \neq r
\end{aligned} \tag{8.8}$$

$$\begin{aligned}
c4 &= d_p(n) < 50 \\
&\forall p = \{(a, b), (b, c), (a, c)\}
\end{aligned} \tag{8.9}$$

Table 8.2: Example of interpreting BC measures for a team-trial

Gaze visitation than 50 pixels	co- less pairs	zero pairs (BC_{0p})	one pair (BC_{1p})	two pairs (BC_{2p})	three pairs (BC_{3p})
Count		2330	375	24	11

where, $c1$ (eqn. 8.6), $c2$ (eqn. 8.7), $c3$ (eqn. 8.8), and $c4$ (eqn. 8.9) accounts for zero-pairs, one-pair, two-pairs and three-pairs having gaze co-occurrences within 50 pixel radius at a certain discrete time n . For example, at any n , if no pairs had euclidean distance less than 50 then GC_{0p} was assigned 1 and the remaining gaze co-visitation possibilities i.e., GC_{1p} , GC_{2p} and GC_{3p} are assigned zero. Similarly, if suppose that there was one pair GC_{0p} , GC_{2p} and GC_{3p} are assigned zero. Finally, the behavioural compliance $\vec{BC}$ of a team-trial was represented using a vector of 4-values such that values in the vector corresponded to number co-visitations observed in each pair-count class as shown in eqn. 8.10. For example, if $\vec{BC}$ of a team-trial was $\{2330, 375, 24, 11\}$ it could be interpreted as:

$$\vec{BC} = \left\{ \sum_n GC_{0p}, \sum_n GC_{1p}, \sum_n GC_{2p}, \sum_n GC_{3p} \right\} \tag{8.10}$$

8.3 RESULTS

8.3.1 Data for Analysis

Following the team-level PB measures, seven measures from IBI-based categorical cognitive states, five measures from IBI-based compliance in cognitive load, gaze fixations and four variables gaze co-occurrences based BC measures, respectively were available for analysis.

At the time of this analysis data of only 17 team were available. As such, IBI of one team and gaze of three teams' data were discarded due to technical issues. Further, observing the trend that initial trials could include learning processes, the remainder of the analyses was performed using

task-performances and PB measures from last two trials. Hence, there were a total of 32 trials of IBI-series and 28 trials of gaze-series and fixations. Overall, this provided a pool of 28-trials PB-based measurements for two different multimodal analysis.

8.3.2 Relationship between PB measures and performance

Pearson correlation was used to analyse the relationship between PB measures against the overall team performance BH.

Relationship between TCS and performance

From Table. 8.3, it could be observed that TCS CS_{16} , CS_{21} and CS_{25} demonstrated a significant relationship with performance. A closer inspection of these cognitive states realises the following:

- CS_{16} accounted for instances where *Alpha* had a high cognitive load, *Bravo* demonstrated the least cognitive load and *Charlie* demonstrated medium cognitive load. This one of the least cognitive-LC state and having a significant positive correlation with performance. On closer inspection of experiment notes revealed that participants who assumed the role of *Alpha*, took in charge of guiding the team. The observation was in line with the theories of ownership and leadership evolution [73]. However, later analysis reveals that this effect did not necessary contribute to the composite index.
- CS_{21} accounted for instances where *Bravo* had a high cognitive load, and *Alpha* and *Charlie* demonstrated medium cognitive load. This is one of the medium-low cognitive-LC state and having a significant positive correlation with performance.
- CS_{25} accounted for instances where *Alpha* and *Bravo* had a high cognitive load, and *Charlie* had a medium cognitive load. Also, this is one of the high-medium cognitive-LC state and having a significant negative correlation with performance. This effect is high likely due to relaxed and complementary behaviour of *Alpha*, *Bravo* and *Charlie*, where as a team they are supporting each other and not having demonstrate more instances of cognitive-LC.

Further, a multiple linear regression was performed to enhance the relationship as shown in Table. 8.4. The result reveals that the combination of CS_{21} & CS_{25} achieves the best performance. A significant regression could be reported as $F(2,29) = 10.79, p = 0.0003$, with an $r^2 = 0.42$ and also an improvement in the adjusted- $r^2 = 0.39$. Team's predicted performance $BH = -226.8254 - 24.83CS_{21} + 17.81CS_{25}$ and both CS_{21} & CS_{25} were significant predictors of team performance. Despite, all three CS-measures

Table 8.3: Linear statistics (mean and standard deviation (SD)) and Pearson correlation (N=32) between IBI-based measures CS , T , LC and task performance (BH)

PB Type	Variable	Mean	SD	pearson-r	p-value	r^2	significance
$\vec{CS}$	CS_0	5.91	2.61	-0.15	0.41288	0.02	ns
	CS_1	11.78	3.60	-0.26	0.14796	0.07	ns
	CS_2	12.03	2.91	0.17	0.34202	0.03	ns
	CS_3	12.16	3.14	0.03	0.84935	0.00	ns
	CS_4	20.75	3.79	-0.24	0.17673	0.06	ns
	CS_5	22.50	4.97	-0.10	0.57176	0.01	ns
	CS_6	21.50	4.03	0.06	0.75757	0.00	ns
	CS_7	4.88	2.15	0.10	0.58997	0.01	ns
	CS_8	5.19	2.99	0.02	0.90556	0.00	ns
	CS_9	4.88	1.93	-0.02	0.90859	0.00	ns
	CS_{10}	10.69	3.31	-0.13	0.47000	0.02	ns
	CS_{11}	9.25	3.44	0.13	0.47729	0.02	ns
	CS_{12}	10.56	3.80	0.06	0.73664	0.00	ns
	CS_{13}	44.12	5.07	-0.00	0.98028	0.00	ns
	CS_{14}	10.16	2.36	0.09	0.62149	0.01	ns
	CS_{15}	9.22	3.23	0.16	0.39559	0.02	ns
	CS_{16}	8.94	2.88	0.45	0.00913	0.21	**
	CS_{17}	4.75	2.41	0.12	0.51288	0.01	ns
	CS_{18}	4.38	2.19	0.00	0.99376	0.00	ns
	CS_{19}	5.69	2.36	-0.03	0.89157	0.00	ns
	CS_{20}	19.91	3.48	0.18	0.32372	0.03	ns
	CS_{21}	20.69	3.34	0.38	0.03118	0.15	*
	CS_{22}	19.81	3.32	-0.06	0.75221	0.00	ns
	CS_{23}	10.12	3.04	-0.06	0.75365	0.00	ns
	CS_{24}	12.84	3.60	-0.08	0.66927	0.01	ns
	CS_{25}	12.19	4.16	-0.58	0.00054	0.33	***
	CS_{26}	7.97	3.45	0.10	0.58147	0.01	ns
$\vec{T}$	T_{-3}	12.81250	4.11125	0.40	0.12522	0.16	ns
	T_{-1}	72.50000	10.05609	-0.10	0.70236	0.01	ns
	T_1	154.81250	12.28423	-0.53	0.03440	0.28	*
	T_3	205.87500	12.96088	0.42	0.10569	0.18	ns
	T_5	153.81250	11.53917	0.34	0.19889	0.11	ns
	T_7	70.31250	7.60525	-0.46	0.07424	0.21	ns
	T_9	15.25000	5.17808	-0.01	0.96132	0.00	ns
$\vec{LC}$	LC_A	58.00000	7.54155	-0.01	0.96227	0.00	ns
	LC_B	49.96875	5.29436	-0.50	0.00383	0.25	*
	LC_C	100.71875	7.66173	-0.20	0.26873	0.04	ns
	LC_D	45.43750	7.24110	0.07	0.71778	0.00	ns
	LC_E	58.81250	6.66351	0.34	0.05669	0.12	ns

collectively showing a significant improvement in performance ($r^2=0.44$), however, the adjusted- r^2 does not show any improvement.

Table 8.4: Multivariate regression using combinations of $CS_{16,21\&25}$ and performance BH

Variable	p-value	r^2	Adjusted r^2	significance
CS_{16}, CS_{21}	(0.094, 0.027*)	0.28	0.23	0.008**
CS_{16}, CS_{25}	(0.012, 0.246)	0.36	0.32	0.0014**
CS_{21}, CS_{25}	(0.001** ,0.038*)	0.427	0.39	0.0003 ***
$CS_{16}, CS_{21} \& CS_{25}$	(0.009** ,0.064* ,0.445)	0.44	0.38	0.0009 ***

Relationship between categorical-TCS and performance

From Table. 8.3, it could be observed that categorical-TCS T_1 and T_7 demonstrated a reasonable negative correlation with performance, although only T_1 demonstrated a significant relationship. Observing the trends of T_1 and T_7 from fig.8.3, it could be inferred that the relative cognitive load experienced by one team-member with respect to the other two members critically impacted performance. Thus, T_1 and 7 accounted for instances where there was higher disparity between team members cognitive involvement resulting in relatively poor performance scores.

It could be argued that patterns in T_5 share resemblance with T_1 and T_7 , however, task conditions that manifested these loads did not impact performance in the current task settings. A post-hoc analysis with bonferroni correction was performed to learn whether, participants favoured T_1 over T_5 and observed a no significance difference. This indicated that there was equally likely possibility of participants encountering T_1 and T_5 during the task-trials, and there were higher chances that participants encountered T_5 compared to T_7 . Thus, providing more evidence that the appearance of team cognitive level was indeed influenced by task conditions.

Another test was performed after summing T_1 and T_7 (T_{17}) to observe if there was any enhancement to their previously observed individual relationships. The test revealed an improvement with significant negative correlation ($r = -0.55$, $p = 0.01$, $N = 28$) indicating that T_1 and T_7 collectively contributed to decline in performance. Overall, the result suggests that higher was the disparity at these specific conditions, results in lower performance scores.

Relationship between cognitive-LC and performance

From Table. 8.3, it could be observed that only the cognitive-LC LC_B demonstrated a significant negative relationship with performance. However, cross analysing with the results in section. 8.3.2 indicate that the significance

could be influenced by TCS CS_{25} . The comparison realises that LC_B does not provide any improvements in performance over its individual predecessor CS_{25} . However, the of relaxed and complementary behaviour of teams is upheld via the significant negative correlation.

8.3.3 Gaze data analysis

Relationship between gaze fixations and performance

A simple linear regression was performed to analyse the relationship between average of overall gaze fixations (f_μ , eqn. 8.2) and performance, and observed a significant positive correlation with performance as shown in table. 8.5. The regression equation is reported as $F(1, 26) = 16.24, p = 0.0004$, with an $r^2 = 0.38$. In this light, team's predicted task performance could be formulated as,

$$y_{BH} = -981.0216 + 0.593f_\mu. \quad (8.11)$$

In addition, a one way analysis of variance was conducted to observe if variation in fixations were result of team cognitive activity. However, no significant effect between (f_μ) and trial progression was observed. This indicates that only good performing teams improved their mean fixation count which enabled them to perform better.

Relationship between gaze co-occurrence and performance

Table 8.5: Linear statistics (mean and standard deviation (SD))and Pearson correlation (N=28) between gaze based f_μ , BC and task performance (BH)

PB Type	Variable	Mean	SD	pearson-r	p-value	r^2	significance
	f_μ	1346.16	193.12	0.62	0.0004	0.38	***
$\vec{BC}$	BC_{0P}	2542.93750	148.00907	-0.56	0.02544	0.31	*
	BC_{1P}	186.56250	135.50045	0.56	0.02406	0.31	*
	BC_{2P}	6.56250	8.22320	0.49	0.05638	0.24	ns
	BC_{3P}	4.18750	4.92720	0.42	0.10283	0.18	ns

Analysing the results of correlation between BC and BH in Table. 8.5, shows that the number of gaze co-occurrence between any pair of team members (BC_{1P}) during a trial demonstrated a significant relationship with task performance. Contrastingly, an opposite effect was observed between zero gaze co-occurrences (BC_{0P}) and performance. This implies that, gaze co-occurrence between a pair provided an indicator of performance. Also, the measure BC_{0P} was an opposite reflection of BC_{1P} indicating that the this

measure significantly dominated two remaining measures in contribution of performance. Further, the analysis also revealed that there was no significant difference between the other two BC measures BC_{2P} and BC_{3P} . On closer inspection it was observed that majority of the BC_{2P} and BC_{3P} had resulted in a relatively low mean and standard deviation (SD) further substantiating its insignificance to performance.

The analysis indicates that, teams with relatively higher gaze co-occurrences performed better in the task. The gaze co-occurrence in the current context was an indicator of how quickly their partners reacted indicating shared understanding, team situational awareness and share mental models. Thus, team with higher gaze co-occurrences could be inferred to have demonstrated a relatively higher team cognition.

Multimodal Analysis-1

From the results above, a multiple linear regression was performed using all of the measures estimated above. Results demonstrated the best relationship with the task performance BH using CS_{25} and f_{mu} . The predicted trial performance (BS Health) can be visualised in Figure. 8.5. A significant regression equation was found ($F(2,25) = 16.9$, $p < 0.0001$), with an r^2 score of 0.57 and an adjusted- r^2 score of 0.54. Both parameters provided a significant contribution such that team's predicted task performance is formulated as,

$$y_{BH} = -340.005 - 26.1405CS_{25} + 0.35f_{\mu} \quad (8.12)$$

The result indicated that performance BH was a function of improvement in overall fixation count, and a relaxed and complementary behaviour of *Alpha*, *Bravo* and *Charlie* as a team.

Multimodal Analysis-2

Another possible relationship was explored from the results, where categorical-TCS measure T_{17} and BC measures BC_{0P} and BC_{1P} had a significant relationship with performance. Hence, the measures BC_{1P} and T_{17} were grouped together, and a multiple linear regression was performed. A significant regression equation was found ($F(2,13) = 11.83$, $p = 0.001$), with an r^2 score of 0.65. The predicted trial performance (BS Health) is shown in eqn. 8.13 and visualised in Figure. 8.6. The performance during a trial improved as 0.28 times pair gaze co-occurrence BC_{1P} with -7.12 lower level of T_{17} instance of team cognition level. Thus both, team cognition level belonging to T_1 and T_{17} following gaze co-occurrence levels of BC_{0P} were significant predictors of performance.

$$y_{BH} = 1298.1983 - 7.30T_{17} + 0.750BC_{1P} \quad (8.13)$$

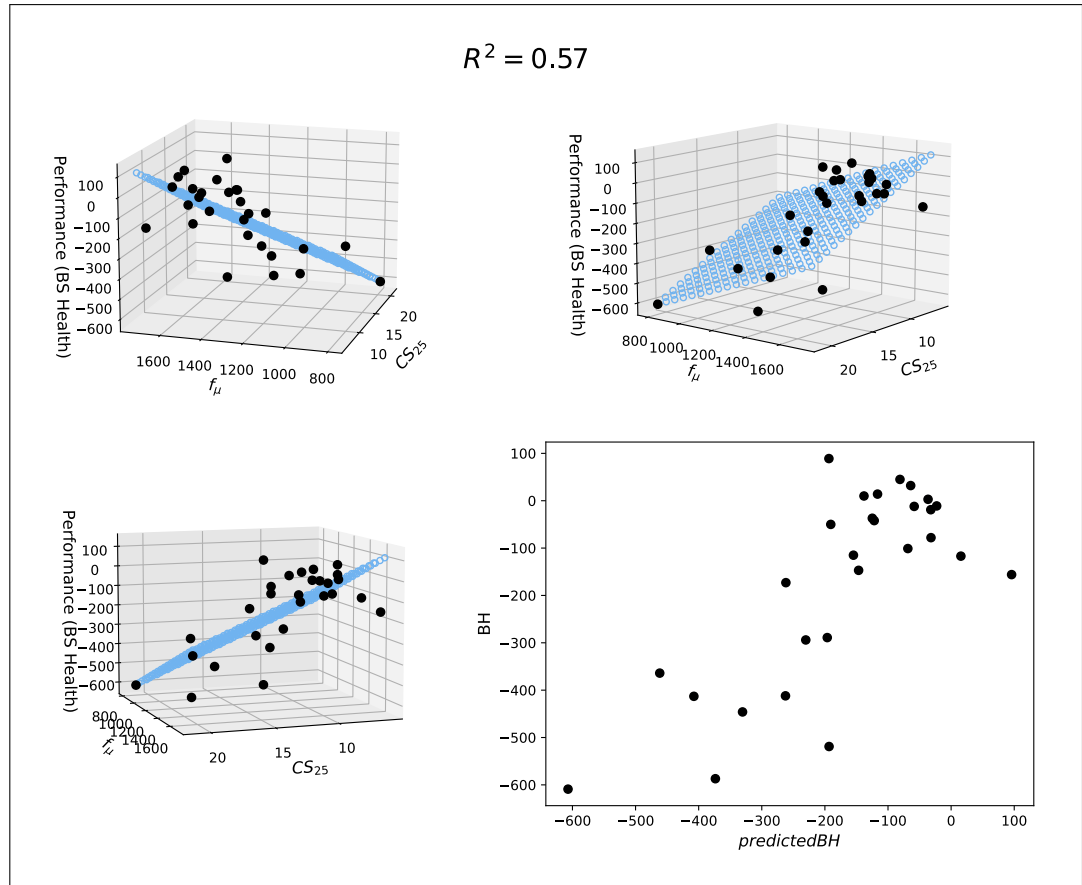


Figure 8.5: A detailed picture of multiple linear regression utilising PB measures CS_{25} and f_{μ} .

