InterFACE: Establishing a Facial Action Unit Input Vocabulary for Hands-Free Extended Reality Interactions, From VR Gaming to AR Web Browsing

Graham Wilson
School of Computing Science
University of Glasgow
Glasgow, United Kingdom
graham.wilson@glasgow.ac.uk

Jamie McCready
School of Computing Science
University of Glasgow
Glasgow, United Kingdom
2459354m@student.gla.ac.uk

Euan Freeman
School of Computing Science
University of Glasgow
Glasgow, United Kingdom
euan.freeman@glasgow.ac.uk

Florian Mathis
University of St. Gallen
St. Gallen, Switzerland
University of Applied Sciences of the Grisons
Chur, Switzerland
florian.mathis@unisg.ch

Harvey Russell
School of Computing Science
University of Glasgow
Glasgow, United Kingdom
2461525r@student.gla.ac.uk

Mark McGill
School of Computing Science
University of Glasgow
Glasgow, Lanarkshire, United Kingdom
mark.mcgill@glasgow.ac.uk

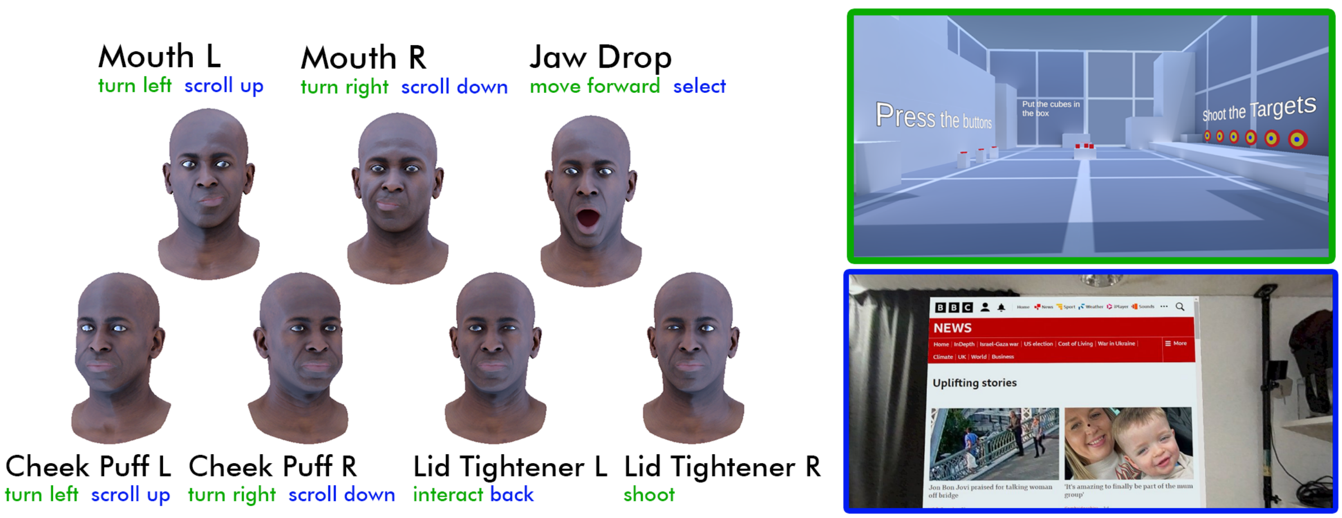

Figure 1: *Left:* The 7 candidate FAUs selected from 53 available FAUs in Study 1 based on their evaluated usability, and their bindings for controlling the VR Game and AR Web Browser in Study 2. *Right:* Screenshots of the two XR applications evaluated in Study 2, where facial input detected on a Meta Quest Pro headset was used exclusively for interaction.

Abstract

Extended Reality (XR) interactions often rely on spatial hand or controller inputs - necessitating dexterous wrist, hand and finger movements including pressing virtual buttons, pinching to select, and performing hand gestures. However, there are scenarios where such dependencies may render XR devices and apps inaccessible to users - from situational/temporary impairments such as encumbrance, to physical motor impairments. In this paper, we contribute to a growing literature considering facial input as an alternative. In a user study (N=20) we systematically evaluate the usability of 53 Facial Action Units in VR, deriving a set of optimal (comfort, effort, performance) FAUs for interaction. We then use these facial inputs to drive and evaluate (N=10) two demonstrator apps: VR locomotion, and AR web browsing, showcasing how close facial interaction can get to existing baselines, and demonstrating that FAUs offer a viable, generalizable input modality for XR devices.

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713694

CCS Concepts

• **Human-centered computing → Virtual reality**; *Empirical studies in HCI*.

Keywords

Virtual Reality, Facial Action Units, Face, Interaction, Impairment

ACM Reference Format:
Graham Wilson, Jamie McCready, Euan Freeman, Florian Mathis, Harvey Russell, and Mark McGill. 2025. InterFACE: Establishing a Facial Action Unit Input Vocabulary for Hands-Free Extended Reality Interactions, From VR Gaming to AR Web Browsing. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 23 pages. https://doi.org/10.1145/3706598.3713694

1 Introduction

Virtual Reality (VR) is an inherently *active* medium, where user head, arm and body movements form the basis of spatial interaction [94], locomotion [23], sense of presence/immersion [80], and embodiment [45] that the technology excels at. Full use of the spatially-tracked controllers (or hand-tracking), that are ubiquitous across headsets, requires not only arm movements for spatial translation and hand/wrist movements for rotation, but also multiple fingers for interaction, including pressing various buttons or performing hand gestures. However, there are a number of scenarios where VR or Extended Reality (XR) technologies may be partially inaccessible. For example, people with situational/temporary impairments [76], physical motor impairments or neuromuscular diseases [50, 57, 87] may face difficulties using controller/hand-based inputs. Others, who may have their hands occupied/encumbered [60], may nonetheless benefit from additional inputs beyond controller/hand input alone. Alternative input modalities/control methods are typically suggested to address impairments, such as voice input [35], gaze input [69], body-based inputs [54] and more. But speech suffers in particular from social acceptability concerns [62], whilst eye-tracking offers a limited range of inputs and is often combined with other modalities such as pinch for selection [67]. And whilst there have been in advances in BCIs to facilitate user interaction [89], such sensing is inherently noisy, invasive (e.g. wearing EEG caps with wet electrodes, or using implantable devices such as Neuralink[1]) and needs to be trained carefully to users. Users could benefit from supplementary interaction modalities that offer a wider range of reliable, directly controlled, generalized inputs that could be used in conjunction with, or in lieu of, voice, gaze, BCI and other inputs.

This absence of available interactions to accommodate for disabilities and impairments has led some authors to propose that VR is in fact an *ableist* technology that "*fails to accommodate the needs of users with a disability*" [24], one that has not necessarily been designed to support those with a range of physical abilities. Despite advancements in VR accessibility, such as tools like SeeingVR, which enhance applications for people with low vision through visual and audio augmentations [97], and adaptive VR interfaces designed for wheelchair users [31], using VR headsets still requires significant physical effort. This includes upper body strength for device support and stability during movement [18], as well as interaction with motion controllers that exclude users with motor impairments who cannot operate these devices effectively.

While existing VR and MR systems are not intentionally designed to exclude disabled people, the examples suggest that many VR/MR applications inherently rely on physical movement, inadvertently creating accessibility barriers. These interaction challenges are particularly problematic given the notable benefits that VR can have for disabled people. For example, for lifelong wheelchair users, VR can provide the experience of walking[83]. Gerling and Spiel explored the inherent assumptions of VR technology in relation to disability studies, showing that VR "*assumes a 'corporeal standard' (i.e., an 'ideal', non-disabled human body), and fails to adequately accommodate disabled people*", relying on "bodily involvement" to be usable, thus excluding those with a minority body [32]. One promising approach to provide potential solutions to some of the existing challenges is to provide an alternative input method is using *facial input*. In a single-user case study, Taheri *et al.* [84, 85] proposed exploiting *Facial Action Units (FAUs)* [25] (visually discernible facial movements) to create a hands-free, facial input-based video game controller for people with quadriplegia. They found that facial input allowed for enjoyable and engaging control of a variety of video game experiences. However, this work examined only a limited, hand-curated set of FAUs for input, bespoke to the single participant both in which FAUs were picked, and what appropriate thresholds were set.

In this paper, we build upon the pioneering work of Taheri *et al.* [87] to explore the potential for establishing a *facial input vocabulary* for universal use. We do so against a landscape where current generation consumer off-the-shelf (COTS) VR headsets have begun to integrate sensing technologies that enable the detection of FAUs, e.g., through lower face cameras [93], eye tracking [36], Electrooculography (EOG) [73] and more. FAUs promise not just to provide an alternate means of control for locomotion and interaction in games, but also to act as another input modality for a range of XR/spatial interactions, able to be robustly detected on a standardised mobile COTS platform and enacted hands-free.

However, to unlock such potential, we need to understand which FAUs are usable for interaction - those that are comfortable, low effort, reliably recognisable, high performance, that could consequently define our facial input vocabulary. Therefore, we conducted a study (N=20) to measure how comfortable, effortful and performant 53 Facial Action Units were to enact, using the face-tracking in-built to the Meta Quest Pro VR headset[2]. Our results identified a set of seven candidate FAUs for interaction (*JawDrop*, *CheekPuffL/R*, *LidTightenerL/R*, *MouthL/R*) that were perceived as highly usable (low fatigue, low effort, high performance) and could be reliably classified in a trained model relative to the user's face resting state (F1 score 0.95, accuracy 97%). We then integrated this model into two proof-of-concept VR Unity applications: the first a VR game that leveraged FAUs to exclusively drive player locomotion and controls (forward, backward, turn left, turn right, select, fire); the second an AR web browser that leveraged FAUs to select, scroll, and move back through browsing history. Through our evaluation (N=10) of these applications, our work demonstrates the viability

[1]https://neuralink.com/, last accessed August 2024

[2]https://www.meta.com/gb/quest/quest-pro/, last accessed August 2024

of facial input as an alternate input modality for consumer XR platforms.

Based on our findings, we reflect on the applications for our identified FAU input vocabulary for VR and XR interaction given the increasing integration of FAU and lower face tracking in COTS hardware. We consider both the accessibility benefits in supporting face-based VR locomotion, as well how such facial inputs might be used alone, or in conjunction with gaze and speech input, to drive a myriad of XR applications. And we reflect on the necessary future work, strongly motivated by our findings, towards a fully realised and in-built accessible interaction modality as standard in VR/XR headsets where facial sensing is available. This would represent a step towards transitioning XR from being ableist to being genuinely accessible to minority bodies that may be unable to physically interact with their hands or controllers, and potentially benefit everyday interactions with spatial computing technology for all users.

1.1 Contribution Statement

This paper makes the following contributions:

- A novel usability evaluation of 53 FAUs sensed on a consumer off-the-shelf VR headset (Quest Pro), resulting in an open dataset (N=20) of objective and subjective FAU performance during a "enact and hold" FAU task;
- A candidate set of 7 FAUs for interaction that exhibit high performance and responsiveness, whilst being perceived as low effort, comfortable and performant;
- A trained model for classifying the binary activation of these FAUs with a 97% classification accuracy;
- Findings from a usability evaluation (N=10) of two VR demonstrator applications where the classifier is used for AR spatial interaction using a web browser, and VR locomotion and interaction in a game - two application areas which combined demonstrate the range of inputs necessary to support facial input as an alternate input modality on consumer XR devices.

2 Related Work

We first discuss situational, temporary and permanent impairments within the context of technology. We then review the accessibility of XR headsets and experiences, especially input modalities for VR interaction. Finally, we review literature on facial input for interaction and the Facial Action Coding System (FACS).

2.1 Situational, Temporary and Permanent Impairments

Many people have *physical impairments* in the mobility, strength or availability of their hands or arms, which may make it difficult to engage in the range of spatial inputs and button presses that VR applications expect [50, 57, 87]. In the UK alone, 24% of the population reported some form of disability, with 25% of impairments reported by disabled people[3] being related to dexterity, and 47% to mobility [29]. Motor impairments result in "*the partial or total loss of function of a body part, usually a limb or limbs. This may result in muscle weakness, poor stamina, lack of muscle control, or total paralysis*"[4]. And traumatic injuries such as spinal cord damage can similarly lead to Quadriplegia - loss of motor and/or sensory function from the neck down. In such cases, the notable benefits of VR - from supporting pain relief [52, 74], well-being and mental health [30, 91, 92], to giving users the freedom to immersively explore virtual worlds and experience free-form walking [83] - are to an extent denied to disabled users of VR headsets, who lack accessible inputs/interactions to give them agency and direct control over their experiences.

However, beyond motor impairments, there are many scenarios in which a user could be unable to use their hands to operate an XR headset for reasons of temporary or situational impairments. For example, Crossan et al. [21] developed a device which utilises foot tapping to deal with situational impairments relating to the usage of a mobile phone. Others, such as Sarsenbayeva et al. [75] performed research into the situational impairment concerning mobile devices. Their research highlighted seven components which cause situational impairment. Of these impairments, the one that is addressed by facial input is encumbrance. Of the solutions which they propose to address these issues, *FatThumb* [16] and *GraspZoom* [55], both approaches seek to still utilise the core principle of hands interacting with the mobile device even if the encumbrance could not accommodate that. If we had a generalised understanding of the usability of facial input, this could also open the door to novel VR interactions addressing encumbrance.

2.2 Accessibility of XR Headsets, Experiences and Interactions

Peck et al. [65] argued that VR technology is biased and *"excludes approximately 95% of the world's population by being primarily designed for male, western, educated, industrial, rich, and democratic populations"*. In fact, VR and XR generally has been repeatedly noted to be an *ableist* [8, 32] and sexist [77, 82] and ageist [13] technology. This is due to the reliance on highly dexterous body/hand movements to drive inputs (e.g. roomscale locomotion, handheld controllers, precise targeting etc.), and the assumptions made about the body morphology (e.g. height, head size, interpupillary distance) and abilities (dexterity, strength, mobility etc.) of the users. However, XR technologies also contain the requisite sensing to be able to uniquely address their existing accessibility limitations *by design* [59]. For example, *All Access Life* [11] demonstrated how a user with cerebral palsy could play VR games without the use of controllers, using computer vision to track the user's arms to interact with the VR playspace. However, the need to support accessibility is often seemingly a secondary consideration commercially. Consequently, accessibility is increasingly becoming a legal requirement worldwide. As part of the *Equality Act 2010* in the UK for instance, reasonable adjustments must be made to accommodate people who are disabled [64]. This means that businesses which utilise XR devices as a tool must make sure that an individual, who otherwise would not be able to use them, has features in place which allow them to interact with these devices. Similar legislation

[3]We use identity-first language to refer to disabled people [79] and also take into account SIGACCESS guidelines (https://www.sigaccess.org/welcome-to-sigaccess/resources/accessible-writing-guide/) throughout the paper. Links last accessed August 2024.

[4]https://www.neuromodulation.com/motor-impairment, last accessed August 2024

is in place in the USA (*Americans with Disabilities Act*) [1] and the EU (*Employment Equality Directive*) [9].

In addition to legislation and promising positive trends in making XR headsets and experiences more accessible, a growing body of literature is emerging that explores the introduction of alternative interactions in XR, some of which accommodate different user needs and preferences. For example, speech input is commonly put forward as an additional input method for XR, with the Apple Vision Pro [3] being one example that supports speech input to drive UI interactions. However, speech input has notable issues, in terms of universal recognition (coverage of accents and languages), social acceptability and discretion (being highly audible, noticeable and betraying the privacy of the user), and disruption. For example, consider the user of VR whilst occupying a bed [90], such as when in hospital or under hospice care. It is rare for these conditions to be completely private, and so it may be considered socially unacceptable to be talking constantly at a level where you can be understood by the microphone and heard by others nearby.

