International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2022)
J.-M. Normand and H. Uchiyama (Editors)

# Exploring EEG-Annotated Affective Animations in Virtual Reality: Suggestions for Improvement

Claudia Krogmeier[†] Christos Mousas[‡]

Department of Computer Graphics Technology, Purdue University, West Lafayette, Indiana 47907, U.S.A.

**Abstract**

*In this work, we recorded brain activity data from participants who viewed 12 affective character animations in virtual reality. Frontal alpha asymmetry (FAA) scores were calculated from electroencephalography (EEG) data to understand objective affective responses to these animations. A subset of these animations were then annotated as either low FAA (meaning they elicited lower FAA responses), or high FAA (meaning they elicited higher FAA responses). Next, these annotated animations were used in a primary $2 \times 2$ study in which we a) examined if we could replicate FAA responses to low FAA and high FAA animations in a subsequent study, and b) investigated how the number of characters in the VR environment would influence FAA responses. Additionally, we compared FAA to self-reported affective responses to the four conditions (one character, low FAA; one character, high FAA; four characters, low FAA; four characters, high FAA). In this way, our research seeks to better understand objective and subjective emotional responses in VR. Results suggest that annotated FAA may not inform FAA responses to affective animations in a subsequent study when more characters are present. However, self-reported affective responses to the four conditions is in line with FAA annotated responses. We offer suggestions for the development of specific affective experiences in VR which are based on preliminary brain activity data.*

**CCS Concepts**

• ***Human-centered computing*** → ***Virtual reality; User studies;*** • ***Computing methodologies*** → ***Perception;*** *Animation;*

## 1. Introduction

Although annotated datasets examining affective properties of animations and virtual character motions exist, fewer studies explore whether previously annotated datasets will elicit the same or expected affective responses to new combinations of characters. In our work, we first show participants 12 affective character animations in virtual reality (VR), and annotate six of these based on frontal alpha asymmetry (FAA) scores calculated from electroencephalography (EEG) data. Next, we examine FAA responses to the same and new combinations of animations in a different experimental design. EEG offers an objective measurement of emotion and a more complete understanding of the user's experience when paired with self-reported data, and has been used to examine emotional responses in numerous VR studies. VR was used to provide an immersive experience for participants, as it may amplify emotions in certain contexts when compared to non-immersive experiences such as movie viewing [KCC*18]. The purpose of our work is to examine emotions elicited in VR using both objective and subjective methods, as VR is growing as a prominent tool in virtual social interactions and mental health applications, both of which could benefit from an improved understanding of emotional responses. A better understanding of emotional brain activity responses in VR may contribute towards the design of targeted affective experiences in brain-computer interface (BCI) applications as well. A primary concern in BCI research is to implement natural brain activity interaction for users [KSL18]. Therefore, understanding how self-reported data aligns with emotional data from brain activity may inform affective BCIs as well.

Previous VR studies have used EEG to investigate mood induction [RRC*15], affective responses to varying levels of presence [ULH20], fear [BMM*20], and virtual threats [GFPRFS14, SEM19]. Multiple VR studies have examined emotional correlates using FAA [SKS*21, KLF*21, RMMH18]: a measurement of the difference in prefrontal cortical activity between the left and right brain hemispheres [HJG18]. According to the valence model of FAA [DES*90], greater left prefrontal cortical activity relative to right prefrontal cortical activity has been associated with positive emotions and approach motivation, while greater relative right prefrontal cortical activity has been associated with negative emotions and withdrawal, as reported in [KLF*21].

In our two-stage work, we use FAA and self-reported questionnaire data to compare brain activity to self-reported data. Ad-

† ckrogmei@purdue.edu
‡ cmousas@purdue.edu



DOI: 10.2312/egve.20221283

ditionally, we compare brain activity responses from preliminary data collection to a primary study to determine if FAA responses can be replicated. Considering the replication crisis in psychology, and the overlap in methodologies between VR and social psychology, it is possible that VR research may suffer from similar issues [LWE*19]. Therefore, this study includes a comparison of brain activity responses to the same animations from one data collection period to the next, as we believed it was important to determine if these responses could be replicated. In preliminary data collection, we measure FAA responses to 12 affective character animations in VR. Salient animations are annotated with *low FAA* or *high FAA* values based on the distribution of values collected during preliminary data collection. These FAA values are used to inform the factor design of a primary study. In our primary study, we examine FAA responses in a $2 \times 2$ repeated measures design, in which *Character Count* and *Annotated FAA* serve as the two factors, each with two levels. Therefore, the four conditions of the primary study are as follows: 1) *OneLow*: *one character*, *low FAA*; 2) *OneHigh*: *one character*, *high FAA*; 3) *FourLow*: *four characters*, *low FAA*; and 4) *FourHigh*: *four characters*, *high FAA*.

We hypothesize that *high FAA* conditions will elicit greater FAA than *low FAA* conditions in the subsequent study, and expect self-reported data to indicate greater positive emotion in line with *high FAA* conditions. The investigation of *Character Count* is exploratory; we do not hypothesize how this may influence FAA responses. Our work contributes to the investigation of objective data for understanding affective experiences in VR. Additionally, this work is a step towards the development of digital stories which are designed to target a specific affective user experience as measured through physiological output and previously annotated datasets.

## 2. Related Work

**Perception of Body Emotion.** Previous research has shown that, similar to humans [MF69], virtual character postures can convey emotions [NLK*13]. Psychological studies have attempted to illustrate the perceptual significance of body motion in expressing emotions [ROCG09]. Researchers have shown that emotions such as happiness and sadness are easier to recognize than disgust, and have found that surprise and fear emotions are difficult to recognize from static poses [Cou04]. A relationship between emotion type and posture characteristics has also been found, indicating that differences in emotions can partially be explained by the activation dimension (i.e., velocities, accelerations, and jerk of body parts) [Wal98]. Atkinson et al. [ADGY04] found that emotions can be recognized from a body motion, and that exaggerations in motions increase emotion recognition accuracy. Sawada et al. [SSI03] observed that dancers varied the speed, force, and directness of their arm movements when conveying joy, sadness, and anger. Lastly, Ennis and Egges [EE12] found that negative compared to positive emotions were more recognizable.

Researchers have investigated the role of bodily information for recognizing a person's affective state, as well as the role of body form and movement in affect perception [ATD07, HH06]. Roether et al. [ROCG09] observed that elbow and hip flexion were essential for anger and fear, while head inclination was significant for recognizing sadness in motions. Overall, the researchers concluded that shape and motion are necessary for perceiving emotions from body expressions. Beyond shape and movement features, researchers have introduced two primary levels of bodily details: 1) high and 2) low-level descriptions [HH06]. The high-level description is often based on the Laban [New93, Zha01] approach, which globally describes body expressions. The low-level description includes more specific features, such as the distance between joints and the angle between body segments. For example, Bernhardt and Robinson [BR07] examined motion in terms of dynamic features, such as velocity, acceleration, and jerk. In our work, we discuss emotions using a high-level approach. In another study, Karg et al. [KKB10] investigated the capability of gait to reveal a person's affective state, and determined that speed, cadence, and stride length are essential factors in correctly discriminating between different human gait expressions. Lastly, Paterson et al. [PPS01] confirmed the role of velocity in the discrimination of affect, finding that speed plays a vital role in the perception of affect.

