

29th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2025)

Using EEG and gaze tracking for verifying procedure usability

Fabio Miranda, Rossitza Setchi*

Research Centre for AI, Robotics and Human-Machine Systems, School of Engineering (IROHMS), Cardiff University, Cardiff CF24 3AA, United Kingdom

Abstract

Efficiency is paramount in industry and can be vastly improved by driving improvements in procedure design. This study explores the feasibility of using electroencephalography (EEG) and gaze tracking for assessing the quality of procedural design. The study hypothesises that EEG and gaze information can be indicative of the difficulties workers face during procedural tasks and therefore used to identify areas for procedural design improvements. Fifteen participants completed a number of origami tasks, designed to contain problem points predicted to stimulate detectable emotional responses. The analysis of the fixation rate and pupil diameter revealed that participants fixated on steps either directly preceding or following the identified problem points, and pupil size increased during the execution of these steps. EEG analysis included power spectral densities (PSD) and event related potentials (ERP), though ERP was found not to be indicative enough for the purpose of this study. Participants provided feedback on challenging steps, which aligned with predictions. Brain activity patterns while undertaking problematic steps compared to base unstimulated brain activity showed that theta activity increased across the whole brain in 77% of recordings, most prominently in the left temporal region; delta activity increased in 65% of recordings, most prominently in the left temporal region; alpha activity decreased in the occipital region in 65% of recordings but increased in the left temporal region in 70%; and beta activity increased in the left frontal region in 74% of recordings. These results validated the hypothesis, as they showed clear trends in the reactions to problem points. Finally, a framework is proposed for a procedure problem point identification using EEG and gaze tracking, and recommendations for further research have been outlined.

© 2025 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the KES International.

Keywords: EEG; electroencephalography; gaze tracking; usability studies; procedure usability;

1. Introduction

Maximising efficiency drives innovation in manufacturing. Well-designed procedures provide clear, concise instructions for manufacturing tasks, but poorly designed procedures increase the risk of errors and inefficiencies.

Problems with the usability of these procedures can be difficult to diagnose if there are language barriers or gulfs between task complexity and production staff skillsets, so it may not always be realistic or practical to rely solely on written or verbal feedback. This raises the question of whether there are tools which can be used for objective measuring of procedure usability.

Electroencephalography (EEG) and gaze tracking are often used together to provide a broader perspective on the relationships between cognition and vision [1]. While considering the current impracticalities and limitation of these technologies, this study aims to determine if there is an opportunity for further research and possible development of EEG and gaze tracking tools for measuring the quality of procedure design in the manufacturing industry.

2. Literature review

2.1. *Electroencephalography and emotions*

EEG is widely considered as a reliable method for detecting human emotions, and research has been conducted using various stimuli to induce these emotions. Pictures and audio stimuli are often used, as there are standardized databases available rating stimuli for valence and arousal. The valence-arousal scale has been employed to classify emotions [2], with high valence/high arousal representing happiness, low valence/low arousal representing sadness, low valence/high arousal representing anger, etc. It has been found that EEG amplitude is greater during negative emotion than positive emotion, making it easier to identify, and the most responsive region of the brain was the front. Alpha, beta and gamma frequencies were found to be the most effective frequencies of EEG in emotion detection [2].

There are many methods of analysing EEG data, depending on the context of its analysis. Almost all data acquired must be pre-processed to remove contamination from non-brain activity, and improving the reliability and usefulness of the data collected [3]. Filtering is used to remove noise, with high-pass filters removing low frequency noise, low-pass removing high frequency noise and notch filters set at the local mains electricity frequency to remove power line noise. Independent component analysis (ICA) separates EEG into statistically independent components, associated with specific types of activity, which can then be identified as neural and non-neural components, and be selected for omission from further analysis; wavelet denoising is used for artifact removal, and referencing can be used to remove biases and noise. When analysing pre-processed data, a number of different frequency-domain and time-domain feature extraction techniques can be used. In user-experience studies, the most commonly used time-domain method is event related potential (ERP), which takes the average of EEG signals from a certain time frame (i.e. a stimulus or event). Power spectral density (PSD) is a commonly used frequency-domain technique which shows the prominence of the different frequencies of EEG in a recording, useful for understanding underlying cognitive behaviours [4].