Other options for accessibility control in XR have some significant drawbacks. Eye tracking alone does not offer as much potential range of control as facial input. Eyes can be difficult to precisely control depending on what the user sees. It is very easy for users to look away from what they are meant to, as shown by Zhang *et al.* [96], with the mode of input effectively interfering with the perception of the visual output of the display. Other alternatives have limited range of inputs available. *Sip-and-puff* is a method by which users can blow or suck a tube to produce an input [6]. While this was originally designed for wheelchairs, it has been adapted by Jones *et al.* to work with the Apple iPod [41]. The interaction offered however only supports pause, play and fast-forward, highlighting the limitations of this device. A device which does not have these limitations but suffers from others is the variety of Accessibility Controllers offered by console manufacturers [4, 5]. However, they are designed for a user who has some manoeuvrability within their arms.

Also underpinning comparisons between available input modalities is the strongly motivated need to explore and support multiple alternate inputs to ensure a broad range of accessibility options to cater for all people. For example, Jimenez-Mixco et al. [39] suggested systems should ensure that *"every interaction is possible with at least two modalities"* [39]. Research has looked at a number of ways by which input could be provided absent the need for hands and body-based interactions, for example using gaze/head-directed input for steering in VR [99]; using gaze interaction exclusively for input as a proxy for hand-based interactions [56]; or coupling gaze input with BCIs to enact discrete input events such as selections [71]. Common here is the challenge that it is not enough to target with gaze alone - there is a need to support a range of input actions, often associated with targeting, to support interaction with UIs, character locomotion etc.

2.3 The *Facial Action Coding System* and Voluntary Control of Facial Action Units

The Facial Action Coding System (FACS) [25] was designed as a method to describe all visually discernible facial movements. Using what is known as *Facial Action Units* or simply Action Units (FAUs/AUs), a comprehensive description of the expression on the face can be mapped and analyzed. It has been used as a basis for representing the face across the technological landscape, from animation studios like Disney Pixar and Dreamworks to Apple implementing it in *ARkit*. A FAU can be described biologically as a tensing or relaxing of one or several facial muscles [20].

Gosselin *et al.* examined voluntary control of individual FAUs [33], looking at adults ability to 'activate' FAUs and the consequential AUs which were also 'activated'. At the time, they did not have access to automated FAU classification approaches, so instead a simple recording set-up was used with a VHS recorder, and as such the term 'activated' does not have an explicit qualifier. It was left to the agreement of 3 'facial coders' meaning the results are somewhat subjective albeit with a consensus reached. The results of the experiments show how other FAUs were activated when a participant was told to perform the 20 chosen action units, and how this varied greatly depending on the chosen AU. Some examples which also feature on the Meta Movement SDK are *LipCornerPuller*, *NoseWrinkler*, and *LidTightener*. Ekman, creator of FACS, along with his peers, also ran a study into the voluntary control of facial action units in children [26]. Gosselin *et al.*'s paper states that because of what is discussed in Provins' work [70] there is a significant difference in the ability to control facial actions correlating with age; being younger means more difficulty. While this may be true, Provin's paper does not feature any study of facial control, instead focusing on larger body parts such as the arms and their grasping ability.

2.4 Facial Input for Interaction

There have been a number of attempts at using *facial input* [12, 63] to drive actions for video games and general UI interactions. Taheri *et al.* [84–86] designed a system for people who are quadriplegic to utilise facial expressions as a virtual controller for games. In a form of single user case study, they worked with their participant to select a set of FAUs and macro expressions - combinations of AUs indicating a commonly understood expression: Happiness (*CheekRaiser*, *LipCornerPuller*); Sadness (*InnerBrowRaiser*, *BrowLowerer*), Disgust (*NoseWrinkler*, *UpperLipRaiser*), Wide Eyes (*OuterBrowRaiser*, *UpperLidRaiser*), Lid & Lid Tightener (*LidTightener*, *LipTightener*), JawDrop (*BrowLowerer*, *LipsPart*). These were triggered by user-defined thresholds, in combination with alternate speech-to-text options for each input to drive gameplay control. They noted that *"the AU combinations used in each game are experimentally determined by Atieh [the user] since it was not possible for them to make every facial expression that an able-bodied person might easily make"*. They found the control scheme had high learnability, was enjoyable and engaging, and ultimately provided their participant a form of input that gave them more agency and capability to control video games, validating that FAUs offer a promising interaction modality for people with motor impairments. However, Taheri's work had limitations - the rationale behind FAU/expression choices was largely based on the individual preference of a single user, impacting generalisability, whilst there was no baseline to compare control performance to - meaning that whilst the control scheme had significant value for the user, it is difficult to more objectively reflect on the extent to which it achieved parity with existing controls/inputs (in terms of usability and performance in particular). In follow-on work, Taheri focused on a combination of voice input and head tilt for facilitating the same participant to control basic locomotion in VR [83], in-part it would seem as facial tracking was not yet available on VR headsets - a technical barrier that is now overcome.

Dey *et al.* [22] examined the use of facial input to drive locomotion in VR across differently affective VR environments. In all environments *smile* and *frown* were used to start and stop on-rails movement, with *clench* for actions. They compared facial expressions vs controllers between subjects, finding significant differences in terms of usability in favour of controller input and varyingly higher presence for controller in select environments, concluding that facial input "is indeed a viable options to navigate and perform actions in virtual environments".

A range of papers have also used specific expressions or facial regions as a form of input, often intended to address assistive technology gaps, most commonly breathing/blowing/mouth gestures e.g. *BreathVR* [81], *VR Blowing* [78], *BlowClick* [98], *LipIO* [40], *GazePuffer* [48]; but also focussing on around-eye inputs [46]; and leveraging innate knowledge of emotional expressions to trigger actions e.g. *Evoker* [37], and more [14, 38]. GazePuffer [48] in particular emphasized the utility of facial inputs to VR and spatial UI inputs generally. Using a Pico 4 headset with facial tracking, they proposed a basic gesture set using FAUs, focussing on combinations of {left puff, right puff, up puff, down puff, omni puff, and suck} and directional mouth movements {left, right, up, down}, assessing their gestures (N=12) for fatigue, effort and psychological resistance (social acceptability), as well as recording the time taken (captured via video recording) to enact and complete a given gesture. Based on this, they rejected *Suck* gestures in particular, then assessed the ease of performing gestures via an online questionnaire (N=74). Finally, they used a combination of *CheekPuff* and gaze input to drive target selection in a Fitts-style targeting task (N=12) as well as general controls for reading, viewing images, and receiving a phone call. However, *GazePuffer* did not systematically benchmark the usability of the available FAUs sensed on an XR device, they focused solely on puff motions. And the choice of candidate gestures was not informed by the quantified performance of those gestures as captured by the target XR device - meaning for example that whilst a gesture may appear responsive based on video footage, the capability of the XR headset and associated facial tracking SDKs to reliably detect said gesture was not taken into account. Moreover, their classification approach introduced up to 260ms of latency to the input. Similarly, *FaceCommands* [51] used Mechanical Turk to survey (N=37) participants on potential facial expressions to be used for input for AR glasses, before coding and mapping suggestions to available FAU groups/expressions, deriving a set of 15 facial gestures for driving various AR interactions (e.g. smiling to answer a call, being sad to decline a call), narrowing this down to 8 candidate gestures (AU12 (*smile*), AU27 (*open mouth*), LP (*lip movements* related to negative emotions), AU1+2 (*lift eyebrows*), AU4 (*frown*), AU43 (*close eyes firmly*), AU46 (*right eye wink*), and AU61 (*eyes left*)), and capturing training data (N=13) to train a classifier for those gestures (mean F1 score of 0.911). And outside of head-mounted displays, there are a number of handheld AR solutions utilizing smartphone cameras. Snapchat, Instagram and many other apps have filters which can track a user's face and even have interactions depending on facial action activation - mostly when a user opens their mouth or raises their eyebrows [2].

2.5 Summary

What is commonly lacking in literature examining facial input for interaction is a broad, comprehensive assessment of the usability of the available inputs - be they treated at the granular facial action unit level, or as groups of FAUs comprising of expressions. The works we examined tended to cherry pick particular FAUs or expressions, based on some negotiated exploration with small numbers of participants, or picking everyday expressions that are assumed to be replicable such as smiling, frowning etc. Moreover, no paper has yet evaluated a facial input-only interface to drive VR gaming *and* interaction - with the closest works using combinations of facial input and speech outside VR [85] or only giving users limited controls in VR (e.g. no turn left/right, toggled forward movement [22]). Consequently, in this paper we set out to focus on facial input at its most granular, systematically assessing the usability of available FAUs on a consumer VR headset, and then applying our findings to a specific use case: facial input driving VR gaming and interaction.

3 Study 1: Identifying Candidate Facial Action Units for Interaction

Given that consumer VR headsets such as the Meta *Quest Pro* now support FAU tracking using inward-facing cameras - a key feature for driving personalised avatar animations for social VR - our aim was to examine which of these FAUs, if any, were viable candidates to be used for short, discrete interactions. The motivation was to find facial inputs that could be employed across a breadth of XR activities, foremost being the use of facial locomotion, potentially supporting users with motor impairments to be able to freely navigate virtual worlds.

3.1 Experimental Design

We evaluated the performance and comfort of 53 FAUs tracked by the Meta Movement SDK[5], see Table 1. There were two types of FAU tracked by this SDK that we did not evaluate: those related to *eye movements* (eyes closed/up/down/left/right), excluded as we ultimately wanted a set of FAUs that did not compromise the VR/XR viewing experience; and those related to *tongue* movements, as at the time of conducting the study the SDK had not yet implemented this support. Both would be valuable additional facial inputs to consider by future work, aided by our study design.

During the study, participants were asked to perform each of the 53 FAUs (repeated 3 times) based on a visual demonstration on a virtual avatar visible ahead of them. FAUs were performed in a random order, except that both sides of two-sided FAUs (e.g., *BrowLowerer* left and *BrowLowerer* right) were done as a sequential pair, in a random order, for 3 trials each. After attempting to perform each FAU, participants were to indicate their subjective experience of performing it, in terms of disomfort, effort and success.

[5]https://developer.oculus.com/documentation/unity/move-overview/, last accessed August 2024

Table 1: The 53 FAUs evaluated in this study. *Directions* indicates variants for performing the FAU towards Left, Right, Top, or Bottom. Textual descriptions are taken directly from the Meta Movement SDK for clarity.

FAU	Directions	Description from *Meta Movement SDK*
BrowLowerer	L/R	Knit and lower the brow area and lower central forehead.
CheekPuff	L/R	Fill the cheeks with air causing them to round and extend outward.
CheekRaiser	L/R	Tighten the outer rings of the eye orbit and squeeze the lateral eye corners.
CheekSuck	L/R	Suck the cheeks inward and against the teeth to create a hollow effect in the cheeks
ChinRaiser	T/B	Push the skin of the chin and the lower lip upward. When the lips are touching, the upward force from the lower lip pushes the up top lip as well.
Dimpler	L/R	Pinch the lip corners against the teeth, drawing them slightly backward and often upward in the process.
InnerBrowRaiser	L/R	Lift the medial brow and forehead area.
JawDrop		Moves the lower mandible downward and toward the neck.
JawSideways	L/R	Moves the lower mandible leftward.
JawThrust		Projects the lower mandible forward.
LidTightener	L/R	Tighten the rings around the eyelids and push the lower eyelid skin toward the inner eye corners.
LipCornerDepressor	L/R	Draw the lip corners downward.
LipCornerPuller	L/R	Draw the lip corners up, back, and laterally.
LipFunneler	L/R × T/B	Fan the lips outward in a forward projection, often rounding the mouth and separating the lips.
LipPressor	L/R	Press the upper and lower lips against one another.
LipPucker	L/R	Draw the lip corners medially causing the lips protrude in the process.
LipStretcher	L/R	Draw the lip corners laterally, stretching the lips and widening the jawline.
LipSuck	L/R × T/B	Suck the lips toward the inside of the mouth.
LipTightener	L/R	Narrow or constrict each lip on a horizontal plane.
LipsToward		Forces contact between top and bottom lips to keep the mouth closed regardless of the position of the jaw.
LowerLipDepressor	L/R	Draw the lower lip downward and slightly laterally.
Mouth	L/R	Pulls the left lip corner leftward and pushes the right side of the mouth toward the left lip corner; pulls the left lip corner leftward and pushes the right side of the mouth toward the left lip corner.
NoseWrinkler	L/R	Lift the sides of the nose, nostrils, and central upper lip area. Often pairs with brow lowering muscles to lower the medial brow tips.
OuterBrowRaiser	L/R	Lift the lateral brows and forehead areas.
UpperLidRaiser	L/R	Pull the top eyelid up and back to widen eyes.
UpperLipRaiser	L/R	Lift the top lip (in a more lateral manner than nose wrinkler)

3.2 Procedure

Whilst wearing a Quest Pro VR headset, we first asked participants to make a neutral resting face and captured this state for a baseline of FAU activation when not actively attempting to perform specific FAUs. Further baseline resting face data was captured at the end of each trial. Then, we randomly proceeded through the 53 FAUs. We conducted an initial practice run where participants were presented with a Meta SDK reference face model animation illustrating what activating the current FAU would look like, and asked participants to practice enacting the given expression until they were comfortable. Participants did not receive visual feedback about their performance, as individuals have varying movement ranges/characteristics, so requiring a set level/constellation of activation may be uncomfortable (or not that individual's preference) After the familiarization phase, we moved to the data collection phase, where participants performed each FAU in Table 1 three times. For each trial, participants were given a short countdown and then prompted to perform and hold the chosen FAU for a period of 3 seconds (i.e. representative of a prolonged discrete input), performing this as comfortably and accurately as possible. After the completion of the trials, they were given an in-headset questionnaire (see Section 3.3).

3.3 Measures

We assessed the objective performance and subjective perception of the available FAUs to make a determination of their suitability for use as discrete inputs in VR.

3.3.1 Subjective Performance. After each set of trials for each FAU, participants received three brief in-headset questions, all on 11-point scales, covering the following topics and questions:

Discomfort: "*How comfortable was the expression?*" from "Very Comfortable" to "Very Uncomfortable";

Effort: "*How much effort did it take to make the expression?*" from "No Effort" to "Maximum Effort", with the scale end points inspired by the Borg rating of perceived exertion [15];

Performance: "*How well would you rate your performance in replicating the expression?*" from "Replicated Perfectly" to "Failed to Replicate".

3.3.2 Objective Performance. We also captured real-time log data at 20Hz for the 3 second duration of the trial (the "holding" phase), and for an additional 1 second post-trial (the "release" phase), capturing the activation state (ranging from 0..1 as provided by the Meta SDK) for every FAU so that we would have a complete picture of what FAUs were activated alongside the AU of focus for the trial. From this, we calculate:

Mean Activation: Referring to the average activation value (0..1) across the trials for a given FAU, based on the 2 second time window of 0.5 to 2.5 seconds during the trial (i.e. ignoring the first 0.5 seconds where the participant reacts to enact the FAU, and the last 0.5 seconds of the trial where fatigue might impact their performance in holding the FAU).

Responsiveness: Referring to the time to reach a given threshold during the trial (in seconds), with the threshold being the lower confidence interval calculated post-hoc based on the per-participant 95% confidence interval of the activation values. This is intended to give a measure of how quickly an FAU could be meaningfully activated given a basic personalised threshold similar to [85].

3.3.3 Overall Usability Score. Finally, as a means of identifying the "best" FAUs available across the range of measures we captured, we calculated an overall usability score for each FAU, normalizing each measure to range between (0—10) where lower was better, providing a rank order taking all available measures into account as follows:

```
Effort + Performance + Discomfort +
(Responsiveness / 2)*10 + (1-Mean
Activation)*10
```

3.4 Participants

Participants were recruited using advertising on social media platforms and internal mailing lists, being paid £10 for their time in Amazon vouchers. 20 participants took part (13 male, 7 female), with mean age = 24.0 years (SD = 5.0), and mean hours of VR usage a week = 1.9 hours (SD = 3.5), with 8 participants not being VR users. 8 participants wore glasses and 1 participant wore a head garment during the study, and 4 had some form of facial hair. The study was approved by our college's ethics committee.