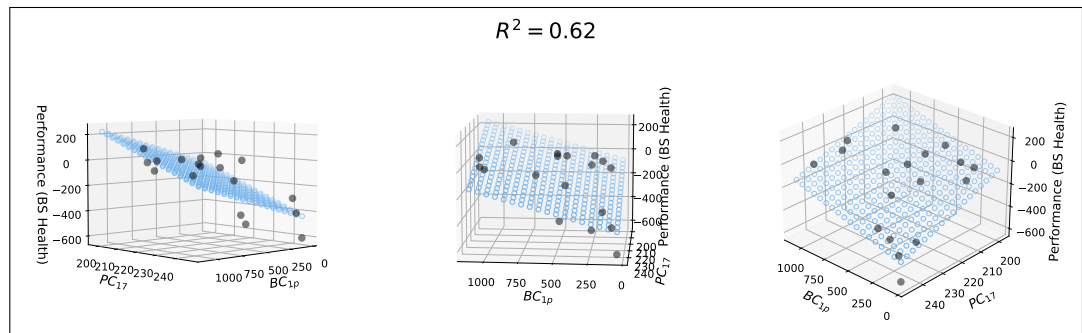


Figure 8.6: A detailed picture of multiple linear regression utilising BC measure BC_{1P} and categorical-TCS measure T_{17} .

8.3.4 Team Cognitive Activity

Markers From IBI-based measures

The IBI-based cardiovascular measures were used to observe markers of cognitive activity as teams progressed through trials. Accordingly, a one way analysis of variance was conducted and observed the following significant differences between the first two trials and the last two trials. As mentioned in section. 5.2.5, every trial comprised of two intensity levels i.e., *Low*

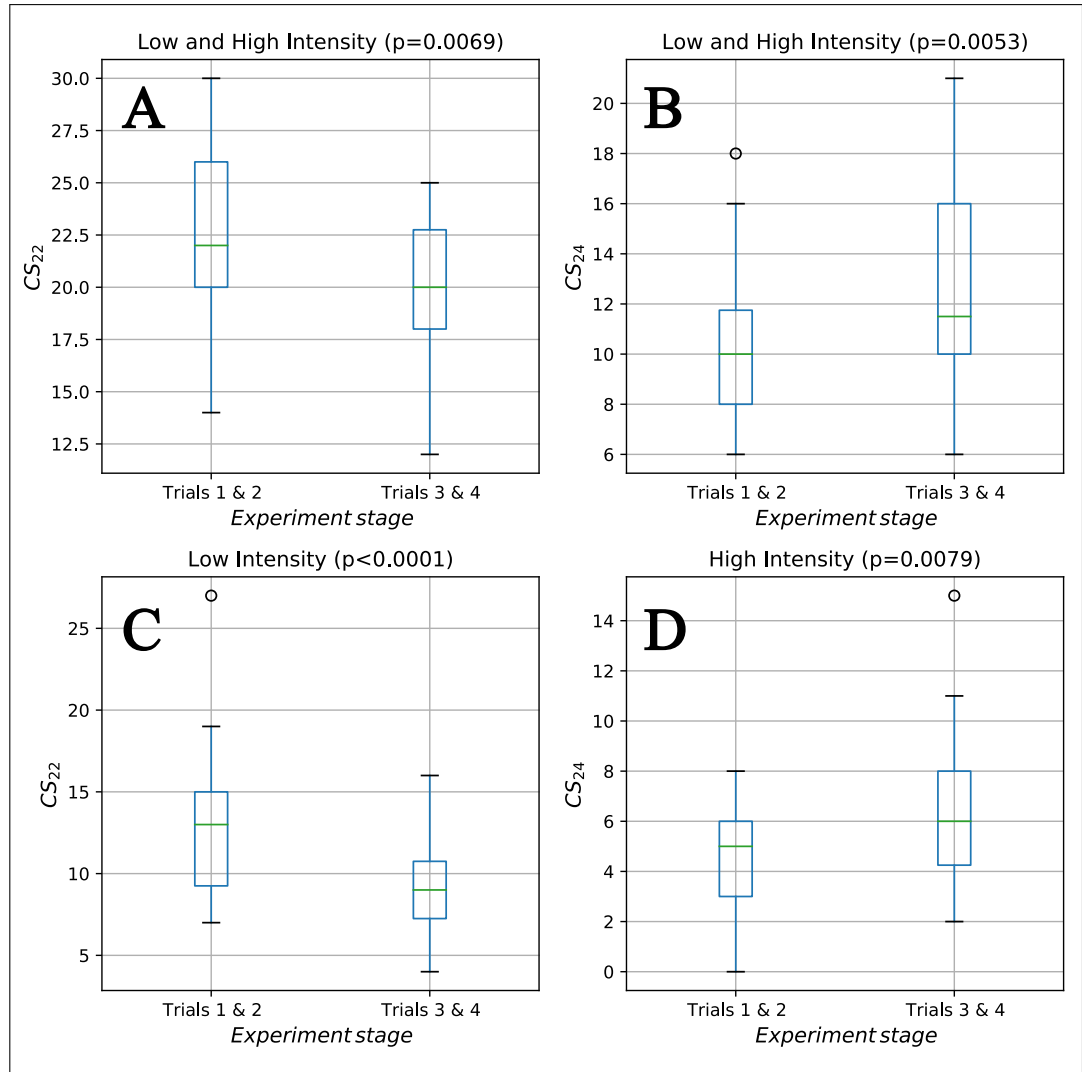


Figure 8.7: Boxplots of TCSs CS_{22} and CS_{24} against different stages of the task showing variations due to teams continuous involvement. Subplots A and B corresponds to counts of CS_{22} and CS_{24} , respectively, observed during the full-length of a task trial. Subplots C and D corresponds to CS_{22} counts during low intensity and CS_{24} counts during the high intensity phase of a task trial.

(starting three waves) and *High* (remaining three waves). As team members collaborated, significant changes were observed in TCS CS_{22} and CS_{24} as shown in Figure 8.7.

Irrespective of the stage of the experiment, figures 8.7 (A and B) demonstrate the variation such that, teams demonstrated a significant low number of CS_{22} state in later trials compared to early trials. An opposite effect was observed in CS_{24} state where later trials demonstrated a significant increase. Additionally, the corresponding variations were significantly evident during the conditions of the task as shown in figures 8.7 (C and B).

On closer inspection of the composition of the cognitive states CS_{22} and CS_{24} , the following could be realised:

- CS_{22} accounted for instances where *Alpha* and *Bravo* experience medium cognitive load, and *Charlie* demonstrated the highest. This one of the medium-low cognitive-LC state and having a negative effect in relation to task trials and low intensity phase of task. This effect could be arising due to non-involvement of *Alpha* and *Bravo* in early trials and demonstrating a better engagement in later trials.
- CS_{24} accounted for instances where *Alpha* and *Charlie* experience the highest cognitive load, and *Bravo* demonstrated the medium. This one of the high-medium cognitive-LC state and having a positive effect in relation to task trials and high intensity phase of task. This effect could be arising due to involvement of *Alpha* and *Charlie* demonstrating a better engagement in later trials compared to early trials.

Markers From Gaze-based measures

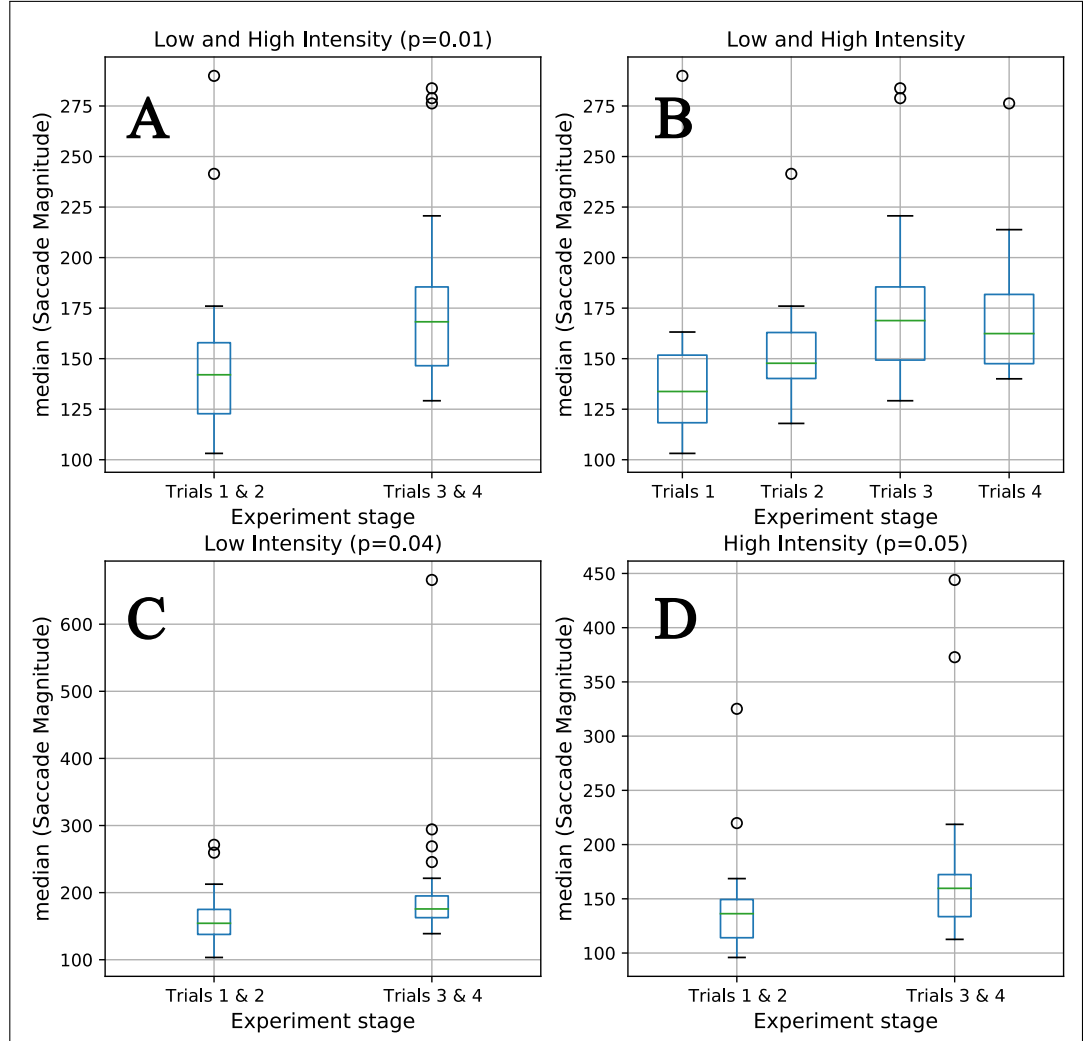


Figure 8.8: Boxplots of gaze-based s_{μ} (eqn. 8.3) against experiment stages and task conditions.

Similar to the previous analysis, gaze-based measures were explored to identify markers of cognitive activity. The team average of median of saccadic magnitudes s_μ demonstrated a significant effect of experiment stage and task conditions as shown in Figure. 8.8. The plot in Figure. 8.8 (B) shows that as teams progressed through trials, s_μ also increased and saturates after the third trial. The plot in Figure. 8.8 (A) shows a significant effect where s_μ significantly elevated in the later stages of the experiment. From Figure. 8.8 (C) and (D) it could be noticed that the effect was more visible at low intensity levels compared to the high intensity level of trials. This is valid because, following the end of a trial all the teams regrouped, analysed and devised the scattering-based defence strategy. Although, the strategic intention was clear, however, the high intensity level created the chaos resulting in small degradation.

8.4 DISCUSSION

This study aimed to explore the relationship between measures from PBR and task performance in a multiparty setting involving three members in a team while they engaged in a cognitively challenging collaborative task. The study also investigated whether an effective unified model could be derived from multiple modalities. The PB techniques in this study were chosen to complement the cognitively demanding nature of the task.

Studies based on neurovisceral integration theory suggest that there exists a connection between cardio-vascular activity and the branch of human brain necessary for regulating cognitive processes [257]. Hence, studies investigating task-exerted cognitive load explored the activity of ANS by analysing HRV from IBI-series. Conventionally, HRV is analysed using time-domain based linear measures (e.g. average heart-rate, standard deviation of normal-normal) [241, 258], frequency-domain based measures (e.g. low and high frequency components) [67] and non-linear based measures (recurrence quantification analysis) [96]. Such methods are often employed on excerpted IBI-series typically of 1 to 5-min durations, thereby, limiting any analyses concerning shorter-latency variations in cognitive activity.

To overcome the above limitations, IBI-series averaged over 2-seconds are used in the current analysis, enabling the capturing of any low-latent variations in teams cognitive load states and levels of cognitive load compliance. Previously, in a pilot study, Dias *et al.* proposed a method for efficiently analysing short-latent variations under a heterogeneous task setting. A similar method is adopted in the current study, however, in a homogeneous task setting. The team cognition levels were individually analysed against task performance demonstrating an advancement in the analysis of team cognitive loads. Traditional team-level cognitive loads were measured based on the assumptions that cognitive demand experienced by teams are aggregations of the demand experience by team-individuals [258]. In this light,

aggregations such as averaging and summing were employed to measure team-cognition level. The method presented in this study extends these concepts into current task setting where, patterns of team cognitive levels were grouped together such that the sum of loads experienced at team-level are same for any team cognitive level group.

A recent cross-domain review [60] emphasised the role of eye-tracking in perceiving team cognitive activity. Team situational awareness is an aggregate of team cognitive processes that could be analysed via gaze behaviour [5]. In early research, videos of participants were employed to analyse gaze-coupling between radar monitoring student-pairs as a measure of team situational awareness [208]. In recent years, gaze-related dynamics from eye-tracking technologies are explored in the analysis of joint attention as proxies of cognitive processes between collaborative pairs. In team settings, gaze-patterns between collaborating pairs are indicators of team's readiness that could have a significant impact on overall team performance.

Concepts of RQA [79,81] and ENA [255] were previously employed to quantitatively analyse gaze coupling. Effective application of these methods requires defined ROI, however, it is less effective when accounting for spatial and temporal co-occurrences. Employment of ROI-based approach is valid when there are fixed objects in the collaborative space, restricting the freedom of exploring gaze coupling outside of ROI boundary. However, when there are multiple objects dynamically moving and distributed around a space, such as in radar display, equipment arrangements and skimming through text documents and programming codes, defining fixed ROIs may have certain drawbacks. Undefined ROIs that could manifest for particular groups and specific to certain scenarios could not be accounted. Such processes are common in training and learning environments because, team cognitive processes develop over a period of members involvement [5,55]. In addition, majority of gaze-based analysis previously observed in literature are in dyadic settings. In triad settings, studies have opted for video analysis depriving any benefits of gaze-dynamics [82,253]. Raw-video analysis could be laborious and time consuming depends human intervention and on some occasions relies on supporting evidence such as communication transcripts.

The gaze co-occurrence estimation methods presented in the current study relies on dynamic ROIs that could occupy any region of the task-space at any time. By reducing the samples and eliminating non-valid gaze data, the influence of gaze co-occurrence on team performance was observed. Further, the analysis presented here are in triad settings and account for temporal gaze coupling between all possible pairs. Thereby, enabling an effective alternative to video-based analysis. It also enables the freedom to explore spatial-temporal gaze-coupling analysis in triadic task settings.

The cross-domain review on measuring team cognition [60] suggested that eye-tracking measures of cognition are rather indirect, and must be collectively analysed with other physiological measures. The current study addresses this requirement via the data fusion model presented in eqn. 8.12 and 8.13 combining gaze-based BC measures with cardio-vascular based CS, T and LC measures. The model represents a composite index accounting team-level compliance in visual cognitive behaviour and cognitive load experienced by members, respectively.

Overall, results elicit the acceptance of all seven set hypothesis. In summary, the team cognitive levels measure via LC of individual IBI-series shared relationship with performance, however, not all measures were effective in the current task settings. Thus, suggesting that not all conditions of the task pose a threat to performance, indicating the presence of certain instances during the task that were more stressful than others.