**Annotated Animation Datasets.** Various animation datasets have been developed over the years. However, we only found a few annotated animation datasets. Ma et al. [MPP06] have collected an extensive database from 30 actors, in which 4,080 motion capture sequences encompass waving and other non-verbal actions displaying four emotion categories: neutral, angry, happy, and sad. Metallinou et al. [MLB*10] developed the USC CreativeIT database that consists of short, scripted actions performed by actors recorded with full-body motion capture. Niewiadomski et al. [NMB*13] developed the Multimodal Multiperson Corpus of Laughter in Interaction (MMLI), in which we find 3D body position information, facial tracking, audio, and video channels, as well as respiration data. In their dataset, all tracks were synchronized, segmented, and annotated to discriminate between laughter and non-laughter. Finally, Volkava et al. [VDLRBM14a] developed a motion capture database consisting of a large set (over 1,400) of natural emotional body expressions typical of monologues performed by amateur actors. Besides performing a physical motion properties analysis and an emotion categorization study, the authors also included in the dataset the intended emotion expression from the actor for each motion sequence to allow for investigations concerning the link between intended and perceived emotions.

**Differences in this Study.** Our research differs from the previously mentioned work in several ways. First, we annotate a variety of animations which include a range of emotions (happy, sad, fearful, surprising, etc) as well as activation dimensions such as velocity (static, jerky), position (center-based, moving forward), and body usage (upper limbs, lower limbs, both), rather than investigate the role of singular parameters on elicited emotion. Additionally, we do not examine user emotion recognition accuracy (or differences between intended and perceived emotions), instead focusing on user feelings measured with brain activity and self-reported data during each high-level affective animation presentation. Further, few animation datasets are annotated with EEG or other physiological data to provide a more objective understanding of perceived emotion.

## 3. Preliminary Data Collection

In this section, we detail our preliminary data collection in which we investigate 12 virtual character animations with FAA values.

### 3.1. Participants, Virtual Environment, Procedure

Participants were invited to participate through email within our university department. The study received approval from the university institutional review board (IRB), and written consent was obtained from participants before their participation. Seventeen participants (eight female; age $M = 25.76$, $SD = 3.72$, all right-handed) were included in our preliminary data collection. Virtual characters were positioned in the center of a white plane in the virtual environment. Lighting in the virtual environment was uniform [LBT*19], so as not to introduce shadows or excessive brightness which could influence mood of the environment. Participants viewed the virtual environment, developed in the Unity game engine, within the Oculus Quest VR headset. This headset has a resolution of 1440 × 1600 per eye, and refresh rate of 72 Hz.

The researcher first positioned the EEG headset and showed the participant a visualization of their brain activity. The researcher asked participants to talk, frown, and clench their jaws as they viewed their brain activity in order to motivate less movement during EEG recording. Participants were instructed to view each animation and think about how they felt. Next, the researcher positioned the VR headset, and ensured that EEG contact quality remained high. During the experiment, participants were comfortably seated in a chair, and looked straight ahead at each animation.

### 3.2. Virtual Character Animations and Presentation

The virtual character model used in this study was downloaded from Adobe Mixamo,[†] and is called "X-bot." Similar research has selected mannequin characters [LBT*19] as well as animated stick figures [VMD*14, VDLRBM14b] when investigating affective properties of animations. Because mannequins do not have facial expressions, these characters can be considered affectively neutral. Additionally, it is thought that bodily expression may be recognized with the same reliability as facial expressions [DG09]. In this way, we could investigate FAA responses to animations with less bias from facial expression or character style.

Animations used with "X-bot" are from Adobe Mixamo and a Unity Asset store package called "Woman Gesture01 Animations." Virtual character animations consisted of: *Excited, Dance, Pointing, Rejected, Terrified, Cry, Dismiss, Clap, Idle, Jump, Laughing* and *Surprise* animations for a total of 12 animations for EEG analysis. All animations investigated in preliminary data collection are presented in Figure 1. All animations can be seen in the video at the link in our data availability statement.

We selected six second stimulus presentation windows, as previous affective datasets have presented affective stimuli for six seconds [LBC*97], and similarly, five seconds [LBT*19]. Each animation was displayed for six seconds, and followed by a five second interstimulus interval (ISI); time between each animation. Participants were instructed to relax their mind during the ISI.

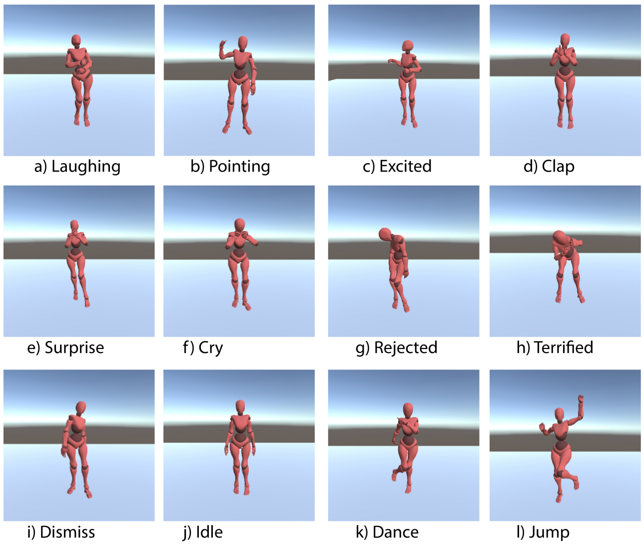


**Figure 1:** *Twelve virtual character animations investigated in preliminary data collection.*

### 3.3. EEG Data Collection and Preprocessing

EEG data collection comprised two identical, 12 minute runs in order to increase participant comfort during this process. Data collected from both runs were appended for each participant for EEG preprocessing and statistical analysis. Each run consisted of five trials. Therefore, 10 trials were available for each participant during data analysis. Within each trial, all animations were presented on screen once. Presentation order of the animations was randomized within each trial. At the start of each run, participants were asked to relax their mind and try to minimize mental wandering for a 45 second period. No virtual characters were present in the environment during this time. The last 30 seconds of this time period served as a baseline period, which we broke into five six-second epochs to match the five six-second animation epochs recorded each run. No participant asked to remove the VR headset between runs. Participants spent 35 minutes in the lab to complete all processes.