3.5 Results

Our full dataset is available on Zenodo (10.5281/zenodo.14756351). For statistical significance testing, an Aligned-Rank Transform (ART) [28] was used to transform non-parametric data prior to conducting a repeated measures ANOVA on the *AU of Focus* factor (the target FAU being deliberately performed by the participant), using the *ARTool* R package [44]. ART enables parametric tests on non-parametric data. When using ART, as noted by Wobbrock et al. [95] *"the response variable may be continuous or ordinal, and is not required to be normally distributed"*, making it well suited to our dataset. Pairwise contrasts for main effects were also conducted [44], however given the number of pairwise comparisons given 53 FAUs (1378 tests) these are not reported. Instead, given the focus of this study was to identify a set of viable FAU candidates to use for VR/XR interaction, we calculated the *pareto optimal* top-k candidate FAUs (using *psel* from the R library *rPref*) based on an optimisation of:

```
high(Mean Activation) * low(Responsiveness)
* low(Discomfort) * low(Effort) *
low(Performance)
```

The result of this is effectively a rank ordering of FAUs (with multiple FAUs per rank), ordered from most to least optimal in balancing these outcome measures.

3.5.1 Objective and Subjective FAU Performance. Figure 2 provides an overview of the objective performance (mean activation, responsiveness) and subjective perception (discomfort, effort, performance). A repeated measures ART ANOVA was performed on *Score*, noting a significant effect ($F(27, 513) = 12.49$, $p < 0.01$). *Post-hoc* pairwise contrasts (using *art.con*) revealed 378 significant contrast results, which would not be meaningful to report - however the breadth of results here indicated that there were clearly strata of varyingly viable FAUs to consider. FAUs with the lowest overall scores. *JawDrop* in particular was very responsive with low discomfort and effort and high performance, but other FAUs scored closely to *JawDrop* e.g. *LidTightener* being slightly higher effort, *CheekPuff* being marginally worse across all measures, and *Mouth* being slightly less responsive.

3.5.2 Top-k FAU Candidates. The calculated optimal FAU candidates largely mirrored the *Score* ordering, with *psel* returning 6 rank levels (see Table 2). In particular, those at Level 1 (*LidTightenerLR*, *JawDrop*) and Level 2 (*UpperLipRaiserLR*, *LipSuckLT/RT*, *JawThrust*, *MouthLR*, *LipSuckLB/RB*, *DimplerLR*, *CheekPuffLR*) were the optimal candidates based on this ordering of measures.

3.5.3 Co-Activations Across FAUs. As FAUs are innately tied to expressions (i.e. raising an eyebrow might pull parts of the cheek and mouth upwards in-turn) we also examined the extent to which other FAUs than the target were activated (where based on the Meta Movement SDK descriptions a value tending towards 1.0 equates to the maximal/strongest expression of that FAU [61]), and indeed whether this activation was correlated with the target FAU interaction.

We can see that our best scoring FAUs commonly featured the strongest expressions in our dataset (top of Figure 3). And there were commonly strong positive correlations between Left/Right

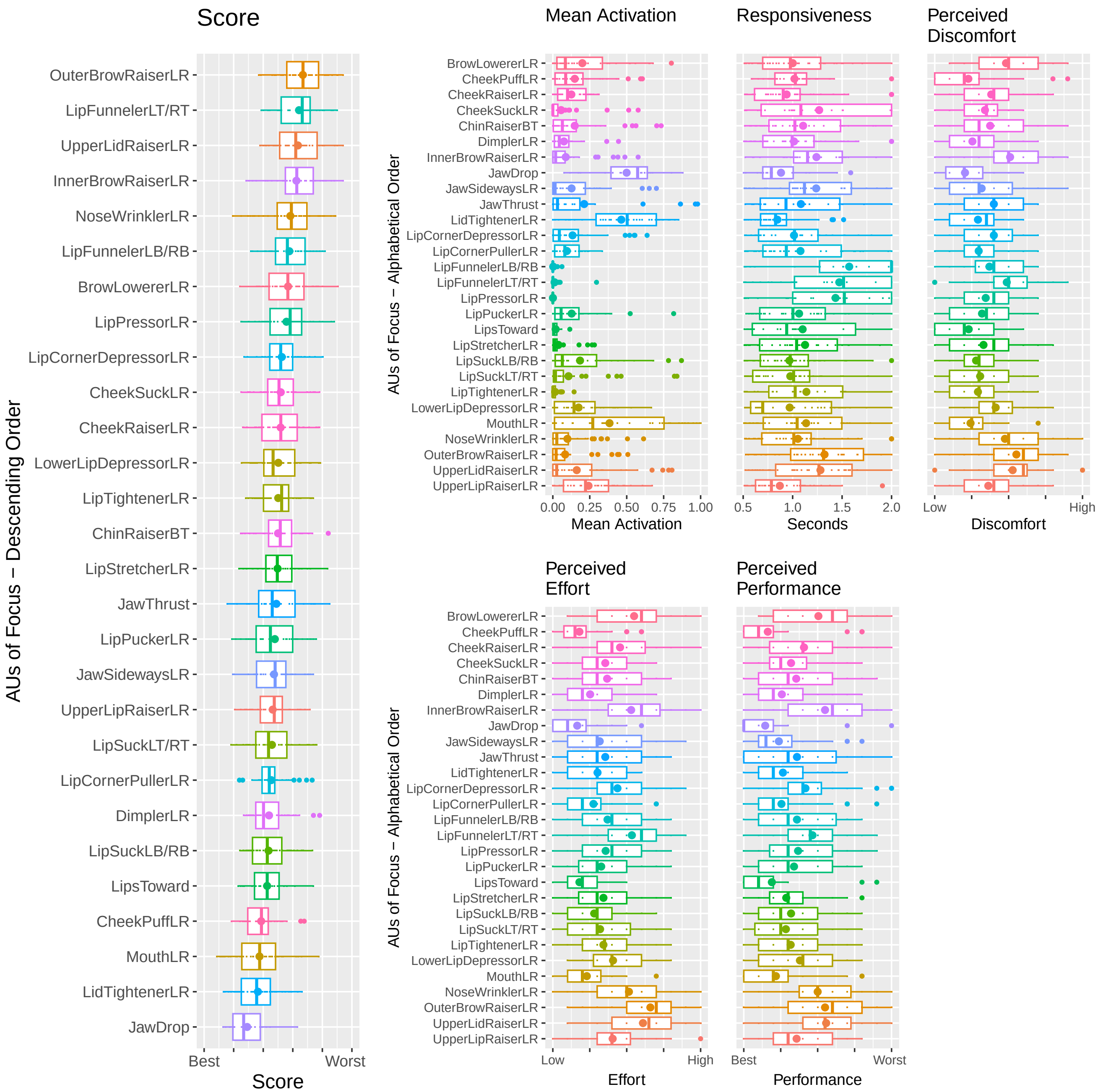

Figure 2: Boxplot of measures. For Mean Activation, Perceived Performance, and overall Score, higher is better; for Responsiveness, Perceived Discomfort, and Perceived Exertion, lower is better. Boxplot lower and upper hinges indicate the 25th and 75th percentiles.

pairs of FAUs (bottom of Figure 3) as well as strong positive correlations within related face regions (e.g. *LipFunneller*, *LipPressor*, *LipPucker*) which would indicate caution should be taken when considering the use of these regions of FAUs in-particular given the risk of false positive activation. This also points to potential problems in using symmetrical FAUs for input given co-activation, however for the majority of our best scoring FAUs despite strong correlations, there appeared to be clear differences in activation magnitudes that would suggest symmetrical actions remain feasible to examine.

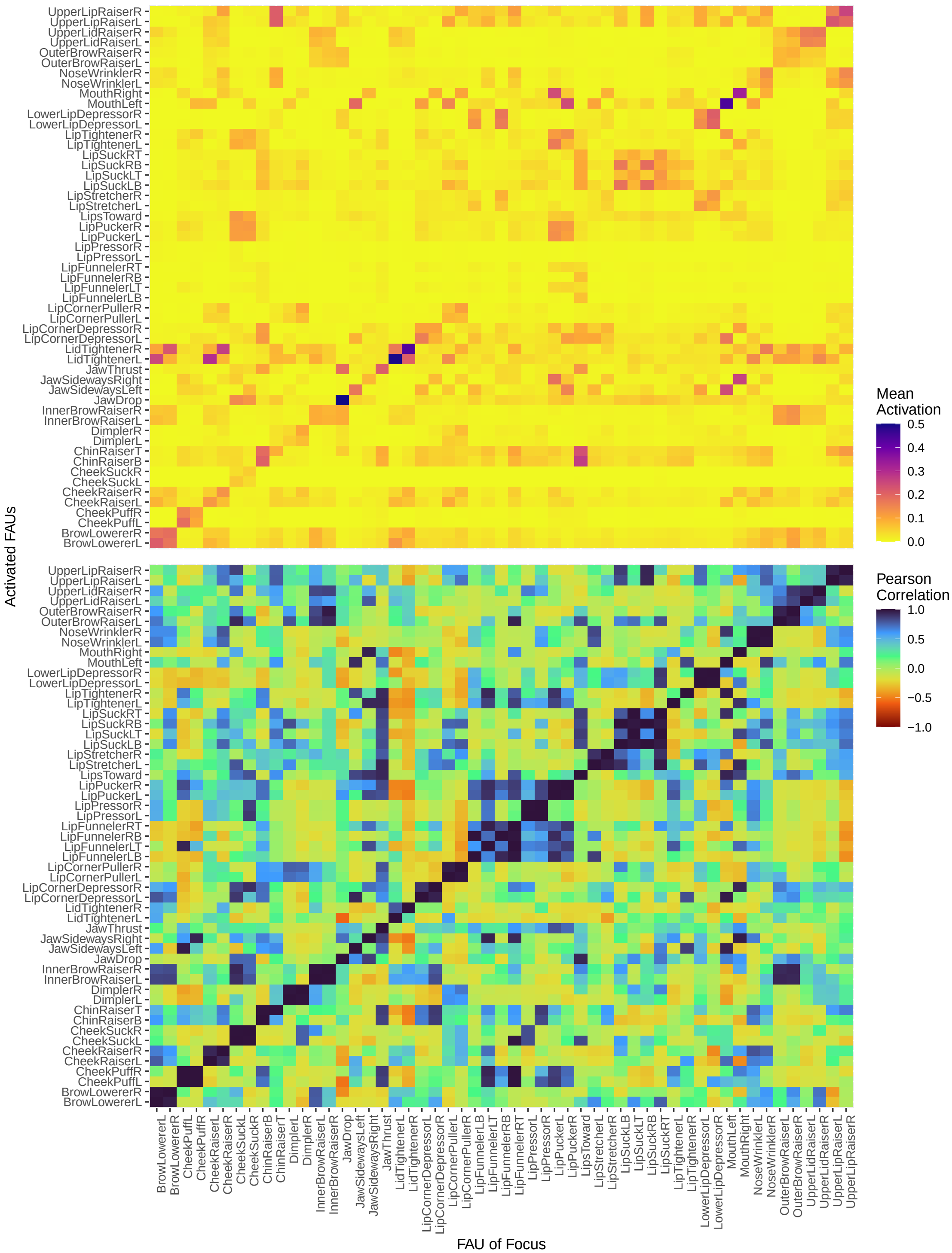

Figure 3: ***Top*: Plot of the mean FAU activation value across Participants and Trials, higher indicates a stronger expression of said FAU, bounded between (0-1). *Bottom*: A correlation matrix where Pearson's correlation has been calculated between the FAU of focus (i.e. the FAU the participant was being asked to perform) and the activation of all available FAUs.**

Table 2: Ranked optimal FAUs (lower is better). Bold FAUs are the candidates we ultimately selected to carry forward to Study 2. Cell colour ranges from green (best rank) to yellow (worst rank). Suffix letters indicate (L)left, (R)ight, (T)op, (B)ottom.

FAU	*Rank*	*FAU*	*Rank*
LidTightenerLR	1.00	CheekRaiserLR	3.00
JawDrop	1.00	LipsToward	3.00
CheekPuffLR	2.00	LipCornerPullerLR	3.00
MouthLR	2.00	BrowLowererLR	3.00
UpperLipRaiserLR	2.00	UpperLidRaiserLR	4.00
LipSuckLT/RT	2.00	NoseWrinklerLR	4.00
JawThrust	2.00	CheekSuckLR	4.00
LipSuckLB/RB	2.00	LipTightenerLR	4.00
DimplerLR	2.00	LipStretcherLR	4.00
LowerLipDepressorLR	3.00	InnerBrowRaiserLR	5.00
LipCornerDepressorLR	3.00	LipFunnelerLT/RT	5.00
JawSidewaysLR	3.00	LipPressorLR	5.00
LipPuckerLR	3.00	LipFunnelerLB/RB	5.00
ChinRaiserBT	3.00	OuterBrowRaiserLR	6.00

3.6 Selecting FAU Interaction Candidate Subset

We selected the two 1 ranked Pareto optimal FAUs (*JawDrop*, *LidTightenerLR*) and the lowest scoring level 2 ranked FAUs (*MouthLR*, *CheekPuffLR*). In practice, this gave us 7 FAUs for subsequent designs, taking into account the available left/right variants for all FAUs bar *JawDrop*.

We stress that this is a generalised selection of candidates, and different preferences for optimisation (e.g. prioritising performance and responsiveness alone for high-performance FAUs; or prioritising comfort and exertion alone for the easiest-to-use FAUs) would be likely to result in subtly different candidate selections for a given interaction context or user group. Importantly, our open dataset allows for such further exploration of FAU interactions, particularly motivated by the fact there were a total of 15 FAUs (including directionality) that were ranked 2 or above and featured a blend of discernible activation, responsiveness, perceived performance and low discomfort and effort. Nonetheless, we believed that this candidate list would give us a selection of viable, usable FAUs that were broadly performant and comfortable to use for which to evaluate multi-FAU interaction.

4 Facial Action Unit Vocabulary and Model

4.1 Vocabulary

We chose a vocabulary of seven facial action units based on our findings in Study 1 (shown in Table 2). This vocabulary includes three pairs of opposite-side action units. This horizontal symmetry is potentially ideal for mapping onto diametric interaction functions, e.g., turning left/right, skipping to previous/next song, decreasing/increasing volume. These pairs also focus on different regions of the face (i.e., eyes, cheeks, mouth), which should help with memorability and increase the robustness of detecting which action unit is being performed. These are complemented by the *JawDrop* action unit, a non-sided face movement, better suited for the primary user interface function, e.g., toggling music playback or moving forward in a game.

4.2 Recogniser

We used the facial action unit measurement data from Study 1 to train and test a neural network for detecting when users were performing the facial action units in our vocabulary. We used Keras [19] to implement the model. All 53 facial action unit measurements were used as input features; we used this full feature set because most of the facial expressions resulted in co-activation of other facial action units (e.g., the *LidTightener* actions result in other measurable eye movements) and so a larger number of features was expected to improve classification accuracy. Model output was a 1×8 tensor representing the estimated probability of the seven facial action units plus a resting face state (i.e., a neutral facial expression).

Data from Study 1 was split 80%/20% into training/testing, respectively. Evaluation of model performance on the test dataset shows an F1 score of 0.95 with classification accuracy of 98%. F1 scores for the individual classes are shown in Figure 4. Analysis of the confusion matrix for the test dataset found that almost all of the incorrect classifications were due to predicting the *Resting* facial state instead of an active facial action unit. From a usability perspective this is ideal because false positives for the resting facial state mean no input actions will occur, versus the less convenient alternative of falsely activating an incorrect action unit.

