# Qualitative exploration of physiological mannerisms of acrophobia during virtual reality heights exposure

1st Saskia Davies
*Computer Science*
*Swansea University*
Swansea, UK
saskia.davies@swansea.ac.uk

2nd Tom Owen
*Computer Science*
*Swansea University*
Swansea, UK
t.owen@swansea.ac.uk

3rd Sean P. Walton
*Computer Science*
*Swansea University*
Swansea, UK
s.p.walton@swansea.ac.uk

***Abstract*—Virtual reality (VR) exposure therapy (ET) has demonstrated invaluable therapeutic outcomes for phobic patients, particularly when supplemented with physiological sensors for internal user insights. Although much of the existing research in VR ET adopts machine learning predictions for incidental categories of affect, we believe that a more intricate and qualitative understanding of direct phobic behavioural patterns has the potential to augment these approaches and drive more informed VR ET development.**

**By gathering a dataset through user studies, we qualitatively explore the fine-grained physiological differences between self-reported acrophobic and non-acrophobic individuals during VR exposure to heights. Through a rigorous, complementary combination of statistical analysis and graphical visualisation, we identify several key distinctions in structural and chronological characteristics between groups, particularly in autonomic responses and facial activations. Notably, mouth movements such as lip funnelling and jaw thrusting emerge as the most practically significant indicators of acrophobic tendencies, with increasing trends and fluctuation patterns over time.**

**Our analysis deeply and intricately explores these key behaviours as potential biomarkers of acrophobia, contributing to a more developed understanding of acrophobia-specific responses by offering insights that extend beyond quantitative measures of affect categorisation. Moving forward, progressive approaches can utilise a more direct and thorough understanding of acrophobia during treatment, which can augment the design and development of existing VR ET methods.**

***Index Terms*—virtual reality, human activity recognition, statistical analysis, data visualisation**

## I. Introduction

Virtual reality (VR) exposure therapy (ET) is an emerging psychological treatment approach which simulates real-life scenarios in a controlled and repeatable manner [1]–[3]. These techniques are employed to safely expose individuals to feared stimuli through gradual and repeated desensitisation to phobic triggers [1], [4], such as heights for acrophobia. By integrating multimodal sensors, affective user responses may be monitored in real time, not only to provide clinicians with a more in-depth understanding of users, but also to dynamically adapt the virtual environment and tailor therapeutic exposure.

Due to the heavily machine learning oriented nature of the affective computing field, existing research often focuses on predictive accuracy for affective states [5], such that qualitative behavioural intricacies – particularly for individual phobias – appear to be lesser known. Tactics of explainability such as SHapley Additive exPlanations (SHAP) [6] are beneficial in highlighting meaningful features throughout model predictions, however the underlying data patterns are represented in limited forms, often denoting thresholds under which feature values correspond to a particular predictive outcome. Important behavioural information may be neglected as a result, such as unique structural characteristics or chronological trends.

Pinpointing acrophobia, we work towards mapping the subtleties in physiological differences between self-reported acrophobic and non-acrophobic individuals when faced with heights in VR. Utilising a combination of statistical tests and graph visualisations, we present a descriptive analysis of behavioural differences between groups. Our findings show that while autonomic and eye-based features demonstrate statistical, distributional and chronological differences, mouth movements – such as increasing and fluctuating lip funnelling or jaw thrusting – reveal the highest practical significance on average. By thoroughly understanding these intrinsic acrophobic signatures, progressive approaches can utilise a more direct and intricate understanding of acrophobia during treatment, which can be pursued alongside incidental categories of affect.

## II. Related Work

Due to the immersive, engaging and adaptable qualities of VR, virtual therapy has demonstrated enhanced therapeutic outcomes [1], [7], and generates more natural user responses than traditional laboratory environments [8]. These advantages are applied to a wide variety of phobic scenarios such as heights [9], confined spaces [10] and public speaking [11], and while these systems may support interactivity, positive results can also be achieved through passive 360° video experiences [12].

VR ET is often supplemented with physiological sensors to provide an otherwise unseen insight into users' internal affective responses. Given the therapeutic importance of psychological fear activation [1], [13], anxiety, stress and fear are commonly targeted detection states. Overlapping modalities primarily include a combination of heart rate, heart rate variability (HRV), skin conductivity, skin temperature, respiration

This work was supported by EPSRC grant number EP/S021892/1.

and motor movement behaviours [14]–[21]. Sophisticated setups alternatively utilise electroencephalogram (EEG) devices, which can achieve accurate results in isolation [22]–[24].

This widespread emphasis on machine learning and predictive results hones its focus on classifying categories of affect, typically as a method of automated exposure adaptation [5]. While explainability techniques such as SHAP feature importance have been used to identify (1) salient modalities and (2) margins in which datapoints contribute towards a certain prediction [6], there is limited qualitative insight on the tangible behavioural intricacies that pertain to individual phobias, particularly over structural patterns and chronological trends. We argue that such insights are additionally valuable in ET research, supporting an enriched human understanding of phobic responses that complements existing approaches to VR therapy.

Leaning away from machine learning analysis and categorical emotion recognition, we examine the physiological mannerisms pertaining specifically to acrophobic individuals when faced with heights in VR. Our statistical-driven approach drives an exploration into patterns and tendencies that are disparate from non-acrophobic users.

## III. Methodology

To explore VR behaviours pertaining to acrophobia, we (1) create a 'VR experiences' prototype that incorporates physiological sensors for data collection, (2) run user studies with this prototype to form a comprehensive dataset, and (3) analyse the physiological variations between users who self-report a high and low fear of heights. A combination of statistical tests and graph visualisations support our qualitative analysis of distinct acrophobic tendencies.

### A. Virtual reality prototype

Our 'VR experiences' prototype was developed for the Meta Quest Pro[1] which, alongside the typical VR capabilities of head and handheld controller tracking, natively supports face and eye tracking. These additional sensors are built into the headset, such that no additional devices are required. We supplemented these modalities with a wireless medical grade finger-clip sensor (eVu-TPS[2]), with in-built tracking capabilities for heart rate, HRV, blood volume, skin conductivity and skin temperature, given their importance across affective computing literature.

