How Human Am I? EEG-based Evaluation of Animated Virtual Characters

Maryam Mustafa
Computer Graphics Lab
TU Braunschweig, Germany
mustafa@cg.cs.tu-bs.de

Stefan Guthe
Graphics Capture and
Massively Parallel Computing
TU Darmstadt, Germany
stefan.guthe@cs.tu-darmstadt.de

Jan-Philipp Tauscher
Computer Graphics Lab
TU Braunschweig, Germany
tauscher@cg.cs.tu-bs.de

Michael Goesele
Graphics Capture and
Massively Parallel Computing
TU Darmstadt, Germany
goesele@cs.tu-darmstadt.de

Marcus Magnor
Computer Graphics Lab
TU Braunschweig, Germany
Dept. Physics and Astronomy
Univ. New Mexico, USA
magnor@cg.cs.tu-bs.de

ABSTRACT

There is a continuous effort by animation experts to create increasingly realistic and more human-like digital characters. However, as virtual characters become more human they risk evoking a sense of unease in their audience. This sensation, called the Uncanny Valley effect, is widely acknowledged both in the popular media and scientific research but empirical evidence for the hypothesis has remained inconsistent. In this paper, we investigate the neural responses to computer-generated faces in a cognitive neuroscience study. We record brain activity from participants *(N = 40)* using electroencephalography (EEG) while they watch videos of real humans and computer-generated virtual characters. Our results show distinct differences in neural responses for highly realistic computer-generated faces such as Digital Emily compared with real humans. These differences are unique only to agents that are highly photorealistic, i.e. the 'uncanny' response. Based on these specific neural correlates we train a support vector machine (SVM) to measure the probability of an uncanny response for any given computer-generated character from EEG data. This allows the ordering of animated characters based on their level of 'uncanniness'.

ACM Classification Keywords

H.1.2. User/Machine Systems: human factors, human information processing

CHI 2017, May 06 - May 11, 2017, Denver, CO, USA

ACM 978-1-4503-4655-9/17/05...$15.00.
DOI: http://dx.doi.org/10.1145/3025453.3026043

Author Keywords

EEG; Computer Graphics; Virtual Humans; Uncanny Valley

INTRODUCTION

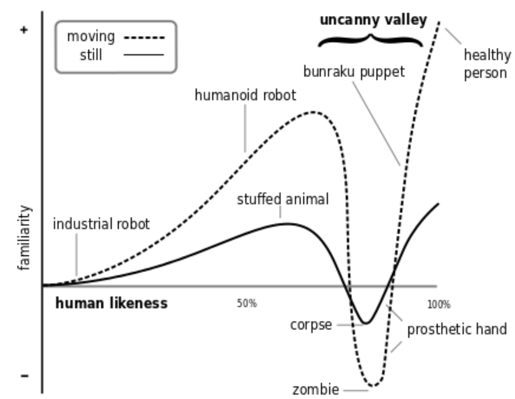


Figure 1: Mori proposed a theory that the more human looking a character becomes the more agreeably it is perceived until the character becomes so nearly human that it causes a response of revulsion, as seen by the valley where the corpse and zombie end up, before the response becomes positive again in response to real humans [30].

In recent years highly realistic computer-generated virtual humans have become ubiquitous in movies, interactive drama video games and as online avatars. Movies such as *The curious case of Benjamin Button* and interactive drama video games such as *Heavy Rain* and *L.A. Noire* have produced highly realistic characters. However, the human reaction to realistic computer-generated (cg) characters is not always a positive one as evidenced by the response to characters in the *The Polar Express* and *Final Fantasy: The Spirits Within* [16]. Both movies were poorly received and the consensus seemed to be that the characters were perceived as disturbing and evoked feelings of discomfort [34]. Intuitively, the more human-like a computer-generated character becomes the more appealing it is to the intended human audience. However, increasing human realism does not necessarily result in more appealing characters. One theory to explain this is the uncanny valley hypothesis put forward in 1970 by Masahiro Mori. He posited that as something starts to look more human it is also perceived as more agreeable until it begins to look so human it evokes a feeling of revulsion, Fig. 1. Mori also hypothesized that the effects would be more pronounced for moving agents as opposed to stationary agents [30]. Although research so far has been inconsistent, most studies have found some evidence in support of Mori's hypothesis [28, 19, 46, 6]. Various theories have been proposed to explain the uncanny valley phenomenon such as disease and threat avoidance [28]. MacDorman and Ishiguro [28] also suggested that the uncanny valley effect might be a result of androids eliciting and violating human expectations about how other humans should look and act. In other words, the more human-like an android or cg human, the more human-like expectations are elicited. The fact that cg humans are incapable of fulfilling these expectations results in a disconnect. Similarly, Saygin et al. [41] discuss the predictive coding hypothesis which is based on a similar idea that the uncanny valley is related to expectation violations in neural computing when the brain encounters human-like agents.

