Monitoring Task Engagement using Facial Expressions and Body Postures

Akilesh Rajavenkatanarayanan
Heracleia Lab
Department of Computer Science and Engineering
The University of Texas at Arlington
akilesh.rajavenkatanarayanan@mavs.uta.edu

Ashwin Ramesh Babu
Heracleia Lab
Department of Computer Science and Engineering
The University of Texas at Arlington
ashwin.rameshbabu@mavs.uta.edu

Konstantinos Tsiakas
Heracleia Lab
Department of Computer Science and Engineering
The University of Texas at Arlington
konstantinos.tsiakas@mavs.uta.edu

Fillia Makedon
Heracleia Lab
Department of Computer Science and Engineering
The University of Texas at Arlington
makedon@uta.edu

ABSTRACT

As more industries adopt the use of robots to increase productivity, there is an increased need for effective human-robot interaction training, especially in the case of heavy and high precision robots. This implies the need for easy assessment methods that ensure accurate and personalized employee training. Most current assessments are done via manual observation and surveys. This paper addresses the need for the design of intelligent systems to assess a user's training needs based on the user's behavior and engagement while performing a vocational task simulation. In this paper, we propose a multi-sensory intelligent system to predict user engagement using facial expression and body posture data while the user performs a task to provide cognitive assessment of the user's capabilities, a critical factor in successful vocational performance using robots.

CCS CONCEPTS

• **Human-centered computing → User interface programming**; **User centered design**; **Information visualization**; • **Computing methodologies → Activity recognition and understanding**; **Neural networks**; *Scene understanding*; *Supervised learning by classification*;

KEYWORDS

Vocational Assessment, Body Pose Estimation, facial expression recognition, Electroencephalography (EEG), Convolutional Neural Networks, Task Engagement, Cognitive assessment, Sequence Learning, Brain Computer Interface (BCI), Industry 4.0

ACM Reference Format:

Akilesh Rajavenkatanarayanan, Ashwin Ramesh Babu, Konstantinos Tsiakas, and Fillia Makedon. 2018. Monitoring Task Engagement using Facial Expressions and Body Postures. In *IWISC 2018: 3rd International Workshop on Interactive and Spatial Computing, April 12–13, 2018, Richardson, TX, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3191801.3191816

IWISC 2018, April 12–13, 2018, Richardson, TX, USA

ACM ISBN 978-1-4503-5439-4/18/04…$15.00
https://doi.org/10.1145/3191801.3191816

1 INTRODUCTION

Assessment of user's performance is a commonly used method by employers in all industries. For example, IT companies measure an employee's performance every year in order to decide on salary raises and bonuses. Vocational assessment methods to assess an employee's capacity to work in a specific domain is done using task simulations. As more and more industries are moving towards the use of intelligent agents and robots (called 'Industry 4.0'), employers look for smart ways to assess suitability of an employee and thus ensure safety and productivity when working with robots. Such assessments use a common task to collect performance data from which an expert analyst can extract information to predict personalized training needs and design a system based on reproducible and repetitive user performance over time. Using machine learning algorithms, the aim is to automate user assessment and evaluation, based on behavior monitoring and analysis of the collected data. In this paper, we focus on developing a system that predicts user's engagement level using facial expressions and body postures while the user is engaged in task sessions.

2 RELATED WORK

Sequence learning (SL) tasks [18, 32] test a person' s ability to arrange thoughts and information in a meaningful order. These have been recognized as important abilities in vocational assessment, especially in the case of human-robot interaction where the user has to exercise attention, good working memory and decision making in interacting with robots safely and efficiently [20]. A great deal of cognitive science research has shown that sequence learning can be used as a tool to assess human behavior towards learning ability, temporary memory and attention [15, 16]. Different SL tasks can be used to predict and assess different abilities: *sequence prediction, sequence generation, sequence learning and sequence recognition.* In this paper we study the sequence learning and working memory of the users. This ability is considered important in industries involving assembly line work for example.

Besides sequence learning, there have been various other methods to enable vocational cognitive assessment [4, 25]. EEG signals have been used for emotion detection [21], stress detection[17], body movement tracking [30] and enjoyment evaluation [1].Research has also been conducted to prove the association of emotion with task engagement [26] and to estimate task engagement from Brain Mapping and neuro imaging data [27]. Most of these techniques involve the participant wearing different types of sensors while being assessed. This shows the need for designing unobtrusive vocational assessment methods to monitor user performance.