One of the needs of employing three members in the task was to advance team studies beyond coupled dynamics. The methods developed in this study provides a substantial groundwork for evaluating performance of teams with two or more participants which are rarely observed in the collaborative learning environment literature. The homogeneous conditions mentioned in the current study enables the developed concepts to be easily adaptable in similar environments such as in academic and training settings. The procedure also provides a mechanism for quantitative analysis.

8.5 SUMMARY

This work explored physiological and behavioural compliance in a three-member team with homogeneous task conditions. Compliance between members physiological and behavioural measures and results provided substantial evidence for future team cognition based analyses. The employment of wireless and non-invasive methods in the form of eye-tracking and chest-strap demonstrate a feasible setup for exploring other team processes related studies. The task conditions, methods and results could be replicable in environments involving groups and multi-party collaborative training. Ultimately, this research offers a solid foundation of evaluating collaborative processes for larger teams.

The next chapter mentions the limitations of the methods and findings of this thesis and provides brief recommendation for future work.

9

Conclusion, Limitations and Future Recommendation

9.1 CONCLUSION

The main purpose of the thesis was to evaluate the state of team and collaborative working research, and endeavoured to contribute via novel enhancement methods and approaches. A careful overview of literature on team research observed that most previous studies approached from a psychophysiological perspective and hence the team processes were the focal point of interest. Contrarily, this thesis approached from an empirical perspective which is inline with the inputs of the team, as well as complies with the cyclic nature of IPO. The thesis starts by briefly organising the abstracts of team performance and constructs of IPO, which were chosen arbitrarily in previous studies. Next, establishes current gaps and motivations from an empirical approach that could augment the research domain. In this light, the thesis realises the need for a library of fundamental components, single modality PBR based approach that concern team size larger than and dyads, team performance prediction and forecasting capabilities, and fusion of data from multi-modalities.

To address the aforementioned demands of the present day team research, chapter. 2 critically examined the relationships between team performance abstracts and the principles of IPO from PBR based empirical perspective. The study led to synthesising a literature on the fundamental components such as PBR, synthetic collaborative tasks, mathematical models for processing and fusion of data, subjective feedback techniques for overseeing the transition. In addition, two preliminary studies were conducted in chapter 3 and 4 to complement the knowledge gained in the review, building a strong foundation for designing research methods for this thesis.

Concepts such as gamification and software developments were utilised to build a state-of-the art experimental platform for collaboratively engaging 3-member teams. Chapter. 5 evaluated the capability, scalability and configurability properties of the environment. Properties such as comprehensively engage participants with minimal physical interference, choice

of PBR techniques that can complement each other, comfortableness in wearing multiple and capabilities of the data and task dynamics in team research was explored. The platform offers expansion for engaging larger teams and future integration of online and real-time predictive models and applications.

Ultimately, in the light of gaps established at the beginning of this thesis, the following summarises novel contributions such as, a 3-member collaborative task environment, an empirical framework that could collect data concurrently from 3-member teams, gaze based findings in terms of convergence, co-occurrence, distribution and compliance, cardiac based cognitive states, and two CM-TLX measures combining gaze and cardiac data. The thesis addressed concepts such as socio-physiological compliance, behavioural compliance, and team cognitive activities. Two studies demonstrated the neural network based predictive capabilities. Overall, contributing to the enhancement of work-team collaborative research methods. The validity of the relationships between CM-TLX and performance could be improved by increasing the sample size. It should be noted that in PBR-based team research employing three or more participants, this thesis collected more team samples than reported in the literature.

9.2 LIMITATIONS

Although, the thesis made a notable and novel contributions, however, several limitations observed during the course of the research is listed below as.

- The configurability offered in the current version of Tank Battle Task is limited, however, is still effective in team research as discussed in the Chapter. The task environment is mostly capable of hosting homogeneous task setting which is commonly observed in training and learning environments. It should be noted that the results are task and domain specific, current configuration only allowed validation of limited aspects of team and collaborative working research.
- The analysis originating from this task and PBR data are in a desktop-based environment, which may not be readily applicable for mobile eye-tracking applications mentioned by Fasold *et al.* [211]. Additionally, due to limited data availability, the neural network model presented here relied on normalised BH scores. The employment of raw BH scores required much larger and deeper NN, hence, was not reported. However, it provided enough evidence for future research to focus on training with raw scores. Future studies preferring to explore RQA-based analysis could employ pair-distances that is a collective representation of (x,y) coordinates, thus, minimising computations.

- The results of gaze measures reported in Chapters 6,7 and 8 originating from the current task environment is based on pixel radius of UGV. Since the task is designed in 3-dimension, when projected on a 2-dimension display (LCD), the radius could vary depending on the location of the UVs on the map. The concept has rich value, however, identifying the radius of dynamic-ROI for other task settings requires certain trial as mentioned in Chapter. 6 or identification of a different logic.
- Future work focuses on enhancing the task environment with a different configuration that could target a broader spectrum of team research. Also, support for hosting heterogeneous task setting will be targeted in the future. A public version of the environment is available to the research community promoting other areas of team research with different configurations. The inclusion of subjective feedback to learn emotion states and boredom that are prevalent in academic settings [259] could be advantageous, however, are not attempted in this study.
- One of the limitations of defining team-cognitive load in Chapter. 8 is the employment of terciles. In the current approach limits 25% and 75% have been used in accordance with interquartile estimation. There are literature suggesting the use of limits 33% and 66% which might provide a different outcome. Further, as opposed to terciles, literature also mentions the use of quartiles enabling more team cognitive levels. Future research could focus specifically on which approach would be more suitable.
- The result of cognitive effects reported in the study has fixed low-intensity followed by high-intensity. Validation and acceptance of this finding requires experimenting with variations in the intensity such as high-intensity phase appearing before low-intensity and then mixture of low and high-intensity.
- Despite the homogeneous nature of the task, heterogeneous were observed indicating the presence of evolution of leadership qualities. However, using IBI and gaze data could not validate this claim could not be substantiated. Processing communication utterances could help in the validation of this result.
- Employing interested and genuine participants could improve the significance of the relationships observed in the results. Although, participants were given clear instructions regarding experiment preparation, however, cannot be verified. It is beyond the controlled environment and must rely on their feedback.

9.3 FUTURE RECOMMENDATIONS

In relation to the above limitations and overall findings of this thesis, the following recommendations for future could significantly augment the field team and collaborative research.

- Future work focuses on enhancing the task environment with a different configuration that could target a broader spectrum of team research. Also, support for hosting heterogeneous task setting will be targeted in the future. A public version of the environment is available to the research community promoting other areas of team research with different configurations. The inclusion of subjective feedback to learn emotion states and boredom that are prevalent academic settings [259] could be advantageous, however, are not attempted in this study.
- The props of the STE i.e., BT, UGV and BS are just simulation agents of current task. Future research could focus on modifying the props that represent other scenarios such as radar surveillance, academic collaborative problem solving, cyber-security, industry equipment operation and health care professional settings. The environment can be easily expanded to host larger teams as well as team of teams which is an area of research currently in demand at an industry/organisation level. Team of teams research is currently primitive and PB based research could significantly revolutionise the research, encouraging an efficient approach.
- The current findings definitely pose avenues for future exploration. For example, identifying additional affect instances of task, such as stress and related PC-levels that could impact teams performance. Thus, jointly addressing the PC-based stressful instances relationship with gaze dynamics would provide solid evidence to team cognitive processes and performance. Further, research on user-experience based measures of reaction duration as task performance and its relationship with PC and BC measures is another goal for future explorations.
- In the light of expanding the usability of the task environment as an embedded training system, future research could also focus on integrating ANN-based predictive models with task environments such as intelligent team tutoring system [218], and advanced embedded training system [32]. Such integration enhances the functionality of training systems enabling real-time performance assessment using gaze physiological dynamics.
- Originally, the thesis aimed to incorporate Google's voice to script conversion to be employed in semantic analysis, word count and utterances [9]. Raw communication data also contains information such

as fundamental frequencies pertaining to emotions and linguistic processes. Future research could incorporate these features for improving the accuracies of performance predictions.

- The findings of Chapter. 4 outlines a range of feature extraction processes for utilising EEG. Additionally, the importance of data from channels and regions were explained. Future research could utilise these resources for exploration of EEG in team settings. In line with the cognitive states established in Chapter. 8, Gorman mentions a similar process for representing EEG signals [45]. Future research could employ such an approach with EEG and HRV data to derive Shannon's entropy for analysing attractor basins of team performance.
- Future research could also explore causal effects on performance and team cognitive processes.
- Most performance related analysis primarily concern the BH score. Synthetically generated performances mentioned in section. 5.5.2 are yet to be explored, and offers an additional avenue for future research.
- Future research could also focus on combined indices of gaze and cognitive states on team cognitive processes.
- Given the time constraint, at this stage it was not possible to completely utilise all of the PBR data collected during the experiment. Specifically, the effects of EEG, communication and recurrent neural networks pose an avenue for future exploration.

References

- [1] D. L. Gladstein, "Groups in context: A model of task group effectiveness," *Administrative science quarterly*, pp. 499–517, 1984.
- [2] S. I. Tannenbaum, R. L. Beard, and E. Salas, "Team building and its influence on team effectiveness: An examination of conceptual and empirical developments," in *Advances in psychology*. Elsevier, 1992, vol. 82, pp. 117–153.
- [3] M. A. Rosen, A. S. Dietz, T. Yang, C. E. Priebe, and P. J. Pronovost, "An integrative framework for sensor-based measurement of teamwork in healthcare," *Journal of the American Medical Informatics Association*, vol. 22, no. 1, pp. 11–18, 2015.
- [4] S. Kazi, S. Khaleghzadegan, J. V. Dinh, M. J. Shelhamer, A. Sapirstein, L. A. Goeddel, N. O. Chime, E. Salas, and M. A. Rosen, "Team physiological dynamics: A critical review," *Human factors*, p. 0018720819874160, 2019.
- [5] N. J. Cooke, J. C. Gorman, C. W. Myers, and J. L. Duran, "Interactive team cognition," *Cognitive science*, vol. 37, no. 2, pp. 255–285, 2013.
- [6] E. Salas, K. C. Stagl, C. S. Burke, and G. F. Goodwin, "Fostering team effectiveness in organizations: Toward an integrative theoretical framework," in *Nebraska symposium on motivation*, vol. 52, 2007, p. 185.
- [7] J. C. Gorman, N. J. Cooke, and P. G. Amazeen, "Training adaptive teams," *Human Factors*, vol. 52, no. 2, pp. 295–307, 2010.
- [8] J. C. Gorman, P. G. Amazeen, and N. J. Cooke, "Team coordination dynamics," *Nonlinear Dynamics, Psychology, and Life Sciences*, vol. 14, no. 3, p. 265, 2010.
- [9] C. Neubauer, J. Woolley, P. Khooshabeh, and S. Scherer, "Getting to know you: a multimodal investigation of team behavior and resilience to stress," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, 2016, pp. 193–200.
- [10] D. P. Baker, R. Day, and E. Salas, "Teamwork as an essential component of high-reliability organizations," *Health services research*, vol. 41, no. 4p2, pp. 1576–1598, 2006.

- [11] R. A. Henning, W. Boucsein, and M. C. Gil, "Social-physiological compliance as a determinant of team performance," *International Journal of Psychophysiology*, vol. 40, no. 3, pp. 221–232, 2001.
- [12] A. N. Elkins, E. R. Muth, A. W. Hoover, A. D. Walker, T. L. Carpenter, and F. S. Switzer, "Physiological compliance and team performance," *Applied ergonomics*, vol. 40, no. 6, pp. 997–1003, 2009.
- [13] E. Salas, N. J. Cooke, and M. A. Rosen, "On teams, teamwork, and team performance: Discoveries and developments," *Human factors*, vol. 50, no. 3, pp. 540–547, 2008.
- [14] E. Salas, T. L. Dickinson, S. A. Converse, and S. I. Tannenbaum, "Toward an understanding of team performance and training." 1992.
- [15] D. R. Ilgen, J. R. Hollenbeck, M. Johnson, and D. Jundt, "Teams in organizations: From input-process-output models to imoi models," *Annu. Rev. Psychol.*, vol. 56, pp. 517–543, 2005.
- [16] B. Bertrand, J.-N. Evain, J. Piot, R. Wolf, P.-M. Bertrand, V. Louys, H. Terrisse, J.-L. Bosson, P. Albaladejo, and J. Picard, "Positive communication behaviour during handover and team-based clinical performance in critical situations: a simulation randomised controlled trial," *British Journal of Anaesthesia*, vol. 126, no. 4, pp. 854–861, 2021, verbal and nonverbal.
- [17] C. La Fleur, B. Hoffman, C. B. Gibson, and N. Buchler, "Team performance in a series of regional and national us cybersecurity defense competitions: Generalizable effects of training and functional role specialization," *Computers & Security*, vol. 104, p. 102229, 2021.
- [18] H.-Y. Liu, "The relationship between swift trust and interaction behaviors on interdisciplinary and non-interdisciplinary teams in nursing education," *Nurse Education in Practice*, vol. 51, p. 102977, 2021.
- [19] G. J. Funke and S. M. Galster, "The effects of cognitive processing load and collaboration technology on team performance in a simulated command and control environment," *International Journal of Industrial Ergonomics*, vol. 39, no. 3, pp. 541–547, 2009.
- [20] G. J. Funke, B. A. Knott, E. Salas, D. Pavlas, and A. J. Strang, "Conceptualization and measurement of team workload: A critical need," *Human Factors*, vol. 54, no. 1, pp. 36–51, 2012.
- [21] A. J. Strang, G. J. Funke, S. M. Russell, A. W. Dukes, and M. S. Middendorf, "Physio-behavioral coupling in a cooperative team task: Contributors and relations." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 40, no. 1, p. 145, 2014.

- [22] E. Salas, D. L. Reyes, and A. L. Woods, "The assessment of team performance: Observations and needs," in *Innovative assessment of collaboration*. Springer, 2017, pp. 21–36.
- [23] D. Dunning, K. Johnson, J. Ehrlinger, and J. Kruger, "Why people fail to recognize their own incompetence," *Current directions in psychological science*, vol. 12, no. 3, pp. 83–87, 2003.
- [24] M. Feldman, E. H. Lazzara, A. A. Vanderbilt, and D. DiazGranados, "Rater training to support high-stakes simulation-based assessments," *Journal of Continuing Education in the Health Professions*, vol. 32, no. 4, pp. 279–286, 2012.
- [25] A. DiMascio, R. W. Boyd, and M. Greenblatt, "Physiological correlates of tension and antagonism during psychotherapy: A study of interpersonal physiology," *Psychosomatic medicine*, vol. 19, no. 2, pp. 99–104, 1957.
- [26] R. W. BOYD and A. DiMASCIO, "A sociophysiologic study1," *The Journal of Nervous and Mental Disease*, vol. 120, no. 3, pp. 207–212, 1954. [Online]. Available: https://journals.lww.com/jonmd/Fulltext/1954/09000/A_Sociophysiologic_Study1.5.aspx
- [27] H. B. Kaplan, N. R. Burch, and S. W. Bloom, "Physiological covariation and sociometric relationships in small peer groups," *Psychobiological approaches to social behavior*, pp. 92–109, 1964.
- [28] R. B. Zajonc, "The requirements and design of a standard group task," 1965.
- [29] R. W. Levenson and J. M. Gottman, "Marital interaction: physiological linkage and affective exchange." *Journal of personality and social psychology*, vol. 45, no. 3, p. 587, 1983.
- [30] N. Wegner and D. Zeaman, "Team and individual performances on a motor learning task," *The Journal of General Psychology*, vol. 55, no. 1, pp. 127–142, 1956.
- [31] L. R. Beach, M. Chi, G. Klein, P. Smith, and K. Vicente, "Naturalistic decision making and related research lines," in *Naturalistic decision making*. Psychology Press, 1997, pp. 77–88.
- [32] W. Zachary, J. A. Cannon-Bowers, P. Bilazarian, D. K. Krecker, P. J. Lardieri, and J. Burns, "The advanced embedded training system (aets): An intelligent embedded tutoring system for tactical team training," *International Journal of Artificial Intelligence in Education (IJAIED)*, vol. 10, pp. 257–277, 1998.