Although this work addresses acrophobia, our full dataset comprises physiological data for a wide range of VR experiences. To invoke a variety of responses, we integrated 10 publicly available 360° videos that are each designed for a particular affective experience. User comfort and attention are managed by clipping all videos to a standardised length of 1:20, and presenting a random subset of just six videos during one interactive session. Two of the total 10 videos are designed around acrophobia, and so our analysis focuses on all instances in which participants encounter these videos.

[1]https://www.meta.com/gb/quest/quest-pro/

[2]https://thoughttechnology.com/evu-tps-package-t4500/

An instructional splash screen prefaces the application during sensor calibration, additionally demonstrating a skip function by which participants can bypass any video that causes discomfort. Sensor data is then captured passively at a fixed rate of 50Hz throughout video playback, and upon completion, instructions to remove the headset and sensor are displayed.

### B. User study

A total of 41 participants took part in the VR user study, primarily recruited by university advertisement. The population comprised various ages (13 aged 18–24; 22 aged 25–34; 6 aged 35+) and genders (28 male; 12 female; 1 non-binary), as well as a diverse range of VR experience and self-reported acrophobia levels. Acrophobia was reported using a five-point Likert scale of heights-related disturbance level (1 = not at all; 5 = very much), based on the Fear Survey Schedule (FSS-III) format [25]. Eighteen participants self-reported a significant fear of heights, scoring either four or five on the Likert scale, while the remaining 23 participants rated their fear as three or below. Of the 18 participants with high fear, 15 experienced at least one acrophobia video, and four experienced both videos. Similarly, 18 participants with low fear experienced at least one video, while eight encountered both.

Studies were held as one-to-one sessions in a quiet, isolated and spacious room. Each session lasted approximately 20 minutes in total, incorporating (1) a demographic questionnaire and study briefing, (2) headset and sensor setup, followed by engagement with the prototype while sensors capture physiological data, and (3) a post-study interview regarding overall experience and emotional responses. This study and its resources were approved by the Research and Ethics Committee at Swansea University.

### C. Analysis

Encompassing all physiological sensors, our dataset contains 123 features for 45 video instances, each containing 4025 sequential datapoints over time. We first reduced the dimensionality of this data, initially by removing all unchanging and non-comportmental variables, and secondly by combining composite features through feature engineering. Specifically, bi-lateral facial features have been averaged into a single value, while 3D movement and rotation are combined into univariate velocity. These approaches condensed the original 123 features into a more manageable and coherent 50, outlined in Table I.

Splitting the dataset into two groups of high (FSS$\geq$4) and low (FSS$<$4) self-reported acrophobia, we perform statistical tests to compare potential variations for every feature. For the features with highest statistical differences and effect sizes, we explore fine-grained behaviours in more detail using box plots and time series graph visualisations. As such, we provide a qualitative exploration of specific mannerisms pertaining to acrophobia, which differ to the non-acrophobic group.

## IV. Results

As many statistical tests assume a normal distribution of the data, we first run a Shapiro-Wilk test [26] on all features in

TABLE I

OVERVIEW OF PHYSIOLOGICAL FEATURES IN THE CONDENSED DATASET, CAPTURED FROM SENSORS IN THE META QUEST PRO AND EVU-TPS.

| Meta Quest Pro | eVu-TPS |
|---|---|
| Headset tracking (position and rotation velocity) | Heart rate (bpm) Heart rate variability (ms) |
| Controller tracking (position and rotation velocity: left and right) | Skin conductance (µS) |
| Facial feature tracking (33 facial activation units) | Skin temperature (°C) |
| Eye tracking (position and rotation velocity: left and right) | Blood volume (SpO2) |

each group to determine which statistical tests to use. Given the naturally skewed pattern of most features – particularly facial action units that rarely deviate from neutral expressions – the test confirms that no features are sufficiently normally distributed, as expected. As such, this data requires a non-parametric test that does not rely on assumptions about the underlying data distribution.

We therefore choose a Kolmogorov-Smirnov (KS) test [27] to compare distributional characteristics between the two groups, such as their shapes and spreads, as well as a Mann-Whitney U (MWU) test [28] to compare central tendencies surrounding the median. Using both techniques provides complementary insights into the behavioural differences between acrophobic and non-acrophobic participants.

## A. Kolmogorov-Smirnov test

The KS test reveals significant differences between the two groups across all physiological features with highly significant p-values ($p<0.05$ [29]), establishing that each feature differs between groups in terms of overall distribution. However, approximated Cohen's d effect sizes indicate that these differences are consistently small, with values ranging from 0.000028 to 0.00066. As a result, the practical significance of these differences is limited due to small effect sizes ($d<0.2$ [30]), despite statistical significance. Notable features containing the largest test statistics and effect sizes include skin temperature, lip and cheek features, jaw thrust and heart rate, summarised in Table II.

TABLE II

TOP 10 KOLMOGOROV-SMIRNOV TEST RESULTS ORDERED BY ESTIMATED EFFECT SIZE

| Feature | Test Statistic | P-Value | Effect Size |
|---|---|---|---|
| temperature | 0.29298 | <0.001 | 0.000660 |
| lipPressor | 0.28791 | <0.001 | 0.000648 |
| lipTightener | 0.25651 | <0.001 | 0.000578 |
| lipFunnelerB | 0.23314 | <0.001 | 0.000525 |
| cheekPuff | 0.21836 | <0.001 | 0.000492 |
| lipFunnelerT | 0.20787 | <0.001 | 0.000468 |
| cheekSuck | 0.19079 | <0.001 | 0.000430 |
| jawThrust | 0.17589 | <0.001 | 0.000396 |
| lowerLipDepressor | 0.17364 | <0.001 | 0.000391 |
| heart rate | 0.16879 | <0.001 | 0.000380 |