The different bands of EEG frequency are classified as delta (1-3 Hz), theta (4-7 Hz), alpha (8-13 Hz), beta (14-30 Hz) and gamma (>30 Hz), each associated with different types of cognitive function. Delta activity is associated with deep sleep and unconscious brain activity [5]. Theta activity is associated with drowsiness and light sleep but also with memory processing in cognitive tasks. Alpha activity is associated with relaxed wakefulness and has been associated with increased attention. Beta activity has been found to be associated with attention, active thinking, problem solving and excitation. Gamma activity is more complex and is associated with higher level cognitive activities such as perception and consciousness [5]. Quantitative EEG, the use of mathematical algorithms to interpret the data visually, facilitates the presentation of frequency-time-domain readings such as colour density spectral arrays (aka spectrograms), which visualise the variation of frequencies through time.

Although not widely explored, it has been shown that psychophysiological tools such as EEG allow a certain level of quantitative measurement of human reaction to design [6]. However, though there have been a number of EEG-based studies evaluating the design process, there is a lack of research investigating multi-level perspectives when dealing with procedural knowledge [6].

2.2. *Gaze tracking*

Gaze tracking allows measurement of movements such as fixations, saccades and smooth pursuit. Feature-based methods are the most predominant estimation methods, using extracted local features such as pupils, cornea reflections, glints etc. to map the location of the user's gaze against the recorded image, with calibration being essential to enabling the system to map input data to gaze direction accurately. Pupil size is generally considered to be an

accurate measure of arousal and attention, with many studies noting the pupils dilating and constricting in response to stimuli. It has been discovered that, with tactile stimulation, pupil size is more sensitive to the physical properties of the stimulus rather than the subjective emotional nature of the stimulus [7]. It has been found that in a lot of cases, negative stimuli cause harsher pupil size increases. When undertaking sorting and assembling tasks in the manufacturing domain, participants' attention varies between instructions and work area as tasks progress [8].

2.3. Challenges of using EEG and gaze tracking in procedure design

It has been proved that EEG and gaze tracking can be used together to find cues for emotion detection, and that participants' emotions could be effectively classified using differential entropy. It was noted that the pupil diameters were highly variable with the instantaneous luminance of the video clips being shown to the participants [9]. Part of gauging usability, especially of a procedure, is to analyze the neurological reaction of the users. In this context, although confusion is a more complex cognitive state, it is likely to be triggered by ambiguity in instructions. It has been observed that in confused states, alpha, beta, delta and theta brain activity increased, and gamma activity decreased, and it was suggested that collecting of eye-tracking data would allow analysis of visual attention processes and identification of confusion sources [10]. Confusion onset by missing information or hidden information in mathematical problems was found to be directly linked with decreased alpha activity, but different areas of the brain show different changes in activity based on whether the information in the problems was hidden or missing [11]. The differences in these findings highlight the challenges in interpreting EEG signals due to the complexity of the brain activity in different parts of the brain. Another gauge of usability is the level of frustration which the users experience. Very little research has been carried out to identify it in EEG signals as, like confusion, it is a complex cognitive state, which has varying levels and features and is difficult to elicit directly.

3. Hypothesis

Existing research indicates that psychophysiological analysis could contribute to improved efficiency in manufacturing. While there has been extensive research using EEG and gaze tracking in many different applications, this is the first study to use these technologies to analyse the usability of procedures.

The study hypothesises that EEG and gaze can be indicative of difficulties participants face during the execution of procedural tasks. By presenting participants with inadequate information at predetermined points in the instructions, it is predicted that they would experience periods of raised levels of frustration and/or confusion, which would be reflected in trends in the EEG and gaze data. Participants would be expected to spend longer looking at the instructions which they thought to be unclear, with the increased intensity of emotions leading to an overall increase in pupil size. It is hypothesised that the brain activity during these periods would change, and patterns would be noticed in these changes, providing clear markers indicating the problems in the instructions.

4. Methods

4.1. Origami tasks

Origami task instructions of differing difficulty levels were tested in this study. These tasks, taking between around two and seven minutes, each comprising of mostly very straightforward steps, but contained at least one instruction predicted to be a problem point, as shown in Figure 1. The problem points were all of different nature, to prevent participants learning between tasks. The task creating a tulip involved the creation of separate stem and flower, with the main problem point being the lack of physical connector for these two parts. The pelican task involved some intricate folding, requiring more spatial thinking to prevent asymmetry/distortion. The frog task has two instructions which did not clearly show which parts of the paper to fold, and the boat task involved turning the paper inside out, with a high risk of ripping the paper. The tasks were presented in this ascending difficulty order, so as to minimise any residual emotions carried through from problem points in previous tasks.