- [33] N. J. Cooke and S. M. Shope, "Synthetic task environments for teams: Certt's uav-ste," in *Handbook of human factors and ergonomics methods*. CRC Press, 2004, pp. 476–483.
- [34] E. Martin, D. R. Lyon, and B. T. Schreiber, "Designing synthetic tasks for human factors research: An application to uninhabited air vehicles," in *Proceedings of the Human Factors and Ergonomics Society annual meeting*, vol. 42, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 1998, pp. 123–127.
- [35] L. Ahonen, B. U. Cowley, A. Hellas, and K. Puolamäki, "Biosignals reflect pair-dynamics in collaborative work: Eda and ecg study of pair-programming in a classroom environment," *Scientific reports*, vol. 8, no. 1, p. 3138, 2018.
- [36] J. Toppi, G. Borghini, M. Petti, E. J. He, V. De Giusti, B. He, L. Astolfi, and F. Babiloni, "Investigating cooperative behavior in ecological settings: an eeg hyperscanning study," *PloS one*, vol. 11, no. 4, p. e0154236, 2016.
- [37] C. L. Stephens, A. R. Harrivel, A. T. Pope, and L. J. Prinzel III, "System and method for training of state-classifiers," Jan. 29 2019, uS Patent 10,192,173.
- [38] A. R. Harrivel, C. L. Stephens, R. J. Milletich, C. M. Heinich, M. C. Last, N. J. Napoli, N. Abraham, L. J. Prinzel, M. A. Motter, and A. T. Pope, "Prediction of cognitive states during flight simulation using multimodal psychophysiological sensing," in *AIAA Information Systems-AIAA Infotech@ Aerospace*, 2017, p. 1135.
- [39] J. K. Olsen, K. Sharma, N. Rummel, and V. Aleven, "Temporal analysis of multimodal data to predict collaborative learning outcomes," *British Journal of Educational Technology*, vol. 51, no. 5, pp. 1527–1547, 2020.
- [40] R. Veerabhadrapa, M. Ul Hassan, J. Zhang, and A. Bhatti, "Compatibility evaluation of clustering algorithms for contemporary extracellular neural spike sorting," *Frontiers in Systems Neuroscience*, vol. 14, p. 34, 2020. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fnsys.2020.00034>
- [41] A. Silva, G. Emmanuel, J. T. McClain, L. Matzen, and C. Forsythe, "Measuring expert and novice performance within computer security incident response teams," in *International Conference on Augmented Cognition*. Springer, 2015, pp. 144–152.
- [42] V. Echeverria, R. Martinez-Maldonado, and S. Buckingham Shum, "Towards collaboration translucence: Giving meaning to multimodal

- group data,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–16.
- [43] L. Eloy, A. EB Stewart, M. Jean Amon, C. Reinhardt, A. Michaels, C. Sun, V. Shute, N. D. Duran, and S. D’Mello, “Modeling team-level multimodal dynamics during multiparty collaboration,” in *2019 International Conference on Multimodal Interaction*, 2019, pp. 244–258.
 - [44] B. Cowley, M. Filetti, K. Lukander, J. Torniainen, A. Henelius, L. Ahonen, O. Barral, I. Kosunen, T. Valtonen, M. Huotilainen *et al.*, “The psychophysiology primer: a guide to methods and a broad review with a focus on human–computer interaction,” *Foundations and Trends® in Human–Computer Interaction*, vol. 9, no. 3-4, pp. 151–308, 2016.
 - [45] J. C. Gorman, M. J. Martin, T. A. Dunbar, R. H. Stevens, T. L. Galloway, P. G. Amazeen, and A. D. Likens, “Cross-level effects between neurophysiology and communication during team training,” *Human factors*, vol. 58, no. 1, pp. 181–199, 2016.
 - [46] C. Neubauer, M. Chollet, S. Mozgai, M. Dennison, P. Khooshabeh, and S. Scherer, “The relationship between task-induced stress, vocal changes, and physiological state during a dyadic team task,” in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, 2017, pp. 426–432.
 - [47] A. D. Walker, E. R. Muth, F. S. Switzer III, and P. J. Rosopa, “Predicting team performance in a dynamic environment: A team psychophysiological approach to measuring cognitive readiness,” *Journal of Cognitive Engineering and Decision Making*, vol. 7, no. 1, pp. 69–82, 2013.
 - [48] G. Murray and C. Oertel, “Predicting group performance in task-based interaction,” in *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, 2018, pp. 14–20.
 - [49] R. Stevens, T. Galloway, C. Berka, and A. Behneman, “A neurophysiologic approach for studying team cognition,” in *Interservice/Industry Training Simulation and Education Conference (I/ITSEC), Paper*, no. 10135, 2010.
 - [50] G. Chanel, J. M. Kivikangas, and N. Ravaja, “Physiological compliance for social gaming analysis: Cooperative versus competitive play,” *Interacting with Computers*, vol. 24, no. 4, pp. 306–316, 2012.
 - [51] S. Park, T. J. Grosser, A. A. Roebuck, and J. E. Mathieu, “Understanding work teams from a network perspective: A review and future research directions,” *Journal of Management*, p. 0149206320901573, 2020.

- [52] S. T. Bell, "Deep-level composition variables as predictors of team performance: a meta-analysis." *Journal of applied psychology*, vol. 92, no. 3, p. 595, 2007.
- [53] X. Cheng, X. Li, and Y. Hu, "Synchronous brain activity during cooperative exchange depends on gender of partner: A fnirs-based hyperscanning study," *Human brain mapping*, vol. 36, no. 6, pp. 2039–2048, 2015.
- [54] J. M. Kivikangas, J. Kätsyri, S. Järvelä, and N. Ravaja, "Gender differences in emotional responses to cooperative and competitive game play," *PloS one*, vol. 9, no. 7, p. e100318, 2014.
- [55] M. A. Marks, J. E. Mathieu, and S. J. Zaccaro, "A temporally based framework and taxonomy of team processes," *Academy of management review*, vol. 26, no. 3, pp. 356–376, 2001.
- [56] J. A. LePine, R. F. Piccolo, C. L. Jackson, J. E. Mathieu, and J. R. Saul, "A meta-analysis of teamwork processes: tests of a multidimensional model and relationships with team effectiveness criteria," *Personnel Psychology*, vol. 61, no. 2, pp. 273–307, 2008.
- [57] K. A. Jehn, L. Greer, S. Levine, and G. Szulanski, "The effects of conflict types, dimensions, and emergent states on group outcomes," *Group Decision and Negotiation*, vol. 17, no. 6, pp. 465–495, 2008.
- [58] H. J. Woodley, M. J. McLarnon, and T. A. O'Neill, "The emergence of group potency and its implications for team effectiveness," *Frontiers in psychology*, vol. 10, p. 992, 2019.
- [59] S. C. Collyer and G. S. Malecki, "Tactical decision making under stress: History and overview." 1998.
- [60] J. L. Wildman, E. Salas, and C. P. Scott, "Measuring cognition in teams: A cross-domain review," *Human factors*, vol. 56, no. 5, pp. 911–941, 2014.
- [61] N. Lehmann-Willenbrock, R. A. Meyers, S. Kauffeld, A. Neininger, and A. Henschel, "Verbal interaction sequences and group mood: Exploring the role of team planning communication," *Small Group Research*, vol. 42, no. 6, pp. 639–668, 2011.
- [62] P. Mitkidis, J. J. McGraw, A. Roepstorff, and S. Wallot, "Building trust: Heart rate synchrony and arousal during joint action increased by public goods game," *Physiology & behavior*, vol. 149, pp. 101–106, 2015.
- [63] W. He and B. Zheng, "Collaborative performance in laparoscopic teams: behavioral evidences from simulation," *Surgical endoscopy*, vol. 30, no. 10, pp. 4569–4574, 2016.

- [64] O. Špakov, H. Istance, K.-J. Rähkä, T. Viitanen, and H. Siirtola, "Eye gaze and head gaze in collaborative games," in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, 2019, pp. 1–9.
- [65] M. Balconi, L. Gatti, and M. E. Vanutelli, "When cooperation goes wrong: brain and behavioural correlates of ineffective joint strategies in dyads," *International Journal of Neuroscience*, vol. 128, no. 2, pp. 155–166, 2018, nIRS.
- [66] S. Wallot, P. Mitkidis, J. J. McGraw, and A. Roepstorff, "Beyond synchrony: joint action in a complex production task reveals beneficial effects of decreased interpersonal synchrony," *PloS one*, vol. 11, no. 12, p. e0168306, 2016.
- [67] E. Montague, J. Xu, and E. Chiou, "Shared experiences of technology and trust: An experimental study of physiological compliance between active and passive users in technology-mediated collaborative encounters," *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 5, pp. 614–624, 2014.
- [68] D. Mønster, D. D. Håkonsson, J. K. Eskildsen, and S. Wallot, "Physiological evidence of interpersonal dynamics in a cooperative production task," *Physiology & behavior*, vol. 156, pp. 24–34, 2016.
- [69] D. Døjbak Håkonsson, J. K. Eskildsen, L. Argote, D. Mønster, R. M. Burton, and B. Obel, "Exploration versus exploitation: Emotions and performance as antecedents and consequences of team decisions," *Strategic Management Journal*, vol. 37, no. 6, pp. 985–1001, 2016.
- [70] R. Stevens, T. Galloway, P. Wang, C. Berka, V. Tan, T. Wohlgemuth, J. Lamb, and R. Buckles, "Modeling the neurodynamic complexity of submarine navigation teams," *Computational and mathematical organization theory*, vol. 19, no. 3, pp. 346–369, 2013.
- [71] R. H. Stevens and T. L. Galloway, "Toward a quantitative description of the neurodynamic organizations of teams," *Social Neuroscience*, vol. 9, no. 2, pp. 160–173, 2014.
- [72] N. Sciaraffa, G. Borghini, P. Aricò, G. Di Flumeri, A. Colosimo, A. Bezarianos, N. Thakor, and F. Babiloni, "Brain interaction during cooperation: Evaluating local properties of multiple-brain network," *Brain sciences*, vol. 7, no. 7, p. 90, 2017.
- [73] E. Salas, D. E. Sims, and C. S. Burke, "Is there a "big five" in teamwork?" *Small group research*, vol. 36, no. 5, pp. 555–599, 2005, leadership skills.

- [74] H. K. Westli, B. H. Johnsen, J. Eid, I. Rasten, and G. Brattebø, "Team-work skills, shared mental models, and performance in simulated trauma teams: an independent group design," *Scandinavian journal of trauma, resuscitation and emergency medicine*, vol. 18, no. 1, pp. 1–8, 2010.
- [75] G. Knoblich and N. Sebanz, "The social nature of perception and action," *Current Directions in Psychological Science*, vol. 15, no. 3, pp. 99–104, 2006.
- [76] N. Warner, M. Letsky, and M. Cowen, "Cognitive model of team collaboration: Macro-cognitive focus," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 49, no. 3. Sage Publications Sage CA: Los Angeles, CA, 2005, pp. 269–273.
- [77] H. H. Clark, "Coordinating with each other in a material world," *Discourse studies*, vol. 7, no. 4-5, pp. 507–525, 2005.
- [78] M. R. Endsley, "Situation awareness global assessment technique (sagat)," in *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference*, May 1988, pp. 789–795 vol.3.
- [79] G. Chanel, M. Bétrancourt, T. Pun, D. Cereghetti, and G. Molinari, "Assessment of computer-supported collaborative processes using interpersonal physiological and eye-movement coupling," in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, 2013, pp. 116–122.
- [80] L. Ahonen, B. Cowley, J. Torniainen, A. Ukkonen, A. Vihavainen, and K. Puolamäki, "Cognitive collaboration found in cardiac physiology: Study in classroom environment," *PloS one*, vol. 11, no. 7, p. e0159178, 2016.
- [81] M. T. Tolston, A. J. Strang, G. J. Funke, B. Miller, R. Brown, and L. Menke, "Evaluating the relationship between team performance and joint attention with longitudinal multivariate mixed models," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 60, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2016, pp. 740–744.
- [82] E. Haataja, J. Malmberg, and S. Järvelä, "Monitoring in collaborative learning: Co-occurrence of observed behavior and physiological synchrony explored," *Computers in Human Behavior*, vol. 87, pp. 337–347, 2018.
- [83] M. Dindar, I. Alikhani, J. Malmberg, S. Järvelä, and T. Seppänen, "Examining shared monitoring in collaborative learning: A case of

- a recurrence quantification analysis approach," *Computers in Human Behavior*, vol. 100, pp. 335–344, 2019.
- [84] R. L. Moreland, "Are dyads really groups?" *Small Group Research*, vol. 41, no. 2, pp. 251–267, 2010.
- [85] S. Khaleghzadegan, S. Kazi, and M. A. Rosen, "Unobtrusive measurement of team cognition: A review and event-based approach to measurement design," *Contemporary Research*, pp. 95–113, 2020.
- [86] K. Stoeffler, Y. Rosen, M. Bolsinova, and A. A. von Davier, "Gamified performance assessment of collaborative problem solving skills," *Computers in Human Behavior*, vol. 104, p. 106036, 2020.
- [87] U. Boehm, D. Matzke, M. Gretton, S. Castro, J. Cooper, M. Skinner, D. Strayer, and A. Heathcote, "Real-time prediction of short-timescale fluctuations in cognitive workload," *Cognitive Research: Principles and Implications*, vol. 6, no. 1, pp. 1–29, 2021.
- [88] "Gazepoint gp3," 01 2009. [Online]. Available: <https://www.gazept.com/>
- [89] R. V. Palumbo, M. E. Marraccini, L. L. Weyandt, O. Wilder-Smith, H. A. McGee, S. Liu, and M. S. Goodwin, "Interpersonal autonomic physiology: A systematic review of the literature," *Personality and Social Psychology Review*, vol. 21, no. 2, pp. 99–141, 2017.
- [90] "emotiv epoc+." [Online]. Available: <https://www.emotiv.com/epoc/>
- [91] R. Veerabhadrappe, M. Ul Hassan, J. Zhang, and A. Bhatti, "Compatibility evaluation of clustering algorithms for contemporary extracellular neural spike sorting," *Frontiers in systems neuroscience*, vol. 14, p. 34, 2020.
- [92] "Unity 2019.4.16f1." [Online]. Available: <https://mirror-networking.com/>
- [93] J. Hackman, "The design of work teams. handbook of organizational behavior. edited by: Lorsch jw. 1987."
- [94] T. McMahan, I. Parberry, and T. Parsons, "Evaluating electroencephalography engagement indices during video game play," in *FDG*, 2015.
- [95] G. Di Flumeri, G. Borghini, P. Aricò, N. Sciaraffa, P. Lanzi, S. Pozzi, V. Vignali, C. Lantieri, A. Bichicchi, A. Simone, and F. Babiloni, "Eeg-based mental workload neurometric to evaluate the impact of different traffic and road conditions in real driving settings," *Frontiers*

- in Human Neuroscience*, vol. 12, p. 509, 2018. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fnhum.2018.00509>
- [96] R. Veerabhadrapa, I. T. Hettiarachchi, and A. Bhatti, "Using recurrence quantification analysis to quantify the physiological synchrony in dyadic ecg data," in *2021 IEEE International Systems Conference (SysCon)*, 2021, pp. 1–8.
 - [97] S. J. Guastello, "Physiological synchronization in a vigilance dual task." *Nonlinear dynamics, psychology, and life sciences*, vol. 20, no. 1, pp. 49–80, 2016.
 - [98] A. Malviya, G. A. Fink, L. Sego, and B. Endicott-Popovsky, "Situational awareness as a measure of performance in cyber security collaborative work," in *2011 Eighth International Conference on Information Technology: New Generations*. IEEE, Conference Proceedings, pp. 937–942.
 - [99] H. Vrzakova, M. J. Amon, A. Stewart, N. D. Duran, and S. K. D'Mello, "Focused or stuck together: multimodal patterns reveal triads' performance in collaborative problem solving," in *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 2020, pp. 295–304.
 - [100] N. J. Cooke, J. C. Gorman, J. L. Duran, and A. R. Taylor, "Team cognition in experienced command-and-control teams." *Journal of Experimental Psychology: Applied*, vol. 13, no. 3, p. 146, 2007.
 - [101] N. J. Cooke and J. C. Gorman, "Interaction-based measures of cognitive systems," *Journal of cognitive engineering and decision making*, vol. 3, no. 1, pp. 27–46, 2009.
 - [102] S. Brams, G. Ziv, O. Levin, J. Spitz, J. Wagemans, A. M. Williams, and W. F. Helsen, "The relationship between gaze behavior, expertise, and performance: A systematic review." *Psychological bulletin*, vol. 145, no. 10, p. 980, 2019.
 - [103] C. Zhang, C. Miller, K. Volkman, J. Meza, and K. Jones, "Evaluation of the team performance observation tool with targeted behavioral markers in simulation-based interprofessional education," *Journal of Interprofessional Care*, vol. 29, no. 3, pp. 202–208, 2015.
 - [104] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Physics reports*, vol. 438, no. 5-6, pp. 237–329, 2007.
 - [105] R. D. Dias, M. A. Zenati, R. Stevens, J. M. Gabany, and S. J. Yule, "Physiological synchronization and entropy as measures of team

- cognitive load," *Journal of biomedical informatics*, vol. 96, p. 103250, 2019.
- [106] N. J. Cooke, "Using team communication to understand team cognition in distributed vs. co-located mission environments," *Applied Psychology*, vol. 20051212, p. 044.
 - [107] D. W. Shaffer, W. Collier, and A. R. Ruis, "A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data," *Journal of Learning Analytics*, vol. 3, no. 3, pp. 9–45, 2016.
 - [108] Z. Swiecki and D. W. Shaffer, *Toward a taxonomy of team performance visualization tools*. International Society of the Learning Sciences, Inc.[ISLS]., 2018, eNA for team performance visualisation.
 - [109] F. E. Ritter, W. G. Kennedy, and B. J. Best, "The best papers from brims 2011: models of users and teams interacting," *Computational and Mathematical Organization Theory*, vol. 19, no. 3, pp. 283–287, 2013.
 - [110] R. A. Henning and K. T. Korbelak, "Social-psychophysiological compliance as a predictor of future team performance," *Psychologia*, vol. 48, no. 2, pp. 84–92, 2005.
 - [111] R. Henning, A. Armstead, and J. Ferris, "Social psychophysiological compliance in a four-person research team," *Applied ergonomics*, vol. 40, no. 6, pp. 1004–1010, 2009.
 - [112] X. Cui, D. M. Bryant, and A. L. Reiss, "Nirs-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation," *Neuroimage*, vol. 59, no. 3, pp. 2430–2437, 2012.
 - [113] P. Barach and M. B. Weinger, "Trauma team performance," *Trauma*, pp. 145–158, 2007.
 - [114] J. Schaveling, S. Blaauw, and K. van Montfort, "Predictors of group performance in a police criminal investigation department: the role of gender homogeneity, leadership and team characteristics," *Journal of Police and Criminal Psychology*, vol. 32, no. 4, pp. 358–368, 2017.
 - [115] J. E. Bono, T. M. Glomb, W. Shen, E. Kim, and A. J. Koch, "Building positive resources: Effects of positive events and positive reflection on work stress and health," *Academy of Management Journal*, vol. 56, no. 6, pp. 1601–1627, 2013.
 - [116] B. D. Lawson, T. W. Britt, A. M. Kelley, J. R. Athy, and S. M. Legan, "Computerized tests of team performance and crew coordination suitable for military/aviation settings," *Aerospace medicine and human performance*, vol. 88, no. 8, pp. 722–729, 2017.