TABLE III

TOP 10 MANN-WHITNEY U TEST RESULTS ORDERED BY EFFECT SIZE

| Feature | Test Statistic | P-Value | Effect Size |
|---|---|---|---|
| lipFunnelerB | $3.0894 \times 10^9$ | <0.001 | -0.27924 |
| lipTightener | $3.2116 \times 10^9$ | <0.001 | -0.25691 |
| lipFunnelerT | $3.2858 \times 10^9$ | <0.001 | -0.24334 |
| cheekPuff | $3.4385 \times 10^9$ | <0.001 | -0.21543 |
| jawThrust | $3.5191 \times 10^9$ | <0.001 | -0.20071 |
| lowerLipDepressor | $3.5379 \times 10^9$ | <0.001 | -0.19726 |
| lipPressor | $3.5983 \times 10^9$ | <0.001 | -0.18623 |
| heart rate | $5.5898 \times 10^9$ | <0.001 | 0.17778 |
| lipsToward | $5.5349 \times 10^9$ | <0.001 | 0.16775 |
| upperLidRaiser | $3.9487 \times 10^9$ | <0.001 | -0.12218 |

TABLE IV

TOP 10 MANN-WHITNEY U TEST RESULTS ORDERED BY TEST STATISTIC

| Feature | Test Statistic | P-Value | Effect Size |
|---|---|---|---|
| heart rate | $5.5898 \times 10^9$ | <0.001 | 0.17778 |
| lipsToward | $5.5349 \times 10^9$ | <0.001 | 0.16775 |
| innerBrowRaiser | $5.2571 \times 10^9$ | <0.001 | 0.11697 |
| eyesLookDown | $5.2002 \times 10^9$ | <0.001 | 0.10657 |
| lipSuckB | $5.1836 \times 10^9$ | <0.001 | 0.10353 |
| temperature | $5.1439 \times 10^9$ | <0.001 | 0.09628 |
| eyesClosed | $4.9983 \times 10^9$ | <0.001 | 0.06966 |
| noseWrinkler | $4.8751 \times 10^9$ | <0.001 | 0.04715 |
| upperLipRaiser | $4.8697 \times 10^9$ | <0.001 | 0.04616 |
| high freq HRV | $4.8548 \times 10^9$ | <0.001 | 0.04344 |

## B. Mann-Whitney U test

The MWU test further confirmed significant differences across most features (all except eye rotation velocities and eyebrow lowering) such that the typical, central values in each group are distinct, with many features reaching highly significant p-values ($p<0.05$ [29]) and large test statistics. While rank-biseral correlation effect sizes generally remained small, 16 demonstrated moderate practical significance ($r \geq 0.1$ [30]). Similarly to the KS test results, lip features appear among the highest effect sizes, along with jaw thrust and cheek puffing. When comparing the highest test statistics that signify the largest group differences, autonomic data, including heart rate, temperature and high-frequency HRV (HF-HRV), emerges more commonly alongside additional lip features and eye behaviours, such as looking down, eyes closing and inner brow raising. Of these features, heart rate and 'lips toward' – upper and lower lip contact – have the highest combination of test statistic and effect size, appearing in both Table III and Table IV.

## C. Graphical analysis

To further investigate the trends between contrasting behaviours, we take the previous statistical results to group a selection of the highest scoring features by autonomic (Figures 1 and 2), mouth (Figures 4 and 5) and eye (Figures 7 and 8) categories, visualising the differences between groups through box plots and time series graphs.

*1) Autonomic features:* Plots comparing temperature indicate a notable difference in central tendencies between the two groups. The box plot shows that acrophobic participants have an exceedingly lower median temperature with a much wider

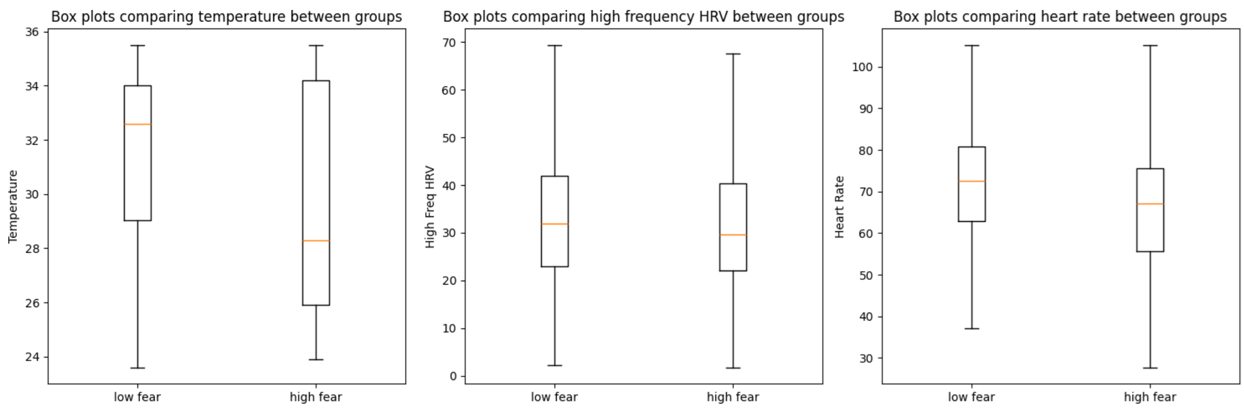


Fig. 1. Box plots comparing the distribution of temperature, high frequency heart rate variability and heart rate between acrophobic and non-acrophobic groups.