The Emotiv Epoc X EEG headset[‡] (Emotiv Systems Inc., San Francisco, CA, USA) was used to collect data from participants. Figure 2 shows the Emotiv EEG headset, a participant wearing the headset, and the headset electrode placements.

The Emotiv Epoc X is a 14 channel headset based on the international 10-20 system which includes electrode sites AF4, AF3, F3, F4, F7, F8, FC5, FC6, O1, O2, P7, P8, T7 and T8. M1 and M2 electrode sites served as the reference and ground electrodes, respectively. The Emotiv Epoc X records data at a 256 Hz sampling rate, and filters data online using a built-in digital 5th order Sinc

[†] https://www.mixamo.com/

[‡] https://www.emotiv.com/epoc-x/

a) b) c)

**Figure 2:** *a) The Emotiv Epoc X EEG headset, b) a participant wearing the VR and EEG headsets, and c) the electrodes in the Emotiv Epoc X System (image courtesy of Emotiv).*

filter with a bandwidth of .16-43 Hz. Notch filters at 50 and 60 Hz remove line noise. F3 and F4 electrodes were used in order to calculate FAA [SRSA17].

Raw EEG data was recorded into Emotiv's software, EmotivPro, and then preprocessed using EEGLAB [DM04] in MATLAB (MathWorks). Data was filtered between 2 Hz and 30 Hz using `eegfiltnew`. Bad channels were rejected and then interpolated. An average reference was applied to the data, and Independent Component Analysis (ICA) was run to remove artifacts in the data [DM04]. Eye components were removed from the data with 70% probability. Continuous EEG data was segmented into six-second epochs corresponding to each animation and baseline. Epochs greater or less than 100 $\mu V$ were rejected; this led to the rejection of three participants who did not have enough high quality EEG data left to analyze after this step. Next, `spectopo` from EEGLAB was used to determine the power spectra in the 8-13 Hz alpha range. Finally, the natural log-transformed alpha power of the left electrode (F3) was subtracted from the natural log-transformed alpha power of the right electrode (F4) using the following formula: $(ln[Right] - ln[Left])$ [SRSA17] to determine FAA for each animation and baseline. Next, all FAA values were normalized between zero and one, as original FAA values included both positive and negative values. Finally, baseline FAA was subtracted from each animation FAA to determine FAA elicited from viewing the virtual character animations.

### 3.4. Evaluation and Annotation

All FAA scores were normally distributed. The distribution of FAA values is shown in Figure 3. The goal of preliminary data collection was to determine pairs of animations which significantly differed in FAA value. In this way, we could identify salient animations to examine in the primary study. One-tailed paired samples t-tests were used to determine statistical significance. Descriptive and statistically significant findings are discussed below. For readability, means and standard deviations (SDs) of all animation FAA values are presented in Table 1. Findings from preliminary data collection indicate a dynamic range of FAA values, with the highest FAA value significantly greater than the lowest FAA value.

Of the 12 animations, *Terrified* elicited the lowest FAA value, while *Jump* elicited the highest FAA value. We determined that *Terrified* was significantly lower than *Jump* ($t[16] = -3.017$, $p = .004$), as well as lower than *Excited* ($t[16] = -1.969$, $p = .033$). While Figure 3 shows that *Terrified* is also lower than both *Laughing* and *Idle*, neither difference reaches statistical significance at ($t[16] = -1.682$, $p = .056$) for *Laughing*, and ($t[16] = -1.666$, $p = .058$) for *Idle*, despite a trend towards significance. The higher standard deviations of *Laughing* and *Idle* compared to *Excited* and *Jump* may indicate higher participant variability in response to the former animations in particular. Please see Figure 4 for mean FAA values with error bars. We additionally determined that *Dance* FAA was significantly lower than *Jump* FAA ($t[16] = -1.986$, $p = .032$). Considering both the distribution of FAA values and the need to provide adequate animation variety for the primary study, six animations were annotated. *Terrified* and *Dance* were annotated as *low FAA* animations, while *Jump*, *Excited*, *Laughing* and *Idle* were annotated as *high FAA* animations.

Increases in FAA can be the result of either an increase in left prefrontal cortical activity, *or* a decrease in right prefrontal cortical activity; both of which can influence an FAA score [SRSA17]. Considering the valence model of affect, which proposes that greater left prefrontal cortical activity relative to right prefrontal cortical activity is associated with positive emotions and approach motivation [DES*90, KLF*21], *Terrified* and *Jump* FAA values are as expected. However, *Dance* was also significantly lower than *Jump* in FAA value. Additionally, we expected *Cry* and *Dismiss*, sad and off-putting animations, respectively, to elicit lower FAA values.

Here, we offer two brief interpretations of these findings. First, it is possible that certain animations were not salient enough to elicit a large emotional response from participants. As many animations elicited similar FAA responses, it is possible that many have similarly small emotional attributes. For example, *Cry* and *Clap* were not statistically significantly different. Considering the valence model of FAA [DES*90], which states that positive emotions may increase FAA, we expected *Clap* to elicit a significantly greater FAA response than *Cry*. Second, while multiple models of FAA exist, no model has been proven to be universally valid [KLF*21]. It is unclear if FAA may more appropriately be understood as an indicator of affective response [DES*90], motivational direction [HJG18], or effortful control of emotions [LNG20]. Approach motivation has also been linked to anger, despite anger being a negatively valenced emotion [HJG18]. Therefore, it is possible that *Dismiss*, for example, may have elicited approach motivation as seen through higher FAA values, as the *Dismiss* animation could be considered evocative of anger. Regardless of complexities in the relationship between affective responses and FAA values, we annotated the most salient animation events with FAA values with the goal of understanding a) if *low FAA* and *high FAA* annotated animations would elicit similar FAA responses in a following investigation (the primary study), and b) how new combinations of animations would influence FAA responses.

## 4. Methods

Details concerning the methodology of the primary study which differ from that of preliminary data collection are presented here.