As advances in computing hardware and software give rise to the widespread use of realistic virtual humans, insights into the perception of cg humans and a methodology to empirically quantify this effect would be invaluable. In recent years there have been several studies aiming to understand and find empirical evidence for the uncanny valley [27, 42, 28]. Most studies exploring the uncanny valley have either used subjective rating methodologies [44, 19, 47], or gaze behaviour [9, 43]. Although these and similar studies have provided valuable insight into the uncanny valley phenomenon, there is as yet no methodology for measuring and predicting this effect. Behavioural studies alone are insufficient to test and quantify the uncanny valley effect particularly because they rely on the explicit (overt) output of the cognitive process. This is particularly problematic since the uncanny valley phenomenon is complex and may involve implicit (covert) cognitive processing [48]. Given the complexity of this phenomenon and the incomplete understanding of it, it is difficult to measure it with a single measurement like familiarity, eeriness or pleasantness [39], instead it requires a more refined and continuous measurement. Neuroimaging methods have been used as alternates for the study of the underlying mechanisms of the uncanny valley. Saygin et al. [41] used functional magnetic imaging (fMRI) to explore neural responses to robots, androids and humans. Similarly, Cheetham et al. [10] used fMRI to investigate the processing of human and non-human categories in the brain. While fMRI provides valuable insight into the underlying mechanisms of the uncanny valley, it is unsuited to our purposes. The fMRI Blood-Oxygen-Level Dependent (BOLD) signal is an indirect measure of neural activity and the temporal response of the blood supply, which is the basis of fMRI, is poor relative to the electrical signals that define neuronal communication. Instead, we use Electroencephalography (EEG) data to examine the neural differences in the perception of real and cg faces. EEG hardware is significantly cheaper than fMRI, has a much better temporal resolution than fMRI (EEG's can detect changes within a millisecond time-frame) and allows monitoring of the information processing during stimulus presentation. EEG has also been used in decades of cognitive neuroscience research to inform us about the processing of facial stimuli in the brain. Most of this work has been focused on facial recognition, exploring the selectivity of face-sensitive ERP responses and in trying to locate the brain regions associated with face processing [31, 17, 23]. We use the accumulated knowledge in these areas to explore the implicit human response to realistic computer generated (cg) characters.

In this paper, we propose the application of EEG to explore and evaluate the perception of realistic computer-generated virtual human faces, with a particular focus on predicting the uncanny valley effect. We use recordings from real humans, virtual humans from the Institute of Creative Technologies Virtual Human Toolkit [18] and state of the art computer-generated humans *Digital Emily* [2] and *Digital Ira* [1]. We also include highly realistic characters from interactive drama video games, 'Kara' from *Detroit: Become Human* [12], 'Ernst' from *Squadron 42* [37] and 'HeadTech' from Janimation [22]. These agents are used to represent varying degrees of human-likeness as determined by a perceptual study where participants (*N=80*) rank the perceived realism of the agents on a Likert scale.

We then use an EEG device to investigate the neural responses to the video of each agent. We intentionally use moving humans given Mori's hypothesis that the uncanny effect is exacerbated with motion [30]. Based on the results of these experiments, we use the neural responses to different characters to train a support vector machine (SVM). Our goal is to create a methodology that would allow the perceptual categorization of a cg character with a single trial of EEG data. In other words, that one participant could watch a clip of a cg human once and we would be able to evaluate it as being perceived as uncanny or not based on neural data.