5 Facial Input Prototype Demonstrators

To demonstrate and test the use of facial inputs in XR in-practice, we decided to focus on the design of two prototypes that exemplified different common forms of interaction with an XR device - *Locomotion* in the form of an interactive VR game; and *Web Browsing* as a form of spatial interface interaction for AR. Across both Tasks, FAUs were triggered using the model trained previously (updated per-frame), with input FAU activation data from the Meta SDK, and both tasks were run from the Unity Editor using Meta Link to ensure performance.

Facial Action Unit	F1 Score
CheekPuffL	0.93
CheekPuffR	0.95
JawDrop	0.94
LidTightenerL	0.97
LidTightenerR	0.98
MouthL	0.95
MouthR	0.93
Resting	0.98
Overall	*0.95*

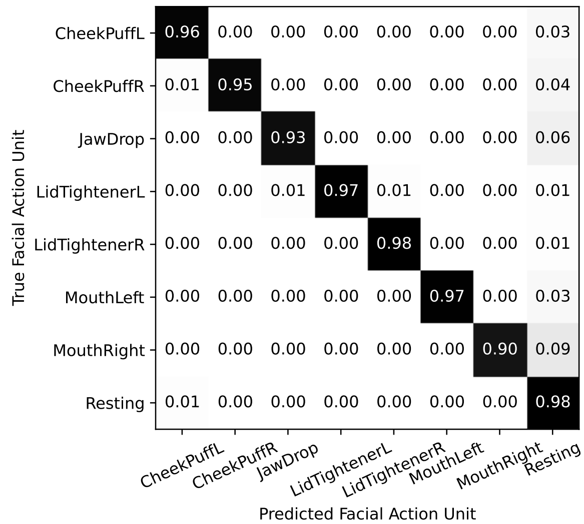

Figure 4: F1 scores for the eight facial action units (left) and confusion matrix showing classification accuracy (right).

5.1 Task 1 - Gaming in VR

This constituted of a small scale free-roaming VR test environment (see Figure 5), where users would have to navigate, interact with virtual objects and buttons, and shoot virtual targets - effectively necessitating common locomotion and interaction controls that drive a host of VR games and experiences.

5.1.1 Environment. We designed a basic environment that consisted of a room (Figure 5 (c)) and a twisting corridor (Figure 5 (d)). This environment contained a series of tasks to be completed, based on exemplars from the Unity XR Interaction Toolkit[6]:

Training Prior to starting the trial, participants were shown a gallery of faces (Figure 5 (b)) which demonstrated the controls, and gave participants feedback on FAU activation, allowing them to practice each FAU input in turn. These were depicted by the Meta Movement SDK *High Fidelity Avatar* face.

Shooting (triggered by Shoot) Shooting a series of 5 static targets.

Object Manipulation (triggered by Interact) Grabbing a series of 3 cubes and then moving them to a nearby container/basket.

Selection (triggered by Interact) Pressing a series of 3 buttons.

Escape the Maze Once the above sub-tasks were completed, a door would open and the user would then have to navigate a small number of left/right turn corridors to arrive at the "exit", ending the trial.

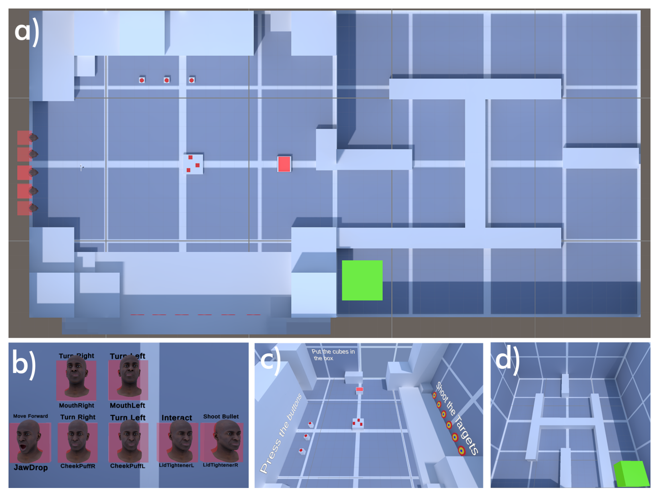

Figure 5: Environment for Task 1. a) Top-down view; b) Starting gallery of faces depicting the FAU controls, used to train participants. The red background would turn green when a given FAU was activated, and participants were instructed to try each FAU until they were comfortable activating them; c) the first area encountered, with the three tasks: button selection left, object manipulaton center, and shooting right; d) the maze participants had to navigate once the three tasks were complete.

5.1.2 Facial Input Control Scheme. Based on our selected FAU candidates for interaction (LidTightener, JawDrop, Mouth, CheekPuff), we designed the following control scheme:

Forward Activated on *JawDrop* as a continuous input.

Turn Left/Right Activated on *Mouth - Left/Right* and *CheekPuff - Left/Right* as a continuous input / continuous rotation. We doubled up on these controls to allow users to experience all of the FAU candidates. This was enacted as an inertial turn to reduce jerkiness - meaning that the turn rate would ramp up/down to a maximum of 3 degrees per second and a decay rate of 0.1 degrees per second.

[6] https://docs.unity3d.com/Packages/com.unity.xr.interaction.toolkit@3.0/manual/index.html, last accessed August 2024

Interact Activated on *LidTightener - Left* as a discrete toggle to enable grabbing virtual objects.

Shoot Activated on *LidTightener - Right* as a discrete input.

This assignment was primarily informed by not assigning frequently used turn controls to *LidTightener* on the assumption they might interfere with viewing, and using the best FAU for what was anticipated to be the most frequently used action (moving forwards). These assignments ensured that participants could experience all of the candidate FAUs available, whilst leveraging directionality for turning left/right to better aid recall. Moreover, whilst we could have used toggles for movement, we elected to retain continuous input to get as close to parity with the controller input, in terms of responsiveness and immediacy of input, as possible

For targeting, we used the head orientation (with a visible reticule placed approximately 1m from the user's *CenterEyeAnchor*) rather than available eye tracking to avoid previously discussed issues with eye tracking input potentially impacting the user's comfort/ability to view the VR scene or detracting from the focus on facial vs controller input. And we differentiated between Shoot and Interact options to introduce a realistic sense of having multiple disparate inputs to enact, despite the fact this design could be simplified to enact both with one input.

We also implemented a *baseline control scheme* for comparison using the VR controllers, with standard controls: thumbstick for forward, turn left, turn right; triggers for select/shoot.

5.2 Task 2 - Web Browsing in AR

The second task was a basic Chrome-based web browser using the Unity Web Browser project[7], with users able to navigate a single website: the BBC's Uplifting Stories page[8].

5.2.1 Environment. This task was to exemplify a spatial interface for AR, and thus we enabled the full-colour passthrough mode of the Meta Quest Pro, and rendered a large exocentric browser window.

[7] https://github.com/Voltstro-Studios/UnityWebBrowser/, last accessed August 2024
[8] https://www.bbc.co.uk/news/topics/cx2pk70323et, last accessed August 2024

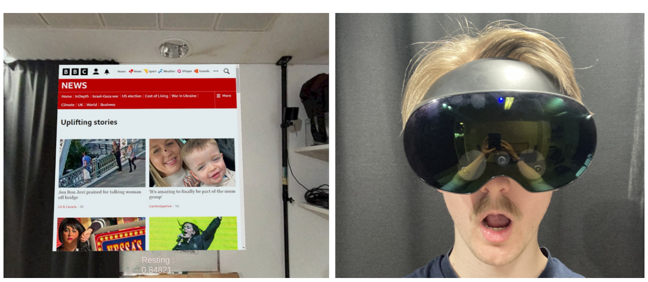

Figure 6: *Left*: Screenshot of the Task 2 AR Web Browser, which featured a large virtual browser window. This was presented on a large 100*100m canvas placed approximately 80m away from the participant, creating the sense of a large, exocentric world-locked display with an effective viewing size comparable to a large 30inch monitor. *Right:* User wearing the Meta Quest Pro headset and performing a *JawDrop* interaction, which for the browser enacted a selection on the current URL being targeted.

5.2.2 Facial Input Control Scheme. Based on our candidate FAUs, we devised the following control scheme:

Select Activated on *JawDrop* as a discrete input.

Scroll Up/Down Activated on *Mouth - Left/Right* and *CheekPuff - Left/Right* as a continuous input / continuous scrolling. Similarly to the Locomotion task we implemented basic inertial scrolling.

Back Activated on *LidTightener - Left* as a discrete action.

To adapt these facial inputs to interacting with an interface, we added a general time delay of 1.5 seconds across browser interactions to prevent accidental successive inputs (e.g. where *JawDrop* was held for too long and would click a link on the next rendered page accidentally). Targeting was again enacted through head orientation, and we implemented a *baseline control scheme*, where the VR controller thumbstick scrolled up/down and enacted back transitions, and triggers were used to select links based on head orientation targeting.

6 Study 2: Facial Input for Locomotion and Interaction in a Virtual Environment

Given our demonstrators, we elected to perform a small-scale formative evaluation to examine whether our candidate FAUs ultimately provided a viable means of interacting with an XR device, and did so in a way that was generalisable i.e. a user being able to interact with a spatial interface or menu, launch a game that necessitated moving around a 3D environment, and then play said game - all using facial input alone.

6.1 Design

Consequently, we had two factors, **Input** and **Task** resulting in four conditions: {*Facial, Controller*} * {*VR Game, AR Web Browser*}. Note that to ensure parity between the control schemes whilst keeping the controls simple and memorable, no other movement functions were provided in the VR game (e.g. no backwards movement, strafing, sprinting etc.). We examined these factors within-subjects.

6.2 Measurements

We provided participants with a questionnaire after each condition prioritising measures that were applicable across Task types, comprising of participants' (1) **perceived workload**, using the 6-item NASA TLX [34], (2) **perceived exertion**, using the 1-item Borg Rating of Perceived Exertion [15], (3) and **simulator sickness**, using the 1-item MISC MIsery Scale [17]. Additionally, we collected **usability insights** using five 1-item 7-point Likert-Type questions covering *ease of use*, *responsiveness*, *reliability*, *error proneness*, and *general usability*, ranging from Strongly Disagree to Strongly Agree, informed by [49]. Finally, we asked for the participants' **social comfort**, using three 1-item 7-point questions asking if participants would be comfortable performing these interactions in 1) public settings (e.g. in a cafe); 2) semi-private settings (e.g. in a hospital or train); 3) private settings (e.g. alone in your home), ranging from Strongly Disagree to Strongly Agree, informed by [47, 54], and their

willingness to use FAUs as input, e.g., *"If there were no alternative controls available (e.g. if you were unable to use your hands), would you use these controls for this kind of activity?"*, ranging from Strongly Disagree to Strongly Agree.

At the end of the study, we then asked participants to rank the FAUs in order of preference of usage for *discrete* (e.g. pressing a button, shooting, selecting a website link), *short continuous* (e.g. quickly turning or moving forward in between other actions in the game), and *long continuous* (e.g. scrolling down a long web page, moving forward for a prolonged period) interactions. We then followed up with interview questions:

- Did you like or dislike facial input, and why?
- Based on your experience of moving and interacting in VR, what aspects of facial input were beneficial to, or superior over, controller input if any? And what was worse compared to the controller?
- Based on your experience of browsing the web in AR, what aspects of facial input were beneficial to, or superior over, controller input if any? And what was worse compared to the controller?
- Would you change anything about how the facial inputs mapped to the different controls in the game or browser, and if so why? Did particular mappings feel right or wrong to you?
- If you could use facial inputs anywhere, anytime and have them detected, where could you imagine using facial input in the future in your everyday life?

6.3 Procedure

This study was approved by our college ethics committee. Participants were first briefed through a plain language statement, and informed consent sought to take part. Then demographics data was captured about their age, gender and existing VR usage.

Participants were first instructed on how to perform each of the facial inputs outside of VR, through demonstration by the experimenter. Participants were then fitted with a Meta Quest Pro headset, absent side light blockers to ensure the face was well lit, and given training on each of the Tasks controls as they encountered them. The *VR Game* always came first for all participants, as this also featured an in-headset training step where participants could enact FAUs and get real-time feedback of their activation. This was also used to filter out any participants for whom FAU did not sufficiently function (see *Participants*). The order of Facial and Controller input was counter-balanced for *VR Game*, with the opposite order used for the *AR Web Browser*.

For the *VR Game*, once training was complete participants completed between 1-2 trials, based on a minimum of 5 minutes of time interacting. For the *AR Web Browser*, participants were instructed to browse the aforementioned BBC site for 4 minutes, and to visit and scroll through at least 3 articles in that time. After each permutation of *Input*Task* participants filled out a post-Condition questionnaire based on the measures mentioned previously, and ranked the FAU inputs and answered interview questions at the close of the study.

6.4 Participant Demographics

Study 2 took place approximately 3 months after Study 1. 13 participants were recruited, including 3 who had previously taken part in Study 1, however 3 dropped out, for reasons of being unable to activate one or more of the FAUs (P3 - *LidTightener* inputs would not work; P4 - only *JawDrop* worked, potentially because they had a more slender face with less pronounced cheeks) or because of experiencing cybersickness during the *VR Game* (P7, VR Game, Controller condition). This meant that 10 participants ultimately completed the study (3 female, 6 male, 1 non-binary) aged 20 to 29 (mean=24 years, sd=2.7 years). With respect to **VR usage**, 2 participants had never used VR before, 3 "several times a year", 2 "several times a month", 2 "several times a week" and 1 "daily".

7 Results

7.1 Analysis

For statistical significance testing between conditions (e.g. for questionnaire data) an Aligned-Rank Transform (ART) [27] was used to transform non-parametric data prior to conducting one or two-way repeated measures ANOVAs, using the *ARTool* R package [42]. For effect size, partial eta squared η_p^2 is reported [58, 72]. Pairwise contrasts for main effects were also conducted [43]. For boxplots, lower and upper hinges indicate the 25th and 75th percentiles, with notches extending 1.58 * IQR / sqrt(n) as an approximate 95% confidence interval. For qualitative data, a single coder performed inductive coding over two passes, establishing an initial codebook on the first pass and then refining into themes on the second pass.

7.2 Quantitative Data

Whilst our focus for this study was on qualitative insights, we first report our descriptive statistics to support interpretation. See Table 3 for an overview of significance testing results.

7.2.1 Workload and Physical Impact (Figure 7). If we consider Workload firstly, we can see that there were significant effects for both *Condition* and *Task* - with the *VR Game* being significantly harder than the *AR Browser*, and *Facial Input* being harder than *Controller Input*. This is not unsurprising, but worth noting is the generally low workload of the Facial Input Browser. And if we examine the physical impact on exertion, we can see similar findings. For nausea/sickness, there were no significant effects.

7.2.2 Usability (Figure 7). Generally facial input was seen as relatively easy to use, usable and responsive - albeit consistently less so than the controller input. Notably however, facial input was perceived as much more error prone and unreliable, with significant effects of *Condition* on all measures.

7.2.3 Social Comfort and Willingness to Use (Figure 8). As could be expected, *Condition* had a significant effect on perceived social comfort for public and semi-private spaces, but not for private spaces, reflecting obvious concerns about large facial gestures. There was also an effect of *Task* for public settings, but this is likely an artefact of the fact the VR game featured no reality awareness mechanism.

For willingness to use, participants were broadly positive towards facial input, with the majority of the responses around agree to slightly agree.