Our approach uses ***Convolutional Neural Networks (CNN)*** to recognize emotions from facial expressions. Since facial expressions are most often connected with body movements, our paper considers both types of data. The assumption is that both face and body data contribute to conveying an emotion of the individual being assessed, and can be combined to enable us to predict the fluctuation of the user's engagement while executing a task [24]. Hence, we can extract features such as the position of hands, head, body orientation, etc. from the person's body postures in order to recognize the user's emotional state while performing a task [11, 12]. Estimating emotions, facial expressions with computer vision has existed for a long time. Research in this area is still an ongoing research topic [2, 28, 31]. Research has also been done to monitor stress levels of a user using facial expressions, head movements and eye movements using Bayesian networks [22]. Research has also been conducted [7, 8, 29] that shows how task engagement is reflected in one's body postures.

3 FACE AND BODY MONITORING SYSTEM

3.1 Sequence Learning Task Assessment

We use the Sequence Learning (SL) task as a use case to evaluate users. The SL task is recognized as an important tool for assessing *cognitive load* and its relation to training by therapists and performance experts [5, 32–34]. The SL task involves listening or seeing a set of character sequences and being able to repeat them correctly in a certain amount of time. The sequences could be delivered via sound (speech) or image on a computer screen. Performance outcomes from SL task can help therapists and other experts to determine what particular treatment or rehabilitation an individual might need to enhance his/her performance in a given domain or application. Although task specific engagement assessment for cognitive tasks, memory tasks has been explored previously [6, 32] using sensors, we explore the possibility of using unobtrusive monitoring methods that include a user's facial expressions and body postures recorded with RGB camera towards task specific engagement in this study.

3.2 System Architecture

We propose an application with a GUI to administer tasks and monitor user performance metrics. Since performance evaluation is a very broad area, we focus on working memory assessment. For this purpose, we used the Sequence Learning experimental setup, explained further in the later sections. Using this GUI, an expert/administrator will be able to administer tasks and view performance metrics such as, current sequence, user response, and engagement value computed from data feed from the MUSE sensor. The MUSE sensor is an off-the shelf, low-cost EEG headband by InteraXan, and it has been used in multiple studies [10, 14, 19]. The GUI in **Figure 1** shows the initial design.

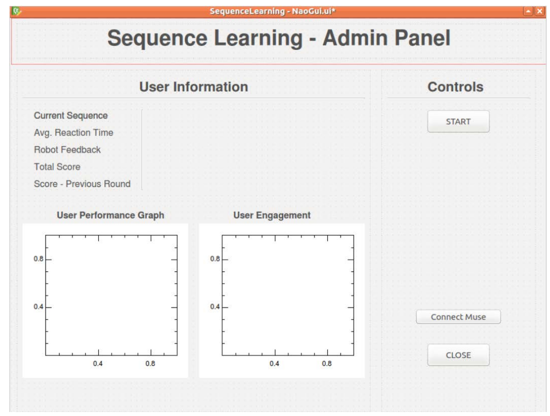

Figure 1: Initial design of proposed system administrator GUI to administer the cognitive assessment.

During the experiment, the user performs the SL task as the socially assistive robot NAO, dictates the sequences. While the user performs the task, the user's EEG data is recorded using the MUSE sensor to measure engagement. Also, an RGB camera is used to monitor user postures and facial expressions based on which the user's current emotional state is predicted. Analysis of the data was done to compare the change in the emotional state of the user, predicted by the vision system and the engagement value computed with the data feed from the MUSE sensor while performing the task. The computed engagement values were used as ground truth. This system's architecture is explained with the help of an architecture diagram in **Figure 2** below.

4 METHODOLOGY

4.1 Experimental Setup

Our experimental setup was inspired from the work described by Tsiakas et. al. [32], with minor modifications. The setup consists of a NAO robot that dictates sequences of random difficulty and gives feedback to the users as they perform the SL task. There are three levels of difficulty for the task, based on the length of the sequence L = [5,7,9] characters. Thus, level 1 has sequence length 5, level 2 has sequence length 7, and level 3 has sequence length 9 and this is the highest difficulty level. At this point of our research, the robot is used only as a system to dictate the sequence. During the task, the users have three buttons in front of them ('A', 'B', 'C') and as the robot dictates the sequences, the user plays the SL task while standing, by pressing the buttons in front of them and repeating the sequence from memory. An RGB camera is used to record the user performing the SL task and a MUSE sensor is used to capture the EEG signals of the user. The data from the camera is used to predict user's emotional state using face recognition and body posture data. A computer is used to run the application GUI through which the administrator can administer the task. The experimental setup is shown in **Figure 3**.