This paper makes two primary contributions: 1) to identify differences in neural components in response to real-human and cg humans ; 2) to provide a comprehensive method for evaluation and categorization of the uncanny response to any computer-generated (cg) virtual human. We propose that our approach would allow designers of animated cg humans to evaluate and measure the 'human-ness' of their characters without the use of multiple participants, or lengthy behavioural studies. It would allow the incremental modification of cg characters to minimize the uncanny valley effect and a better understanding of how cg faces are perceived may help to improve the realism in computer-generated content for movies and games.

RELATED WORK

Face perception has been studied extensively in the cognitive sciences over the last decades. Similarly, achieving completely realistically rendered humans has been a defining goal for graphics researchers. Our work builds on developments and research from both computer graphics and the cognitive sciences.

Computer Graphics and The Uncanny Valley

The pursuit of the perfect virtual human has been the holy grail for graphics researchers for the last few decades [38, 4, 49, 29]. One major difficulty in producing realistic rendered human faces is the uncanny valley effect. In recent years studies have explored this hypothesis using robots and images. MacDorman et.al. [27] conducted studies where a cg character's facial proportions, skin texture, and level of detail were varied to examine the resulting effect on perceived eeriness, human likeness and attractiveness. They evaluated the presented human faces by asking participants to rate each image on an 11-point semantic differential scale. All their studies used manipulations to one base model of a 30-year old male. In contrast, our work looks at the neural responses to a variety of male and female faces in video sequences. We also make a comparison between real faces and computer generated faces. Similarly, Burleigh et al. [6] looked at the relationship between human likeness and eeriness using digital human faces. They observed that stimuli defined by a near-perfect resemblance to humans do not appear to cause a greater negative effect when compared to stimuli with perfect human likeness or no human likeness. Another study by Zell et al. [50] looked at the question of stylization and explored the effect of shape and material on perceived realism. The results showed that shape is the more important factor when rating realism and expression intensity, while material is the main determiner for appeal. Our work is similar in terms of exploring realism in computer generated characters but differs in that we look at very realistic computer generated characters in motion and use an EEG for exploring biological differences in how humans perceive animated cg faces.

Given that emotion is a central aspect of the uncanny experience many studies have explored facial expression of emotion as related to the uncanny valley. Tinwell et al. [46] investigated if 'uncanniness' is increased for a character with a perceived lack of facial expression in the upper parts of the face. Their results indicate that even fully and expertly animated characters are rated as more uncanny than humans and that, in virtual characters, a lack of facial expression in the upper parts of the face during speech exaggerates the uncanniness. Another study conducted by McDonnell et al. [29] explored the uncanny valley from the perspective of perceived trust and rendering style. The study used deception as a basis for the experiments to investigate the UV theory. Their results showed that participants felt a subconscious untrustworthiness towards a high quality virtual character. In contrast Kim et al. [24] looked specifically at gaze behaviour to compare real human perceptions of a virtual human (VH) with their expectancy of the virtual human's gaze behaviour. Our work on the other hand aims to exploit this subconscious processing to predict the human response to virtual characters.

Recently, Fan et al. [14] conducted studies to explore the processes involved in visual realism perception of faces. Their facial stimuli consisted of real face images, scrambled face images, inverted face images, images that show different parts of faces and images of faces with top and bottom misaligned. Their work shows that both holistic and piecemeal processing are involved in visual perception of faces. They also suggest that shading information is more important than colour for holistic processing. Their work focuses on the actual mechanism of perception of faces whereas we focus specifically on the difference in how we perceive real and computer generated faces. Our goal is to be able to find a method for quantifying the feeling of 'uncanniness'.

There have not been many studies within a graphics context that use modalities such as EEG to understand the perception of computer-generated characters. Urgen et al. [48] used EEG data to explore neural components of the perception of motion between androids, robots and humans. Their work focused on verifying that motion exacerbates the uncanny valley effect. Their results show that the event related brain potential N400, which has been associated with violation of predictions, is greater for a moving android than for a stationary android. In contrast we focus on highly realistic virtual humans and the evaluation of cg humans as a whole. Similarly, Saygin et al. [42] used functional magnetic resonance imaging (fMRI) to explore the selectivity of the human action perception (APS) system for the appearance and motion of a human, a robot and an android. Their study found distinctive responses to the mismatch between appearance and motion, where suppression effects for the human and robot were similar to each other but were stronger for the android. Although interesting, this work does not focus specifically on human realism as related to uncanniness, instead focusing on the motion aspect of uncanniness. Another recent study by Mustafa et al. [32] explored the neural response to still images of real, virtual and abstract faces. Their results show a distinct difference in the brains response to different face categories (real, virtual and abstract). In contrast our work focuses on highly realistic animated characters.