7.2.4 Rankings (Figure 8). Finally, we asked participants to consider the appropriateness of the selected FAUs for different generalised types of Interaction, to see if there were any emergent preferences

Measure	Main Effects Factor	DoF	*F*	*p*	f^2	Mean (SD)	
				Workload & Physical Impact			
TLX - Workload	*Input*	1, 28	19.34	<0.01	0.41		
	Task	1, 28	5.66	0.02	0.17	**Cont-Brow:** 6.92 (8.78)	**Cont-Game:** 18.41 (12.14)
	*Input*Task*	1, 28	0.2	0.66	0.01	**Face-Brow:** 23.17 (12.39)	**Face-Game:** 31.36 (18.89)
MISC Misery Scale	*Input*	1, 28	0.49	0.49	0.02		
	Task	1, 28	1.27	0.27	0.04	**Cont-Brow:** 0.6 (1.26)	**Cont-Game:** 1.82 (2.64)
	*Input*Task*	1, 28	0.19	0.66	0.01	**Face-Brow:** 0.7 (0.95)	**Face-Game:** 1.18 (1.72)
Borg RPE - Exertion	*Input*	1, 28	8.86	0.01	0.24		
	Task	1, 28	5.99	0.02	0.18	**Cont-Brow:** 0.5 (0.94)	**Cont-Game:** 1.41 (1.59)
	*Input*Task*	1, 28	0.11	0.74	0	**Face-Brow:** 1.45 (1.28)	**Face-Game:** 1.86 (1.4)
				Usability			
Ease of Use	*Input*	1, 28	32	<0.01	0.53		
	Task	1, 28	3.06	0.09	0.1	**Cont-Brow:** 2.9 (0.32)	**Cont-Game:** 1.91 (1.45)
	*Input*Task*	1, 28	0.16	0.69	0.01	**Face-Brow:** 1.1 (1.73)	**Face-Game:** 0.91 (1.7)
Responsive	*Input*	1, 28	24.36	<0.01	0.47		
	Task	1, 28	1.83	0.19	0.06	**Cont-Brow:** 2.7 (0.67)	**Cont-Game:** 1.73 (1.19)
	*Input*Task*	1, 28	1.22	0.28	0.04	**Face-Brow:** 0.8 (1.87)	**Face-Game:** 0.55 (1.75)
Reliable	*Input*	1, 28	48.37	<0.01	0.63		
	Task	1, 28	0.38	0.54	0.01	**Cont-Brow:** 2.7 (0.67)	**Cont-Game:** 2.18 (0.98)
	*Input*Task*	1, 28	1.03	0.32	0.04	**Face-Brow:** -0.1 (1.6)	**Face-Game:** 0.36 (1.69)
Error Prone	*Input*	1, 28	32.86	<0.01	0.54		
	Task	1, 28	1.06	0.31	0.04	**Cont-Brow:** -2.4 (0.97)	**Cont-Game:** -1.09 (1.7)
	*Input*Task*	1, 28	3.21	0.08	0.1	**Face-Brow:** 0.8 (1.93)	**Face-Game:** 0.64 (1.36)
Usable	*Input*	1, 28	13.38	<0.01	0.32		
	Task	1, 28	1.01	0.32	0.03	**Cont-Brow:** 2.6 (0.52)	**Cont-Game:** 2 (1.1)
	*Input*Task*	1, 28	0.27	0.61	0.01	**Face-Brow:** 1.3 (1.7)	**Face-Game:** 1.27 (1.42)
				Social Acceptability			
Social Comfort - Public	*Input*	1, 28	52.08	<0.01	0.65		
	Task	1, 28	6.54	0.02	0.19	**Cont-Brow:** 1.9 (1.85)	**Cont-Game:** 0.55 (1.86)
	*Input*Task*	1, 28	3.67	0.07	0.12	**Face-Brow:** -1.8 (1.69)	**Face-Game:** -2.18 (1.25)
Social Comfort - Semi-Private	*Input*	1, 28	54.85	<0.01	0.66		
	Task	1, 28	1.32	0.26	0.05	**Cont-Brow:** 2 (1.83)	**Cont-Game:** 0.91 (1.81)
	*Input*Task*	1, 28	2.19	0.15	0.07	**Face-Brow:** -1.1 (1.66)	**Face-Game:** -1 (1.67)
Social Comfort - Private	*Input*	1, 28	1.88	0.18	0.06		
	Task	1, 28	0.49	0.49	0.02	**Cont-Brow:** 2.6 (0.52)	**Cont-Game:** 2.27 (1.01)
	*Input*Task*	1, 28	0.25	0.62	0.01	**Face-Brow:** 1.9 (1.85)	**Face-Game:** 2 (0.89)

Table 3: Breakdown of statistical testing by measure, including *post hoc* tests. Cells in green highlight rows with a significant main effect. For means and standard deviations, shorthand labels denote Cont: Controller Input; Face: Facial Input; Brow: AR Web Browser; and Game: VR Game.

based on their experience. Interestingly, there was a strong preference for *LidTightener* for short discrete interactions (6/10 users), whilst preferences moved towards *CheekPuff* and *JawDrop* for long continuous interactions, and *Mouth* was largely not preferred.

7.3 Qualitative Data - Interviews

7.3.1 Sentiment Towards Facial Input.

Hands-Free Interactions and Usability. 7/10 participants noted liking Facial Inputs (FI) for a variety of reasons. Foremost of these were discussions around the perceived usability of FI, and in particular the benefits of a hands-free interaction modality. 6/10 participants noted some aspect regarding usability, with FI perceived as being "*meaningful*" (P0), "*fun*" (P6/P9), "*surprisingly smooth*" (P8), "*not... that physically demanding*" and "*responsive*" (P1), "*convenient*" (P3), easy to use and "*very fluid*" (P9). 6/10 participants reflected on the benefits in terms of freeing up hands for other tasks, e.g.:

> P0: "*it was nice to not have to use the controller. It was nice to be able to read things and then kind of just move my face as I'm reading*";
> P3: "*would be great in situations where I can't use my hands... I actually have freedom with my hands. I can*

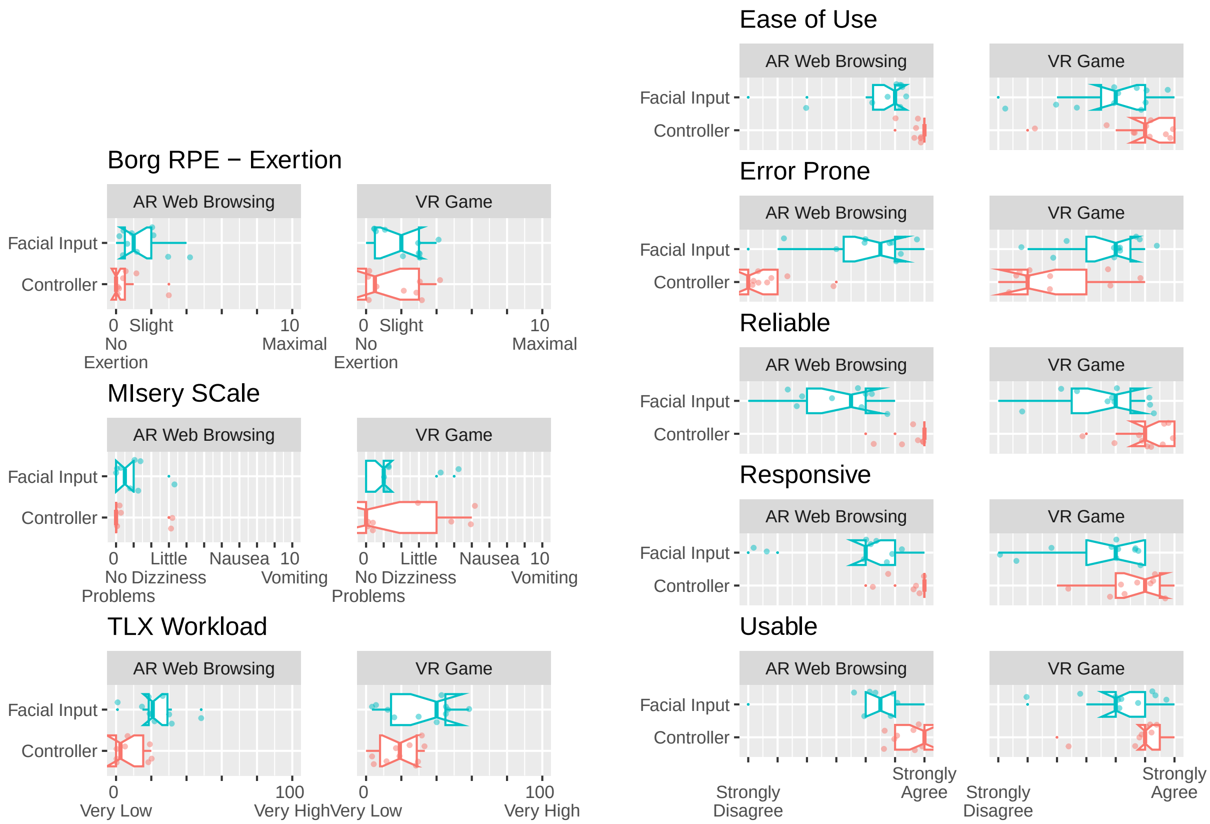

Figure 7: *Left*: Boxplot of per-condition questionnaire for exertion, sickness and workload - higher is worse. *Right*: Boxplot of per-condition questionnaire for ease of use, errors, reliability, responsiveness and general usability. Higher is better for all measures except error prone.

> *do whatever, juggle balls!*".
> P12: "*If I was navigating a web page while walking with AR glasses and I'd be able to navigate it as my face. I think that's something I could see myself using it in... If I didn't want to be waving my hands about to do gesture-based, hand-based interaction with it in public when walking down the streets, I could do something more subtle.*"

Indeed, P3 elaborated that, for example, when cooking a meal and unable to wash their hands, FI would offer a valuable alternate input modality.

Multi-Tasking and Mobile Interaction. Similarly, FI was also noted to be potentially beneficial whilst mobile or multi-tasking, e.g.:

> P6: "*I would only probably use it walking about if I was, like, desperately needed something to know and be like, like, pull this up*";
> P8: "*Definitely for multitasking if you're busy using your hands. For example, I do art on an iPad... but I can use my eyebrow to, like, skip a song*".
> P11: "*I would use it if I'm multitasking*"

Reliability and Fatigue. But FI was not without difficulties, with 3/10 participants noting they expressly did not like FI, e.g. P10: "*It made me feel uncomfortable*". 4/10 participants remarked on their struggles with **reliability**:

> P3: "*when using the eyes to try and shoot, I somewhat struggled to get this shot actually going in. So that wasn't necessarily very reliable*".
> P12: "*I couldn't do the left eyelid properly. I could do everything else, but that was the one that I struggled with. I worked most of the time, but just not consistently.*"

Fatigue and effort were commonly noted drawbacks for 6/10 participants, particularly for prolonged usage:

> P9: "*it's quite hard, especially if I'm just trying to scroll for ages, just keeping one cheek puffed. I felt like I was running out of air, and I was kind of, like, gasping in between.*";
> P5: "*It was still like a bit tiring holding your face in certain ways";*
> *P6: "the only thing I disliked is how sore my face is now*".
> P11: "*I think long term, yeah, you don't want to stay on a continuous facial expression. It's tiring.*"

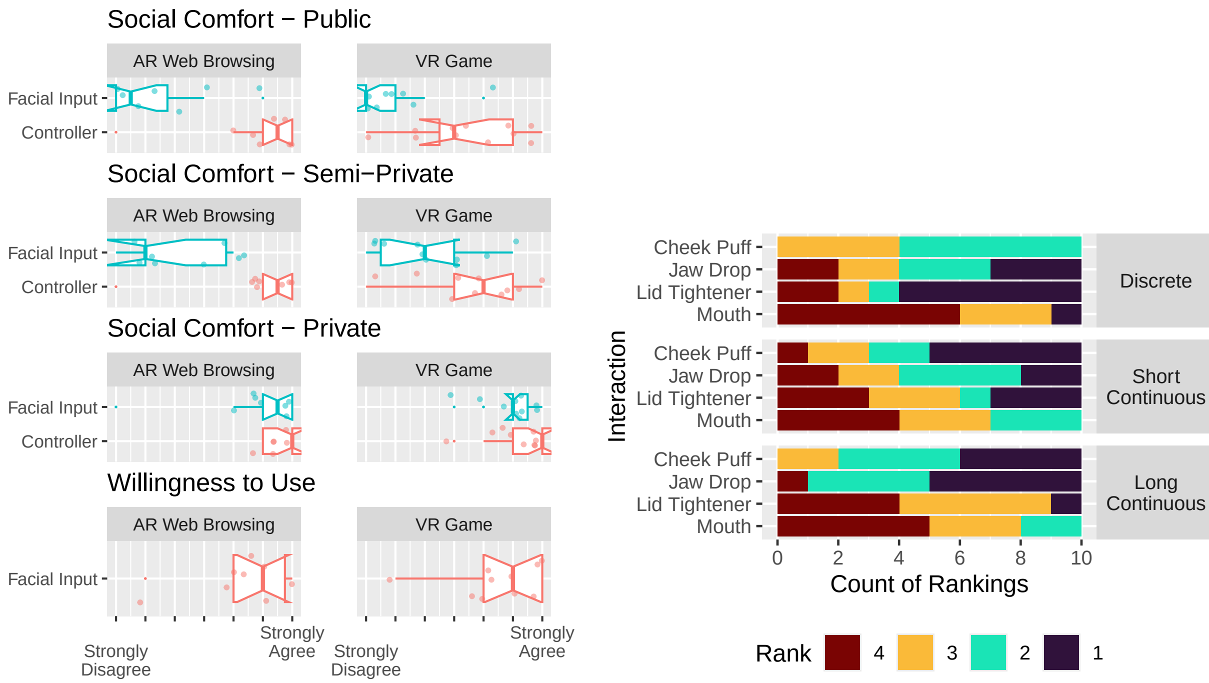

Figure 8: *Left*: Boxplot of per-condition questionnaire for social comfort. Higher is better for all measures except error prone. *Right*: Stacked barchart of preference rankings across *FAU* and *Interaction*, lower rank is better.

In contrast, the controller was perceived as more reactive or easier (3/10 participants) and supported greater input flexibility through concurrent inputs (2/10 participants)

7.3.2 Facial Input Challenges.

Differing Abilities. 5/10 participants reflected on the differing abilities of users to be able to enact certain FI poses, for example in moving eyebrows independently: "*Some people are very good at moving their eyebrows. They can make, like, McDonald's signs even with it. But I struggle with that*" (P0); or being unable to move upper regions of the face (P3) or puff cheeks sufficiently (P6/P9).

False Positive Inputs. Two participants (P3, P12) noted potential frustration with false positive activations, particularly when considering *JawDrop* and talking to others in the background:

> P3: "*It might be inappropriate in a situation where I'm playing the game [and my] mouth is open and someone's talking with me and I'm talking with them*".

Direction and (A)Symmetry. 2 participants noted some confusion around how the directionality of the FI corresponded to the interaction technique. For example, when reflecting on the Browser task:

> P10: "*To scroll up and down you had to move your left or right cheek. But to select something you had to move your jaw. For me, moving the jaw is a downward task rather than a selection task.*"
> P0: "*Forward and back's a funny one, because you kind of have to... I think forward and back shouldn't be a left-right thing*"

P0 suggested that for Left/Right inputs such as turning in-game, their preference was for "*anything you can replicate on the left and right. So, cheek puffs, eyebrows*". Asymmetry was also raised by the same participants, in the sense of being confused when having functionally different inputs mapped to left/right FAUs e.g. shooting and selection in the Game task, or back in the Browsing task.

Need for Customisation. 5/10 participants remarked specifically on the need for customising the FI control scheme per-application, partly to address differing abilities as previously noted, but also simply preferences such as around directionality and asymmetry. Notably, all participants exhibited some form of indirect feedback around subtle changes to mappings or the responsiveness of FI detection that provoked a need for personalization/customization.