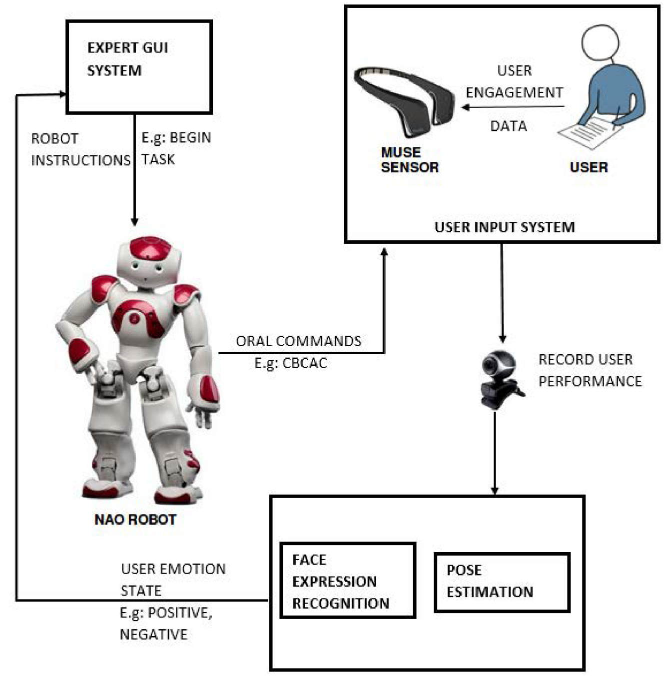

Figure 2: The proposed system for monitoring task engagement. The administrator GUI initiates the task. User's performance is recorded with the MUSE sensor and a camera records data for face recognition and body posture estimation.

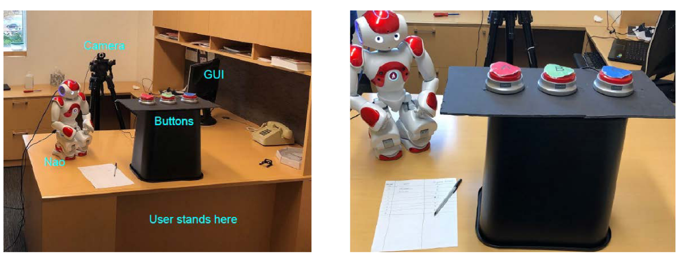

Figure 3: Experimental Setup

4.2 Experimental Procedure

For this experiment, we collected data from 11 male participants all of them graduate students in our university. We were very careful in choosing participants for this study because, we did not want participants who had previously come across this task or have played this task to avoid artificial facial expression and body postures.

At the beginning of the experiment, each user was asked to stand in front of the buttons and wear the MUSE EEG sensor. The task administrator then ensures the correct placement of the EEG sensor. The NAO robot then greets the user and explained the SL task. After this, the participants were asked to play the task which consisted of 12 turns per person and each turn with a random difficulty level. For example, the robot generated the sequence [c,b,b,a,a,c,c] randomly, which is considered a level 2 sequence. The task was intentionally designed to have long character sequences so that it is difficult to remember and respond.

Based on the correctness of the user's response, the robot gave feedback to users at random intervals. There were two types of feedback positive and negative. Some of the positive feedback used were: "very good!! Keep going." or "Oh!! you missed it. But keep going". Some of the negative feedback used were: "Maybe that was too easy." or "Looks like you are not paying any attention.". These feedback from the robot helped produce facial expression and body pose change that we could capture. We also recorded their performance in terms of success or failure and EEG data from MUSE for each turn.