Cognitive Studies of Face Perception

Cognitive researchers have conducted many studies using an EEG, in an attempt to understand the exact mechanisms of face perception in the human brain. For example, Bentin et al. [3] studied the event related potentials (ERPs) associated with face perception in comparison to human hands, animal faces and furniture. They also looked at ERPs associated with upright faces, inverted faces, whole faces and isolated face components. Their studies showed that human faces elicited a negative potential called the N170, which was absent from the ERPs evoked by other animate and inanimate stimuli. They further showed that the N170 was delayed for inverted faces but its amplitude did not change. They hypothesize that N170 may reflect the underlying mechanism tuned to face detection. In general, it has been shown by multiple studies that human faces elicit larger N170s than other object categories [5, 21, 40]. These studies provide us with a basis for our work for measuring the evoked potentials in response to rendered faces.

A similar study by George et al. [17] explored the neural basis for normal and scrambled face processing. The stimuli used were three faces produced using one pair of eyes, one pair of noses and a pair of mouths. The scrambled faces were produced by reversing the positions of the eyes and nose only. Their study found no difference between the vertex P2 evoked by faces and scrambled faces, although temporal ERP's between the two conditions were different.

Another study conducted by Jefferys et al. [23] also analysed evoked potentials in response to faces and objects such as shoes, cars and planes. This study specifically looks at the response properties of a distinct scalp potential called the 'vertex positive peak' (VPP). Their results showed that the VPP responds preferentially to suddenly presented faces as opposed to objects. The focus of this study is an exploration of the physiological processes of face perception.

EXPERIMENT DESIGN

Our goal is to develop an EEG-based approach for virtual human evaluation. We conduct two studies ; One is an online perceptual study to determine the varying degrees of human-likeness of our selected agents. The second study was an EEG-based experiment to determine the neural response to the uncanny valley.

Participants

The perceptual study was placed online and open to allow the maximum number of participants to determine how realistic each presented virtual and real human is. The total number of participants in the study were 80.

For the EEG experiment forty right-handed adults (22 female and 18 male; average age = 24) with normal or corrected-to-normal vision and no history of neurological disorders participated. Informed consent was obtained from each participant and participants were paid per hour or received course credit.

The participants for both the studies were distinct and unique.

Stimuli

The stimuli consists of 5-second video clips of real and virtual human agents speaking (Fig. 7). Since we want to focus on the visual response to agents, the audio is not included in the actual stimulus presentation. The real human videos are recorded in our lab (Fig. 7b). Each actor is given the same monologue to recite in a neutral way so as to reduce differences in facial expression.

The most simplistic agent we use is an animated 'comic' character (Fig. 7a). We also use 4 virtual humans (Fig. 7d) from the Virtual Humans Toolkit from the Institute of Creative Technologies [18]. Given that we also want to analyse the neural response to highly realistic cg humans we use 5 characters that are highly realistic in terms of how human-like they appear (Fig. 7d,7e). Digital Emily and Digital Ira are considered state-of-the-art in terms of real-human character animation [1, 2]. We also use realistic game characters 'Kara' from *Detroit: Become Human* [12], 'Ernst' from *Squadron 42* [37] and 'HeadTech' from Janimation [22].

The video clips are chosen for the degree of realism and the lack of emotional expression. Since we are interested in the perception of virtual humans we choose to exclude emotionally heavy content. We refer to these agents as real humans (Fig. 7b) , virtual humans (VH) (Fig. 7c), comic (Fig. 7a) and Emily and Ira human (Fig. 7e). None of the participants had prior experience with the presented stimuli.

Although each agent has a a different background, previous work has shown that the dynamics of face processing identified using ERPs also applies to faces seen in complex, naturalistic scenes [7, 32]. We therefore do not expect the background to affect how the different agents are perceived.