- [117] C. Best and B. R. FLTLT, "Science and technology enablers of live virtual constructive training in the air domain," *Air & Space Power Journal*, vol. 32, no. 4, pp. 59–73, 2018.
- [118] J. R. Keebler, D. DiazGranados, and D. C. Smith, "Learning team theories and measurement through the game pandemic," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 58, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2014, pp. 442–446.
- [119] B. Maurer, M. Lankes, B. Stiglbauer, and M. Tscheligi, "Eyeco: Effects of shared gaze on social presence in an online cooperative game," in *International Conference on Entertainment Computing*. Springer, 2016, pp. 102–114.
- [120] S. Järvelä, J. M. Kivikangas, J. Kätsyri, and N. Ravaja, "Physiological linkage of dyadic gaming experience," *Simulation & Gaming*, vol. 45, no. 1, pp. 24–40, 2014.
- [121] M. M. Spapé, J. M. Kivikangas, S. Järvelä, I. Kosunen, G. Jacucci, and N. Ravaja, "Keep your opponents close: social context affects eeg and fmg linkage in a turn-based computer game," *PLoS One*, vol. 8, no. 11, p. e78795, 2013.
- [122] S. van der Haar, M. Koeslag-Kreunen, E. Euwe, and M. Segers, "Team leader structuring for team effectiveness and team learning in command-and-control teams," *Small group research*, vol. 48, no. 2, pp. 215–248, 2017.
- [123] M. D. McNeese, P. Bains, I. Brewer, C. Brown, E. S. Connors, J. Tyrone Jefferson, J. Rashaad E.T. Jones, and J. Ivanna Terrell, "The neocities simulation: Understanding the design and experimental methodology used to develop a team emergency management simulation," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 49, no. 3, pp. 591–594, 2005. [Online]. Available: <https://doi.org/10.1177/154193120504900380>
- [124] J. M. Morales, J. F. Ruiz-Rabelo, C. Diaz-Piedra, and L. L. Di Stasi, "Detecting mental workload in surgical teams using a wearable single-channel electroencephalographic device," *Journal of surgical education*, vol. 76, no. 4, pp. 1107–1115, 2019.
- [125] E. de Visser, T. Shaw, A. Mohamed-Ameen, and R. Parasuraman, "Modeling human-automation team performance in networked systems: Individual differences in working memory count," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 54, no. 14. SAGE Publications Sage CA: Los Angeles, CA, 2010, pp. 1087–1091.

- [126] M. D. McNeese, P. Bains, I. Brewer, C. Brown, E. S. Connors, T. Jefferson Jr, R. E. Jones Jr, and I. Terrell Jr, "The neocities simulation: Understanding the design and experimental methodology used to develop a team emergency management simulation," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 49, no. 3. SAGE Publications Sage CA: Los Angeles, CA, 2005, pp. 591–594.
- [127] G. J. Funke, A. M. Bennett, W. T. Nelson, and S. M. Galster, "Collaboration technologies and the supervisory control of ucavs in tactical c2: Effects on performance and workload," in *2007 International Symposium on Aviation Psychology*, 2007, p. 216.
- [128] J. C. Walliser, P. R. Mead, and T. H. Shaw, "The perception of teamwork with an autonomous agent enhances affect and performance outcomes," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 61, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2017, pp. 231–235.
- [129] W. Bezdek, J. Maleport, and R. Olshan, "Live, virtual & constructive simulation for real time rapid prototyping, experimentation and testing using network centric operations," in *AIAA Modeling and Simulation Technologies Conference and Exhibit*, 2008, p. 7090.
- [130] D. D. Hodson, B. L. Esken, A. J. Gutman, and R. R. Hill, "Quantifying radar measurement errors in a live–virtual–constructive environment to determine system viability: a case study," *The Journal of Defense Modeling and Simulation*, vol. 11, no. 2, pp. 115–124, 2014.
- [131] J. Hannay, O. Mevassvik, A. Skjeltorp, and K. Brathen, "Live, virtual, constructive (lvc) simulation for land operations training: Concept development & experimentation (cd&e)," in *Proc. NATO Modelling and Simulation Group Symp. on Integrating Modelling & Simulation in the Defence Acquisition Lifecycle and Military Training Curriculum (STO-MP-MSG-126)*, 2014.
- [132] F. Reer and N. C. Krämer, "Underlying factors of social capital acquisition in the context of online-gaming: Comparing world of warcraft and counter-strike," *Computers in Human Behavior*, vol. 36, pp. 179–189, 2014.
- [133] S. Andrist, A. R. Ruis, and D. W. Shaffer, "A network analytic approach to gaze coordination during a collaborative task," *Computers in Human Behavior*, vol. 89, pp. 339–348, 2018.
- [134] P. Sariñana-González, Á. Romero-Martínez, and L. Moya-Albiol, "Cooperation between strangers in face-to-face dyads produces more cardiovascular activation than competition or working alone," *Journal of Psychophysiology*, 2018.

- [135] —, “Cooperation induces an increase in emotional response, as measured by electrodermal activity and mood,” *Current Psychology*, vol. 36, no. 2, pp. 366–375, 2017.
- [136] F. Ringeval, A. Sonderegger, J. Sauer, and D. Lalanne, “Introducing the recola multimodal corpus of remote collaborative and affective interactions,” in *2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG)*. IEEE, 2013, pp. 1–8.
- [137] T. Nguyen, C. P. Lim, N. D. Nguyen, L. Gordon-Brown, and S. Nahavandi, “A review of situation awareness assessment approaches in aviation environments,” *IEEE Systems Journal*, vol. 13, no. 3, pp. 3590–3603, 2019.
- [138] K. Van De Merwe, H. Van Dijk, and R. Zon, “Eye movements as an indicator of situation awareness in a flight simulator experiment,” *The International Journal of Aviation Psychology*, vol. 22, no. 1, pp. 78–95, 2012.
- [139] M. M. Rodgers, V. M. Pai, and R. S. Conroy, “Recent advances in wearable sensors for health monitoring,” *IEEE Sensors Journal*, vol. 15, no. 6, pp. 3119–3126, 2014.
- [140] N. Hajari, I. Cheng, B. Zheng, and A. Basu, “Determining team cognition from delay analysis using cross recurrence plot,” in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2016, pp. 3482–3485.
- [141] S. Guastello, “Group dynamics: Adaptability, coordination, and leadership emergence,” *Chaos and complexity in psychology: Theory of non-linear dynamical systems*, pp. 402–433, 2009.
- [142] H. González, O. Infante, and C. Lerma, “Response to active standing of heart beat interval, systolic blood volume and systolic blood pressure: recurrence plot analysis,” in *Translational Recurrences*. Springer, 2014, pp. 109–123.
- [143] L. Cramer, I. T. Hettiarachchi, and S. Hanoun, “Effects of dynamic resilience on the reactivity of vagally mediated heart rate variability,” *Frontiers in Psychology*, vol. 11, p. 3998, 2021. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fpsyg.2020.579210>
- [144] I. Hettiarachchi, S. Hanoun, D. Nahavandi, J. Iskander, M. Hossny, and S. Nahavandi, “Towards more accessible physiological data for assessment of cognitive load - a validation study,” in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2018, pp. 3045–3050.

- [145] I. Hettiarachchi, S. Hanoun, D. Nahavandi, and S. Nahavandi, "Validation of polar oh1 optical heart rate sensor for moderate and high intensity physical activities." *PLoS ONE*, vol. 14, no. 5.
- [146] L. Claudia, I. Oscar, P.-G. Héctor, and V. J. Marco, "Poincaré plot indexes of heart rate variability capture dynamic adaptations after haemodialysis in chronic renal failure patients," *Clinical physiology and functional imaging*, vol. 23, no. 2, pp. 72–80, 2003.
- [147] S. M. Pikkujämsä, T. H. Mäkikallio, L. B. Sourander, I. J. Räihä, P. Puukka, J. Skyttä, C.-K. Peng, A. L. Goldberger, and H. V. Huikuri, "Cardiac interbeat interval dynamics from childhood to senescence: comparison of conventional and new measures based on fractals and chaos theory," *Circulation*, vol. 100, no. 4, pp. 393–399, 1999.
- [148] M. I. SHARIF, M. A. KHAN, and A. A. S. KHAN, "Comparative study on mental stress assessment from ecg signal using detrended fluctuation and recurrence quantification analysis."
- [149] L. Ahonen, B. Cowley, J. Torniainen, A. Ukkonen, A. Vihavainen, and K. Puolamäki, "Analysis of electrodermal activity recordings in pair programming from 2 dyads," 2016. [Online]. Available: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0159178>
- [150] H. A. Colibri, "Colibri-r package github." [Online]. Available: <https://github.com/bwrc/colibri/>
- [151] I. Daskalov and I. Christov, "Improvement of resolution in measurement of electrocardiogram rr intervals by interpolation," *Medical engineering & physics*, vol. 19, no. 4, pp. 375–379, 1997.
- [152] J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm." *IEEE Transactions on Biomedical Engineering, Biomedical Engineering, IEEE Transactions on, IEEE Trans. Biomed. Eng*, vol. BME-32, no. 3, pp. 230 – 236, 1985.
- [153] H. Akima, "A new method of interpolation and smooth curve fitting based on local procedures," *Journal of the ACM (JACM)*, vol. 17, no. 4, pp. 589–602, 1970.
- [154] S. Wallot and D. Mønster, "Calculation of average mutual information (ami) and false-nearest neighbors (fnn) for the estimation of embedding parameters of multidimensional time series in matlab," *Frontiers in psychology*, vol. 9, p. 1679, 2018.
- [155] C. L. Webber Jr and J. P. Zbilut, "Recurrence quantification analysis of nonlinear dynamical systems," *Tutorials in contemporary nonlinear methods for the behavioral sciences*, vol. 94, no. 2005, pp. 26–94, 2005.

- [156] N. Marwan, "How to avoid potential pitfalls in recurrence plot based data analysis," *International Journal of Bifurcation and Chaos*, vol. 21, no. 04, pp. 1003–1017, 2011.
- [157] J. P. Zbilut, N. Thomasson, and C. L. Webber, "Recurrence quantification analysis as a tool for nonlinear exploration of nonstationary cardiac signals," *Medical engineering & physics*, vol. 24, no. 1, pp. 53–60, 2002.
- [158] P. I. Good, *Resampling methods*. Springer.
- [159] M. Jafari and N. Ansari-Pour, "Why, when and how to adjust your p values?" *Cell Journal (Yakhteh)*, vol. 20, no. 4, p. 604, 2019.
- [160] U. Hasson and C. D. Frith, "Mirroring and beyond: coupled dynamics as a generalized framework for modelling social interactions," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 371, no. 1693, p. 20150366, 2016.
- [161] I. T. Hettiarachchi, T. T. Nguyen, and S. Nahavandi, "Motor imagery data classification for bci application using wavelet packet feature extraction," in *International Conference on Neural Information Processing*. Springer, 2014, pp. 519–526.
- [162] T. Nguyen, S. Nahavandi, A. Khosravi, D. Creighton, and I. Hettiarachchi, "Eeg signal analysis for bci application using fuzzy system," in *2015 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2015, pp. 1–8.
- [163] I. T. Hettiarachchi, T. T. Nguyen, and S. Nahavandi, "Multivariate adaptive autoregressive modeling and kalman filtering for motor imagery bci," in *2015 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2015, pp. 3164–3168.
- [164] D. Zhang, L. Yao, X. Zhang, S. Wang, W. Chen, R. Boots, and B. Benattallah, "Cascade and parallel convolutional recurrent neural networks on eeg-based intention recognition for brain computer interface," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.
- [165] S. Jerri, M. Murugappan, R. Nagarajan, and K. Wan, "Physiological signals based human emotion recognition: a review," in *2011 IEEE 7th International Colloquium on Signal Processing and its Applications*. IEEE, 2011, pp. 410–415.
- [166] T. Nguyen, I. Hettiarachchi, A. Khatami, L. Gordon-Brown, C. P. Lim, and S. Nahavandi, "Classification of multi-class bci data by

- common spatial pattern and fuzzy system," *IEEE Access*, vol. 6, pp. 27 873–27 884, 2018.
- [167] T. Song, W. Zheng, P. Song, and Z. Cui, "Eeg emotion recognition using dynamical graph convolutional neural networks," *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 532–541, 2018.
- [168] Y. Wang, S. Qiu, J. Li, X. Ma, Z. Liang, H. Li, and H. He, "Eeg-based emotion recognition with similarity learning network," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 1209–1212.
- [169] R. Veerabhadrapa, I. T. Hettiarachchi, and A. Bhatti, "Using recurrence quantification analysis to quantify the physiological synchrony in dyadic ecg data," in *2021 IEEE International Systems Conference (SysCon)*. IEEE, 2021, pp. 1–8.
- [170] I. T. Hettiarachchi, S. Hanoun, R. Veerabhadrapa, D. Jia, S. Hosking, and A. Bhatti, "Performance quantification and heart rate analysis in a repeated-trial simulation-based training task," in *2021 IEEE International Systems Conference (SysCon)*. IEEE, 2021, pp. 1–7.
- [171] 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.
- [172] T. Nguyen, I. Hettiarachchi, A. Khosravi, S. M. Salaken, A. Bhatti, and S. Nahavandi, "Multiclass eeg data classification using fuzzy systems," in *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, 2017, pp. 1–6.
- [173] M. Attia, I. Hettiarachchi, M. Hossny, and S. Nahavandi, "A time domain classification of steady-state visual evoked potentials using deep recurrent-convolutional neural networks," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 766–769.
- [174] Y. Luo, G. Wu, S. Qiu, S. Yang, W. Li, and Y. Bi, "Eeg-based emotion classification using deep neural network and sparse autoencoder," *Frontiers in Systems Neuroscience*, vol. 14, p. 43, 2020.
- [175] R. S. Lewis, N. Y. Weekes, and T. H. Wang, "The effect of a naturalistic stressor on frontal eeg asymmetry, stress, and health," *Biological psychology*, vol. 75, no. 3, pp. 239–247, 2007.