Concurrent Inputs. In contrast to the controller, 2 participants questioned whether they could adequately enact concurrent controls with the default control scheme we provided for the Game task, with *CheekPuff* and *JawDrop* being perceived as being difficult to enact together:

> P3: "*Sometimes my plan was to have mouth open and then cheek puff, so move and at the same time turn. But I don't know if that was an actual option available. But that was one thing I had tried to do*".

Social Acceptability and Subtle Interactions. 5/10 participants noted some form of concern around social acceptability or the need for more subtle interactions or less exaggerated movements:

> P9: *"If you're just sitting there with your mouth open, every few seconds you do look a bit strange... I certainly wouldn't want to, like, be making weird mouth movements in the middle of a cafe."*;
> P3: "*If I'm in a public setting, I don't want to be opening my mouth constantly like that. It's kind of weird. I'd get weird stares*";
> P5: "*You could even do something more subtle for, like, selection, like biting your lip or something*".

7.3.3 Envisioned Use Cases.

Casual Interaction Benefits for Spatial Interfaces. The Browser demo exemplified potential benefits for everyday interactions with spatial interfaces for 6/10 participants:

> P0: "*For scrolling through the news, doom scrolling*";
> P1: "*Especially for like more mundane tasks like scrolling or just like more 2D tasks of selecting stuff*";
> P3: "*Basically any scenario where I just cannot use my hands"; P6: "I found it extremely easy to use and it picked up a bit better as well with scrolling so yeah I thought it was quite good to use*".
> P12: "*I felt quite lazy almost rather than using a controller to navigate with it. The mouth actions... those worked really well, so I was quite happy to do that. But it wasn't a preference to use that over a controller, it was just an alternative, but that was more of an equal footing.*"

A Viable, But Non-Preferable, Primary and Alternate Modality for Games. Whilst it is clear from the feedback that Controller input was, unsurprisingly, more usable and performant, the FI mapping used here was seemingly viable, and even enjoyable for 3/10 participants:

> P9: "*So for the game one, I quite liked it because, I mean, I don't do VR often, but it was quite fun navigating like that and it felt very fluid*";
> P6: "*I thought it was a much more fun way to control a game... it controlled really well which i wasn't expecting, I was surprised that [it] went so smooth[ly]. I liked it for how uh easy it was and yeah kind of fun [using] your face instead of a controller*".

Indeed, P8 raised the idea that they might use this as an additional supplementary input e.g. for enacting shooting despite having a controller: "*I am terrible at aiming but if I could just like look at a target and aim that would be a lot smoother*". And P12 noted that if they "*only had access to one or two buttons on a controller, a third button as a facial action might be quite useful*" e.g. to bring up a map in-game.

Interestingly, there were also certain input mappings that were seen not just as usable, but as *appropriate* or in some way enhancing the interaction in the game. Using *LidTightner*, P11 noted for the browser that "*I liked going back. It's quick with the eyes*", whilst shooting in the Game task was a common point noted by 4/10 participants:

> P10: "*I felt like when you shoot an object, when you use your facial expressions, they were a bit more like you feel the effect more because you feel like oh, I just shot this at the target"; P1: "Just aiming at the target and shooting at them I think it's fast cool I think it's fast I think it's cool it's fun*";
> P6: "*[For] turning left and right... the interaction was completely fine, it was fantastic and so was the shooting and the moving forward*".

But for two participants the controls were sufficiently problematic that they would not use these for games:

> P1: "*Of I'm playing this for 20 minutes eventually I'll just be too tired and fatigued. Also no way for competitive gaming to be feasible, especially if you're up against controller players or people with mouse and keyboard*".

Accessibility. Finally, 6/10 participants noted some perceived benefit for accessibility, suggesting that FI was seen as sufficiently viable to enable the interactions we prototyped:

> P10: "*I understand why it's useful. It's more, you know, for accessibility reasons*";
> P3: "*Imagine I'm a patient on the bed and I can't even move. Being able to have my XR or whatever kind of screen and I'm lying down, that would be very useful and especially [if] I can't even use my hands*";
> P5: "*I think, again, my brain just goes to... if you've got some sort of disability or something that... would be perfect*".
> P11: "*I'm thinking of my cousin who can't use her arms. So definitely useful for [her].*"

8 Discussion

Across two studies, we have explored the application and utility of Facial Action Units as an alternate input modality for XR devices. In Study 1, we identified the most usable FAUs that can be sensed by a current cutting-edge consumer XR device (Meta Quest Pro), with LidTightener and JawDrop representing the two 1-ranked FAUs. We then trained a basic classifier for seven of the most usable FAUs from Study 1, achieving an overall F1 score of 0.95. Then, using this classifier, for Study 2 we built and evaluated two demonstrator XR apps that could be used entirely with only facial input: a *VR game*, and an *AR web browser*. In a small scale evaluation, we validated that these apps were generally usable and did not overly exert users, and that users saw value in a range of use cases based on their experience, from accessibility to everyday hands-free interaction – but that the reliability and responsiveness of facial input remains to be improved, and for the VR game in particular, the controller unsurprisingly remained the most performant option.

8.1 Facial Input as an Alternate Input Modality for XR

We demonstrated the feasibility of having an XR platform where facial input could be offered as a generalized alternate input modality

– enabling the user to interact with spatial interfaces and immersive games without additional devices or body movements. There has been a rich history of examining facial input in HCI, and we are not alone in exploring facial controls for games or UIs before. However, our work is novel in that it focused on the end-to-end challenge of identifying suitable FAUs for interaction, classifying their activation, and then demonstrating their effectiveness in interacting with spatial interfaces and controlling locomotion, all on a consumer off-the-shelf XR headset. What this ultimately demonstrates is that facial input offers a viable alternate input modality, that offers benefits not just in terms of accessibility (as previously validated by Taheri [84, 86, 87] in particular), but also has applications for hands-free input for spatial and immersive computing across a range of situational and temporary impairments.

XR accessibility has remained a sore point, a result of a focus on embodied interaction and assumptions around a 'corporeal standard' that has been repeatedly challenged [32] and yet has seen slow progress in leading platforms. Take the *Meta Quest* range for example, where accessibility is largely addressed in terms of visual impairments but not interaction inequality[9]. Yet these devices continue to add sensing that is entirely capable of facilitating generalized, accessible interactions. Any UI that can be interacted with through a controller-based raycast could similarly be interacted with through an eye/head raycast and facial input to trigger selection. Recently released devices exemplify these possibilities, with the Apple Vision Pro for example using eye-tracking as the primary targeting modality, whilst research has repeatedly extolled the benefits of gaze-based interaction [66, 68].

Similarly, any roomscale VR experience that requires of the user that they can freely navigate, shoot, and interact can trivially be mapped to usable FAU controls. Whilst this description skips over some of the complexities of a generalized input (e.g., games designed for a wide variety of controls, or with a reliance on hand-tracked interactions), there is nonetheless an opportunity to provide a facial controller that emulates many of the features of handheld controllers, and could provide a new level of access and control to users with a breadth of motor impairments in particular for whom XR usage has largely been denied.

However, there are widely-held concerns about the social acceptability of performing FAUs in any non-private setting, which may hamper the utility a user can get from it, as well as hamper the wider adoption, at least in the short-term while familiarity grows. Per-user calibration and choice of FAUs may allow for the use of discreet, less visible inputs for public settings, or potentially some form of mode-switching, where one set of FAUs is preferred when private (e.g., full set, large activation range) and another when in public (e.g., smaller set, smaller activation range, more eye-based FAUs etc).

8.2 Interaction Design Recommendations

Our limitations (discussed below) should be considered carefully when interpreting and applying our formative findings. However, it is clear that LidTightenerLR and JawDrop were both considered widely usable and low fatigue by participants and could be detected by the Quest Pro reliably, whilst there was a range of secondary FAUs (CheekPuffLR, MouthLR, UpperLipRaiserLR, LipSuckLT/RT, JawThrust, LipSuckLB/RB, DimplerLR) that had slightly degraded perceived performance (more effort, less responsive, but generally still comfortable and with good perceived performance) and could be considered as candidates for interaction. When they worked reliably as evidenced in Study 2, JawDrop, LidTightenerLR, CheekPuffLR, and MouthLR provided a reasonable input vocabulary.

A user with good lip muscle dexterity might find value in some of the more varied Lip (e.g. LipSuckLR variants across top and bottom lips) and Jaw (e.g. JawThrust, JawSidewaysLR) inputs available. However, based on our Study 1 findings, Lip-based inputs appear to have some notable challenges in false positive co-activations/co-occurrence which, given the number of FAUs available here, may limit their practical utility, or at least restrict users to employing only a subset. Similarly, some of the more varied non-lip inputs utilizing Cheek and Brow movements had less discernible activation and degraded perceived performance and/or discomfort, and thus might be less generally useful for facial input.

Regarding input mappings, JawDrop was useful both as an asymetrical input (meaning there was no potential interpretation of directionality around its use) and easy to hold in the "activated" state with low fatigue. This makes it an excellent candidate both for discrete and continuous 1D input. Of the remaining candidate FAUs we tested in Study 2 in particular, they all had symmetrical pairs and implied directionality (left/right). For the AR Browser, this introduced challenges around interpretation and mapping (e.g. should a left activation scroll up or down?). And intriguingly the LidTightener inputs seemed particularly appropriate for targeting actions/selection given their implicit association with gaze. And across the symmetrical FAUs in Study 2, prolonged continuous usage was a problem for a number of participants, suggesting that future research (discussed in more detail later in this section) needs to consider how we can lessen the need to continuously hold a facial input e.g. by using twitch-based toggles or short bursts of interaction (e.g. a CheekPuff twitch instigating a yaw snap turn rather than a continuous turn).

In summary, absent further evidence, for now we would recommend that:

(1) *JawDrop* is used for any discrete binary or continuous 1D input.
(2) *LidTightenerLR* is used for any binary selection or action that has association with gaze or head orientation-based targeting.
(3) *CheekPuffLR* and *MouthLR* are used for short continuous interactions, but that if interaction is going to be for a prolonged period their use should be minimized through toggles, bursts etc.
(4) Where a symmetrical left/right FAU is mapped to a symmetrical input on another dimension (e.g. scrolling up/down in a browser) this mapping should be user customisable.

These recommendations are formative however, and we would expect future work to further refine them (e.g. trialling other rank 2 FAUs from study 1, examining extent of personalization required, evaluating individual and population differences regarding what

[9] https://www.meta.com/en-gb/help/quest/articles/in-vr-experiences/oculus-features/accessibility-features/, last accessed August 2024

constitutes a viable, comfortable facial input etc.) as we move towards advocating for XR facial input as a standard accessibility option. Moreover, if applying our findings to a specific user population (e.g. users with chronic fatigue), we would recommend returning to the Study 1 dataset in particular to reconsider our scoring metric (e.g. upweighting fatigue) and consider replicating Study 1 to benchmark with said population specifically.

8.3 Limitations

8.3.1 Headset Sensing Constraints. Whilst we used the Meta Quest Pro for these studies, there are other headsets on the market with FAU tracking. The HTC Vive pro has a facial tracking device, albeit an attachable peripheral and only capable of tracking 38 blend-face shapes on the exposed lower portion of the face [7], whilst Pico headsets now feature comparable FAU tracking. Crucially, the models underpinning FAU detection are constantly being updated on these platforms, for example the Quest Pro added tongue tracking [10] after we had conducted this study. This also means that the findings of this study need to be carefully considered relative to potential advancements in the recognition models being used on these devices - an FAU which featured a low and difficult-to-discern activation state may for example have improved performance in later iterations. However, our subjective assessment of performing FAUs would remain valid. Future work should also consider replicating our methodology to assess additional FAUs[10] and indeed composite FAUs/expressions, to further expand the available facial input vocabulary. Nonetheless, the Meta Quest Pro arguably covers the majority of FAUs that could be feasibly sensed from a VR-integrated head-facing camera.

8.3.2 FAU Input Space and Use of Singular FAUs versus Expressions. Our novel focus was on benchmarking and exploring the use of singular FAUs for interaction. It would not have been feasible to examine combinations of the 53 FAUs without a significantly expanded study, but future work would be assisted here by our dataset, for example considering which FAUs *did not* co-occur with a given target FAU, as potential FAUs that might be selectively combined by users.

Similarly, we did not look at typical gestalt expressions of e.g. happiness, sadness, surprise etc. These have been previously explored in research (e.g. [85]), but more crucially by starting with multi-FAU expressions this would have greatly limited our understanding of the potential available input space. And, by utilising individual FAUs, it leaves gestalt/holistic expressions unused as dedicated inputs, meaning that spontaneous emotional expression is less likely to be inadvertently taken as input (or can be used solely by the headset for e.g. detecting and conveying social signals in VR). We envision that future work will need to consider how to discern between targeted FAU interaction and false-positive activation through expression where emotive expressions are likely to occur during interaction (e.g. when playing an immersive game).

8.3.3 Sample Size and Population. Our sample size was smaller in Study 2 (N=13, 3 dropouts) compared to Study 1 (N=20), however the focus for Study 2 was on qualitative experience of FAU interaction, and we nonetheless reached a point of saturation where interviews were repeating similar obvious, notable challenges and benefits. Whilst this means that, in broad strokes, we know the challenges posed by using FAUs for input, further study will be required (more participants drawn from different populations using different tasks/applications, in different contexts such as mobile, standing, seated etc.) to develop a broader and more nuanced picture of how FAU interaction might be integrated into everyday life and computing more generally.

There is also the consideration of whether populations that experience some form of impairment (situational, temporary, permanent) might exhibit different results, as we evaluated with people without disclosed disabilities or reported impairments when using XR. We can speculate that a person that is quadriplegic might see significant merit in FAU interaction, mirroring the work of Taheri *et al.* [84–86], whereas a person with muscular dystrophy might nonetheless have difficulty with certain FAUs or prolonged FAU activation if using impacted muscle groups. Realistically, developing such a nuanced understanding requires extensive further research in considering potential populations that could benefit from FAU use, and benchmarking with them, replicating the methodology we have laid out.

8.3.4 Duration of FAU Testing. We did not conduct a controlled study of prolonged FAU activation. Performing the same action repeatedly would inevitably lead to additional discomfort and fatigue; this is an inherent aspect of all movement-based input techniques. It would not have been practical (in terms of experimental time) or ethical for us to ask users to perform all FAUs repeatedly to a level that caused fatigue or failure. However, the comfort and effort ratings of FAUs we captured could be used to extrapolate which would also be more uncomfortable/fatiguing to perform repeatedly, and future work should consider how to replicate our benchmarking to safely examine repeated and prolonged use of singular FAUs.

8.3.5 Left/Right Confusion. In some symmetrical FAU pairs in Study 1, it was evident that there was some unexpected activation co-occurrence - e.g. CheekPuffL/R. Our difficulty here is that we cannot discern whether this is a result of mirror confusion or participant error, a valid co-occurrence (e.g. when activating the left FAU also to a degree activated the right FAU), or indeed an artifact of the black box Meta FAU classifier that underlies the API we relied on (e.g. interpreting some movement of the FAU as an errant activation value increase). However, regarding methodological considerations, we recommend that any future replications of Study 1 consider introducing real-time validation that the correct FAU is being targeted by the participants. Whilst we deliberately did not provide activation feedback so as not to bias perceptions, in retrospect some binarized feedback could warn if a participant was potentially activating the "wrong" side for a given trial.

8.4 Refinements and Future Research

Some of the design decisions we made were in service of demonstrating the end-to-end potential of providing a generalized facial input for XR. Consequently, there is significant scope for refinement of our approach in-practice.

[10]The full FACS set has approximately 90 codes, see https://www.paulekman.com/facial-action-coding-system/

8.4.1 Additional Sensed Inputs. If we consider Study 1, we examined 53 FAUs based on what the Meta Quest Pro can currently sense. But these excluded a set of inputs for time that merit further investigation, i.e., eye-based gestures, and tongue tracking, both also now supported by the Quest Pro. Furthermore, if we were to desire more FAU inputs (e.g. for more complex interactions that get closer to completely mapping every available controller button along), then we would need to re-examine our dataset, in particular establishing the extent to which FAUs destructively interfere with each other, or indeed lead to concurrent false positive activations – an avenue of future research that would be accelerated by our dataset and analysis.