**Table 1:** *Preliminary data collection: FAA values.*

| Animation Event | Mean | SD |
|---|---|---|
| Excited | .0199 | .1801 |
| Dance | -.0189 | .2231 |
| Point | -.0022 | .1779 |
| Rejected | .0171 | .1841 |
| Terrified | -.0230 | .1659 |
| Cry | -.0027 | .1616 |
| Dismiss | .0222 | .2006 |
| Clap | .0047 | .1762 |
| Idle | .0417 | .2196 |
| Jump | .0551 | .1991 |
| Laughing | .0282 | .2308 |
| Surprise | .0185 | .2115 |

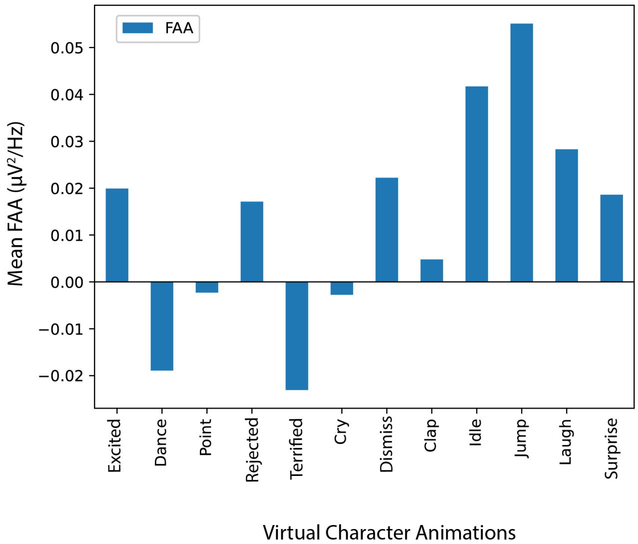


**Figure 3:** *Mean FAA scores from twelve virtual character animations presented in virtual reality.*

### 4.1. Primary Study

Twenty-six students participated in the primary study. Three participants were removed from the statistical analyses due to poor EEG data quality, likely from a poor EEG headset fit. Therefore, 23 students were included in this study (four female; age $M = 24.09$, $SD = 4.66$, two left-handed, two ambidextrous). *Character Count* (one or four characters), and *Annotated FAA* (low or high previously annotated FAA) served as the two factors. We tested the difference between four conditions using a $2 \times 2$ repeated measures Analysis of Variance (ANOVA), a medium effect size .25 (partial $\eta^2 = .06$) [Coh88], and an alpha of .05. G*Power3 [FELB07] was used to conduct an *a priori* power analysis. This analysis showed that 23 participants were necessary to achieve a power of .80. Participants experienced one experimental run in this study. Therefore, each condition was displayed five times for a total of five trials per condition. The four conditions are as follows: **OneLow** (*one character*, annotated with *low FAA*; *Terrified* animation), **OneHigh** (*one character*, annotated with *high FAA*; *Jump* animation), **FourLow** (*four characters*, annotated with *low FAA*; one *Terrified* and three *Dance* animations), and **FourHigh** (*four characters*, annotated with *high FAA*; one *Jump*, *Excited*, *Laughing* and *Idle* animation). These are shown in Figure 5.

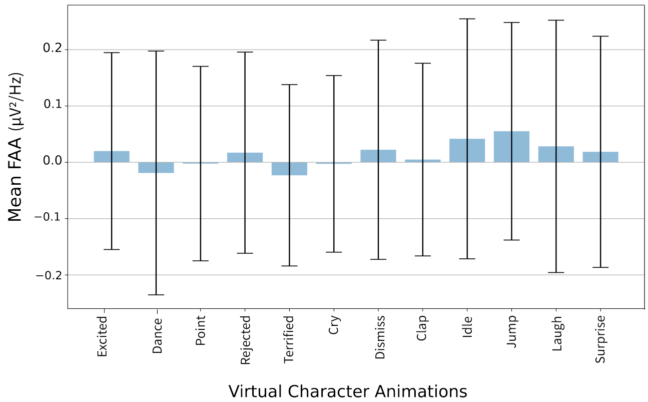


**Figure 4:** *Mean FAA scores from twelve virtual character animations, with error bars with a 95% confidence limit. Note the difference in scale between Figure 3 and Figure 4.*

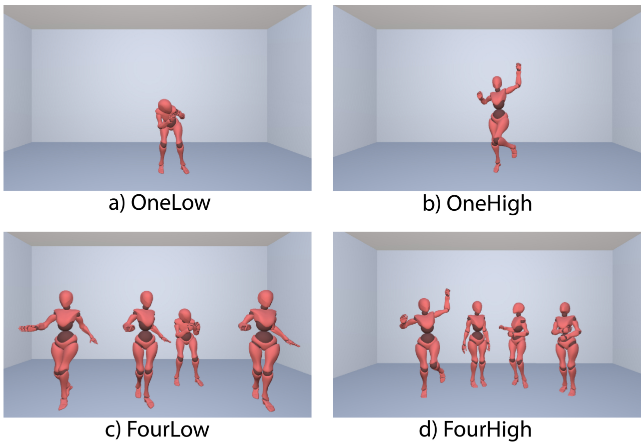


**Figure 5:** *Four conditions used in the primary study.*

### 4.2. Questionnaire

After the experiment, the researcher removed both headsets and the participant filled out the questionnaire. Four video clips corresponding to the four conditions were included on the questionnaire, and were randomized within the survey system. Participants rated the positive quality of each condition after viewing each video. The word choice descriptor for the question was taken from the Positive and Negative Affect Schedule (PANAS) scale [WCT88]. For each video, participants were asked to use a slider to select how *Enthusiastic/Positive* the scene made them feel. Sliders could be dragged between one, for a low positive value, and seven, for a high positive value. We also used a one-item measurement of presence, defined

**Table 2:** *Primary work: FAA and self-reported values.*

| Condition | Measurement | Mean | SD |
|---|---|---|---|
| OneLow | FAA | .1037 | .1387 |
| OneHigh | FAA | .0386 | .1329 |
| FourLow | FAA | .0672 | .1189 |
| FourHigh | FAA | .0866 | .1405 |
| OneLow | Self-reported | 1.78 | 1.278 |
| OneHigh | Self-reported | 5.78 | 1.536 |
| FourLow | Self-reported | 4.13 | 2.007 |
| FourHigh | Self-reported | 4.74 | 1.573 |

as "the participant's sense of 'being there' in the virtual environment" [SUS94], as emotional responses in VR can be influenced by presence [DAP*15]. Lastly, participants had the option to provide written commentary about their experience.

## 5. Results

In this section, we present our results concerning FAA and self-reported affective responses to the four conditions. All data is provided at the link in our data availability statement.

### 5.1. FAA Data

We investigated OneLow, OneHigh, FourLow and FourHigh conditions. We used Greenhouse-Geisser corrections when Mauchley's test of sphericity was violated. All data was normally distributed as assessed by the Shapiro-Wilk test. A two-way repeated measures Analysis of Variance (ANOVA) was used to explore differences in FAA between the two levels of *Character Count* and *Annotated FAA*. Means and standard deviations (SDs) are shown in Table 2.

We determined a significant interaction effect between *Character Count* and *Annotated FAA* ($F[1,22] = 7.725$, $p = .011$, $\eta_p^2 = .260$), shown in Figure 6. Simple effects analysis showed that the *low FAA* condition, *Terrified*, elicited higher FAA than the *high FAA* condition, *Jump*, when *one character* was presented ($F[1,22] = 13.483$, $p = .001$, $\eta_p^2 = .380$). No other simple effects results were statistically significant. Therefore, we determined no significant differences in FAA responses between *one character* and *four character* conditions, nor differences in FAA responses between *low FAA* and *high FAA* during *four character* presentation.