5 IMPLEMENTATION

5.1 Administrator GUI

Figure 4 shows the final *administrator/expert GUI* that we designed, which an administrator can use to administer the task. This GUI shows the administrator, the participant's name which he entered at the beginning of the experiment, the user's response sequence, robot's feedback for the user's response if any and a score value. We can also visualize user performance and user engagement computed with data from the MUSE headband. The scores are computed at each turn as the user performs. As mentioned in section 4.2 there are 12 turns that each user had to play. For each turn, at level 1, the user gets a +2 for correct response or -2 for wrong, +3 or -3 for level 2, +4 or -4 for level 3. For example, in Figure 4, the score value -17 corresponds to the user score at turn 11. This can be computed based on above logic using the success or failure information available in the user performance graph in the same figure.

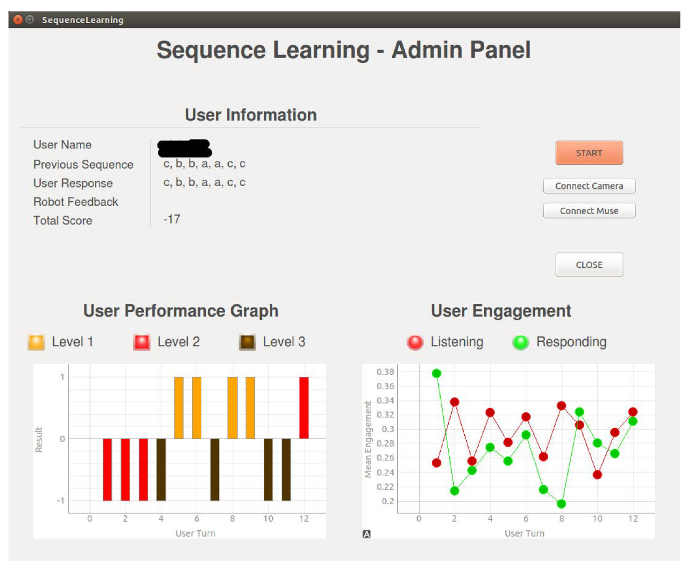

Figure 4: Administrator GUI with visualization

5.1.1 *User performance and engagement.*

Three metrics for user performance will be displayed for the administrator: the score value, the user performance graph and the engagement graph for each turn. For the user performance graph, success is considered as +1 and failure is considered -1 for the purpose of visualization and plotted accordingly. The user engagement value is calculated from the alpha, beta and the theta waves captured by the MUSE headband. The MUSE headband records the alpha, beta and theta waves from the frontal lobe at 20 Hz through four sensors on the headband. We took the mean value of these four sensors on the headband and calculated the mean engagement value using the formula $beta/alpha + theta$ [23]. We calculated the mean engagement value for users while listening and responding to the sequence. Figure 4 clearly depicts the Administrator GUI with the user performance graph and user engagement graph.

5.2 Computer Vision System

Two modules have been built for the emotion recognition, one from facial expression and the other from the upper body posture. The data from both the modules are used to make a prediction of user's engagement while performing the SL task.

5.2.1 Emotion Recognition Based on Facial Expressions using CNN. It is more likely for a person to express his/her emotion through facial expression. To predict emotions from facial expressions, a CNN was built based on the architecture in [3]. The system was trained and evaluated with FER 2013 dataset. It is a convolutional neural network with 4 residual depth-wise separable convolutions. Each of the convolution is followed by a batch normalization operation and a Rectified Linear Unit(ReLU) activation function. The last layer of the neural network is with a global average pooling and a softmax activation function. The architecture has approximately 60,000 parameters. The complete pipeline consists of a face detection module and emotion classification module. The system classified the frames to one of the three classes: positive, negative and neutral. **Figure 5** shows the architecture of the CNN.

5.2.2 Performance of the facial expression recognition system. The system when tested with labeled images with different facial expressions, achieved an accuracy level of 89% to 92% depending on factors such as angle of the image taken and amount of pixels that actually contribute in describing the class.

5.2.3 Emotion Recognition using body postures. We used *body pose* estimation to provide an additional way, besides facial expressions, to recognize certain emotions. The assumption here is that the human mind and body are connected and that the human body, with its intricate movements and many degrees of freedom, provides many alternative ways to express emotions. This means, any influence to the mind will be reflected in the body. Prior research [13] has shown that negative emotions such as disgust, fear, anger and sadness, are expressed using body motions, in totally different ways than they are expressed with facial expressions. In other words, emotions can be expressed through the face, or through the body or with both. For example, in case of body postures, some negative emotions includes moving the hands near the ears, over the face or near the chin. They may also include having the head hang down, the body lean forward, the head pulled back, or the hands crossing etc.