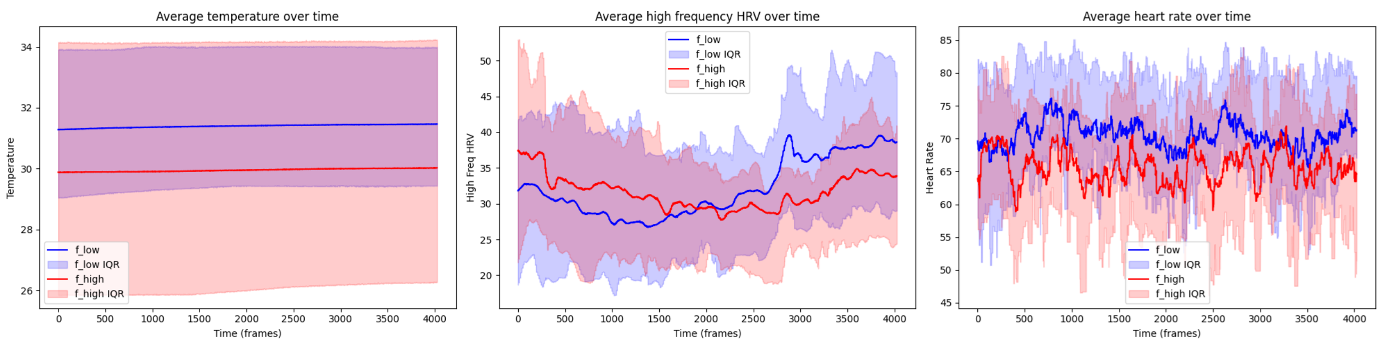


Fig. 2. Time series graphs comparing the average temperature, high frequency heart rate variability and heart rate between acrophobic and non-acrophobic groups, where the shaded areas denote interquartile range.

interquartile range (IQR) and therefore greater variability, leaning towards lower temperatures. Similarly, the average temperature remains consistently higher for non-acrophobic participants in the average time series graph, with a smaller IQR that indicates more stable temperatures compared to the large variation of the acrophobic group. These graphs correlate with statistical test results, where significant differences were found for both distribution and centric values.

The box plot comparing HF-HRV shows minor differences in the median and quartiles, where the acrophobic group has lower HF-HRV overall. The time series graph shows more distinctive results, where acrophobic participants' HF-HRV tends to lower over time with more acrophobia exposure. This line is surpassed by the non-acrophobic group around halfway through, who start with lower HF-HRV on average and gradually increase. As the MWU test produced more significant results than the KS test, these graphs highlight that there are more noticeable differences in central tendencies compared to overall distributions.

In contrast, the respective box plot and time series graphs denote a higher overall heart rate in non-acrophobic participants, regarding both central tendencies and group distribution as seen in the statistical test results. However, corresponding to the time series graph of HF-HRV results, the acrophobic group has much more fluctuation in their average heart rate over time compared to non-acrophobic participants. Of all autonomic features, heart rate had the highest effect sizes, implying these contrasts are realistic in practice.

*2) Mouth features:* Various mouth-related features were common across the top significant statistical results, with top and bottom lip funnelling, as well as lip tightening, demonstrating the highest MWU practical significance. Minimal activation of these features is shown through box plots, where the median of both groups are near zero for each feature. However, while the non-acrophobic group has no IQR, data from the acrophobic group contains many more instances of lip movement. This difference is clear when graphing average values over time, particularly for upper lip movement and lip tightening, which are consistently higher in the acrophobic group with more frequent and extreme fluctuations. The acrophobic group additionally trends upwards throughout the latter half of exposure, whereas the non-acrophobic group has little rate of change outside of fluctuations. Lower lip funnelling has similar tendencies, however they are on a much smaller scale.

Jaw thrust activity visualisations also demonstrate results similar to the statistical tests, with clear differences in both distribution and central tendencies between groups. As with the previous mouth features, acrophobic participants have a larger IQR, showing a greater tendency for jaw movements despite

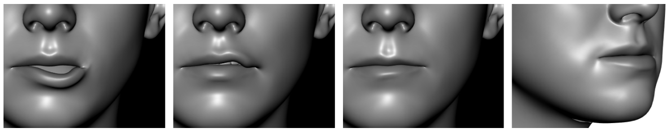

Fig. 3. Illustrations showing left-hand behaviours of lipFunnelerB, lipFunnelerT, lipTightener and jawThrust respectively.

the median of both groups surrounding zero. Additionally, this group has a consistently higher average across the time series graph, with a larger range of fluctuations compared to the non-acrophobic group. As the IQR surrounds this average, higher values are centrally more common in acrophobic participants.

*3) Eye features:* Plots for inner brow raising show a slightly higher central tendency in non-acrophobic participants. The box plot denotes a marginally higher median with a substantial IQR and maximum, while the time series graph indicates a consistently higher average over the duration of exposure. In contrast, the acrophobic group typically exhibits less expressive inner brow raising on average despite some repeated areas of accelerated activation in the second half, with a smaller IQR that rarely surpasses the mean and therefore suggests more stationary expressions in comparison.

While there is little difference in the distribution and patterns of eye closure between each group, non-acrophobic participants appear to have slightly higher values overall, with a higher median, IQR and time series trend. However, non-acrophobic participants exhibit a marginally wider, positive range of downward gazes, with a slightly higher median compared to the near-zero median of the acrophobic group. This difference in central tendency is further supported by the time series graph, in which non-acrophobic averages are slightly higher over time with periods of increased lower quartiles.

While the box plot for upper eyelid raising shows a small median value for both groups, the acrophobic group has a distinctly higher IQR, therefore suggesting more frequent eye widening behaviours. This tendency is also clearly demonstrated when visualising averages over time, with consistently larger values and ranges, as well as stronger, distinctive peaks compared to the more stable trends found in the non-acrophobic group. While not appearing in Table III's top 10 effect sizes, each of these eye-related features additionally attained MWU practical significance, which suggests the identified differences are applicable in practice.

## V. Discussion

Overall, we examined a number of physiological behaviour responses to VR height exposure, comparing self-reported acrophobic individuals to those without a distinct fear of heights. Results of the KS test indicate that, for every feature, the two groups differ significantly in terms of their overall distributional characteristics, while complementary results from the MWU test further support a statistically significant difference in central tendencies for almost every feature. Although effect sizes and therefore the magnitude of these differences are typically small, 16 features produced substantial practical MWU significance between groups, highlighting potential signatures associated with acrophobia. We explored categorisable subsets of these most significant features, within groups representing autonomic, mouth and eye activations.