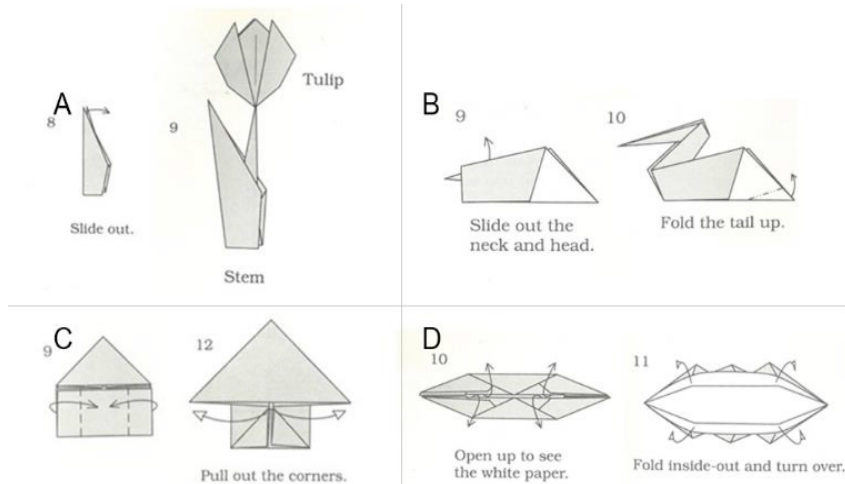


Figure 1 - Origami tasks (A-D) and some of their problem point steps.
A - Tulip, B - Pelican, C - Frog, D - Boat [12]

4.2. Equipment and data acquisition

The EEG equipment used in this study is the Brain Products actiCAP set up with LiveAmp wireless amplifier (Figure 2a), recorded with BrainVision Recorder (Brain Products GmbH, 2021). 32 channels are used for recording, positioned according to the 10/20 system, referenced at FCz and grounded at Fpz. The good impedance check threshold is 25 k Ω , and the recordings are at a frequency of 500 Hz.

Tobii Glasses 3 (Figure 2b) are used for the gaze tracking (Tobii Pro AB, 2021), recorded on its accompanying recording software then analysed using Tobii Pro Lab (Tobii Pro AB, 2024). The equipment is calibrated using the Glasses 3 one-point calibration with a calibration card positioned 0.5-1 metres from the participant. This option is chosen over the system's automatic calibration option as it is more accurate. The images are 1920 x 1080 resolution, and the gaze is sampled at 100 Hz.

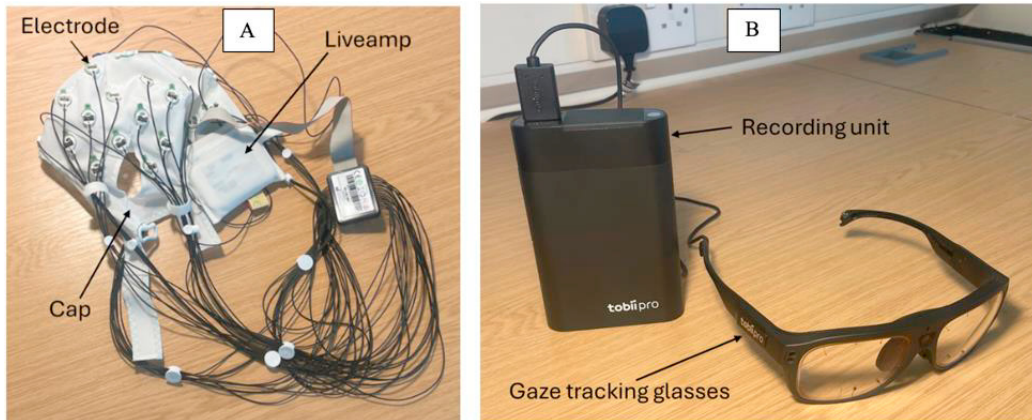


Figure 2 - Equipment used in the study:
A - Brain Products actiCAP with LiveAmp, B - Tobii Glasses 3

4.3. Participants

Prior to completion of the experiment, each participant completed a questionnaire to provide information about their background to inform the researcher in case of inconsistencies in the results. Fifteen participants took part, all students at Cardiff University. The participants were a mixed-gender group (11 male, 4 female) with a diverse range of academic backgrounds, most prominently engineering and psychology. The experience levels declared were mostly one or two out of ten, with the maximum experience level seven out of ten.