And our paper exclusively focused on facial input to see how far we could get with just this modality alone. But in practice we would expect interaction designers to exploit the full range of sensed inputs available on COTS XR hardware – from speech, to head and eye tracking, to facial input and more – to give users a range of options to express their intents fully. This could be particularly important to support more complex, and more varied control schemes that facilitate greater degrees of movement and interaction (e.g. walk, run, sprint; strafe left/right; move backwards; crouch).

8.4.2 Customization. We examined only one mapping of usable FAUs to available control inputs - but our qualitative feedback was clear that users would desire more customization/personalisation of controls. We could imagine offering users the full range of FAU inputs to choose from, or indeed the range of level 1/2 ranked FAUs identified from Study 1, with guidance around their fatigue, usability, and potential destructive interference. In doing do, users could make informed decisions, trial different bindings, and ultimately select FAUs most appropriate to their capabilities, facial morphologies, and appearance (e.g. facial hair).

8.4.3 More Training Data. Regarding our classifier, 20 participants is obviously an insufficient dataset to robustly generalise across a population. The fact we had 2 drop-outs related to FAU use in Study 2 emphasizes that more extensive data capture is needed here - with a more controlled capture across different facial morphologies advisable. This is an endeavour that XR platforms, rather than academia, would be ideally suited to, given the resources they have available.

8.4.4 Classifier vs Thresholding. Moreover, our classifier is an unusual case, in that we are effectively classifying FAU activation by the expressions associated with maximally activating an FAU, based (presumably) on inputs provided by another black-box classifier model running on Meta devices processing the lower face and eye tracking imagery. There is a debate as to whether more simple heuristics such as thresholding may be good enough here. Simpler approaches provide a promising avenue for others to further explore this interaction modality without the need for complex recognition models - though these may indeed be more relevant for commercial everyday use of facial input. Moreover, with a simple classifier we have already demonstrated numerous benefits, and challenges, with moving towards generalised facial input for XR.

8.4.5 Preventing False Positives. Our approach does in theory, however, lend itself to creating inputs potentially robust to false positive activation. Consider that a generalised, viable solution would need to prove itself robust to false positives as a result of a) affective expressions (e.g. smiling, laughing, disgust etc), and b) speech (resulting in jaw, cheek and mouth movements in particular). For the former, we might further classify individual affective expressions precisely to avoid false positive inputs when these expressions are activated, rather than as they have previously been used as input. And for the latter, we might examine what FAUs are activated given specific vocalizations as detected by the in-built microphone, and again attempt to negate false positives. Such advances would be necessary to enable *social* XR experiences with voice communication, or indeed to support users to fully engage with immersive affective experiences.

8.4.6 Fatigue and Health Impact Over Time. Our examination of facial input was largely restricted to short periods of interaction. Prolonged usage of an XR device in-the-wild, and prolonged *interaction* with said device, could pose unique problems to the usability of specific FAUs, which could have unintended consequences. Consider for example the use of *JawDrop* for a prolonged period of time resulting in *Temporomandibular Joint Dysfunctions* (TMJ) such as a locked jaw. Or prolonged use of *CheekPuffLR* or *MouthLR* resulting in fatigue that impacts speech. In proposing new input modalities (or rather new applications that could make extensive use of them), we also risk exposing users to increased health risks, much as the use of keyboards and mice resulted in an uptick of carpal tunnel syndrome [88], and usage of single-display monitors resulted in more musculoskeletal disorders of the neck and spine [53].

8.4.7 Concurrent Input. One notable limitation was that our multi-class classifier identified a single activated FAU at-a-time - meaning that compared to the controller there was no capacity for concurrent input. This didn't meaningfully impact the AR Web Browser task in our view, but would have notable implications in terms of restricting the freedom of expression of users in the VR game, where games often rely on concurrent inputs (moving, targeting, shooting etc.) to create challenging and engaging gameplay. Further research would be needed to identify which FAUs, if any, can be concurrently and independently controlled e.g. using multi-label classification to drive facial actions 'chords'. If there are few FAUs where concurrent input can be supported, then for more complex games, concurrency across modalities (e.g. FAU and speech, FAU and eye gaze) might offer a viable alternative. Failing that, accommodations may need to be made *temporally* e.g. slowing gameplay down to allow switching between multiple FAU inputs.

9 Conclusion

Across two studies we have explored the use of facial action units as an input modality for controlled extended reality interfaces and experiences. Study 1 (N=20) assessed the range of FAUs currently sensed by consumer XR hardware, and then narrowed this set down to 7 FAUs perceived as the most usable by our participants. We then trained a model to classify the currently enacted FAU, and used this to drive two prototype XR applications: a VR Game, and an AR Web Browser. Through a formative exploration of facial input for XR (N=10), we finally demonstrated the viability of XR platforms supporting facial input as a generalisable alternative to existing hand/controller inputs, reflecting on the potential benefits for hands-free spatial interactions and accessibility, and the necessary next steps to incorporate facial input into XR as an additional first-class interaction modality for everyday use.

Acknowledgments

This research was supported by UK Research and Innovation (UKRI) under the UK Government's Horizon Europe funding guarantee (AUGSOC) [EP/Z000068/1]. Dr. Graham Wilson's time was funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (#835197, *ViAjeRo*).

References

[1] [n. d.]. The Americans with Disabilities Act. https://www.ada.gov/
[2] [n. d.]. Face Expressions | Docs. https://docs.snap.com/lens-studio/references/templates/face/face-expressions
[3] [n. d.]. Get started with accessibility features on Apple Vision Pro. https://support.apple.com/en-gb/guide/apple-vision-pro/tan426c48bdc/visionos
[4] [n. d.]. Microsoft accessibility controller. https://www.xbox.com/en-GB/accessories/controllers/xbox-adaptive-controller
[5] [n. d.]. Playstation access contoller. https://www.playstation.com/en-gb/accessories/access-controller/
[6] [n. d.]. Understanding Assistive Technologies: What Are Sip-and-Puff Systems? https://www.boia.org/blog/understanding-assistive-technologies-what-are-sip-and-puff-systems
[7] [n. d.]. Vive facial tracker. https://business.vive.com/uk/product/vive-focus-3-facial-tracker/
[8] [n. d.]. XR access. https://xraccess.org/
[9] 2000. Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation. http://data.europa.eu/eli/dir/2000/78/oj/eng Volume: 303.
[10] 2023. Meta Quest Pro Now Tracks Your Tongue. https://www.uploadvr.com/quest-pro-tongue-tracking/
[11] All Access Life. 2022. How to Make VR Accessible to People with Disabilities Beat Saber on Oculus Quest. https://youtu.be/FHeVIJib-mk
[12] Balasubramanian Anand, Bilva Bhalchandra Navathe, Sudha Velusamy, Hariprasad Kannan, Anshul Sharma, and Viswanath Gopalakrishnan. 2012. Beyond touch: Natural interactions using facial expressions. In *2012 IEEE Consumer Communications and Networking Conference (CCNC)*. 255–259. https://doi.org/10.1109/CCNC.2012.6181097 ISSN: 2331-9860.
[13] Steven Baker, Jenny Waycott, Elena Robertson, Romina Carrasco, Barbara Barbosa Neves, Ralph Hampson, and Frank Vetere. 2020. Evaluating the use of interactive virtual reality technology with older adults living in residential aged care. *Information Processing & Management* 57, 3 (2020), 102105. https://doi.org/10.1016/j.ipm.2019.102105
[14] Nikolaus Bee, Bernhard Falk, and Elisabeth André. 2009. Simplified facial animation control utilizing novel input devices: a comparative study. In *Proceedings of the 14th international conference on Intelligent user interfaces (IUI '09)*. Association for Computing Machinery, New York, NY, USA, 197–206. https://doi.org/10.1145/1502650.1502680
[15] Gunnar Borg. 1998. Borg's perceived exertion and pain scales. *Human Kinetics* (1998), 104 vii. https://doi.org/10.1097/00005768-199809000-00018 ISBN: 0-88011-623-4 (Paperback).
[16] Sebastian Boring, David Ledo, Xiang 'Anthony' Chen, Nicolai Marquardt, Anthony Tang, and Saul Greenberg. 2012. The fat thumb: Using the thumb's contact size for single-handed mobile interaction. In *Proceedings of the 14th international conference on human-computer interaction with mobile devices and services (MobileHCI '12)*. ACM, New York, NY, USA, 39–48. https://doi.org/10.1145/2371574.2371582 Number of pages: 10 Place: San Francisco, California, USA Type: conference tex.acceptance: 25% - 54/212 tex.acmid: 2371582 tex.date-added: 2014-01-11 20:51:34 +0000 tex.date-modified: 2014-01-11 20:54:23 +0000.
[17] Jelte E. Bos, Scott N. MacKinnon, and Anthony Patterson. 2005. Motion Sickness Symptoms in a Ship Motion Simulator: Effects of Inside, Outside, and No View. *Aviation, Space, and Environmental Medicine* 76, 12 (Dec. 2005), 1111–1118.
[18] Marco Antonio Martínez Cano, Carolina Padilla Velasco, and Steve Bakos. 2024. Making Virtual Reality (VR) Accessible for People with Disabilities. In *Encyclopedia of Computer Graphics and Games*. Springer, 1104–1109.
[19] François Chollet et al. 2015. Keras. https://keras.io.
[20] Elizabeth A. Clark, J'Nai Kessinger, Susan E. Duncan, Martha Ann Bell, Jacob Lahne, Daniel L. Gallagher, and Sean F. O'Keefe. 2020. The Facial Action Coding System for Characterization of Human Affective Response to Consumer Product-Based Stimuli: A Systematic Review. *Frontiers in Psychology* 11 (2020). https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00920
[21] Andrew Crossan, Stephen Brewster, and Alexander Ng. 2010. Foot tapping for mobile interaction. *Proceedings of HCI 2010 24* (2010), 418–422. https://www.scienceopen.com/hosted-document?doi=10.14236/ewic/HCI2010.49
[22] Arindam Dey, Amit Barde, Bowen Yuan, Ekansh Sareen, Chelsea Dobbins, Aaron Goh, Gaurav Gupta, Anubha Gupta, and Mark Billinghurst. 2022. Effects of interacting with facial expressions and controllers in different virtual environments on presence, usability, affect, and neurophysiological signals. *International Journal of Human-Computer Studies* 160 (April 2022), 102762. https://doi.org/10.1016/j.ijhcs.2021.102762
[23] Massimiliano Di Luca, Hasti Seifi, Simon Egan, and Mar Gonzalez-Franco. 2021. Locomotion Vault: the Extra Mile in Analyzing VR Locomotion Techniques. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–10. https://doi.org/10.1145/3411764.3445319
[24] John Dudley, Lulu Yin, Vanja Garaj, and Per Ola Kristensson. 2023. Inclusive Immersion: a review of efforts to improve accessibility in virtual reality, augmented reality and the metaverse. *Virtual Reality* 27, 4 (Dec. 2023), 2989–3020. https://doi.org/10.1007/s10055-023-00850-8
[25] Paul Ekman and Wallace V Friesen. 1978. Facial action coding system. *Environmental Psychology & Nonverbal Behavior* (1978).
[26] Paul Ekman, Gowen Roper, and Joseph C. Hager. 1980. Deliberate Facial Movement. *Child Development* 51, 3 (1980), 886–891. https://doi.org/10.2307/1129478
[27] Lisa A Elkin, Paul G Allen, Matthew Kay, James J Higgins, and Jacob O Wobbrock. 2021. An aligned rank transform procedure for multifactor contrast tests; an aligned rank transform procedure for multifactor contrast tests. 15, 21 (2021). https://doi.org/10.1145/3472749.3474784 ISBN: 9781450386357 Publisher: Virtual Event.
[28] Lisa A. Elkin, Matthew Kay, James J. Higgins, and Jacob O. Wobbrock. 2021. An Aligned Rank Transform Procedure for Multifactor Contrast Tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 754–768. https://doi.org/10.1145/3472749.3474784
[29] Esme Kirk-Wade. 2023. *Research Briefing: UK disability statistics: Prevalence and life experiences*. Technical Report. https://researchbriefings.files.parliament.uk/documents/CBP-9602/CBP-9602.pdf
[30] D. Freeman, S. Reeve, A. Robinson, A. Ehlers, D. Clark, B. Spanlang, and M. Slater. 2017. Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychological Medicine* 47, 14 (2017), 2393–2400. https://doi.org/10.1017/S003329171700040X
[31] Kathrin Gerling, Patrick Dickinson, Kieran Hicks, Liam Mason, Adalberto L. Simeone, and Katta Spiel. 2020. Virtual Reality Games for People Using Wheelchairs. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3313831.3376265
[32] Kathrin Gerling and Katta Spiel. 2021. A Critical Examination of Virtual Reality Technology in the Context of the Minority Body. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3411764.3445196
[33] Pierre Gosselin, Mélanie Perron, and Martin Beaupré. 2010. The Voluntary Control of Facial Action Units in Adults. *Emotion (Washington, D.C.)* 10 (April 2010), 266–71. https://doi.org/10.1037/a0017748
[34] Sandra G. Hart and Lowell E Staveland. 1988. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In *Human mental workload*. https://doi.org/10.1016/s0166-4115(08)62386-9
[35] Daniel Hepperle, Yannick Weiß, Andreas Siess, and Matthias Wölfel. 2019. 2D, 3D or speech? A case study on which user interface is preferable for what kind of object interaction in immersive virtual reality. *Computers & Graphics* 82 (Aug. 2019), 321–331. https://doi.org/10.1016/j.cag.2019.06.003
[36] Steven Hickson, Nick Dufour, Avneesh Sud, Vivek Kwatra, and Irfan Essa. 2019. Eyemotion: Classifying Facial Expressions in VR Using Eye-Tracking Cameras. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 1626–1635. https://doi.org/10.1109/WACV.2019.00178 ISSN: 1550-5790.
[37] Seokhyeon Hong, Yeonsoo Choi, Youjin Sung, Yurhee Jin, Young Yim Doh, and Jeongmi Lee. 2022. Evoker: Narrative-based Facial Expression Game for Emotional Development of Adolescents. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3491101.3516486
[38] Mirja Ilves, Yulia Gizatdinova, Veikko Surakka, and Esko Vankka. 2014. Head movement and facial expressions as game input. *Entertainment Computing* 5, 3 (Aug. 2014), 147–156. https://doi.org/10.1016/j.entcom.2014.04.005
[39] Viveca Jimenez-Mixco, Rafael de las Heras, Juan-Luis Villalar, and María Teresa Arredondo. 2009. A New Approach for Accessible Interaction within Smart Homes through Virtual Reality. In *Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments*, Constantine Stephanidis

(Ed.). Springer, Berlin, Heidelberg, 75–81. https://doi.org/10.1007/978-3-642-02710-9_9

[40] Arata Jingu, Yudai Tanaka, and Pedro Lopes. 2023. LipIO: Enabling Lips as both Input and Output Surface. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3544548.3580775

[41] Michael Jones, Kevin Grogg, John Anschutz, and Ruth Fierman. 2008. A Sip-and-Puff Wireless Remote Control for the Apple iPod. *Assistive Technology* 20, 2 (June 2008), 107–110. https://doi.org/10.1080/10400435.2008.10131937

[42] Matthew Kay, Lisa A. Elkin, James J. Higgins, and Jacob O. Wobbrock. 2021. ARTool: Aligned Rank Transform. https://CRAN.R-project.org/package=ARTool

[43] Matthew Kay, Lisa A. Elkin, and Jacob O. Wobbrock. [n. d.]. Contrast tests with ART. https://cran.r-project.org/web/packages/ARTool/vignettes/art-contrasts.html

[44] Matthew Kay and Jacob O Wobbrock. 2016. Package 'ARTool'. *CRAN Repository* 2016 (2016), 1–13.