Initial Study

To get an estimate of the perceptual 'human-ness' of each agent we conduct an initial perceptual study and ask participants *(N = 80)* to rate how 'real' each agent appears on a Likert scale from 1 (least human) to 6 (most human) (Fig. 2). We chose a 6-point likert scale because we did not want to allow participants to choose a neutral answer or be undecided. As can be seen the cg agents form a scale in terms of their realism with comic being the least human-like (M = 1, SD = 0) and Emily being the most realistic (M = 5.28, SD = 0.76) in relation to real humans (M = 5.69, SD = 0.75). The questionnaire asked the participants how 'real' each agent is. This is not the same as 'uncanny'. Our use of an initial questionnaire to determine degree of human-ness is similar to a strategy used by Strait et al., to study the uncanny valley [45].

Procedure

Prior knowledge can affect judgements of artificial agents and so each participant is given exactly the same introduction to the experiments and the same exposure to the videos. The participants are also asked at the end of the study if any of the agents were familiar to them. EEG is recorded as participants watch video clips of the agents. The experiment consists of 10 blocks and each video is shown once in each block. All videos are presented in randomized order while ensuring that a video is not repeated on two consecutive trials. Each participant experiences a different randomized stimuli sequence. To prevent an erroneous visual evoked potential at the beginning of each video onset a gray screen with a white fixation cross is displayed. Participants are instructed to fixate the cross at the center of the screen. To focus their attention participants are asked to identify the agent in the video as either real or a virtual human.

EEG Recording and Data Processing

EEG measures the electrical activity of a large number of neurons close to the brain surface. Traditional EEG systems require anywhere from 32 to 64 electrodes to be fitted to the head of a participant at specific locations (Fig. 3). This is usually achieved with a cap of attached electrode positions that is pulled over the head. To ensure conductivity between the electrodes and the scalp, contact gel needs to be applied to the electrodes. We use an EEG with 32 electrodes attached according to the international 10-20 system (Fig 3) [15]. The raw EEG data is low-pass filtered with a stop-band frequency of 50Hz to remove power line noise (5 lobe Lanzcos filter). The data is also high-pass filtered with a stop-band frequency of 0.5Hz to remove baseline drift (3 lobe Lanzcos filter) and re-referenced to average mastoid electrodes. Then the data is

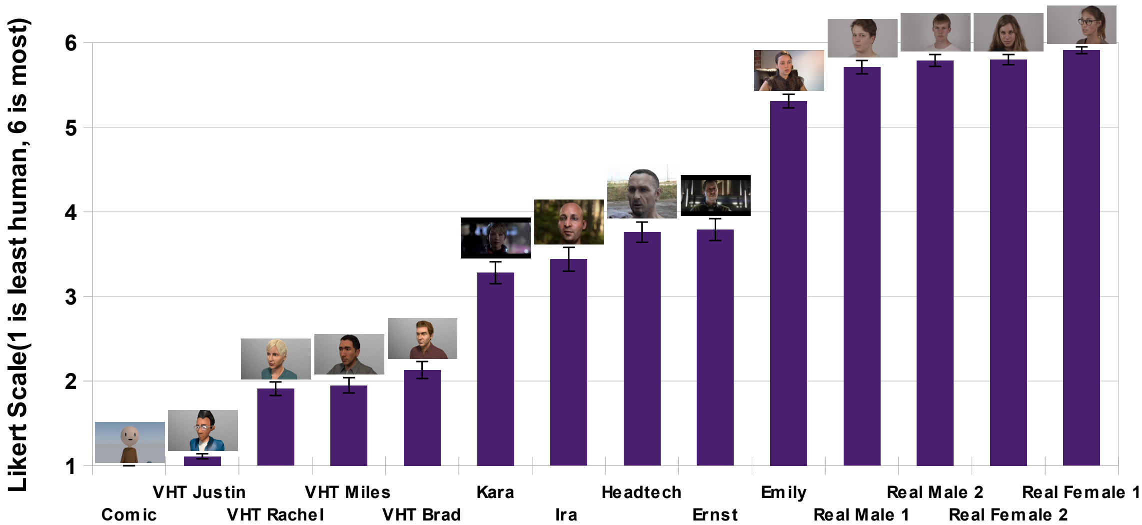


Figure 2: Results of perceptual study to determine how realistic each agent appears *(N = 80)*. Comic was determined to be the least human-like and Emily the closest to photorealistic human.