Regarding autonomic features, the acrophobic group exhibited decreasing HF-HRV over time, alongside lower, more variable skin temperatures compared to the non-acrophobic group. As HF-HRV is consistent with heightened parasympathetic activity and a reduced sympathetic response [31]–[33], lower ratings are typically associated with stress and difficulty managing the body's 'fight-or-flight' response. Conversely, higher ratings reflect better emotion regulation and decision-making capabilities [32], [33]. The acrophobic group's HF-HRV decrease over time may therefore suggest a worsening struggle to cope with exposure, as opposed to the successful regulatory responses of non-acrophobic participants.

Reduced skin temperatures may be similarly representative of such reactions. Peripheral vasoconstriction commonly occurs during fear, redirecting blood flow to increase core temperature and consequently decrease external temperature [34]. These reactions manifest as sensations of 'chills' and 'cold sweat', and are often indicative of psychological fear responses [35].

While the heart rate of acrophobic participants appeared lower on average, visualisations show less stability and much more fluctuation over time. Erratic heart rate can be a causal factor of decreasing HF-HRV, further suggesting a dysregulation of the 'fight-or-flight' autonomic balance caused during exposure [32].

Facial activation units were also highlighted to portray significant and practical differences between groups, especially in mouth-related features. Acrophobic participants demonstrated more dynamic and irregular lip and jaw movements, with noticeable intermittent peaks and increased activity during the latter stages of exposure. Upper lip funnelling, lip tightening and jaw thrusting were particularly salient over time, in comparison to the more static behaviours of the non-acrophobic group.

Lip tightening and jaw thrusting are commonly affiliated with anger [36], [37], aggression and concentration [38] – such that a more strained expression indicates physical struggle, or efforts of emotional regulation – while upper lip funnelling has been associated with disgust and aversive responses [39], [40], often reflecting a snarl or grimace. As with HF-HRV, the latter upward trend may represent an increasing struggle with psychological and biological 'fight-or-flight' symptoms, with twitching, fluctuating mouth movements presenting as an indication of frustration or difficulty.

Of the eye-related behaviours that produced interesting distinctions between groups, non-acrophobic individuals exhibited more expressive inner brow movements, typically associated with sympathetic social engagement such as surprise and curiosity [36], whereas acrophobic participants are less likely to gaze downwards, and more commonly express widened eyes. Alongside wide eyes being a key characteristic

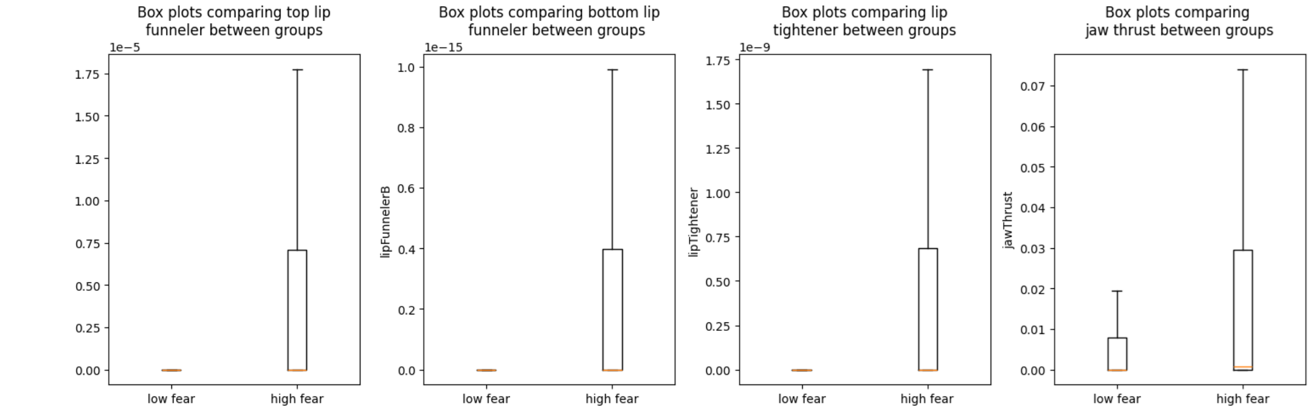


Fig. 4. Box plots comparing the distribution of lipFunnellerT, lipFunnellerB, lipTightener and jawThrust between acrophobic and non-acrophobic groups.

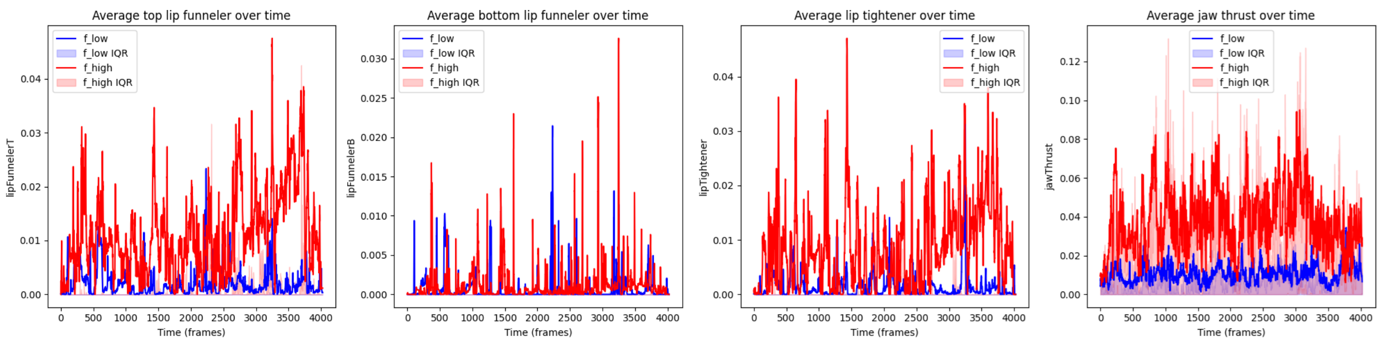


Fig. 5. Time series graphs comparing the average lipFunnellerT, lipFunnellerB, lipTightener and jawThrust between acrophobic and non-acrophobic groups, where the shaded areas denote interquartile range.