### 5.2. Self-Reported Data

Data concerning the two *four character* conditions were normally distributed. However, self-reported data concerning the two *one character* conditions were not normally distributed according to the Shapiro-Wilk test. Therefore, non-parametric analyses, Wilcoxson Signed Rank tests, were used to understand how *Character Count* and *Annotated FAA* influenced self-reported emotion. First, we compared *one character* and *four character* conditions with *low FAA*. A Wilcoxon Signed-Rank two-tailed test indicated that OneLow (mean rank = 8.75) was rated lower than FourLow (mean rank = 11.78), $Z = -3.559, p < 0.001$. OneHigh (mean rank = 10.04) was rated higher than FourHigh (mean rank = 5.63),

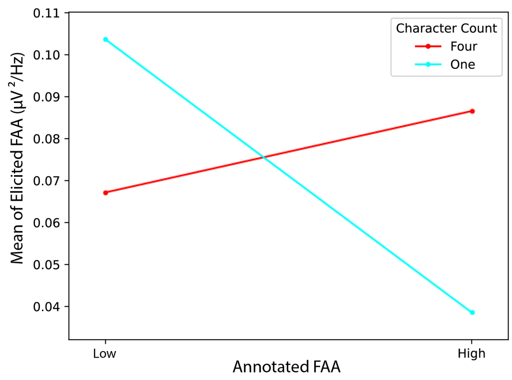


**Figure 6:** *When one character is presented, the low FAA condition elicits higher FAA than the high FAA.*

$Z = -2.590, p = .010$. These results suggest that *one character* conditions may have produced more consistent affective responses than *four character* conditions. Comparing the *one character* conditions, OneHigh (mean rank = 12.39) was rated higher in positive affect than OneLow annotated (mean rank = 3.50), as expected, $Z = -4.120, p < .001$. Lastly, FourLow (mean rank = 8.29) and FourHigh (mean rank = 10.27) did not differ in positive affect rating, $Z = -1.216, p = .224$. Please see Table 2 for Means and Standard Deviations (SDs). Participants indicated they felt present in the virtual environment, with presence scores at ($M = 3.04$, $SD = 0.71$) on a four-point Likert scale.

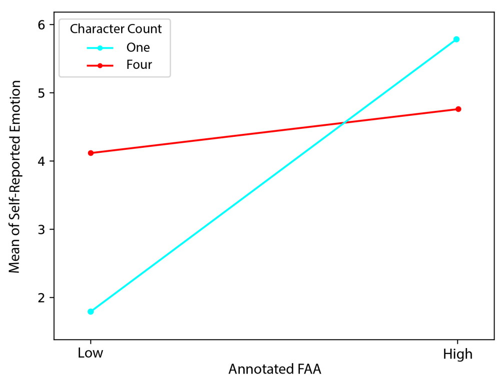


**Figure 7:** *When one character is presented, the low FAA condition elicits lower self-reported positive emotion than the high FAA condition.*

## 6. Discussion

From preliminary data collection, we selected six animations to use in our primary $2 \times 2$ study investigating *Character Count* (*one character* and *four characters*), and *Annotated FAA* (*low FAA* and *high FAA* annotations). In a primary study, we investigated both

FAA and self-reported affective responses to these four conditions. We determined an interaction effect between *Character Count* and *Annotated FAA* concerning FAA. Our findings indicate that when *one character* was presented, the *low FAA* condition (*Terrified*) elicited higher FAA than the *high FAA* condition (*Jump*); OneLow > OneHigh. Therefore, we reject our hypothesis concerning brain activity data. Additionally, no significant differences in FAA responses when *four characters* were presented, nor significant differences in FAA between *one character* and *four character* conditions were found.

Unlike brain activity data, self-reported data is in line with our expectations: OneHigh was rated significantly more positively than OneLow. Additionally, OneLow was rated significantly lower in positive affect than FourLow. We also found that OneHigh was rated more positively than FourHigh. Similar to the FAA data, there was no difference in self-reported emotion between FourLow and FourHigh. Both FAA and self-reported data also suggest that conditions with *one character* rather than *four characters* elicited greater emotional responses, regardless of the direction of valence, perhaps indicative of greater response variability to the *four character* groups of different animations.

Despite selecting the most salient animations from preliminary data collection, it is possible that these six annotated animations did not elicit a sufficiently dynamic range of affective responses, as mentioned in Section 3.4. Having a small range of affective responses may have contributed to similarities in emotional responses to the *four character* conditions. Simplicity of both the virtual environment and characters, combined with repeated condition trials and lack of an active participant task may have led to an altogether disengaging experience for the participant. In particular, one participant mentioned it was "...difficult to connect emotionally given the style of the avatars," referring to the plain mannequin characters. Considering that user preferences for virtual character rendering style may depend on the user's task [RUB14], it is likely that users would have preferred a more life-like humanoid character for the task of considering how they felt. However, results suggest that some emotion was elicited, as certain conditions differed in FAA and self-reported responses. Although an animated character may elicit a greater emotional response than a static character [WBA*14], emotional responses would have likely been stronger with humanoid characters with which participants could connect.

Participant commentary on the questionnaire seems to inform our findings that *one character* conditions may have been easier to respond to emotionally than *four character* conditions. For example, one participant stated that "the scenes with *one character* made it easier to feel the emotions of the character." This comment is interesting for two reasons. First, the comment is in line with our results, in that FAA responses significantly differed when *one character* was present, but not when *four character* were present. Other comments such as "the scene where a character was cowering alone (*Terrified*) was the most impactful in my opinion," "the person jumping up and down (*Jump*) was very clearly excited," as well as "the jumping one (*Jump*) is more emotional," all highlight the salience of both the *Terrified* and *Jump one character* conditions specifically. Secondly, the first comment also suggests that the *one character* condition allowed him to *feel the emotions of the character*, suggesting empathy rather than a consideration of his personal feelings during the experience. User expectations about a character's behavior increase as virtual character realism increases [POU*22]. Therefore, it is likely that participants expected little from the characters in this study. Additionally, virtual character facial display is thought to be an important channel for affect conveyance [BFRS10]. Despite potentially low expectations and limited affect derived from virtual character facial expressions, participant commentary reveals potential empathy strategies used to interact with characters during the study.

Research has shown that higher empathetic responding may be associated with greater right frontal cortical activity and lower FAA [AHT20, THJI12]. If we consider that the *Jump* condition may have elicited greater feelings of empathy than *Terrified*, FAA responses could reflect empathy. Additionally, if participants felt guilt concerning the *Terrified* character, FAA scores may have been higher, as guilt can elicit increased approach motivation concerning correcting a negative situation [HJG18]. However, these points do not explain the reversal of FAA responses from preliminary data collection to the primary study concerning *one character* conditions. Concerning the *four character* conditions, both FAA responses and self-reported data showed no difference in emotion between *four character* conditions. *Four character* conditions may have had a larger range of interpretation, and therefore, affective perception, as clearly mentioned by one participant: "The scenes where multiple figures were present seemed more confusing than anything else, as they seemed to convey different emotions depending on which character I happened to notice first."