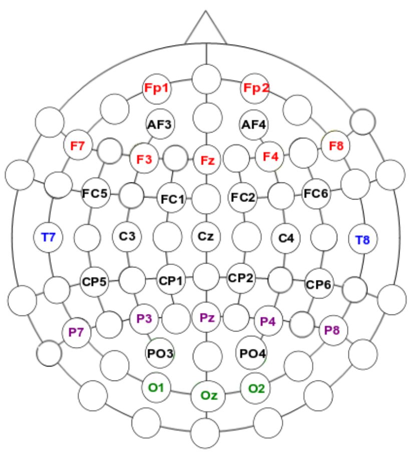


Figure 3: Electrode layout in the 10-20 system. Each electrode captures neural signals from the underlying brain region.

epoched ranging from 100 ms before video to 1000 ms after video onset, and are time-locked to the onset of the video clip. Automatic epoch rejection is based on a 4 channel EOG and the mastoids. We look at only the first 1-second of the neural response as we are interested in investigating the N400 event-related potential (ERP) which occurs approx. 300ms to 600ms after stimulus presentation.

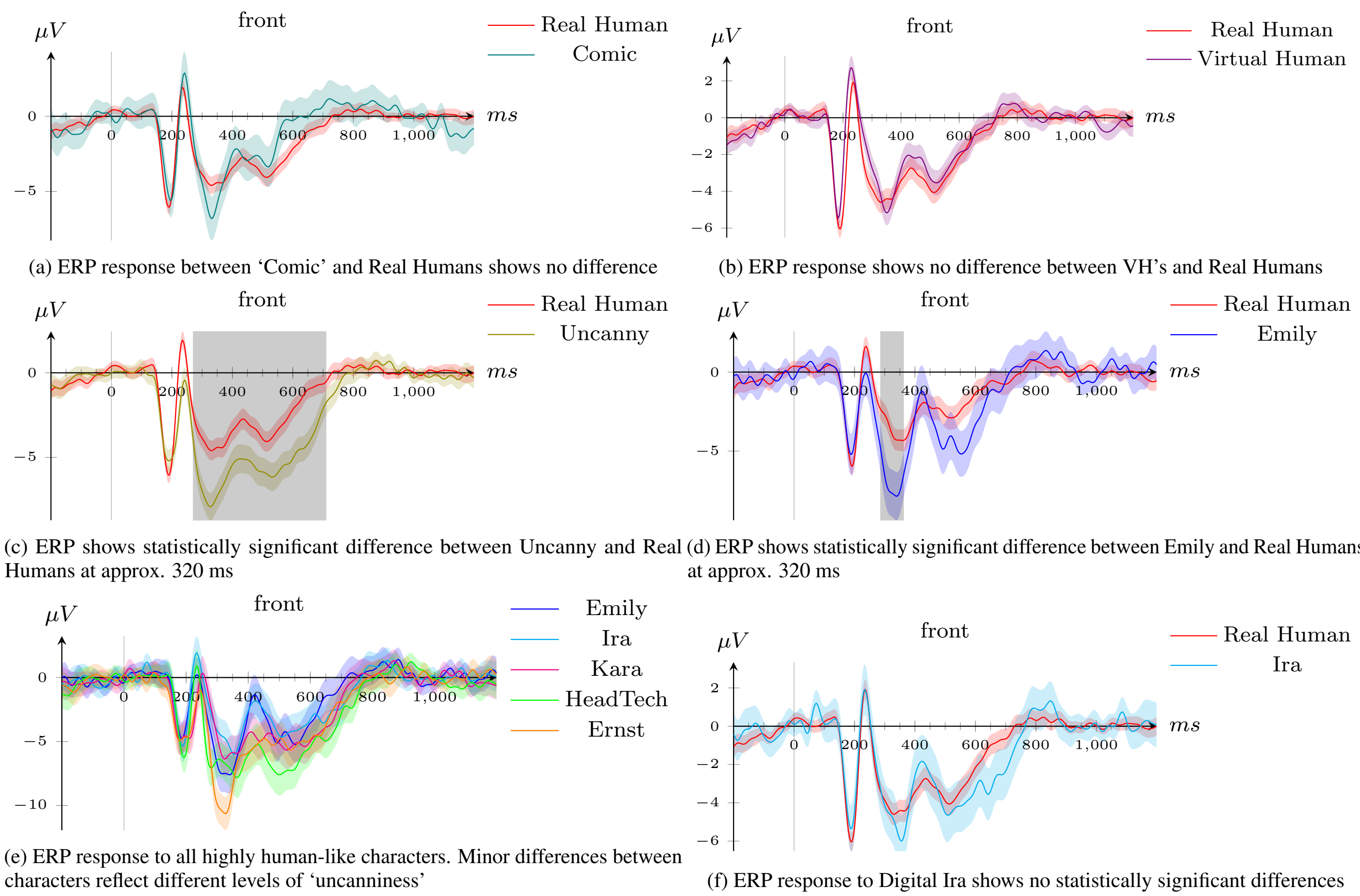
After pre-processing, grand average event-related brain potentials (ERP) for all participants and all trials, were computed and plotted for each character (Fig. 4).