[45] Konstantina Kilteni, Raphaela Groten, and Mel Slater. 2012. The Sense of Embodiment in Virtual Reality. *Presence: Teleoperators and Virtual Environments* 21, 4 (Nov. 2012), 373–387. https://doi.org/10.1162/pres_a_00124

[46] Jinhyuk Kim, Jaekwang Cha, and Shiho Kim. 2020. Hands-Free User Interface for VR Headsets Based on In Situ Facial Gesture Sensing. *Sensors* 20, 24 (Jan. 2020), 7206. https://doi.org/10.3390/s20247206 Number: 24 Publisher: Multidisciplinary Digital Publishing Institute.

[47] Marion Koelle, Swamy Ananthanarayan, and Susanne Boll. 2020. Social Acceptability in HCI: A Survey of Methods, Measures, and Design Strategies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–19. https://doi.org/10.1145/3313831.3376162

[48] Yunfei Lai, Minghui Sun, and Zhuofeng Li. 2024. GazePuffer: Hands-Free Input Method Leveraging Puff Cheeks for VR. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 331–341. https://doi.org/10.1109/VR58804.2024.00055 ISSN: 2642-5254.

[49] James R. Lewis, Brian S. Utesch, and Deborah E. Maher. 2015. Measuring Perceived Usability: The SUS, UMUX-LITE, and AltUsability. *International Journal of Human–Computer Interaction* 31, 8 (Aug. 2015), 496–505. https://doi.org/10.1080/10447318.2015.1064654 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10447318.2015.1064654.

[50] Georgia Belle Von Kampen Loewen. 2023. Accessible Gaming Wearables: Opportunities for Players with Upper Limb Motor Disabilities. Publisher: Carleton University.

[51] Katsutoshi Masai, Kai Kunze, Daisuke Sakamoto, Yuta Sugiura, and Maki Sugimoto. 2020. Face Commands - User-Defined Facial Gestures for Smart Glasses. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 374–386. https://doi.org/10.1109/ISMAR50242.2020.00064 ISSN: 1554-7868.

[52] Marta Matamala-Gomez, Tony Donegan, Sara Bottiroli, Giorgio Sandrini, Maria V. Sanchez-Vives, and Cristina Tassorelli. 2019. Immersive Virtual Reality and Virtual Embodiment for Pain Relief. *Frontiers in Human Neuroscience* 13 (2019). https://doi.org/10.3389/fnhum.2019.00279

[53] Mark Mcgill, Aidan Kehoe, Euan Freeman, and Stephen Brewster. 2020. Expanding the Bounds of Seated Virtual Workspaces. *ACM Transactions on Computer-Human Interaction* 27, 3 (May 2020), 13:1–13:40. https://doi.org/10.1145/3380959

[54] Daniel Medeiros, Romane Dubus, Julie Williamson, Graham Wilson, Katharina Pohlmann, and Mark McGill. 2023. Surveying the social comfort of body, device, and environment-based augmented reality interactions in confined passenger spaces using mixed reality composite videos. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (Aug. 2023). https://eprints.gla.ac.uk/304940/ Publisher: Association for Computing Machinery.

[55] Takashi Miyaki and Jun Rekimoto. 2009. GraspZoom: Zooming and scrolling control model for single-handed mobile interaction *(MobileHCI09 - The 11th International Conference on Human-Computer Interaction with Mobile Devices and Services)*. https://doi.org/10.1145/1613858.1613872

[56] Noha Mokhtar and Augusto Esteves. 2024. Hand Me This: Exploring the Effects of Gaze-driven Animations and Hand Representations in Users' Sense of Presence and Embodiment. In *Proceedings of the 2024 Symposium on Eye Tracking Research and Applications (ETRA '24)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3649902.3656362

[57] Giovanni Morone, Antonia Pirrera, Antonio Iannone, and Daniele Giansanti. 2023. Development and Use of Assistive Technologies in Spinal Cord Injury: A Narrative Review of Reviews on the Evolution, Opportunities, and Bottlenecks of Their Integration in the Health Domain. *Healthcare* 11, 11 (Jan. 2023), 1646. https://doi.org/10.3390/healthcare11111646 Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

[58] Peter E. Morris and Catherine O. Fritz. 2013. Effect sizes in memory research. *http://dx.doi.org/10.1080/09658211.2013.763984* 21, 7 (Oct. 2013), 832–842. https://doi.org/10.1080/09658211.2013.763984 Publisher: Routledge.

[59] Martez Mott, Edward Cutrell, Mar Gonzalez Franco, Christian Holz, Eyal Ofek, Richard Stoakley, and Meredith Ringel Morris. 2019. Accessible by Design: An Opportunity for Virtual Reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 451–454. https://doi.org/10.1109/ISMAR-Adjunct.2019.00122

[60] Alexander Ng, Stephen A. Brewster, and John H. Williamson. 2014. Investigating the effects of encumbrance on one- and two- handed interactions with mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 1981–1990. https://doi.org/10.1145/2556288.2557312

[61] Oculus Developers. [n. d.]. Face Tracking for Movement SDK for Unity. https://developer.oculus.com/documentation/unity/move-face-tracking/?locale=en_GB

[62] Laxmi Pandey, Khalad Hasan, and Ahmed Sabbir Arif. 2021. Acceptability of Speech and Silent Speech Input Methods in Private and Public. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3411764.3445430

[63] Igor Sunday Pandžić, Prem Kalra, Nadia Magnenat Thalmann, and Daniel Thalmann. 1994. Real-time facial interaction. *Displays* 15, 3 (July 1994), 157–163. https://doi.org/10.1016/0141-9382(94)90004-3

[64] Expert Participation. [n. d.]. Equality Act 2010. https://www.legislation.gov.uk/ukpga/2010/15/contents Type: Text.

[65] Tabitha C. Peck, Kyla A. McMullen, and John Quarles. 2021. DiVRsify: Break the Cycle and Develop VR for Everyone. *IEEE Computer Graphics and Applications* 41, 6 (2021), 133–142. https://doi.org/10.1109/MCG.2021.3113455

[66] Ken Pfeuffer, Hans Gellersen, and Mar Gonzalez-Franco. 2024. Design Principles and Challenges for Gaze + Pinch Interaction in XR. *IEEE Computer Graphics and Applications* 44, 3 (May 2024), 74–81. https://doi.org/10.1109/MCG.2024.3382961 Conference Name: IEEE Computer Graphics and Applications.

[67] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze + pinch interaction in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 99–108. https://doi.org/10.1145/3131277.3132180

[68] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze + pinch interaction in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction (SUI '17)*. Association for Computing Machinery, New York, NY, USA, 99–108. https://doi.org/10.1145/3131277.3132180

[69] Alexander Plopski, Teresa Hirzle, Nahal Norouzi, Long Qian, Gerd Bruder, and Tobias Langlotz. 2022. The Eye in Extended Reality: A Survey on Gaze Interaction and Eye Tracking in Head-worn Extended Reality. *ACM Comput. Surv.* 55, 3 (March 2022), 53:1–53:39. https://doi.org/10.1145/3491207

[70] K.A. Provins. 1997. The Specificity of Motor Skill and Manual Asymmetry: A Review of the Evidence and Its Implications. https://doi.org/10.1080/00222899709600832

[71] G S Rajshekar Reddy, Michael J Proulx, Leanne Hirshfield, and Anthony Ries. 2024. Towards an Eye-Brain-Computer Interface: Combining Gaze with the Stimulus-Preceding Negativity for Target Selections in XR. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3613904.3641925

[72] John T.E. Richardson. 2011. Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review* 6, 2 (2011), 135–147. https://doi.org/10.1016/j.edurev.2010.12.001

[73] Soha Rostaminia, Alexander Lamson, Subhransu Maji, Tauhidur Rahman, and Deepak Ganesan. 2019. W!NCE: Unobtrusive Sensing of Upper Facial Action Units with EOG-based Eyewear. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 1 (March 2019), 23:1–23:26. https://doi.org/10.1145/3314410

[74] Yuko Sano, Akimichi Ichinose, Naoki Wake, Michihiro Osumi, Masahiko Sumitani, Shin-ichiro Kumagaya, and Yasuo Kuniyoshi. 2015. Reliability of phantom pain relief in neurorehabilitation using a multimodal virtual reality system. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2482–2485. https://doi.org/10.1109/EMBC.2015.7318897

[75] Zhanna Sarsenbayeva, Niels van Berkel, Chu Luo, Vassilis Kostakos, and Jorge Goncalves. 2017. Challenges of situational impairments during interaction with mobile devices. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction (OzCHI '17)*. Association for Computing Machinery, New York, NY, USA, 477–481. https://doi.org/10.1145/3152771.3156161

[76] Sidas Saulynas, Lawrence E. Burgee, and Ravi Kuber. 2017. All situational impairments are not created equal: A classification system for situational impairment events and the unique nature of severely constraining situational impairments. *iConference 2017 Proceedings* (2017). https://hdl.handle.net/2142/96688 Publisher: iSchools.

[77] Kelsea Schulenberg, Guo Freeman, Lingyuan Li, and Catherine Barwulor. 2023. "Creepy Towards My Avatar Body, Creepy Towards My Body": How Women Experience and Manage Harassment Risks in Social Virtual Reality. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 236 (oct 2023), 29 pages. https://doi.org/10.1145/3610027

[78] MinYeong Seo, KyungEun Kang, and HyeongYeop Kang. 2024. VR Blowing: A Physically Plausible Interaction Method for Blowing Air in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 30, 7 (July 2024), 3680–3692.

https://doi.org/10.1109/TVCG.2023.3238478 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[79] Ather Sharif, Aedan Liam McCall, and Kianna Roces Bolante. 2022. Should I Say "Disabled People" or "People with Disabilities"? Language Preferences of Disabled People Between Identity- and Person-First Language. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility* (Athens, Greece) *(ASSETS '22)*. Association for Computing Machinery, New York, NY, USA, Article 10, 18 pages. https://doi.org/10.1145/3517428.3544813

[80] Mel Slater. 2009. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 1535 (Dec. 2009), 3549–3557. https://doi.org/10.1098/rstb.2009.0138 tex.pmcid: PMC2781884.

[81] Misha Sra, Xuhai Xu, and Pattie Maes. 2018. BreathVR: Leveraging Breathing as a Directly Controlled Interface for Virtual Reality Games. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173914

[82] Kay Stanney, Cali Fidopiastis, and Linda Foster. 2020. Virtual Reality Is Sexist: But It Does Not Have to Be. *Frontiers in Robotics and AI* 7 (2020). https://doi.org/10.3389/frobt.2020.00004

[83] Atieh Taheri, Arthur Caetano, and Misha Sra. 2024. Virtual Steps: The Experience of Walking for a Lifelong Wheelchair User in Virtual Reality. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 168–178. https://doi.org/10.1109/VR58804.2024.00040 ISSN: 2642-5254.

[84] Atieh Taheri and Ziv Weissman. 2021. Designing a First Person Shooter Game for Quadriplegics. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3411763.3451850

[85] Atieh Taheri, Ziv Weissman, and Misha Sra. 2021. Design and Evaluation of a Hands-Free Video Game Controller for Individuals With Motor Impairments. *Frontiers in Computer Science* 3 (Dec. 2021). https://doi.org/10.3389/fcomp.2021.751455

[86] Atieh Taheri, Ziv Weissman, and Misha Sra. 2021. Design and Evaluation of a Hands-Free Video Game Controller for Individuals With Motor Impairments. *Frontiers in Computer Science* 3 (Dec. 2021). https://doi.org/10.3389/fcomp.2021.751455 Publisher: Frontiers.

[87] Atieh Taheri, Ziv Weissman, and Misha Sra. 2021. Exploratory Design of a Hands-free Video Game Controller for a Quadriplegic Individual. In *Proceedings of the Augmented Humans International Conference 2021 (AHs '21)*. Association for Computing Machinery, New York, NY, USA, 131–140. https://doi.org/10.1145/3458709.3458946

[88] Jane F. Thomsen, Fred Gerr, and Isam Atroshi. 2008. Carpal tunnel syndrome and the use of computer mouse and keyboard: A systematic review. *BMC Musculoskeletal Disorders* 9, 1 (Oct. 2008), 134. https://doi.org/10.1186/1471-2474-9-134

[89] Bram van de Laar, Hayrettin Gürkök, Danny Plass-Oude Bos, Mannes Poel, and Anton Nijholt. 2013. Experiencing BCI Control in a Popular Computer Game. *IEEE Transactions on Computational Intelligence and AI in Games* 5, 2 (June 2013), 176–184. https://doi.org/10.1109/TCIAIG.2013.2253778 Conference Name: IEEE Transactions on Computational Intelligence and AI in Games.

[90] Thomas Van Gemert, Kasper Hornbæk, Jarrod Knibbe, and Joanna Bergström. 2023. Towards a Bedder Future: A Study of Using Virtual Reality while Lying Down. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–18. https://doi.org/10.1145/3544548.3580963

[91] Nadine Wagener, Tu Dinh Duong, Johannes Schöning, Yvonne Rogers, and Jasmin Niess. 2021. The role of mobile and virtual reality applications to support well-being: An expert view and systematic app review. In *Human-Computer Interaction–INTERACT 2021: 18th IFIP TC 13 International Conference, Bari, Italy, August 30–September 3, 2021, Proceedings, Part IV 18*. Springer, 262–283.

[92] Nadine Wagener, Arne Kiesewetter, Leon Reicherts, Paweł W. Woźniak, Johannes Schöning, Yvonne Rogers, and Jasmin Niess. 2024. MoodShaper: A Virtual Reality Experience to Support Managing Negative Emotions. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) *(DIS '24)*. Association for Computing Machinery, New York, NY, USA, 2286–2304. https://doi.org/10.1145/3643834.3661570

[93] Lihang Wen, Jianlong Zhou, Weidong Huang, and Fang Chen. 2022. A Survey of Facial Capture for Virtual Reality. *IEEE Access* 10 (2022), 6042–6052. https://doi.org/10.1109/ACCESS.2021.3138200 Conference Name: IEEE Access.

[94] Graham Wilson, Mark McGill, Daniel Medeiros, and Stephen Brewster. 2023. A Lack of Restraint: Comparing Virtual Reality Interaction Techniques for Constrained Transport Seating. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (May 2023), 2390–2400. https://doi.org/10.1109/TVCG.2023.3247084 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[95] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963

[96] Harry Zhang, Matthew R. H. Smith, and Gerald J. Witt. 2006. Identification of Real-Time Diagnostic Measures of Visual Distraction With an Automatic Eye-Tracking System. *Human Factors* 48, 4 (Dec. 2006), 805–821. https://doi.org/10.1518/001872006779166307

[97] Yuhang Zhao, Edward Cutrell, Christian Holz, Meredith Ringel Morris, Eyal Ofek, and Andrew D. Wilson. 2019. SeeingVR: A Set of Tools to Make Virtual Reality More Accessible to People with Low Vision. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300341

[98] Daniel Zielasko, Sebastian Freitag, Dominik Rausch, Yuen C. Law, Benjamin Weyers, and Torsten W. Kuhlen. 2015. BlowClick: A Non-Verbal Vocal Input Metaphor for Clicking. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction (SUI '15)*. Association for Computing Machinery, New York, NY, USA, 20–23. https://doi.org/10.1145/2788940.2788953

[99] Daniel Zielasko, Yuen C. Law, and Benjamin Weyers. 2020. Take a Look Around – The Impact of Decoupling Gaze and Travel-direction in Seated and Ground-based Virtual Reality Utilizing Torso-directed Steering. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 398–406. https://doi.org/10.1109/VR46266.2020.00060 ISSN: 2642-5254.