Many participants asked about the meaning of the *four character* scenes. Some participants asked why one of the characters stood still while the others were happy, referring to the *Idle* character in the four character *high FAA* annotated scene. *Idle*, as well as the other animations may have taken on a new meaning when presented next to the other three animations, despite all being previously annotated as *high FAA*. In a study examining user's preferences for personal space, it was found that participants kept a larger distance between themselves and virtual characters showing angry emotions rather than happy emotions, but *only when one virtual character was present, rather than a group* [BRO*18]. The authors determined no effect of emotion for the group of virtual characters. Therefore, it is possible that groups of virtual characters may influence participant emotions differently than single characters alone.

When considering the complex relationship between motivation and affect, in which affect may motivate behavior [LNG20], it appears that seemingly small changes, such as the number of characters present in the virtual environment and the addition of animation combinations, may have had significant influence on scene context, and in turn, affective properties of the scenes. Clearly, the context which may have arisen from the combination of the *four character* conditions may have been more influential than previous FAA annotations for FAA elicited in the primary study. For example, one participant mentioned trying to "make a story of the animations... the addition of the story helped me understand the emotion of the virtual characters," again also suggesting empathy.

Considering potential differences in contextual interpretation, certain scenes may have had longer carryover effects than others. Although prolonging the duration of the study, a longer ISI may have been helpful for understanding responses to each condition separately. If confusion from the *four character* conditions bled into *one character* conditions, participants may have tried to relate the *Terrified* single animation to the *four character* condition in which a *Terrified* character is surrounded by three *Dance* characters. FAA may have been higher during *Terrified*, as participants may have exerted significant effort to understand the meaning of this animation. Reappraisal, or changing how one thinks about something emotional, can increase FAA during negative image viewing [CSMW16], which could explain FAA results concerning our *Terrified one character* condition in the primary study.

One participant suggested that the *four character* conditions were anxiety-inducing, while the *one character* scene was not, stating that "lots of movements in scenes made [her] feel more anxious...slight movements seemed less anxiety-inducing." Research suggests a relationship between anxiety and FAA, with greater right frontal lateralization in anxious participants [MWHK08]. One study [VMD*14] found that the speed and number of peaks in the movement trajectory of an animation could explain much of the variation in participant's responses to the animations. Considering the sudden trigger of the *Terrified* animation, this animation may have been particularly stimulating due to a large motion peak at the beginning of this six second animation. In summary, results may have been significantly influenced by the *four character* conditions, especially if emotional effects from their arising contextual interpretations carried over into *one character* conditions.

### 6.1. Future Directions

Considering our results and lack thereof, we provide multiple suggestions for future research which seeks to elicit affective responses based on previously annotated affective data. First, we recommend investigating affective data which is highly engaging, rather than isolating single factors, as the latter may result in homogeneous responses across different factors. It will be important to consider that groups of characters may elicit different emotional responses than single characters alone, and may change contextual and affective properties of the environment in unexpected ways. We suggest exploring affective responses using human virtual characters with facial expressions in a fully furnished virtual environment. Because FAA may increase or decrease due to a wide variety of factors, future research may benefit by first exploring dyadic virtual character interactions, and then moving towards groups of three or more characters. Differences in animation motion properties such as speed should be considered prior to annotation.

While FAA can be considered a correlate of emotion, its relationship with emotion is not fully understand, as there exists no universally accepted model concerning FAA [KLF*21]. In future research, it may be more successful to use emotion classification techniques [SMT20, NWSL11] for EEG-annotated emotional stimuli. The inclusion of cortisol samples from participants may also help inform measures of emotion from EEG data [HJC*18]. While previous work suggests differences in FAA responses between male and female participants [DSPB76, ABH*16], many FAA studies do not report sex differences [RA18]. Because males and females may process emotion differently [SFA*15], it would be important to include an equal amount of males and females and investigate potential sex differences in FAA responding in future work as well.

Ensuring participant homogeneity in anxiety and depression between preliminary and primary data collection periods is suggested if working with FAA data. Including an active behavioral task may also help to increase emotional engagement across participants. While it remains difficult to balance the measurement of enough high-quality EEG data with participant comfort and engagement, factors such as the lack of an active task may negatively influence both attention and emotion of participants.

As the ability to be introspective about one's emotional state varies from person to person, including a small self-reported measurement of emotion after each condition trial in the VR headset may be helpful in focusing participants on the task of thinking about how they felt. While this may also allow for better temporal accuracy concerning self-reported data, it may add significant time to the experiment. Continued work towards acquiring objective and subjective emotional response data simultaneously is necessary. Most interesting and unexpected in our findings is the reversal of FAA responses for *Terrified* and *Jump one character* animations from preliminary data collection to the primary study. Clearly, replication studies which take our suggestions and shortcomings into account will be informative.

## 7. Conclusion

In this paper, we discuss both preliminary data collection in which we annotate six animations as *low FAA* and *high FAA*, and a primary study investigating the influence of *Character Count* (*one character* vs. *four characters*) and *Annotated FAA* (*low FAA* vs. *high FAA*) on elicited FAA and self-reported affect through a $2 \times 2$ repeated measures design. We reject our hypothesis concerning brain activity data, as results indicate that the *one character low FAA* condition, *Terrified*, elicited significantly higher FAA than the *one character high FAA* condition, *Jump* in the primary study. Self-reported data shows that during *one character* conditions, positive affect is as expected; the *high FAA* condition is rated higher in positive affect than the *low FAA* condition. Both FAA responses and self-reported data indicate *four character* conditions may have a larger variability in their perception than *one character* conditions. Limitations of our study which may have contributed to these findings include the simplicity of the virtual characters and environment, unknown contextual effects arising from combinations of animations, and low participant engagement. To address these concerns in future work, we suggest researchers annotate animation events which are highly engaging, include additional means of affect conveyance, such as facial expressions, and consider how seemingly small changes (such as character count) may have unexpected influences on perception and emotional responses. It is our hope that this paper may be informative for researchers who seek to elicit specific affective responses based on previously annotated data.