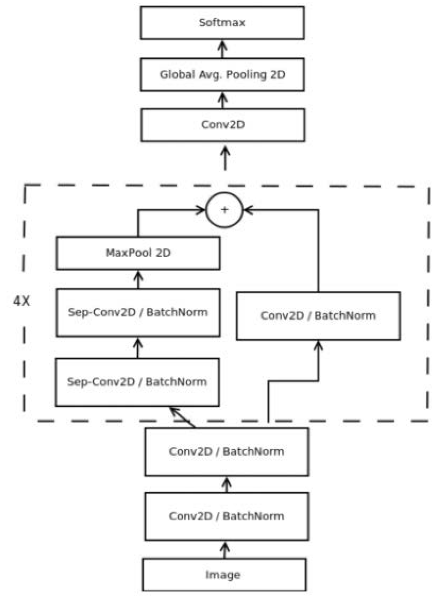

Figure 5: Architecture for Emotion recognition from Facial Expression

In our work, we use a deep CNN explained by Cao et. al. [9] to estimate the body joint positions. Initially the VGG19 network was built to extract different features from the image. The extracted features were used to detect the key-points. The feed-forward of the network builds a confidence matrix of different body joints points which is a 2D vector. In Parallel, a 2D vector field of part affinities were built which is used to find the association between the joint positions. The results from the joint estimation and the affinity matrix were concatenated to predict the joint positions of the individuals. With the joint positions given, an association is made to connect to the respective person. With the recognized joint positions, traditional computer vision algorithm was built to compute the class(negative, positive, neutral) based on the information provided by [11, 12]. **Figure 6** explains the association of the body postures with their emotional state. The joint positions are processed for every frame that is captured but the results are computed for every second which is 20 frames per second. The previous state of the emotion is checked to fix frame-wise errors and every second is assigned to an emotion that contributes to 60% in that second. Likewise, the previous state of the second is also considered to estimate the current state.

6 ACCURACY AND RESULTS

Data was collected from 11 participants. The emotion prediction from the face recognition module and the emotion prediction from the body postures module were used to predicts the emotional state of the user for every frame collected throughout the task. **Figure 7** shows how the facial expression module and emotion from the body postures system module were combined to predict emotions for each frame. Based on the emotion information extracted from the frames for every sequence, the change in engagement when

	Abdomen twist*	Chest bend**	Head bend**	Shoulder ad/abduct***	Shoulder swing**	Elbow bend**	Weight transfer	No. of postures
Anger	0	20, 40	−20, 25	−60, −80	45, 90	50, 110	Forwards	32
Disgust	−25, −50	−20, 0	−20	−60, −80	−25, 45	0, 50	Backwards	32
Fear	0	20, 40	25, 50, −20	−60	45, 90	50, 110	Backwards Neutral	24
Happiness	0	0, −20	0, −20	50	0, 45	0, 50	Forwards Neutral	32
Sadness	0, −25	0, 20	25, 50	−60, −80	0	0	Backwards Neutral	32
Surprise	0	−20	25, 50	50	−25, 0, 45	0, 50	Backwards	24

Figure 6: Association of Body postures association with Emotions based on work by based on the work by Mark Coulson [11].

compared to the previous sequence is predicted. The way the current system works such is that, if either of the system predicts negative emotion, it is considered as a drop in engagement. This is because the initial attempt considered each of the modalities as a weak classifier and the probability of the modality predicting correct is computed with the initial data collected as the ground truth. Figure 8 shows the accuracy plot of the behavior monitoring system(Vision system) versus the EEG signals(MUSE data) which produces an accuracy of 71% combined for all participants.

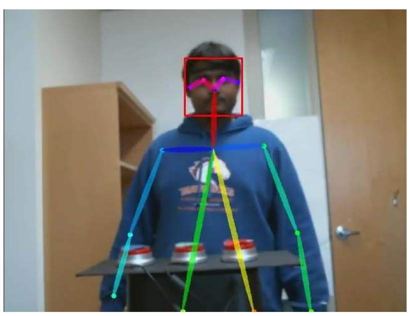

Figure 7: Predicting the emotion from face and the body joint positions by the vision system