5. Experimentation

5.1. Experimental procedure and data collection

The participants were briefed before the experiment with what actions to avoid during the experiment which might disrupt the readings. These were mainly to combat any potential disruptions in the EEG readings, which would be sensitive to any kind of muscle movement, especially of the head where it may end up physically moving the electrodes, but to also minimise risk of misaligned calibration or other errors in the gaze tracking analysis. Participants were asked to avoid any kind of verbal activity.

The tasks were selected due to their perceived differences in difficulty and were presented to the participants in ascending order, to minimise the amount of residual frustration carried through the experiment. .

Whilst each task was being carried out, the researcher monitored the live EEG and the live footage from the gaze tracking. Timestamps were noted of any unusual activity in the EEG (e.g. disturbances) as well as timestamps of any moments where it seemed like the participants were encountering issues in their tasks. Participants were asked to feed back at the end of each task, noting any moments during which they felt as if they had experienced any particular moments of positive or negative emotion. This feedback was also recorded.

5.2. Data processing

The gaze recordings had to be processed so that there were areas of interest (AOIs) assigned to each step in the instruction, using Tobii Pro Lab [13]. All recordings which logged gaze samples of less than 75% were rejected from the statistical analysis, as there would be insufficient data to analyse them accurately [14]. Times of interest (TOIs) were determined based on the feedback and the observations recorded during each experiment. Base and stimulus TOIs were assigned to each recording, which were respectively used as the basis for comparison between each participant's regular state while performing straightforward steps and stimulated state while at a problem point.

Statistics were gathered for the number of fixations registered by each participant in each area of interest for each experiment. Seeing as each participant took a different amount of time for each task, the data was standardised by normalising it against time taken in the task (fixations per minute). These statistics were analysed using Microsoft Excel. Pupil diameter was also measured and analysed, with the average pupil size compared between base and stimulus TOIs.

The EEG data was processed using EEGLAB [15]. The first step was to use EEGLAB's 'Clean rawdata' function, removing bad channels and applying a 1 Hz high-pass filter. The data was then re-referenced against the average and a 40 Hz low-pass filter was applied. Independent component analysis (ICA) was used to remove components with >90% probability of being eye or muscle artifacts. Batch processing on all datasets was automated using MATLAB.

6. Results

6.1. Gaze analysis

6.1.1. Pupil size

Table 1 shows the percentage of recordings in which the average pupil diameter is larger in the stimulus than the base time of interest, and shows the average change in size.

Table 1 - Changes in pupil diameter from base to stimulus

Task	Proportion of recordings in which average pupil diameter increased (%)	Mean change in pupil diameter (%)
Tulip	83	+3.8
Pelican	71	+1.9
Frog	77	+3.5
Boat	71	+3.3
Overall	75	+3.1

The standard deviation in average change in pupil diameter is 5.3%, and outliers are considered to be values outside of two standard deviations either side of the mean. Three outliers have been removed (out of 56 data points). The data clearly shows a trend in which average pupil size was greater during the stimulus TOIs than the base TOIs.

6.1.2. Fixations: comparison with predictions and participant feedback

Table 2 shows how the predicted problem points in each experiment compare with: A- the step of each task on which the participants fixated the most and B- the responses of the collated feedback given by the participants after completion of each task, showing the steps most flagged by participants to be problem points. Seeing as the times it took each participant to complete each task are all unique, the fixations data is normalised to fixations per minute. Proportion of fixations (number of fixations on step / total number of fixations) was initially considered, but this was deemed to be misleading, as participants registered varying overall fixation rates and fixations on areas outside of the instructions were not considered.

Note: for the Tulip task, instructions are split into instructions for the stem (S1-S9) and for the flower (T1-T6).

Table 2 - Fixation results comparison with predictions and participant feedback

Task	Prediction	Most fixations per minute	Most flagged steps (number of flags)
Tulip	S8, Connection	T6, S8	Connection (10), S7 (6)
Pelican	9, 10	8	9 (11), 10 (7)
Frog	9, 12	8	9 (12), 12 (10)
Boat	9, 10	10	9 (11), 10 (10)

The steps with the most fixations per minute for the pelican and the frog are the steps that immediately precede the steps most flagged to have been problem points, the tulip's step S8 directly follows the second most flagged step. The other most fixated on step in the tulip task is the finished flower, step T6, with the other most flagged step being the connection. The most fixated on step for the boat task matches the second most flagged step.