### Data Availability Statement

The datasets for this study can be found online at: https://osf.io/c59xv/?view_only=3606d1c636be4fae8dc820e993ccfee2

## References


[ABH*16] ARNS M., BRUDER G., HEGERL U., SPOONER C., PALMER D. M., ETKIN A., FALLAHPOUR K., GATT J. M., HIRSHBERG L., GORDON E.: Eeg alpha asymmetry as a gender-specific predictor of outcome to acute treatment with different antidepressant medications in the randomized ispot-d study. *Clinical Neurophysiology 127*, 1 (2016), 509–519. 8

[ADGY04] ATKINSON A. P., DITTRICH W. H., GEMMELL A. J., YOUNG A. W.: Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception 33*, 6 (2004), 717–746. 2

[AHT20] ALIMARDANI M., HERMANS A., TINGA A. M.: Assessment of empathy in an affective vr environment using eeg signals. *arXiv preprint arXiv:2003.10886* (2020). 7

[ATD07] ATKINSON A. P., TUNSTALL M. L., DITTRICH W. H.: Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures. *Cognition 104*, 1 (2007), 59–72. 2

[BFRS10] BICKMORE T. W., FERNANDO R., RING L., SCHULMAN D.: Empathic touch by relational agents. *IEEE Transactions on Affective Computing 1*, 1 (2010), 60–71. 7

[BMM*20] BĂLAN O., MOISE G., MOLDOVEANU A., LEORDEANU M., MOLDOVEANU F.: An investigation of various machine and deep learning techniques applied in automatic fear level detection and acrophobia virtual therapy. *Sensors 20*, 2 (2020), 496. 1

[BR07] BERNHARDT D., ROBINSON P.: Detecting affect from non-stylised body motions. In *International conference on affective computing and intelligent interaction* (2007), Springer, pp. 59–70. 2

[BRO*18] BÖNSCH A., RADKE S., OVERATH H., ASCHÉ L. M., WENDT J., VIERJAHN T., HABEL U., KUHLEN T. W.: Social vr: How personal space is affected by virtual agents' emotions. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2018), IEEE, pp. 199–206. 7

[Coh88] COHEN J.: *Statistical Power Analysis for the Behavioral Sciences*, 2 ed. Lawrence Erlbaum Associates, New Jersey, USA, 1988. 5

[Cou04] COULSON M.: Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. *Journal of nonverbal behavior 28*, 2 (2004), 117–139. 2

[CSMW16] CHOI D., SEKIYA T., MINOTE N., WATANUKI S.: Relative left frontal activity in reappraisal and suppression of negative emotion: Evidence from frontal alpha asymmetry (faa). *International Journal of Psychophysiology 109* (2016), 37–44. 8

[DAP*15] DIEMER J., ALPERS G. W., PEPERKORN H. M., SHIBAN Y., MÜHLBERGER A.: The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Frontiers in psychology 6* (2015), 26. 6

[DES*90] DAVIDSON R. J., EKMAN P., SARON C., SENULIS J., FRIESEN W.: Emotional expression and brain physiology i: Approach/withdrawal and cerebral asymmetry. *Journal of Personality and Social Psychology 58*, 2 (1990), 330–41. 1, 4

[DG09] DE GELDER B.: Why bodies? twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences 364*, 1535 (2009), 3475–3484. 3

[DM04] DELORME A., MAKEIG S.: Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods 134*, 1 (2004), 9–21. 4

[DSPB76] DAVIDSON R. J., SCHWARTZ G. E., PUGASH E., BROMFIELD E.: Sex differences in patterns of eeg asymmetry. *Biological Psychology 4*, 2 (1976), 119–137. 8

[EE12] ENNIS C., EGGES A.: Perception of complex emotional body language of a virtual character. In *International conference on motion in games* (2012), Springer, pp. 112–121. 2

[FELB07] FAUL F., ERDFELDER E., LANG A.-G., BUCHNER A.: G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods 39*, 2 (2007), 175–191. 5

[GFPRFS14] GONZÁLEZ-FRANCO M., PECK T. C., RODRÍGUEZ-FORNELLS A., SLATER M.: A threat to a virtual hand elicits motor cortex activation. *Experimental brain research 232*, 3 (2014), 875–887. 1

[HH06] HIRAI M., HIRAKI K.: The relative importance of spatial versus temporal structure in the perception of biological motion: an event-related potential study. *Cognition 99*, 1 (2006), B15–B29. 2

[HJC*18] HWANG S., JEBELLI H., CHOI B., CHOI M., LEE S.: Measuring workers' emotional state during construction tasks using wearable eeg. *Journal of Construction Engineering and Management 144*, 7 (2018), 04018050. 8

[HJG18] HARMON-JONES E., GABLE P. A.: On the role of asymmetric frontal cortical activity in approach and withdrawal motivation: An updated review of the evidence. *Psychophysiology 55*, 1 (2018), e12879. 1, 4, 7

[KCC*18] KIM A., CHANG M., CHOI Y., JEON S., LEE K.: The effect of immersion on emotional responses to film viewing in a virtual environment. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2018), IEEE, pp. 601–602. 1

[KKB10] KARG M., KÜHNLENZ K., BUSS M.: Recognition of affect based on gait patterns. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 40*, 4 (2010), 1050–1061. 2

[KLF*21] KISKER J., LANGE L., FLINKENFLÜGEL K., KAUP M., LABERSWEILER N., TETENBORG F., OTT P., GUNDLER C., GRUBER T., OSINSKY R., ET AL.: Authentic fear responses in virtual reality: A mobile eeg study on affective, behavioral and electrophysiological correlates of fear. *Frontiers in Virtual Reality* (2021), 106. 1, 4, 8

[KSL18] KEROUS B., SKOLA F., LIAROKAPIS F.: Eeg-based bci and video games: a progress report. *Virtual Reality 22*, 2 (2018), 119–135. 1

[LBC*97] LANG P. J., BRADLEY M. M., CUTHBERT B. N., ET AL.: International affective picture system (iaps): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention 1*, 39-58 (1997), 3. 3

[LBT*19] LAMMERS S., BENTE G., TEPEST R., JORDING M., ROTH D., VOGELEY K.: Introducing acass: an annotated character animation stimulus set for controlled (e) motion perception studies. *Frontiers in Robotics and AI* (2019), 94. 3

[LNG20] LACEY M. F., NEAL L. B., GABLE P. A.: Effortful control of motivation, not withdrawal motivation, relates to greater right frontal asymmetry. *International Journal of Psychophysiology 147* (2020), 18–25. 4, 7

[LWE*19] LANIER M., WADDELL T. F., ELSON M., TAMUL D. J., IVORY J. D., PRZYBYLSKI A.: Virtual reality check: Statistical power, reported results, and the validity of research on the psychology of virtual reality and immersive environments. *Computers in Human Behavior 100* (2019), 70–78. 2

[MF69] MEHRABIAN A., FRIAR J. T.: Encoding of attitude by a seated communicator via posture and position cues. *Journal of Consulting and Clinical Psychology 33*, 3 (1969), 330. 2

[MLB*10] METALLINOU A., LEE C.-C., BUSSO C., CARNICKE S., NARAYANAN S., ET AL.: The usc creativeit database: A multimodal database of theatrical improvisation. *Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality* (2010), 55. 2

[MPP06] MA Y., PATERSON H. M., POLLICK F. E.: A motion capture library for the study of identity, gender, and emotion perception from biological motion. *Behavior research methods 38*, 1 (2006), 134–141. 2