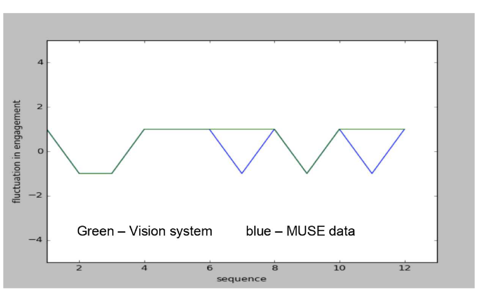

Figure 8: Accuracy of the MUSE data versus Behavioral Monitoring System

7 DISCUSSION AND FUTURE WORK

The accuracy and result discussed above provides us satisfactory results for this study but, this method is not ready to be implemented in the real world. It justifies further research in many areas of the proposed system. In particular, we plan on conducting a user study which will study the system's usability in real world applications. In addition, we will use different types of user performance visualizations to study the system's effectiveness on vocational training experts. Since the difficulty levels generated for the users were hardcoded to be randomly generated, future studies will incorporate the participant's performance at each session or turn, in generating a sequence of specific difficulty. This is currently not done in real practice. The sequences generated for each turn must adjust to user performance in the prior turn of the assessment and thus enable dynamic tracking of user performance. In addition, we will incorporate the tracking of other human factors in the assessment, besides memory. Furthermore, we plan user studies to predict the emotion using more accurate EEG data, rather than the alpha, beta and theta waves data from the MUSE sensor. Finally, we plan to consider additional constraints for the pose estimator to improve accuracy and to combine the MUSE data as an additional modality in order to predict the outcome of a sequence. The user's engagement prediction with face recognition and body posture is currently done offline. That is, the result of this prediction is not available in the administrator GUI. We are working on a real-time system that will display the result of prediction directly in the administrator GUI.

8 ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under award numbers 1565328, 1719031 and 1338118. Also, the authors would like to thank Michalis Papakostas, Maher Abujelala for their help throughout this work.

REFERENCES

[1] Maher Abujelala, Cheryl Abellanoza, Aayush Sharma, and Fillia Makedon. 2016. Brain-ee: Brain enjoyment evaluation using commercial eeg headband. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, 33.

[2] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. 2006. Face description with local binary patterns: Application to face recognition. *IEEE transactions on pattern analysis and machine intelligence* 28, 12 (2006), 2037–2041.

[3] Octavio Arriaga, Matias Valdenegro-Toro, and Paul Plöger. 2017. Real-time Convolutional Neural Networks for Emotion and Gender Classification. *arXiv preprint arXiv:1710.07557* (2017).

[4] Ashwin Ramesh Babu, Akilesh Rajavenkatanarayanan, Maher Abujelala, and Fillia Makedon. 2017. Votre: A vocational training and evaluation system to compare training approaches for the workplace. In *International Conference on Virtual, Augmented and Mixed Reality*. Springer, 203–214.

[5] Maria Bannert. 2002. Managing cognitive loadâĂŤrecent trends in cognitive load theory. *Learning and instruction* 12, 1 (2002), 139–146.

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

[7] Nadia Bianchi-Berthouze. 2013. Understanding the role of body movement in player engagement. *Human–Computer Interaction* 28, 1 (2013), 40–75.

[8] Nadia Bianchi-Berthouze, Paul Cairns, Anna Cox, Charlene Jennett, and Whan Woong Kim. 2006. On posture as a modality for expressing and recognizing emotions. In *Emotion and HCI workshop at BCS HCI London*.

[9] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2016. Realtime multi-person 2d pose estimation using part affinity fields. *arXiv preprint arXiv:1611.08050* (2016).

[10] Raymundo Cassani, Hubert Banville, and Tiago H Falk. 2015. MuLES: An open source EEG acquisition and streaming server for quick and simple prototyping and recording. In *Proceedings of the 20th International Conference on Intelligent User Interfaces Companion*. ACM, 9–12.

[11] Mark Coulson. 2004. Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Dependence. *Journal of Nonverbal Behavior* 28, 2 (01 Jun 2004), 117–139. https://doi.org/10.1023/B:JONB.0000023655.25550.be

[12] Beatrice De Gelder. 2006. Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience* 7, 3 (2006), 242–249.

[13] Kristin Fraser, Irene Ma, Elise Teteris, Heather Baxter, Bruce Wright, and Kevin McLaughlin. 2012. Emotion, cognitive load and learning outcomes during simulation training. *Medical education* 46, 11 (2012), 1055–1062.