- [176] L.-C. Shi, Y.-Y. Jiao, and B.-L. Lu, "Differential entropy feature for eeg-based vigilance estimation," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2013, pp. 6627–6630.
- [177] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for eeg-based emotion classification," in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2013, pp. 81–84.
- [178] 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.
- [179] X.-h. Wang, T. Zhang, X.-m. Xu, L. Chen, X.-f. Xing, and C. P. Chen, "Eeg emotion recognition using dynamical graph convolutional neural networks and broad learning system," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2018, pp. 1240–1244.
- [180] Y. Li, L. Wang, W. Zheng, Y. Zong, L. Qi, Z. Cui, T. Zhang, and T. Song, "A novel bi-hemispheric discrepancy model for eeg emotion recognition," *IEEE Transactions on Cognitive and Developmental Systems*, 2020.
- [181] R. Veerabhadrapa, A. Bhatti, C. P. Lim, T. T. Nguyen, S. J. Tye, P. Monaghan, and S. Nahavandi, "Statistical modelling of artificial neural network for sorting temporally synchronous spikes," in *International Conference on Neural Information Processing*. Springer, 2015, pp. 261–272.
- [182] J. Zhang, T. Nguyen, S. Cogill, A. Bhatti, L. Luo, S. Yang, and S. Nahavandi, "A review on cluster estimation methods and their application to neural spike data," *Journal of neural engineering*, vol. 15, no. 3, p. 031003, 2018.
- [183] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Sjtu eeg emotion dataset <http://bcmi.sjtu.edu.cn/home/seed/seed.html>," 01 2013. [Online]. Available: <http://bcmi.sjtu.edu.cn/home/seed/seed.html>
- [184] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.
- [185] I. Hossain, A. Khosravi, I. Hettiarachchi, and S. Nahavandi, "Multiclass informative instance transfer learning framework for motor imagery-based brain-computer interface," *Computational intelligence and neuroscience*, vol. 2018, 2018.

- [186] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: <https://www.tensorflow.org/>
- [187] S. J. Reddi, S. Kale, and S. Kumar, "On the convergence of adam and beyond," *arXiv preprint arXiv:1904.09237*, 2019.
- [188] C. E. Nehme, "Modeling human supervisory control in heterogeneous unmanned vehicle systems," MASSACHUSETTS INST OF TECH CAMBRIDGE DEPT OF AERONAUTICS AND ASTRONAUTICS, Tech. Rep., 2009.
- [189] "Unity mirror-networking." [Online]. Available: <https://unity3d.com/get-unity/download>
- [190] K. L. Felmingham, C. Rennie, B. Manor, and R. A. Bryant, "Eye tracking and physiological reactivity to threatening stimuli in post-traumatic stress disorder," *Journal of anxiety disorders*, vol. 25, no. 5, pp. 668–673, 2011.
- [191] A. Pérez, M. Carreiras, and J. A. Duñabeitia, "Brain-to-brain entrainment: Eeg interbrain synchronization while speaking and listening," *Scientific reports*, vol. 7, no. 1, pp. 1–12, 2017.
- [192] S. Laborde, M. Raab, and N. P. Kinrade, "Is the ability to keep your mind sharp under pressure reflected in your heart? evidence for the neurophysiological bases of decision reinvestment," *Biological psychology*, vol. 100, pp. 34–42, 2014.
- [193] D. Buczyłowska and F. Petermann, "Age-related commonalities and differences in the relationship between executive functions and intelligence: Analysis of the nab executive functions module and wais-iv scores," *Applied Neuropsychology: Adult*, vol. 24, no. 5, pp. 465–480, 2017.
- [194] J. M. Bugg, N. A. Zook, E. L. DeLosh, D. B. Davalos, and H. P. Davis, "Age differences in fluid intelligence: Contributions of general slowing and frontal decline," *Brain and cognition*, vol. 62, no. 1, pp. 9–16, 2006.

- [195] C. R. De Luca, S. J. Wood, V. Anderson, J.-A. Buchanan, T. M. Proffitt, K. Mahony, and C. Pantelis, "Normative data from the cantab. i: development of executive function over the lifespan," *Journal of clinical and experimental neuropsychology*, vol. 25, no. 2, pp. 242–254, 2003.
- [196] F. L. da Silva, "Eeg and meg: relevance to neuroscience," *Neuron*, vol. 80, no. 5, pp. 1112–1128, 2013.
- [197] G. G. Knyazev, "Eeg delta oscillations as a correlate of basic homeostatic and motivational processes," *Neuroscience & Biobehavioral Reviews*, vol. 36, no. 1, pp. 677–695, 2012.
- [198] T. Gruber and M. M. Müller, "Oscillatory brain activity in the human eeg during indirect and direct memory tasks," *Brain research*, vol. 1097, no. 1, pp. 194–204, 2006.
- [199] M. P. Noonan, N. Adamian, A. Pike, F. Printzlau, B. M. Crittenden, and M. G. Stokes, "Distinct mechanisms for distractor suppression and target facilitation," *Journal of Neuroscience*, vol. 36, no. 6, pp. 1797–1807, 2016.
- [200] J. Tiihonen, M. Kajola, and R. Hari, "Magnetic mu rhythm in man," *Neuroscience*, vol. 32, no. 3, pp. 793–800, 1989.
- [201] 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.
- [202] B. Merker, "Cortical gamma oscillations: the functional key is activation, not cognition," *Neuroscience & Biobehavioral Reviews*, vol. 37, no. 3, pp. 401–417, 2013.
- [203] R. D. Dias, M. A. Zenati, R. Stevens, J. M. Gabany, and S. J. Yule, "Physiological synchronization and entropy as measures of team cognitive load," *Journal of biomedical informatics*, vol. 96, p. 103250, 2019.
- [204] R. Gilgen-Ammann, T. Schweizer, and T. Wyss, "Rr interval signal quality of a heart rate monitor and an ecg holter at rest and during exercise," *European journal of applied physiology*, vol. 119, no. 7, pp. 1525–1532, 2019.
- [205] C. A. Rothkopf, D. H. Ballard, and M. M. Hayhoe, "Task and context determine where you look," *Journal of vision*, vol. 7, no. 14, pp. 16–16, 2007.
- [206] E. H. Hess and J. M. Polt, "Pupil size in relation to mental activity during simple problem-solving," *Science*, vol. 143, no. 3611, pp. 1190–1192, 1964.

- [207] J. Beatty, B. Lucero-Wagoner *et al.*, "The pupillary system," *Handbook of psychophysiology*, vol. 2, no. 142-162, 2000.
- [208] G. Hauland, "Measuring individual and team situation awareness during planning tasks in training of en route air traffic control," *The International Journal of Aviation Psychology*, vol. 18, no. 3, pp. 290-304, 2008.
- [209] K. Huang, T. Bryant, and B. Schneider, "Identifying collaborative learning states using unsupervised machine learning on eye-tracking, physiological and motion sensor data." *International Educational Data Mining Society*, 2019.
- [210] M. Villamor and M. M. Rodrigo, "Predicting successful collaboration in a pair programming eye tracking experiment," in *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*, 2018, pp. 263-268.
- [211] F. Fasold, A. Nicklas, F. Seifriz, K. Schul, B. Noël, P. Aschendorf, and S. Klatt, "Gaze coordination of groups in dynamic events—a tool to facilitate analyses of simultaneous gazes within a team," *Frontiers in Psychology*, vol. 12, p. 816, 2021.
- [212] C.-M. Huang, S. Andrist, A. Saup্পé, and B. Mutlu, "Using gaze patterns to predict task intent in collaboration," *Frontiers in psychology*, vol. 6, p. 1049, 2015.
- [213] A. Goswami, G. Walia, M. McCourt, and G. Padmanabhan, "Using eye tracking to investigate reading patterns and learning styles of software requirement inspectors to enhance inspection team outcome," in *Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*. ACM, 2016, p. 34.
- [214] M. I. Coco and R. Dale, "Cross-recurrence quantification analysis of categorical and continuous time series: an r package," *Frontiers in psychology*, vol. 5, p. 510, 2014.
- [215] D. C. Richardson, R. Dale, and J. M. Tomlinson, "Conversation, gaze coordination, and beliefs about visual context," *Cognitive Science*, vol. 33, no. 8, pp. 1468-1482, 2009.
- [216] H. Andersen, C. Pedersen *et al.*, "Using eye tracking data to indicate team situation awareness," in *9th International Conference on Human-Computer Interaction*. Lawrence Erlbaum Associates, Inc., 2001, pp. 1318-1322.

- [217] P. Chopade, S. M. Khan, D. Edwards, and A. von Davier, "Machine learning for efficient assessment and prediction of human performance in collaborative learning environments," in *2018 IEEE International Symposium on Technologies for Homeland Security (HST)*. IEEE, 2018, pp. 1–6.
- [218] K. M. Ouverson, A. G. Ostrander, J. Walton, A. Kohl, S. B. Gilbert, M. C. Dorneich, E. Winer, and A. M. Sinatra, "Analysis of communication, team situational awareness, and feedback in a three-person intelligent team tutoring system," *Frontiers in Psychology*, vol. 12, 2021.
- [219] C. Prince, E. Ellis, M. T. Brannick, and E. Salas, "Measurement of team situation awareness in low experience level aviators," *The International Journal of Aviation Psychology*, vol. 17, no. 1, pp. 41–57, 2007.
- [220] P. Brennan, C. Holden, G. Shaw, S. Morris, and R. Oeppen, "Leading article: What can we do to improve individual and team situational awareness to benefit patient safety?" *British Journal of Oral and Maxillofacial Surgery*, vol. 58, no. 4, pp. 404–408, 2020.
- [221] M. Cherubini, M.-A. Nüssli, and P. Dillenbourg, "Deixis and gaze in collaborative work at a distance (over a shared map) a computational model to detect misunderstandings," in *Proceedings of the 2008 symposium on Eye tracking research & applications*, 2008, pp. 173–180.
- [222] F. Fasold, B. Noël, F. Wolf, and S. Hüttermann, "Coordinated gaze behaviour of handball referees: a practical exploration with focus on the methodical implementation," *Movement & Sport Sciences-Science & Motricité*, no. 102, pp. 71–79, 2018.
- [223] Y. Liu, P.-Y. Hsueh, J. Lai, M. Sangin, M.-A. Nussli, and P. Dillenbourg, "Who is the expert? analyzing gaze data to predict expertise level in collaborative applications," in *2009 IEEE international conference on Multimedia and Expo*. IEEE, 2009, pp. 898–901.
- [224] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Advances in psychology*. Elsevier, 1988, vol. 52, pp. 139–183.
- [225] C. C. Knott, W. T. Nelson, M. K. McCroskey, and B. T. Miller, "Analysis of eye movements in change detection with teams using a simulated tactical situation display," NATIONAL RESEARCH COUNCIL WASHINGTON DC, Tech. Rep., 2008.
- [226] S. Rubio, E. Díaz, J. Martín, and J. M. Puente, "Evaluation of subjective mental workload: A comparison of swat, nasa-tlx, and workload profile methods," *Applied psychology*, vol. 53, no. 1, pp. 61–86, 2004.

- [227] P. Hoonakker, P. Carayon, A. P. Gurses, R. Brown, A. Khunlertkit, K. McGuire, and J. M. Walker, "Measuring workload of icu nurses with a questionnaire survey: the nasa task load index (tlx)," *IIE transactions on healthcare systems engineering*, vol. 1, no. 2, pp. 131–143, 2011.
- [228] G. J. Funke, S. M. Russell, B. A. Knott, and B. T. Miller, "What can a multidimensional measure of stress tell us about team collaborative tools?" in *2009 International Symposium on Aviation Psychology*, 2009, p. 551.
- [229] M. Frostling-Henningsson, "First-person shooter games as a way of connecting to people: "brothers in blood"," *Cyberpsychology & behavior*, vol. 12, no. 5, pp. 557–562, 2009.
- [230] J. Jansz and M. Tanis, "Appeal of playing online first person shooter games," *Cyberpsychology & behavior*, vol. 10, no. 1, pp. 133–136, 2007.
- [231] F. Biocca, C. Harms, and J. K. Burgoon, "Toward a more robust theory and measure of social presence: Review and suggested criteria," *Presence: Teleoperators & virtual environments*, vol. 12, no. 5, pp. 456–480, 2003.
- [232] M. Hudson and P. Cairns, "The effects of winning and losing on social presence in team-based digital games," *Computers in Human Behavior*, vol. 60, pp. 1–12, 2016.
- [233] R. Schroeder, "Copresence and interaction in virtual environments: An overview of the range of issues," in *Presence 2002: Fifth international workshop*. Citeseer, 2002, pp. 274–295.
- [234] F. Biocca and C. Harms, "Defining and measuring social presence: Contribution to the networked minds theory and measure," *Proceedings of PRESENCE*, vol. 2002, pp. 1–36, 2002.
- [235] F. Biocca, C. Harms, and J. Gregg, "The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity," in *4th annual international workshop on presence, Philadelphia, PA*, 2001, pp. 1–9.
- [236] Y. A. De Kort, W. A. IJsselsteijn, and K. Poels, "Digital games as social presence technology: Development of the social presence in gaming questionnaire (spgq)," *Proceedings of PRESENCE*, vol. 195203, pp. 1–9, 2007.
- [237] G. Riva, J. Waterworth, and D. Murray, "6 measuring social presence in team-based digital games," in *Interacting with Presence*. De Gruyter Open Poland, 2014, pp. 83–101.

- [238] S. W. Kozlowski and B. S. Bell, "Team learning, development, and adaptation," in *Work group learning*. Psychology Press, 2007, pp. 39–68.
- [239] L. Argote, C. A. Insko, N. Yovetich, and A. A. Romero, "Group learning curves: the effects of turnover and task complexity on group performance 1," *Journal of Applied Social Psychology*, vol. 25, no. 6, pp. 512–529, 1995.
- [240] H. Vrzakova, M. J. Amon, A. E. Stewart, and S. K. D'Mello, "Dynamics of visual attention in multiparty collaborative problem solving using multidimensional recurrence quantification analysis," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–14.
- [241] I. T. Hettiarachchi, S. Hanoun, R. Veerabhadrapa, D. Jia, S. Hosking, and A. Bhatti, "Performance quantification and heart rate analysis in a repeated-trial simulation-based training task," in *2021 IEEE International Systems Conference (SysCon)*. IEEE, Conference Proceedings, pp. 1–7.
- [242] M. Dindar, I. Alikhani, J. Malmberg, S. Järvelä, and T. Seppänen, "Examining shared monitoring in collaborative learning: A case of a recurrence quantification analysis approach," *Computers in Human Behavior*, vol. 100, pp. 335–344, 2019.
- [243] J. Malmberg, S. Järvelä, J. Holappa, E. Haataja, X. Huang, and A. Siipo, "Going beyond what is visible: What multichannel data can reveal about interaction in the context of collaborative learning?" *Computers in Human Behavior*, vol. 96, pp. 235–245, 2019.
- [244] G. F. Wilson and C. A. Russell, "Real-time assessment of mental workload using psychophysiological measures and artificial neural networks," *Human factors*, vol. 45, no. 4, pp. 635–644, 2003.
- [245] S. G. Hart, "Nasa-task load index (nasa-tlx); 20 years later," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, no. 9. Sage publications Sage CA: Los Angeles, CA, 2006, pp. 904–908.
- [246] R. Mallubhatla, K. R. Pattipati, D. L. Kleinman, and Z.-B. Tang, "A model of distributed team information processing under ambiguity," *IEEE transactions on systems, man, and cybernetics*, vol. 21, no. 4, pp. 713–725, 1991.
- [247] W. B. Rouse, J. A. Cannon-Bowers, and E. Salas, "The role of mental models in team performance in complex systems," *IEEE transactions on systems, man, and cybernetics*, vol. 22, no. 6, pp. 1296–1308, 1992.