[MWHK08] MATHERSUL D., WILLIAMS L. M., HOPKINSON P. J., KEMP A. H.: Investigating models of affect: relationships among eeg alpha asymmetry, depression, and anxiety. *Emotion 8*, 4 (2008), 560. 8

[New93] NEWLOVE J.: *Laban for actors and dancers: putting Laban's movement theory into practice: a step-by-step guide*. Theatre Arts Books, 1993. 2

[NLK*13] NORMOYLE A., LIU F., KAPADIA M., BADLER N. I., JÖRG S.: The effect of posture and dynamics on the perception of emotion. In *Proceedings of the ACM symposium on applied perception* (2013), pp. 91–98. 2

[NMB*13] NIEWIADOMSKI R., MANCINI M., BAUR T., VARNI G., GRIFFIN H., AUNG M. S.: Mmli: Multimodal multiperson corpus of laughter in interaction. In *International Workshop on Human Behavior Understanding* (2013), Springer, pp. 184–195. 2

[NWSL11] NIE D., WANG X.-W., SHI L.-C., LU B.-L.: Eeg-based emotion recognition during watching movies. In *2011 5th International IEEE/EMBS Conference on Neural Engineering* (2011), IEEE, pp. 667–670. 8

[POU*22] PARMAR D., OLAFSSON S., UTAMI D., MURALI P., BICKMORE T.: Designing empathic virtual agents: manipulating animation, voice, rendering, and empathy to create persuasive agents. *Autonomous Agents and Multi-Agent Systems 36*, 1 (2022), 1–24. 7

[PPS01] PATERSON H. M., POLLICK F. E., SANFORD A. J.: The role of velocity in affect discrimination. In *Proceedings of the annual meeting of the cognitive science society* (2001), vol. 23. 2

[RA18] REZNIK S. J., ALLEN J. J.: Frontal asymmetry as a mediator and moderator of emotion: An updated review. *Psychophysiology 55*, 1 (2018), e12965. 8

[RMMH18] RODRIGUES J., MÜLLER M., MÜHLBERGER A., HEWIG J.: Mind the movement: Frontal asymmetry stands for behavioral motivation, bilateral frontal activation for behavior. *Psychophysiology 55*, 1 (2018), e12908. 1

[ROCG09] ROETHER C. L., OMLOR L., CHRISTENSEN A., GIESE M. A.: Critical features for the perception of emotion from gait. *Journal of vision 9*, 6 (2009), 15–15. 2

[RRC*15] RODRÍGUEZ A., REY B., CLEMENTE M., WRZESIEN M., ALCAÑIZ M.: Assessing brain activations associated with emotional regulation during virtual reality mood induction procedures. *Expert Systems with Applications 42*, 3 (2015), 1699–1709. 1

[RUB14] RING L., UTAMI D., BICKMORE T.: The right agent for the job? In *International Conference on Intelligent Virtual Agents* (2014), Springer, pp. 374–384. 7

[SEM19] STOLZ C., ENDRES D., MUELLER E. M.: Threat-conditioned contexts modulate the late positive potential to faces—a mobile eeg/virtual reality study. *Psychophysiology 56*, 4 (2019), e13308. 1

[SFA*15] SPALEK K., FASTENRATH M., ACKERMANN S., AUSCHRA B., COYNEL D., FREY J., GSCHWIND L., HARTMANN F., VAN DER MAAREL N., PAPASSOTIROPOULOS A., ET AL.: Sex-dependent dissociation between emotional appraisal and memory: a large-scale behavioral and fmri study. *Journal of Neuroscience 35*, 3 (2015), 920–935. 8

[SKS*21] SCHÖNE B., KISKER J., SYLVESTER R. S., RADTKE E. L., GRUBER T.: Library for universal virtual reality experiments (luvre): A standardized immersive 3d/360 picture and video database for vr based research. *Current Psychology* (2021), 1–19. 1

[SMT20] SUHAIMI N. S., MOUNTSTEPHENS J., TEO J.: Eeg-based emotion recognition: A state-of-the-art review of current trends and opportunities. *Computational intelligence and neuroscience 2020* (2020). 8

[SRSA17] SMITH E. E., REZNIK S. J., STEWART J. L., ALLEN J. J.: Assessing and conceptualizing frontal eeg asymmetry: An updated primer on recording, processing, analyzing, and interpreting frontal alpha asymmetry. *International Journal of Psychophysiology 111* (2017), 98–114. 4

[SSI03] SAWADA M., SUDA K., ISHII M.: Expression of emotions in dance: Relation between arm movement characteristics and emotion. *Perceptual and motor skills 97*, 3 (2003), 697–708. 2

[SUS94] SLATER M., USOH M., STEED A.: Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments 3*, 2 (1994), 130–144. 6

[THJI12] TULLETT A. M., HARMON-JONES E., INZLICHT M.: Right frontal cortical asymmetry predicts empathic reactions: Support for a link between withdrawal motivation and empathy. *Psychophysiology 49*, 8 (2012), 1145–1153. 7

[ULH20] UHM J.-P., LEE H.-W., HAN J.-W.: Creating sense of presence in a virtual reality experience: Impact on neurophysiological arousal and attitude towards a winter sport. *Sport Management Review 23*, 4 (2020), 588–600. 1

[VDLRBM14a] VOLKOVA E., DE LA ROSA S., BÜLTHOFF H. H., MOHLER B.: The mpi emotional body expressions database for narrative scenarios. *PloS one 9*, 12 (2014), e113647. 2

[VDLRBM14b] VOLKOVA E., DE LA ROSA S., BÜLTHOFF H. H., MOHLER B.: The mpi emotional body expressions database for narrative scenarios. *PloS one 9*, 12 (2014), e113647. 3

[VMD*14] VOLKOVA E. P., MOHLER B. J., DODDS T. J., TESCH J., BÜLTHOFF H. H.: Emotion categorization of body expressions in narrative scenarios. *Frontiers in psychology 5* (2014), 623. 3, 8

[Wal98] WALLBOTT H. G.: Bodily expression of emotion. *European journal of social psychology 28*, 6 (1998), 879–896. 2

[WBA*14] WU Y., BABU S. V., ARMSTRONG R., BERTRAND J. W., LUO J., ROY T., DAILY S. B., DUKES L. C., HODGES L. F., FASOLINO T.: Effects of virtual human animation on emotion contagion in simulated inter-personal experiences. *IEEE transactions on visualization and computer graphics 20*, 4 (2014), 626–635. 7

[WCT88] WATSON D., CLARK L. A., TELLEGEN A.: Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of personality and social psychology 54*, 6 (1988), 1063. 5

[Zha01] ZHAO L.: *Synthesis and acquisition of laban movement analysis qualitative parameters for communicative gestures*. University of Pennsylvania, 2001. 2