[14] Leo Galway, Paul McCullagh, Gaye Lightbody, Chris Brennan, and David Trainor. 2015. The potential of the brain-computer interface for learning: a technology review. In *Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on*. IEEE, 1554–1559.

[15] Susan E Gathercole and Alan D Baddeley. 2014. *Working memory and language*. Psychology Press.

[16] Stephen Grossberg and Lance R Pearson. 2008. Laminar cortical dynamics of cognitive and motor working memory, sequence learning and performance: toward a unified theory of how the cerebral cortex works. *Psychological review* 115, 3 (2008), 677.

[17] Martijn Haak, Steven Bos, Sacha Panic, and LJM Rothkrantz. 2009. Detecting stress using eye blinks and brain activity from EEG signals. *Proceeding of the 1st driver car interaction and interface (DCII 2008)* (2009), 35–60.

[18] Larry E Humes and Shari S Floyd. 2005. Measures of working memory, sequence learning, and speech recognition in the elderly. *Journal of Speech, Language, and Hearing Research* 48, 1 (2005), 224–235.

[19] Thrasyvoulos Karydis, Filipe Aguiar, Simmie L Foster, and Andreas Mershin. 2015. Performance characterization of self-calibrating protocols for wearable EEG applications. In *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, 38.

[20] Azfar Khalid, Pierre Kirisci, Zied Ghrairi, Jürgen Pannek, and Klaus-Dieter Thoben. 2017. Safety Requirements in Collaborative Human–Robot Cyber-Physical System. In *Dynamics in Logistics*. Springer, 41–51.

[21] Mu Li and Bao-Liang Lu. 2009. Emotion classification based on gamma-band EEG. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE, 1223–1226.

[22] Wenhui Liao, Weihong Zhang, Zhiwei Zhu, and Qiang Ji. 2005. A real-time human stress monitoring system using dynamic Bayesian network. In *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*. IEEE, 70–70.

[23] Timothy McMahan, Ian Parberry, and Thomas D Parsons. 2015. Evaluating Electroencephalography Engagement Indices During Video Game Play.. In *FDG*.

[24] Hanneke KM Meeren, Corné CRJ van Heijnsbergen, and Beatrice de Gelder. 2005. Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences of the United States of America* 102, 45 (2005), 16518–16523.

[25] Ziad S Nasreddine, Natalie A Phillips, Valérie Bédirian, Simon Charbonneau, Victor Whitehead, Isabelle Collin, Jeffrey L Cummings, and Howard Chertkow. 2005. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society* 53, 4 (2005), 695–699.

[26] Reinhard Pekrun and Lisa Linnenbrink-Garcia. 2012. Academic emotions and student engagement. In *Handbook of research on student engagement*. Springer, 259–282.

[27] Russell A Poldrack. 2006. Can cognitive processes be inferred from neuroimaging data? *Trends in cognitive sciences* 10, 2 (2006), 59–63.

[28] Ashok Samal and Prasana A Iyengar. 1992. Automatic recognition and analysis of human faces and facial expressions: A survey. *Pattern recognition* 25, 1 (1992), 65–77.

[29] Jyotirmay Sanghvi, Ginevra Castellano, Iolanda Leite, André Pereira, Peter W McOwan, and Ana Paiva. 2011. Automatic analysis of affective postures and body motion to detect engagement with a game companion. In *Proceedings of the 6th international conference on Human-robot interaction*. ACM, 305–312.

[30] J Stastny, Pavel Sovka, and A Stancak. 2001. EEG signal classification. In *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, Vol. 2. IEEE, 2020–2023.

[31] Yingli Tian, Takeo Kanade, and Jeffrey F Cohn. 2011. Facial expression recognition. In *Handbook of face recognition*. Springer, 487–519.

[32] Konstantinos Tsiakas, Cheryl Abellanoza, Maher Abujelala, Michalis Papakostas, Tasnim Makada, and Fillia Makedon. [n. d.]. Towards Designing a Socially Assistive Robot for Adaptive and Personalized Cognitive Training. ([n. d.]).

[33] Jeroen JG Van Merrienboer and John Sweller. 2005. Cognitive load theory and complex learning: Recent developments and future directions. *Educational psychology review* 17, 2 (2005), 147–177.

[34] Bernard Weiner. 1972. Theories of motivation: From mechanism to cognition. (1972).