6.2. EEG analysis

Spectrograms were generated for each recording, but it was not possible to discern any differences between any of the recordings. Finding a suitable and consistent range of powers for the colour axis proved unfeasible, as the ranges for each recording were different.

For each section of the brain, the percentage of recordings in which the average power spectral density (PSD) for each band increases from base to stimulus is shown in Table 3.

Table 3 - Proportions of PSD increases per frequency band

Area of Brain	Proportion of recordings in which PSD increased (%), (Average change in PSD)			
	Delta	Theta	Alpha	Beta
Whole brain	65 (+63%)	77 (+13%)	53 (-2%)	58 (-1%)
Left-brain	67 (+26%)	65 (+14%)	47 (-1%)	58 (+2%)
Right-brain	67 (+35%)	63 (+14%)	60 (+2%)	63 (+4%)
Frontal	65 (+24%)	63 (+15%)	53 (+2%)	67 (+5%)
Left frontal	60 (+31%)	65 (+18%)	53 (+7%)	74 (+8%)
Right frontal	67 (+36%)	65 (+18%)	56 (+2%)	63 (+6%)
Parietal	65 (+19%)	56 (+5%)	58 (+3%)	67 (+10%)
Left parietal	58 (+22%)	53 (+7%)	56 (+1%)	67 (+12%)
Right parietal	65 (+20%)	58 (+4%)	56 (+4%)	60 (+7%)
Left temporal	70 (+46%)	74 (+25%)	70 (+13%)	60 (+9%)
Right temporal	65 (+66%)	70 (+21%)	58 (+12%)	53 (+5%)
Occipital	65 (+83%)	56 (+13%)	35 (-2%)	44 (+1%)

Theta activity is seen to have the most convincing trend, with 77% of the recordings showing an increase in theta band PSD and increases in all parts of the brain. This is especially prominent in the temporal regions, where 74% of recordings increase at an average of +25% in the left, and 70% of recordings at an average of +20% in the right. Delta activity is also seen to increase in all areas of the brain, by a significantly larger average proportion. The average increase in the occipital region is 83%, with increases in 65% of the recordings. 70% of recordings increase in delta band PSD, averaging an increase of +46%. Alpha activity exhibits particularly noteworthy results. In the occipital regions it decreases in 65% of the recordings, but increases in 70% of the recordings in the left temporal area. This being said, the average change in the occipital region is only -2%, while the average change in the left temporal region is much more prominent at +13%. The frontal and parietal areas experience more increases in beta activity, while the occipital area experiences more decreases. The data shows that the changes in beta activity are especially prominent in the left frontal and parietal regions, with 74% of the left frontal and 67% of the left parietal datasets showing increases.

In summary, the increase from base to stimulus delta activity is greatest in the left temporal region. Theta activity increases, especially in the temporal regions, alpha activity decreases in the occipital region but more prominently increases in the left temporal region, and beta activity increases most in the left frontal and left parietal regions. The same methodology was used to obtain the differences between event related potentials (ERP) in base and stimulus times of interest. The results are presented in Table 4.

Table 4 - ERP comparisons

Area of Brain	Proportion of recordings in which mean ERP increased (%)	Average increase in ERP (%)
Whole brain	53	3
Left-brain	53	2
Right-brain	55	2
Frontal	45	2
Left frontal	47	3
Right frontal	53	5
Parietal	53	2
Left parietal	56	3
Right parietal	55	1
Left temporal	60	6
Right temporal	49	2
Occipital	53	2

The left temporal region has the most significant results considering the changes in ERP, where it increases from base to stimulus in 60% of the recordings with an average increase of 6%. In other areas, the data shows that there is minimal change in ERP between base and stimulus recordings.

6.3. Proposed framework for a procedure problem point identification model

The hypothesis states the expectation of patterns to be noticed in the differences between readings in base and stimulus times, and for the participants to spend longer looking at steps which are problem points. For the pelican, frog and boat tasks, the most fixated on steps were all either the most flagged step or the step directly before or after the most flagged step. This pattern indicates that this would be a good starting point to quickly identify approximate sections of the instructions to be further investigated.

By collecting the timestamps of all fixations on a section, they could be statistically manipulated to exclude accidental/erroneous fixations, narrowing down the time window during which the step was being undertaken. This time window could then be used to extract relevant sections of EEG, which could then be analysed to see if the changes in activity met certain threshold criteria for flagging to the procedure designers, who could then provide more detail

or redesign that part of the procedure. The findings from this study suggest that if the EEG showed temporal region increases in delta activity by 40%, theta activity by 20% and alpha activity by 10%, this would be a suitable set of criteria to identify the time analysed as a problem point. This model has been visualised in Figure 3. To define a more robust set of threshold criteria, further data collection and analysis would be required, as more data is needed to fully corroborate the criteria suggested in this study.