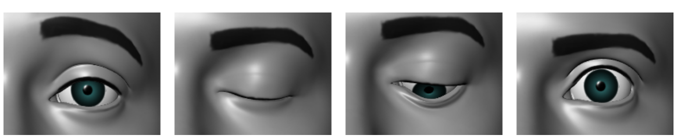

Fig. 6. Illustrations showing the behaviours of innerBrowRaiser, eyesClosed, eyesLookDown and upperLidRaiser respectively.

of heightened vigilance and fear [36], [38], gaze results support psychological research that demonstrates restricted eye movements and reductions in visually explorative behaviours in acrophobic individuals faced with heights [41], [42]. As such, these findings suggest that acrophobic individuals engage in avoidance behaviours, thereby mitigating their exposure to height-related stimuli.

Among all observed features, subtle mouth-related behaviours exhibited the highest levels of practical significance according to MWU effect sizes, and warrant further investigation into how their central tendencies may serve as potential biomarkers for acrophobic responses. Our qualitative exploration of behavioural intricacies details these structural and chronological characteristics of acrophobia, supporting an enriched human understanding of physiological phobic responses that complements existing approaches to VR therapy. Alongside prevalent works, which utilise affect classification as a means to adapt exposure, our findings augment a direct and intricate understanding of acrophobia responses during treatment. In practice, developments may be designed with these telltale responses in mind, for instance, reducing exposure intensity when fluctuating lip movements start to increase over time, as this appears to represent unsuccessful management of the body's 'fight-or-flight' response.

We only presented this analysis over a small number of features in our dataset, based on a subset of the most significant differences. However, other features – including those with small practical significance – produced interesting characteristics and distinctions between users. As such, there are further opportunities to investigate alternative physiological tendencies of acrophobia, such as in head and hand motions.

Additional research must conclusively discern whether such behaviours are driven by phobic implications, or reflect alternative psychological mechanisms. While our sample size and range of acrophobic VR scenarios limit the certainty of generalisability in our findings, future work should seek to recruit larger cohorts that include clinically diagnosed acrophobic individuals, implementing a broader array of heights-related VR environments that accurately represent therapeutic contexts. There is also scope to thoroughly investigate individual users through a wider range of visualisation and analytic tools, as we have amalgamated participant data into two groups for introductory purposes. The distinct contribution of our findings instead work towards a more tangible and practical knowledge

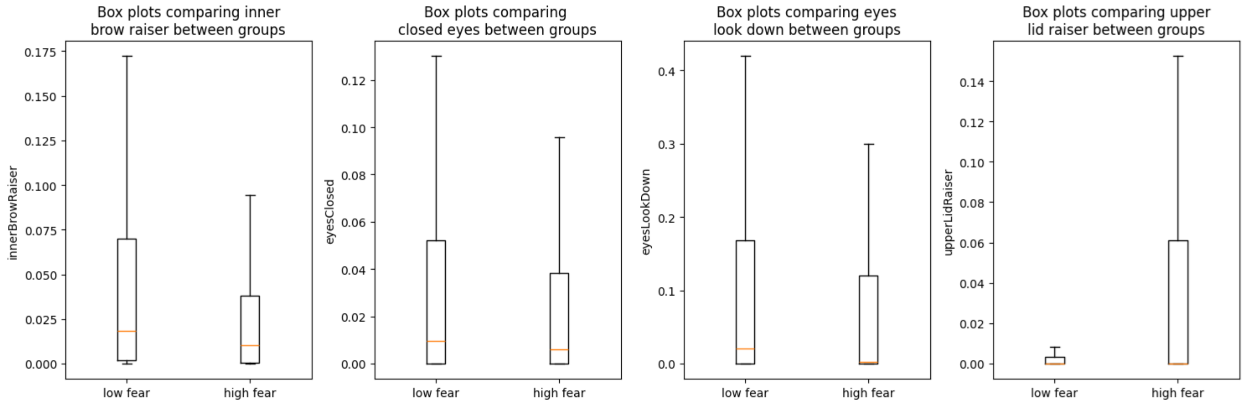


Fig. 7. Box plots comparing the distribution of innerBrowRaiser, eyesClose, eyesLookDown and upperLidRaiser between acrophobic and non-acrophobic groups.

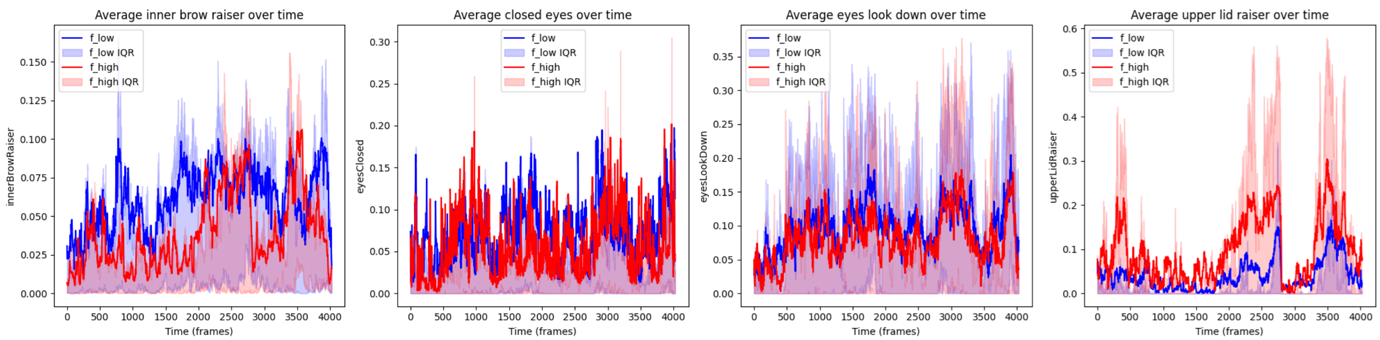


Fig. 8. Time series graphs comparing the average innerBrowRaiser, eyesClose, eyesLookDown and upperLidRaiser between acrophobic and non-acrophobic groups, where the shaded areas denote interquartile range.

of the nuanced behavioural cues of acrophobia, prompting future work to adopt a more qualitative understanding to drive the development of affect-sensitive ET systems. Moving forward, we believe that VR acrophobia exposure may benefit from a focus on autonomic and facial dynamics in order to distinguish direct acrophobic traits, in addition to incidental affect classification.