- [248] A. Ollero and I. Maza, *Multiple heterogeneous unmanned aerial vehicles*. Springer, 2007, vol. 37.
- [249] K. Zheng, G. Chen, M. Herschel, K. Y. Ngiam, B. C. Ooi, and J. Gao, "Pace: Learning effective task decomposition for human-in-the-loop healthcare delivery," in *Proceedings of the 2021 International Conference on Management of Data*, 2021, pp. 2156–2168.
- [250] M. M. Villamor and M. M. T. Rodrigo, "Gaze collaboration patterns of successful and unsuccessful programming pairs using cross-recurrence quantification analysis," *Research and Practice in Technology Enhanced Learning*, vol. 14, no. 1, pp. 1–22, 2019.
- [251] T. D. Phan, "Housing price prediction using machine learning algorithms: The case of melbourne city, australia," in *2018 International Conference on Machine Learning and Data Engineering (iCMLDE)*. IEEE, 2018, pp. 35–42.
- [252] W. Yoong, S. Patra-Das, N. Jeffers, M. Nauta, and W. Lodhi, "Developing situational awareness ('helicopter view')," *The Obstetrician & Gynaecologist*, vol. 23, no. 1, pp. 60–66, 2021.
- [253] H. J. Pijera-Díaz, H. Drachsler, S. Järvelä, and P. A. Kirschner, "Sympathetic arousal commonalities and arousal contagion during collaborative learning: How attuned are triad members?" *Computers in Human Behavior*, vol. 92, pp. 188–197, 2019.
- [254] K. Stoeffler, Y. Rosen, M. Bolsinova, and A. A. von Davier, "Gamified performance assessment of collaborative problem solving skills," *Computers in Human Behavior*, vol. 104, 2020.
- [255] S. Andrist, W. Collier, M. Gleicher, B. Mutlu, and D. Shaffer, "Look together: analyzing gaze coordination with epistemic network analysis," *Frontiers in Psychology*, vol. 6, no. 1016, 2015.
- [256] F. Hermens, R. Flin, and I. Ahmed, "Eye movements in surgery: A literature review," *Journal of Eye movement research*, vol. 6, no. 4, 2013.
- [257] J. Hausser and K. Strimmer, "Entropy inference and the james-stein estimator, with application to nonlinear gene association networks." *Journal of Machine Learning Research*, vol. 10, no. 7, 2009.
- [258] C. Bowers and F. Jentsch, "Team workload. handbook of human factors and ergonomic methods," 2005.
- [259] J. Camacho-Morles, G. R. Slemp, L. G. Oades, R. Pekrun, and L. Morrish, "Relative incidence and origins of achievement emotions in computer-based collaborative problem-solving: A control-value approach," *Computers in Human Behavior*, vol. 98, pp. 41–49, 2019.

A

Processing ECG to extract IBI and HRV measures

A.1 ECG DATA

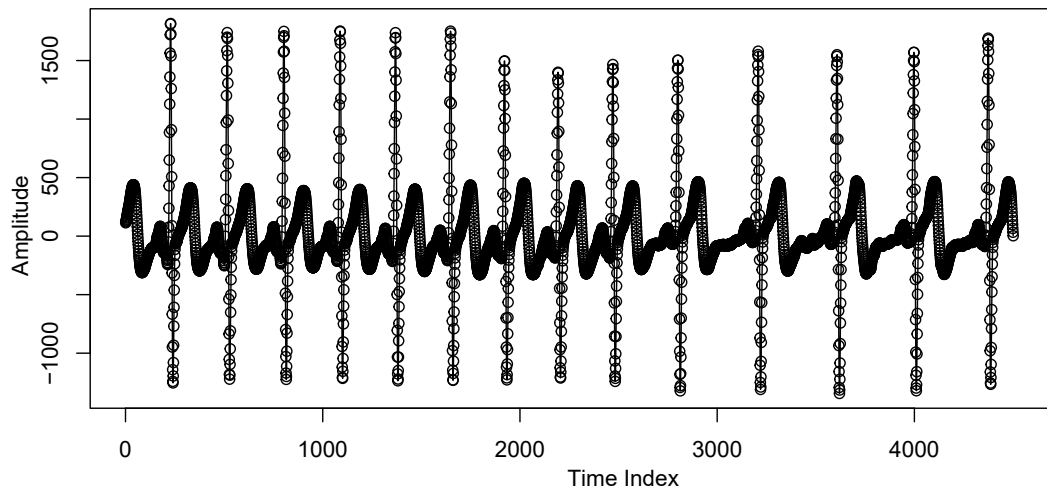


Figure A.1: Example of ECG data recorded at 500Hz sampling frequency

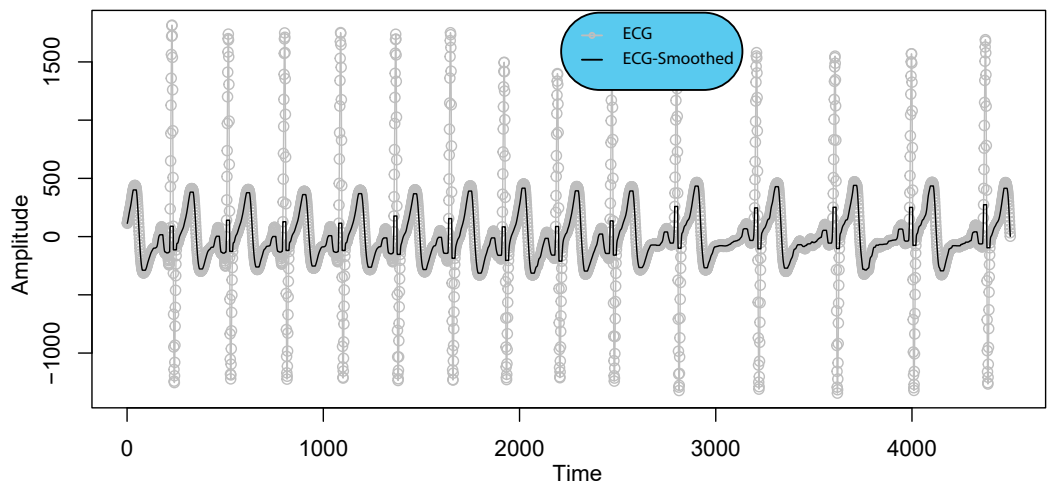


Figure A.2: Using running median filter to smoothen ECG signal

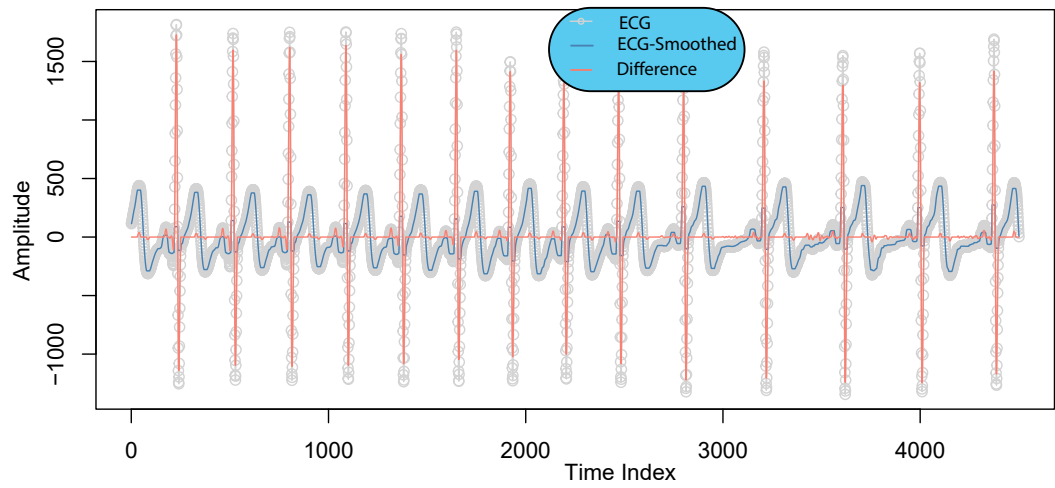


Figure A.3: Subtracting ECG with smoothened ECG to obtain spikes

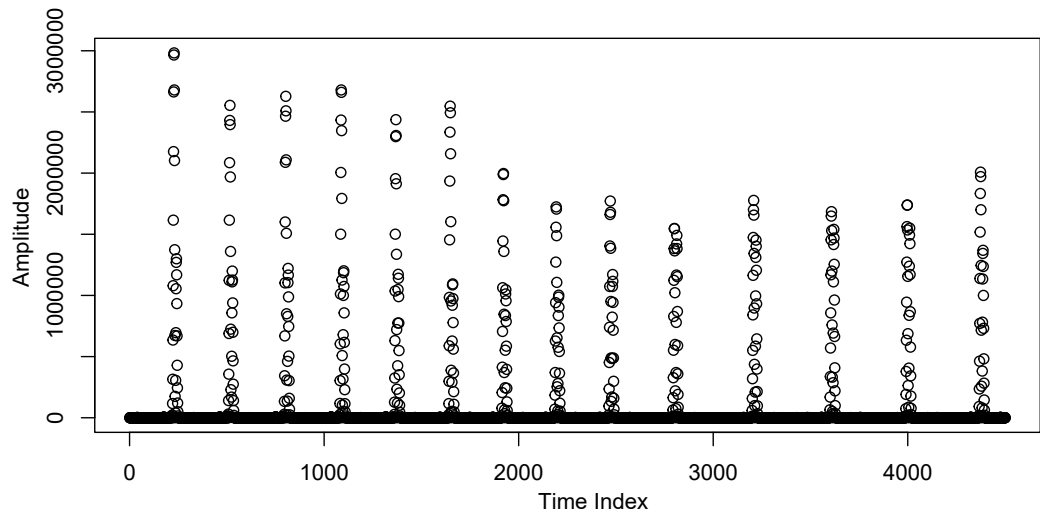


Figure A.4: Square transformation to eliminate negative peaks

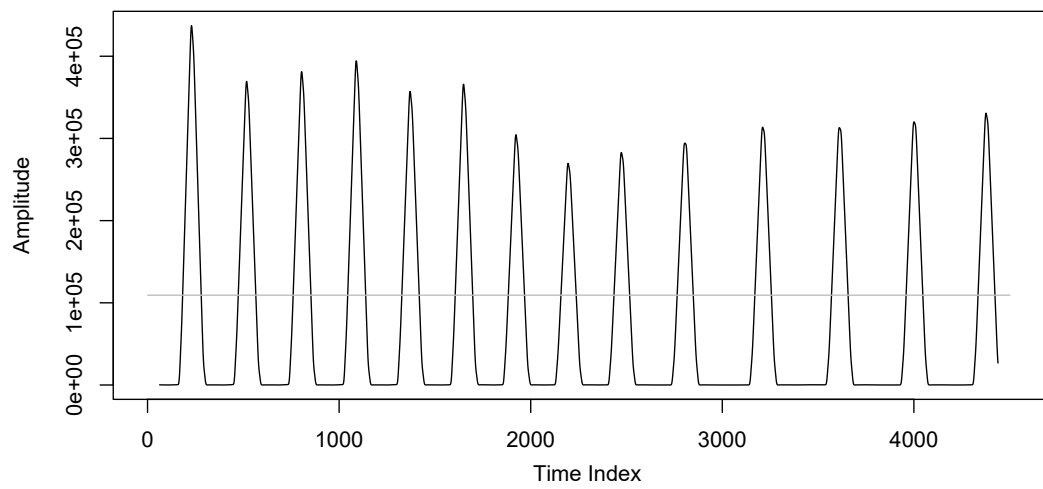


Figure A.5: Using threshold to detect area around R-peaks

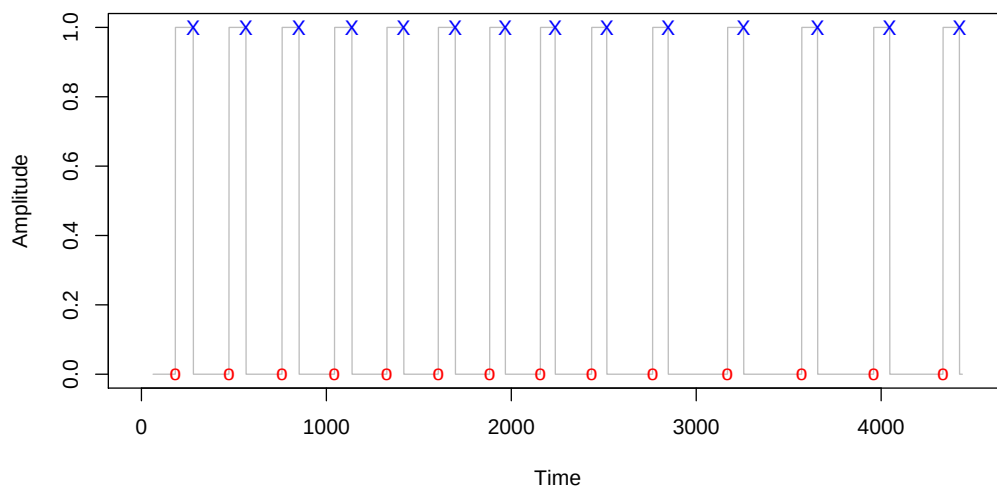


Figure A.6: Binary transformation and detection of rising edges and falling edges

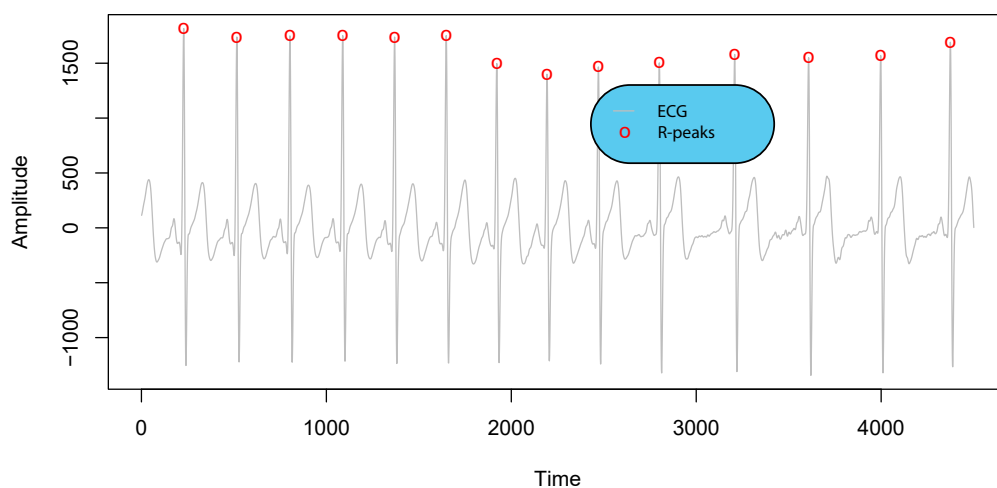


Figure A.7: R-peaks are identified by selecting the maximum ECG value between a pair of rising and falling edges

B

Recurrence Quantification Analysis

B.1 PREPARING DATA FOR RQA

RQA was used to explore recurrence (*RR*), determinism (*Det*), entropy (*H*) and laminarity (*Lam*) properties IBI data. An example of recurrence quantification analysis employed in this study is detailed below.

Since, IBI-series are not evenly spaced, it was first linearly interpolated to four times the number of samples, and then extrapolated such that there was 1 sample/second as shown in Figure. B.1. This ensures an even number of samples, allowing mathematical comparisons (such as correlation) of IBI-series between two participation. Thus, there were approximately 3600 samples in the IBI-series of any participant.

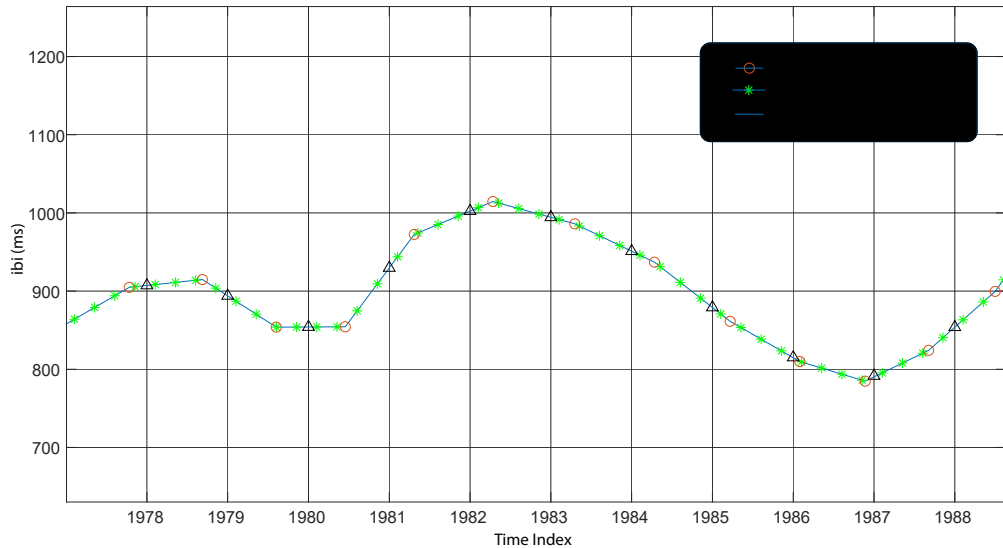


Figure B.1: A plot differences between raw IBI-series, after interpolating with four times the number of samples and extrapolating to exactly 1 sample/second.

B.2 RECURRENCE PLOTS

The first step is to extract RP to represent points recurring within a certain distance. Suppose, x_k represented an IBI-series, then a symmetric matrix $R_{m,n}(\varepsilon)$ was generated as follows:

$$R_{m,n}(\varepsilon) = \Theta(\varepsilon - \|x_m - x_n\|), \quad m, n = 1, 2, \dots, N \quad (\text{B.1})$$

where N is the total number of samples in x_k , ε is the minimum threshold distance (ε in this example was 0.01), $\Theta(\cdot)$ is Heaviside function such that if $a < 0$, $\Theta(a) = 0$ and $a \geq 0$, $\Theta(a) = 1$ and $\|\cdot\|$ is the Euclidean distance.