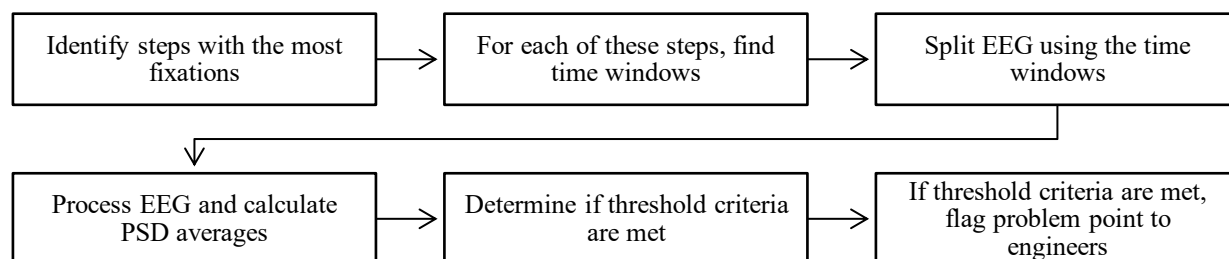


Figure 3 - Framework for problem point recognition system

7. Discussion

7.1. Comparison of hypothesis and results

The background research informed the hypothesis that the pupil size of participants would be higher on average during the stimulus times of interest due to the increase in emotion intensity. In the experiments, the participants' hands and the paper they were folding were positioned closer to their faces than the instructions. If a person meets a problem point in a procedure, they are likely to look more attentively at the instructions to see whether they have missed any key pieces of information. Pupil size is known to decrease when focusing on objects nearer to it and vice versa [16], so the strong trend in increasing pupil size could be due to the change in focus rather than the change in emotional state. This does not change the relevancy of the trend in this instance, however, but may become more difficult to interpret in conditions with less controlled light and focal distances.

The study hypothesises that there would be certain steps which would be more likely to challenge the participants for each of the tasks, due to the lack of clarity. For all but the tulip task, the most flagged steps in each task match perfectly with the predictions, proving the effectiveness of the instructions in triggering a response.

One of the most fixated on steps in the tulip task is the finished flower. The most flagged step is the connection between the flower and the stem, so this suggests that the participants finished the task, then looked back at that step to check if they had missed something. The boat task's results match the hypothesis that the most fixated on areas would be the problematic instructions themselves. The rest of the most fixated on steps are either directly preceding or following the most flagged steps. This suggests that the participants were more focused on either checking the steps before to establish whether they had executed that correctly, or the subsequent steps, to find out how the origami should look following the completion of the problematic step. While not completely matching with the hypothesis, the fixation results provide useful insight into the participants' reactions to the problem points.

The EEG findings validate the hypothesis of notable consistent changes in brain activity when encountering problem points, as found with the PSD. The average changes in delta activity are much higher than any of the other frequency bands, though its trend of whether it increased or decreased was less prominent. Base delta activity is far more varied between participants than any of the other frequency bands, with a coefficient of variation of 149%, compared to between 45% and 60% for theta, alpha and beta (considering all channels). This could be purely down to neurological differences between participants, or it could suggest that there were low frequency interferences which varied between days of recording.

7.2. Suitability and compatibility of EEG and gaze tracking in assessing procedure usability

The findings from this study suggest that EEG and gaze tracking are well suited for joint use, measuring the usability of a procedure by diagnosing its problem points. Using gaze tracking allows for the approximation of problem point locations using fixation analysis, while EEG can be used to validate and refine the diagnoses. In the model proposed in this study, the EEG analysis relies on the time indicators provided by the gaze recordings, and the unsuccessful attempt in this study to generate spectrograms means that no gaze tracking independent EEG technique

is presented. Further research would be required, with a larger pool of data, for problem points to be identified purely from patterns in brain activity. However, gaze tracking can still be used independently. The video feed makes it a lot easier to identify behavioural reactions, and the estimates from the fixation measurements would give a good rough indication of any problem points.