## VI. Conclusion

This study introduced a prototype 'VR experiences' application, which incorporated two acrophobia scenarios alongside physiological data capture through integrated multimodal sensors. User studies with 41 participants revealed statistically significant differences between acrophobic and non-acrophobic individuals for the majority of features, both in terms of distributional characteristics and central tendencies. Practical significance was observed in several attributes – notably mouth movements such as lip funnelling, lip tightening and jaw thrusting – which were intermittently and increasingly frequent in the acrophobic group, while more static in the non-acrophobic group.

Additional distinctions were noted in autonomic responses and eye behaviours, advocating these measures as potential indicators of acrophobic reactions. These findings provide introductory insights into how acrophobia itself – surpassing incidental categories of affect – manifests in VR contexts, laying the foundations for qualitative intricacies to be considered in future ET research. Expanding the dataset, both in participant numbers and VR scenarios, could improve the generalisability of findings, working to deepen our understanding of acrophobia responses and better inform the development of adaptive, behavioural-driven VR therapy.

## Ethical Impact Statement

Our research involved human participants who engaged in various VR experiences. Participants were provided with a detailed information sheet and verbal briefing, and gave informed consent via signature to evidence an understanding of the study procedure, our research, data usage and their rights. Demographic data was anonymised and aggregated, while physiological data was securely stored on university-approved systems after anonymisation. All resources and methods were approved by the Research and Ethics Committee at Swansea University.

Given the sensitive nature of phobia treatment, several ethical considerations must be addressed. To enhance reliability, generalisability and fairness, our findings must be validated on larger, representative sample sizes across diverse populations and backgrounds, including a range of VR experiences that accurately reflect clinical contexts. Transparency in methodology is vital to avoid or identify potential biases and strengthen the validity of the results.

Deployment of VR therapy must be handled responsibly, (1) prioritising the privacy and security of data and clinical outcomes, (2) ensuring that therapeutic interventions are effective, validating the reliability of findings to prevent misdiagnosis or harmful exposure, and (3) considering the broader clinical implications for both patients and healthcare providers when integrating such technology into practical workflows.

## REFERENCES


[1] Matthew Price, Natasha Mehta, Erin B. Tone, and Page L. Anderson. Does engagement with exposure yield better outcomes? components of presence as a predictor of treatment response for virtual reality exposure therapy for social phobia. *Journal of Anxiety Disorders*, 25(6):763–770, 2011.

[2] Thomas D Parsons and Albert A Rizzo. Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of behavior therapy and experimental psychiatry*, 39(3):250–261, 2008.

[3] M. Slater, D.-P. Pertaub, and A. Steed. Public speaking in virtual reality: facing an audience of avatars. *IEEE Computer Graphics and Applications*, 19(2):6–9, 1999.

[4] Dan Opdyke, James S Williford, and Max North. Effectiveness of computer-generated (virtual reality) graded exposure in the treatment of acrophobia. *Am J psychiatry*, 1(152):626–628, 1995.

[5] Oana Bălan, Gabriela Moise, Alin Moldoveanu, Marius Leordeanu, and Florica Moldoveanu. An investigation of various machine and deep learning techniques applied in automatic fear level detection and acrophobia virtual therapy. *Sensors*, 20(2), 2020.

[6] Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 4768–4777, Red Hook, NY, USA, 2017. Curran Associates Inc.

[7] Jonathan B Grayson, Edna B Foa, and G Steketee. Habituation during exposure treatment: Distraction vs attention-focusing. *Behaviour Research and Therapy*, 20(4):323–328, 1982.

[8] Galina Menshikova, Yuriy Bayakovski, Elizaveta Luniakova, Maxim Pestun, and Denis Zakharkin. Virtual reality technology for the visual perception study. In *Transactions on Computational Science XIX*, pages 107–116. Springer, 2013.

[9] Daniel S Gonzalez, Angelo D Moro, Christian Quintero, and Wilson J Sarmiento. Fear levels in virtual environments, an approach to detection and experimental user stimuli sensation. In *2016 XXI Symposium on Signal Processing, Images and Artificial Vision (STSIVA)*, pages 1–6. IEEE, 2016.

[10] Chai-Fen Tsai, Shih-Ching Yeh, Yanyan Huang, Zhengyu Wu, Jianjun Cui, and Lirong Zheng. The effect of augmented reality and virtual reality on inducing anxiety for exposure therapy: a comparison using heart rate variability. *Journal of Healthcare Engineering*, 2018(1):6357351, 2018.

[11] Oswald D Kothgassner, Anna Felnhofer, Helmut Hlavacs, Leon Beutl, Rupert Palme, Ilse Kryspin-Exner, and Lisa M Glenk. Salivary cortisol and cardiovascular reactivity to a public speaking task in a virtual and real-life environment. *Computers in human behavior*, 62:124–135, 2016.

[12] Tibor Bosse, Charlotte Gerritsen, Jeroen de Man, and Marco Stam. Inducing anxiety through video material. In Constantine Stephanidis, editor, *HCI International 2014 - Posters' Extended Abstracts*, pages 301–306, Cham, 2014. Springer International Publishing.

[13] Edna B Foa and Michael J Kozak. Emotional processing of fear: exposure to corrective information. *Psychological bulletin*, 99(1):20, 1986.

[14] Andreas Halbig and Marc Erich Latoschik. A systematic review of physiological measurements, factors, methods, and applications in virtual reality. *Frontiers in Virtual Reality*, 2, 2021.