ERP Results

We investigate the EEG data with respect to a specific dependent measure, the N400 event-related potential (ERP) component [25]. The N400 is a negative-going ERP which peaks between 300 ms - 600 ms after stimulus onset and is maximal in fronto-central regions of the human scalp i.e., electrodes Fp1, Fp2, AF3, AF4, F7, F3, Fz, F4, F8 , FC5, FC1, FC2, FC6 (Fig. 3). The N400 is elicited in response to any meaningful stimulus, however its amplitude is greater for anomalous stimuli (i.e., stimuli that violate expectations) [48, 33, 25]. This makes the N400 ideally suited to evaluate the prediction error hypothesis of the uncanny valley [48]. Given that the N400 depicts violations of participants expectations the amplitude should be higher for agents that are more human-like in appearance but are actually not human as compared to the amplitude for real humans.

Fig. 4 shows the event related potentials averaged over all participants and all trials for each agent. The signals used are from the frontal electrodes Fp1, Fp2, AF3, AF4, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6 averaged since previous research shows the N400 component is strongest over these electrodes [25]. The shaded area around each ERP is the 95% confidence interval and shows the deviation between participants and trials. The grey shaded rectangular area shows statistically significantly different components between categories based on this interval.

Our results indicate that observation of all agents elicits an N400 component regardless of the agent type (Fig. 4). Fig. 4a shows the difference in ERP for category comic versus real humans. Since category 'comic' is immediately recognizable as being not real-human and scores low on realism (Fig. 2) there is no expectation of human-ness and hence no violation of expected behavior. The ERP between humans and comic does not show a statistically significant difference in the N400 component (approx. 320 ms).



(a) ERP response between 'Comic' and Real Humans shows no difference

(b) ERP response shows no difference between VH's and Real Humans

(c) ERP shows statistically significant difference between Uncanny and Real Humans at approx. 320 ms

(d) ERP shows statistically significant difference between Emily and Real Humans at approx. 320 ms

(e) ERP response to all highly human-like characters. Minor differences between characters reflect different levels of 'uncanniness'

(f) ERP response to Digital Ira shows no statistically significant differences

Figure 4: ERP responses to the CG and real agents averaged over frontal electrodes. The 'Uncanny' category encompasses 'Kara', 'Ernst' and 'HeadTech' (Fig. 7). The ERPs are time-locked to the start of the videos. The shaded area along each curve shows the standard deviation for the ERP. The grey-shaded rectangular boxes show areas of statistically significant differences between categories.

Similarly, the Virtual Human (VH) stimuli are rated low on realism (Fig. 2) and as expected the ERP response does not show a statistically significant difference in the N400 ERP component (approx.320 ms) as compared to the real human responses.