7.3. Limitations of the Study and Possible Sources of Error

Some inconsistencies in the gaze tracking data could be the result of errors in the gaze mapping onto the snapshots. This is because when analysing the gaze data, the areas of interest have been manually assigned to snapshots for each recording. This was necessary, as the instructions were not positioned consistently between participants, so the gaze data would not be mapped as accurately if all recordings for each task had been mapped from just one snapshot. This means that there are likely to be inconsistencies in the sizing and positioning of the AOIs for each recording. Additionally, parallax leads to the instructions at the bottom of the page having larger projected areas than those on the top of the page. It is likely that the fixation results would be more prominent if the instructions are presented to the participants in fixed positions on the wall in front of them, as the parallax effect would be diminished and the AOIs would be uniform.

The study also relies on the accuracy of Tobii Pro Lab's assisted mapping system, which is dependent on the stability of the video recording and would inevitably have mis-mapped some gaze points. It is assumed that the glasses were calibrated correctly, but any badly calibrated recordings would lead to inaccurate gaze location statistics (though if the calibration was only slightly off, the gaze location registered would likely land within correct areas of interest for most fixations). Four recordings were omitted from the gaze tracking analysis because of having an insufficient number of gaze samples (<75%), but most of the recordings registered over 85% sample collection. While patterns were observed in this study, a sample size of greater than 15 would be required to fully validate the findings.

EEG is a very sensitive tool, with a distinct lack of guidelines on exactly how to treat the signals measured. As such there are several potential sources of error which may affect the accuracy of the results obtained in this study. One flaw with EEG is that it can only measure the activity from the surface of the head, so cannot possibly measure the activity from all parts of the brain, as the cortex is folded and some areas lie deep within the skull. This means that any findings can only point to rough areas of the brain without being able to pinpoint exactly the origin of the signals received.

All of the EEG data has been pre-processed in the exact same way to ensure consistency between recordings. This being said, it is difficult to know what improvements could be made to the way that the data is processed, as there are so many different combinations of processes to apply. It must be noted that some of the pre-processing removed entire sections of data or removed channels which were deemed to have recorded bad readings, so there may therefore be some inconsistencies between datasets. The lack of the FC5 electrode may also contribute to any asymmetries between left-brain and right-brain data, though brain lateralisation is likely to play a large part [17]. When considering the occipital, left temporal and right temporal regions only three electrodes were used. Any of these datasets in which two out of the three channels were rejected in pre-processing, leaving only one channel, are not considered in the results analysis. This accounts for 15% of the occipital datasets.

The trends noticed in the ERP data are not very strong, so the reliability of these findings is uncertain. ERP data is usually collected using short epochs, predefined as a set amount of time after each stimulus onset. In this study, the lengths of time considered for the ERPs are different for each recording, depending on how long each participant spent on the respective base and problem steps highlighted. This may be one of the reasons why there is not a strong trend in ERP change, as the brain activity over that longer period is possibly too varied to provide a feature which could be easily identifiable when searching for problem points.

8. Conclusion

It was found that the average pupil diameter was larger while encountering a problem point. The instruction steps on which the participants fixated the most were generally identified to be within one step either before or after the step which was predicted to be the problem point in each task, suggesting that fixation rate is a good basis to identify problematic steps. It was found, through PSD analysis, that general brain activity was elevated while encountering a problem point, most prominently theta activity in the left and right temporal regions of the brain. Increases were also noticed in delta activity across all areas of the brain. Alpha activity remained largely unaffected, with only the temporal

regions experiencing any significant increase. Beta activity was found to have increased in the frontal and left parietal areas. There was insufficient evidence of any major changes in ERP.

This study provides a novel insight into how EEG and gaze tracking technology could be used in conjunction to identify problem points in procedure designs, and a framework for a procedure problem point identification model is proposed. To further develop this model, a clear set of threshold criteria should be defined from a much wider pool of data. With a large enough pool of data, more explicit emotion detection criteria could be defined. This would allow for more nuanced analysis of procedure usability and thus further optimisation of these procedures. Further controlling light in the environment and/or using a light reflex model to account for pupil size variation with luminance could facilitate the introduction of pupil size into the criteria. This study used origami because of its simplicity and accessibility within the scope of this project. Testing the technology in a more applied context such as a factory or workshop, with more participants, would help further determine its suitability as a tool in practice. Introducing facial recognition alongside EEG and gaze tracking could also provide interesting insight, though further research would need to be carried out to determine its practicality.