Since there were 3600 samples in the IBI-series, the resulting matrix is of size 3600×3600 as shown in figure. B.2. Hence, mapping such a matrix is not ideal for visual presentation. To provide a detailed picture, section of plots are zoomed to two different levels and shown in the figures B.3 and B.4, respectively.

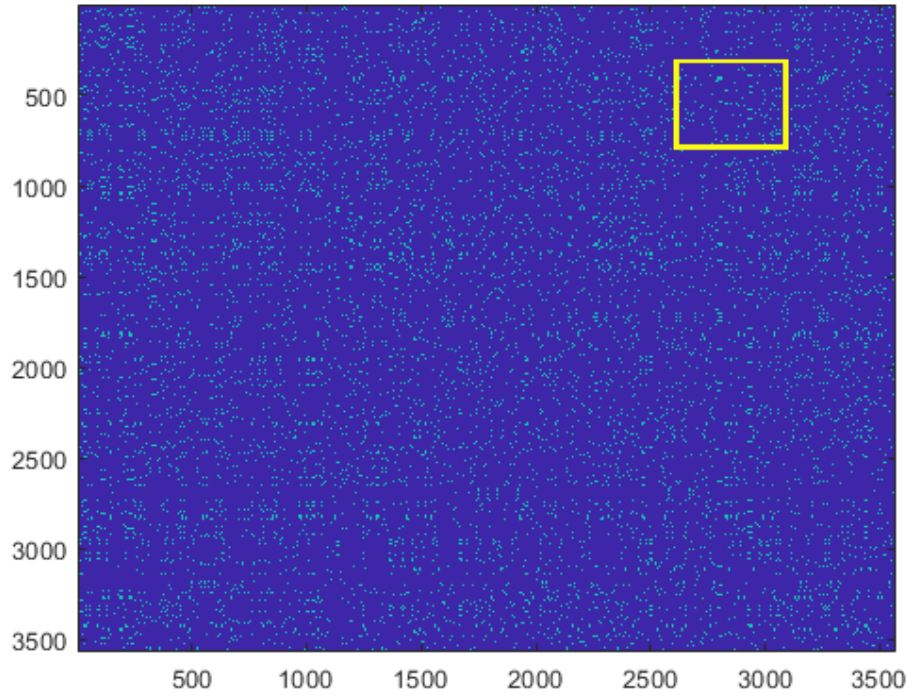


Figure B.2: Example showing recurrence plot of IBI-series x_k . For more detail, a section the highlighted section is further zoomed in figures. B.3 and B.4

Then RP is used derive the above mentioned RQA measures. Because RP is symmetric (i.e., $R_{m,n} \equiv R_{n,m}$), the measures are estimated only from the upper triangle of the RP.

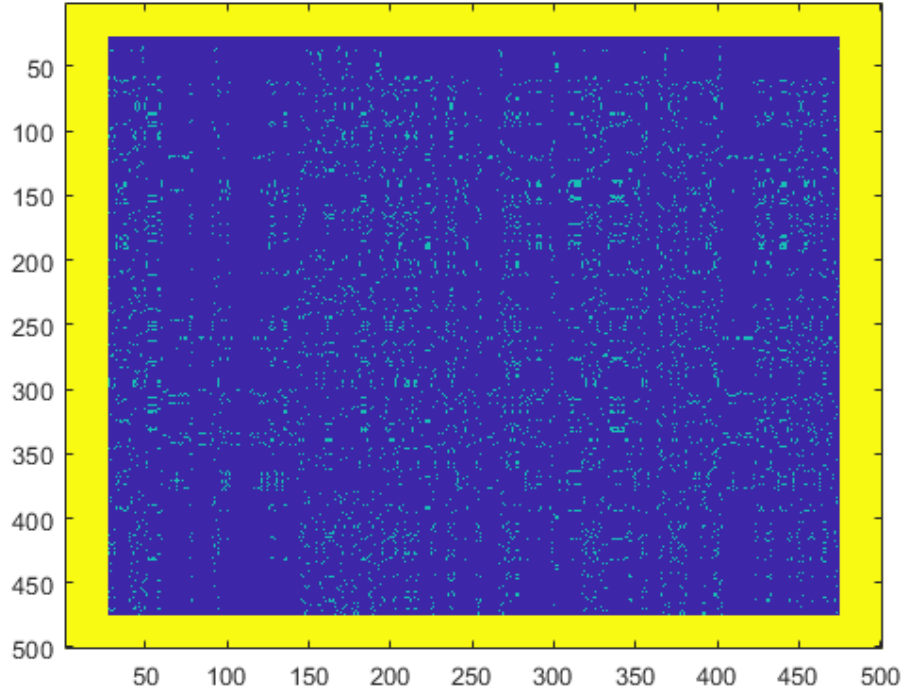


Figure B.3: Example showing zoomed version of the highlighted section of Figure. B.2. For more detail, the highlighted section is further zoomed in Figure. B.4

b.2.1 Recurrence Rate

This section demonstrates calculation of RR using eqn. B.2 from RP (Figure. B.2) i.e., $R(\epsilon)$. The RR measures density of recurring points in the upper triangle of recurrence matrix $R_{m,n}(\epsilon)$ (eqn. B.1) and for a sequence of T data points RR is estimated as,

$$RR = \frac{1}{T^2} \sum_{m=1}^T \sum_{n=m+1}^T R_{m,n}(\epsilon) \quad (\text{B.2})$$

Accordingly, number of data points $T=3600$ and the numerator estimates a total number of ones in the recurrence plot which was equivalent to 716766. Applying in eqn. B.2 gives $RR=0.0562$.

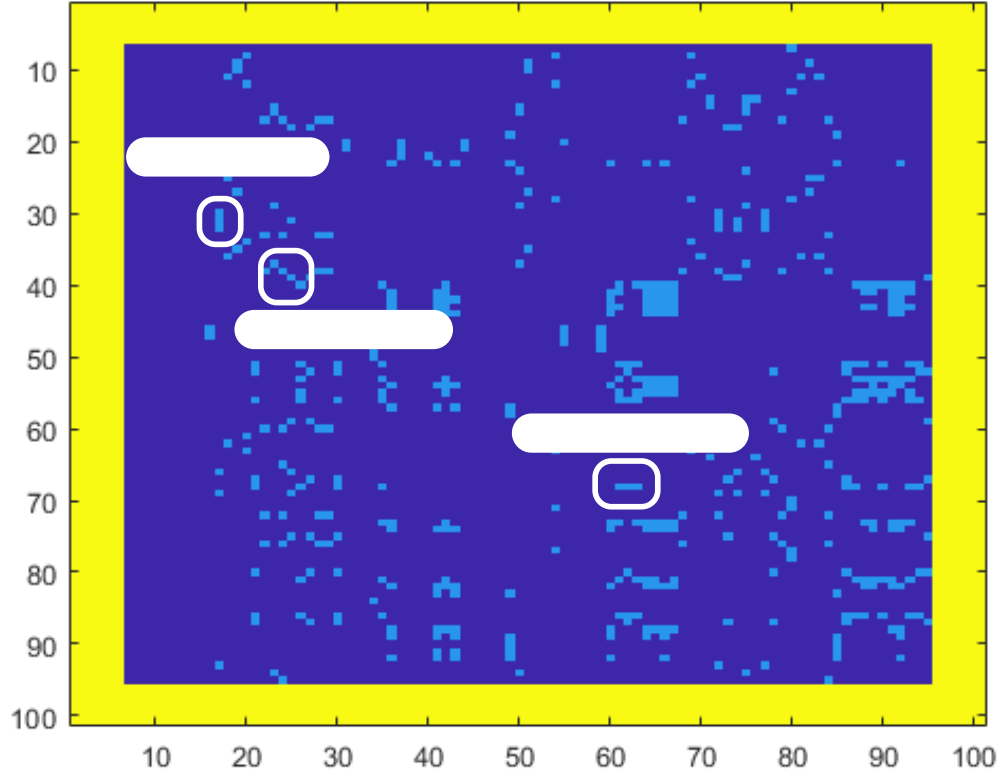


Figure B.4: Example showing zoomed version of the highlighted section of Figure B.3. Also shows what constitutes a vertical, horizontal and diagonal lines from RP.

B.3 DIAGONAL PROFILE MEASURES

Here we demonstrate calculating determinism Det using eqn. B.4. The first step is to estimate the histogram of all possible diagonal lines $hist(\epsilon, d)$. The eqn. B.3 sweeps diagonally the upper triangle of $R(\epsilon)$ accounting for every diagonal line of length= d . This process is repeated for all integer lengths ranging between $d=1, 2, 3, \dots, N$. Since, in this study RP is symmetric, simply multiplying $hist(\epsilon, d)$ by 2 gives histogram for entire RP, and excluding the centre diagonal.

For current example, $hist(\epsilon = 0.01, d)$ resulted in 10 different diagonal lines such that there were 278308 individual points i.e., $d_{l=1} = 278308$, $d_{l=2} = 30345$, $d_{l=3} = 4856$ and so on till $d_{l=10} = 1$ (one longest diagonal line).

b.3.1 Determinism

Then determinism (eqn. B.4) is simply the ratio of sum of all points starting with a minimum length=2 (numerator) to the sum of all possible points

(denominator). As shown in Figure. B.4, the numerator of the equation equal to 36379 and the denominator = 314687, resulting in $det = 0.1156$

$$hist(d) = \sum_{m,n=1}^T (1 - R_{m-1,n-1}(\varepsilon))(1 - R_{m-d,n-d}(\varepsilon)) \prod_{k=0}^{d-1} (R_{m+k,n+k}(\varepsilon)) \quad (B.3)$$

In the current analysis, a minimum of 2 consecutive points defined a line. Thus, from the histogram $hist(d)$ the *Det* property of a RP can be estimated as,

$$Det = \frac{\sum_{d=d_{min}}^T d \, hist(d)}{\sum_{d=1}^T d \, hist(d)} \quad (B.4)$$

In the present context, entropy measures the probability $p(d) = hist(d)/T_d$ of finding a diagonal line of certain length. Accordingly, entropy H is estimated as

$$H = \sum_{d=d_{min}}^T p(d) \log_n(p(d)) \quad (B.5)$$

B.4 VERTICAL-HORIZONTAL PROFILE MEASURES

Similar to the concept of diagonal profile analysis, horizontal and vertical profiles of RP were analyses. Accordingly, the histogram of vertical lines with a minimum length v can be estimated as

$$hist(v) = \sum_{m,n=1}^T (1 - R_{m,n}(\varepsilon))(1 - R_{m,n+v}(\varepsilon)) \prod_{k=0}^{v-1} (R_{m,n+k}(\varepsilon)) \quad (B.6)$$

Then the laminarity Lam associated with $R_{m,n}(\varepsilon)$ can be estimated as,

$$Lam = \frac{\sum_{v=v_{min}}^T v \, hist(v)}{\sum_{v=1}^T v \, hist(v)} \quad (B.7)$$

As shown in Figure. B.4, if minimum of two points appear diagonally consecutively then that condition qualifies as a diagonal line. Similarly, if two points appear horizontally and vertically consecutively, then it forms a horizontal and vertical line, respectively.

C

Tank Battle Task Environment

C.1 GAME MAIN MENU



Figure C.1: Selection menu of tank battle task.

C.2 PARAMETER CONFIGURATION WINDOW

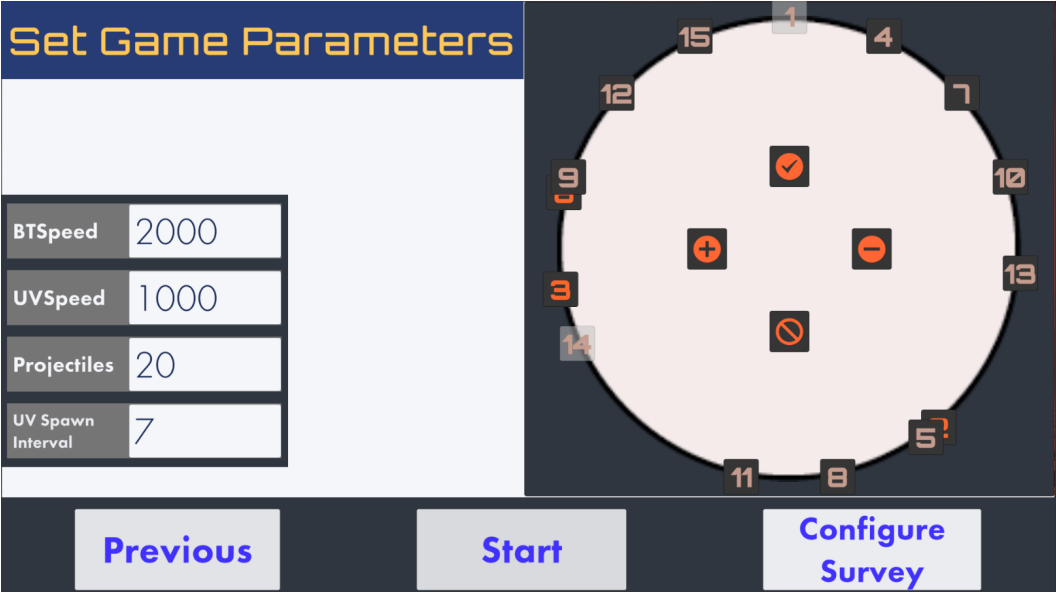


Figure C.2: Task parameters and the circular window for setting UGV spawn locations.

C.3

SURVEY CONFIGURATION WINDOW

Configure Survey

NASA_TLX

Edit

Demographic

Edit

Slide Number of Survey's

4 Surveys

CGQ

Edit

Task

Edit

Previous

Host

Figure C.3: Window showing all questionnaire configurable property of the task..

Survey Name

Previous

Choice

Slide

Input

Min 4

Done

Temporal Demand: How much time pressure did you feel due to the rate or pace at which the tasks occurred? Was the pace slow and leisurely or rapid and frantic?

0

10

Performance: How successful do you think you were in accomplishing the goals of the tasks? How satisfied were you with your performance in accomplishing these goals?

0

10

Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance ?

0

10

Frustration: How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

0

10

Figure C.4: Window showing editing and formatting of surveys and questions.

C.4

SURVEY ATTEMPTING WINDOW

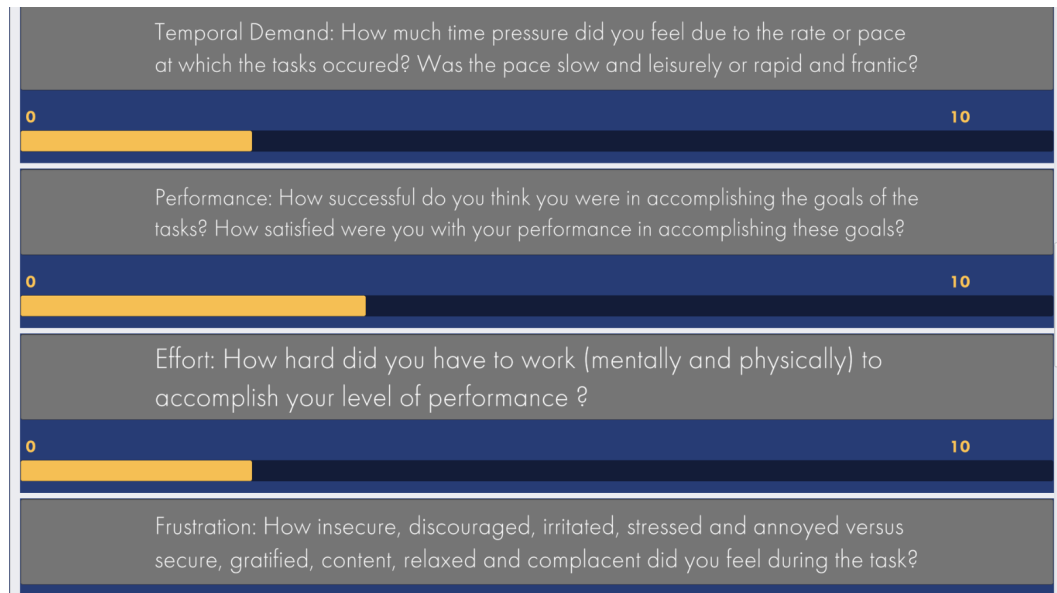


Figure C.5: Window showing participants answering a questionnaire.

C.5 TASK PERFORMANCE FEEDBACK WINDOW

Base Health			206%	
PlayerID	Hits	Miss	Bonus Lost	Bonus Gained
Alpha	41	19	40	
Bravo	46	7	0	
Charlie	56	4	0	
Total	143	30	40	14200

Figure C.6: Performance feedback window showing each members task scores and the based health at the end of a trial.