[15] Paul M.G. Emmelkamp and Katharina Meyerbröker. Virtual reality therapy in mental health. *Annual Review of Clinical Psychology*, 17(Volume 17, 2021):495–519, 2021.

[16] Dongrae Cho, Jinsil Ham, Jooyoung Oh, Jeanho Park, Sayup Kim, Nak-Kyu Lee, and Boreom Lee. Detection of stress levels from biosignals measured in virtual reality environments using a kernel-based extreme learning machine. *Sensors*, 17(10):2435, 2017.

[17] Emmanuel Munguia Tapia, Stephen S. Intille, William Haskell, Kent Larson, Julie Wright, Abby King, and Robert Friedman. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. In *2007 11th IEEE International Symposium on Wearable Computers*, pages 37–40, 2007.

[18] Paul Porter, Fleur Muirhead, Joanna Brisbane, Brooke Schneider, Jennifer Choveaux, Natasha Bear, Jennie Carson, Kym Jones, Desiree Silva, and Cliff Neppe. Accuracy, clinical utility, and usability of a wireless self-guided fetal heart rate monitor. *Obstetrics and Gynecology*, 137(4):673, 2021.

[19] Paolo Melillo, Marcello Bracale, and Leandro Pecchia. Nonlinear heart rate variability features for real-life stress detection. case study: students under stress due to university examination. *Biomedical engineering online*, 10:1–13, 2011.

[20] Hye-Geum Kim, Eun-Jin Cheon, Dai-Seg Bai, Young Hwan Lee, and Bon-Hoon Koo. Stress and heart rate variability: a meta-analysis and review of the literature. *Psychiatry investigation*, 15(3):235, 2018.

[21] Patrice Robitaille and Michael J. McGuffin. Increased affect-arousal in vr can be detected from faster body motion with increased heart rate. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, I3D '19, New York, NY, USA, 2019. Association for Computing Machinery.

[22] Qiaoxiu Wang, Hong Wang, and Fo Hu. Combining eeg and vr technology to assess fear of heights. In *2018 9th International Conference on Information Technology in Medicine and Education (ITME)*, pages 110–114, 2018.

[23] Yi Ding, Neethu Robinson, Qiuhao Zeng, Duo Chen, Aung Aung Phyo Wai, Tih-Shih Lee, and Cuntai Guan. Tsception:a deep learning framework for emotion detection using eeg. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7, 2020.

[24] Nazmi Sofian Bin Suhaimi, James Mountstephens, and Jason Teo. Emotional state classification with distributed random forest, gradient boosting machine and naïve bayes in virtual reality using wearable electroencephalography and inertial sensing. In *2020 IEEE 10th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, pages 12–17, 2020.

[25] Joseph Wolpe and Peter J Lang. A fear survey schedule for use in behaviour therapy. *Behaviour Research and Therapy*, 2(1):27–30, 1964.

[26] Samuel Sanford Shapiro and Martin B Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591–611, 1965.

[27] D. A. Darling. The kolmogorov-smirnov, cramér-von mises tests. *The Annals of Mathematical Statistics*, 28(4):823–838, 1957.

[28] H. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1):50–60, 1947.

[29] Ronald Aylmer Fisher. Statistical methods for research workers. In *Breakthroughs in statistics: Methodology and distribution*, pages 66–70. Springer, 1970.

[30] Lee A Becker. Effect size (es). 2000.

[31] Theodore P Beauchaine and Julian F Thayer. Heart rate variability as a transdiagnostic biomarker of psychopathology. *International journal of psychophysiology*, 98(2):338–350, 2015.

[32] Mara Mather and Julian F Thayer. How heart rate variability affects emotion regulation brain networks. *Current Opinion in Behavioral Sciences*, 19:98–104, 2018. Emotion-cognition interactions.

[33] Fred Shaffer, Rollin McCraty, and Christopher L Zerr. A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. *Frontiers in psychology*, 5:1040, 2014.

[34] Katherine A Herborn, James L Graves, Paul Jerem, Neil P Evans, Ruedi Nager, Dominic J McCafferty, and Dorothy EF McKeegan. Skin temperature reveals the intensity of acute stress. *Physiology & behavior*, 152:225–230, 2015.

[35] Laura Maruskin, Todd Thrash, and Andrew Elliot. The chills as a psychological construct: Content universe, factor structure, affective composition, elicitors, trait antecedents, and consequences. *Journal of personality and social psychology*, 103:135–57, 05 2012.

[36] Eva Krumhuber and Klaus Scherer. Affect bursts: Dynamic patterns of facial expression. *Emotion (Washington, D.C.)*, 11:825–41, 06 2011.

[37] Bahman Zohuri and Shiva Dalili. Understanding anger and effective anger management techniques (a short review). *Management Studies*, 11(4):236–244, 2023.

[38] Paul Zeichner. Nose-raising, nose-lengthening and grimacing: Expressions of arousal, vigilance, confusion, aversion and aggression.

[39] Patrick Stewart and Bridget Waller. Presidential speechmaking style: Emotional response to micro-expressions of facial affect. *Motivation and Emotion*, 33:125–135, 06 2009.

[40] Robert D. Walter, Brian J. Goodacre, Charles J. Goodacre, W. Patrick Naylor, and Udochukwu Oyoyo. A comparison of gingival display with a requested smile, duchenne smile, grimace of disgust, and funnel-shaped expression. *The Journal of Prosthetic Dentistry*, 112(2):220–227, 2014.

[41] Thomas Brandt, Günter Kugler, Roman Schniepp, Max Wuehr, and Doreen Huppert. Acrophobia impairs visual exploration and balance during standing and walking. *Annals of the New York Academy of Sciences*, 1343(1):37–48, 2015.

[42] Günter Kugler, Doreen Huppert, Erich Schneider, and Thomas Brandt. Fear of heights freezes gaze to the horizon. *Journal of Vestibular Research*, 24(5-6):433–441, 2014.