This study demonstrates the potential of EEG and gaze tracking as tools for assessing procedure design usability, by identifying problem points within procedural tasks. The findings highlight consistencies in psychophysiological responses to these problem points, thus informing a framework for a procedure usability analysis model. The study introduces an exciting new approach to enhancing manufacturing efficiency, though extensive further research is required to verify the findings and provide further information for precise problem point diagnosis.

References

- [1] Jamal, S., Cruz, M., Chakravarthy, M., Wahl, C., & Wimmer, H. (2023). Integration of EEG and Eye Tracking Technology: A Systematic Review. *Proceedings of IEEE Southeastcon*, 2023, 209-216. doi:10.1109/SoutheastCon51012.2023.10115167
- [2] Shashi Kumar, G., Sampathila, N., & Martis, R. (2023). Classification of Human Emotional States Based on Valence-Arousal Scale Using Electroencephalogram. *Journal of Medical Signals and Sensors*, 13(2), 173-182. doi:10.4103/jmss.jmss_169_21
- [3] Jacobsen, N. S., Kristanto, D., Welp, S., Inceler, Y. C., & Debener, S. (2025). Preprocessing Choices for P3 Analyses with Mobile EEG: A Systematic Literature Review and Interactive Exploration. *Psychophysiology*, 62(1), 14743. doi:10.1111/psyp.14743
- [4] Bellos, C., K., S., Tzallas, A., Stergios, G., & Tsipouras, M. (2025). Methods and Approaches for User Engagement and User Experience Analysis Based on Electroencephalography Recordings: A Systematic Review. *Electronics*, 14(2), 251. doi:10.3390/electronics14020251
- [5] Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2007). *Handbook of Psychophysiology*. New York: Cambridge University Press.
- [6] Kim, N., Chung, S., & Kim, D. I. (2022). Exploring EEG-based Design Studies: A Systematic Review. *Archives of Design Research*, 35(4), 91-113. doi:10.15187/adr.2022.11.35.4.91
- [7] Van Hooijdonk, R., Mathot, S., Schat, E., Spencer, H., & Van der Stigchel, S. (2019). Touch-Induced Pupil Size Reflects Stimulus Intensity, Not Subjective. *Experimental Brain Research*, 237(1), 201-210. doi:10.1007/s00221-018-5404-2
- [8] Pathmanathan, N., Rau, T., Yang, X., Calepso, A. S., Amtsberg, F., Menges, A., Sedlmair, M., Kurzhals, K. (2024). Eyes on the Task: Gaze Analysis of Situated Visualization for Collaborative Tasks. 2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR) (pp. 785-795). Orlando, USA: IEEE. doi:10.1109/VR58804.2024.00098
- [9] Zheng, W. L., Dong, B. N., & Lu, B. L. (2014). Multimodal Emotion Recognition Using EEG and Eye Tracking Data. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 5040-5043). Chicago, USA: IEEE. doi:10.1109/EMBC.2014.6944757
- [10] Xu, T., Wang, J., Zhang, G., Zhang, L., & Zhou, Y. (2023). Confused or Not: Decoding Brain Activity and Recognizing Confusion in Reasoning Learning Using EEG. *Journal of Neural Engineering*, 20(2), 26018. doi:10.1088/1741-2552/acbf0
- [11] Liang, Y., Liu, X., Qiu, L., & Zhang, S. (2017). An EEG Study of a Confusing State Induced by Information Insufficiency during Mathematical Problem-Solving and Reasoning. *Computational Intelligence and Neuroscience*, 2018, 1-13. doi:10.1155/2018/1943565
- [12] Montroll, J. (1992). *Easy Origami*. Mineola: Dover Publications, Inc.
- [13] Tobii Pro AB. (2024). Lab(24.21.435 (x64))
- [14] Romano Bergstrom, J. C., Olmsted-Hawala, E. L., & Bergstrom, H. C. (2016). Older Adults Fail to See the Periphery in a Website Task. *Universal Access in the Information Society*, 15(2), 261-270. doi:10.1007/s10209-014-0382-z
- [15] EEGLAB. (2025) Retrieved from EEGLAB: <https://eeglab.org/>
- [16] Mathur, A., Gehrmann, J., & Atchison, D. A. (2014). Influences of Luminance and Accommodation Stimuli on Pupil Size and Pupil Center Location. *Visual Psychophysics and Physiological Optics*, 55(4), 2166-2172. doi:10.1167/iops.13-13492
- [17] Rogers, L. (2024). Lateralization of Brain Function. *Oxford Research Encyclopaedia of Psychology*. doi:10.1093/acrefore/9780190236557.013.728