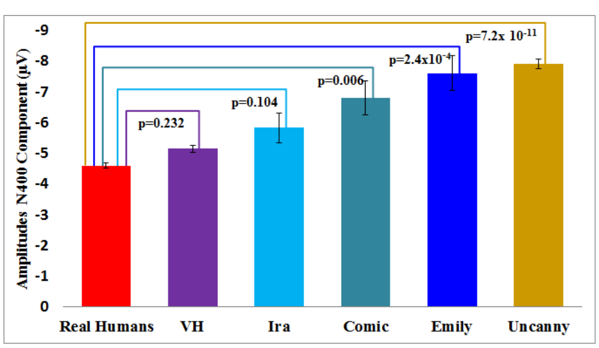


Figure 5: Average amplitudes of N400 component of ERPs from Fig. 4 along with the p-values between agent categories. The amplitudes show an ordering of the characters in terms of 'uncanniness'

However, the uncanny stimuli which were rated in the perceptual study as being close to real humans show a significant difference in the N400 peaking with the greatest difference at approx. 320 ms, as can be seen from Fig. 4c. The ERP responses for all characters within the 'uncanny' category show only minor differences between each other Fig. 4f.

If we look at the N400 response to the Emily stimulus (Fig. 4d), which is the most highly rated in terms of human-ness in our perceptual study (score of 5.2/6, Fig. 2), there is a distinct and significant difference in the N400 component peaking at approx. 320 ms . This is in-line with the previous research which provides evidence that amplitude changes in N400 are associated with the semantic incongruity caused by mismatched expectations [48].

Fig. 5 shows the average peak amplitudes of the N400 component in response to each agent category along with the p-values for the comparison. The peak amplitudes show an ordering of the stimuli based on the N400 component with Emily being the most 'uncanny' and real humans the least. The p-values also show that there are statistically significant differences between the uncanny category and real humans and between Emily and real humans.

SUPPORT VECTOR MACHINE CLASSIFICATION

We use a standard support vector machine [8] for all classification tasks, using a C-SVM with a radial basis function (RBF) classifier and a set of fixed parameters. For the statistical analysis, we performed a standard 5-fold cross-correlation test to avoid over-fitting. The data is split randomly into 5 groups of 228 trials. The training is done with 1140 trials and testing is done with 2037 trials which were not part of the training set. The support vector machine is trained with the peak time and voltage of the N400 component of each of the 13 frontal electrodes. The N400 component is extracted by band-pass filtering the raw EEG data between 1 Hz to 50 Hz. The data is then timelocked to 300 ms after the start of the stimulus (earlier ERP analysis of the N400 shows the exact time where the component peaks), and the closest local minimum found. The exact time and amplitude is then extracted. Both, the cross-correlation test and the peak time and voltage extraction are used to avoid over-fitting the SVM to the training data.

According to Mori's uncanny valley hypothesis [30], and as verified by previous behavioral studies [29, 6], a highly realistic human evokes an uncanny valley response. However, since there is little empirical evidence on what the neural response to 'uncanny' agents looks like, we take the EEG response to Digital Emily to train the SVM on what a uncanny response looks like given that Emily scores the highest on our perceptual study (most realistic cg character). This difference in the response to Emily versus real humans is also seen from the ERP data (Fig. 4d). The neural responses to real humans are used to train the SVM to learn the neural responses to non-uncanny agents. The binary classification is then turned into a probability function using Platt scaling [36]. This allows the SVM to probabilistically determine if there is an uncanny response to a cg character. Figure 6 shows the probabilities for an uncanny response as determined by the SVM plotted against the realism scores from the perceptual study (1 to 6). The probabilities from the SVM correlate with the realism ratings in that the more realistic a virtual agent is, the the higher the probability of an uncanny neural response. Comic is the least human-like in appearance (Likert score of 1) and the SVM also, based on the N400 component, classifies the response with a low probability of being uncanny. As we move higher up the realism scale, the neural responses to the cg characters are classified with higher probabilities of being uncanny. The VH characters are rated by participants on the Likert scale with an average of 1.9, given this, we would expect the neural response to these agents to be classified with a lower probability of being uncanny than Emily or Kara which are higher in terms of realism scores. From Fig. 6 we can see that the classifier categorizes VH characters with a lower probability of being an uncanny response as compared to the cg characters that scored high on the perceptual study like Digital Emily, 'Kara' and 'HeadTech'. Correspondingly, the response to cg characters 'Kara' and 'HeadTech' are categorized with a higher level of 'uncanniness'. This correlates to their perceived realism as measured by the perceptual study (Fig.2) where these characters were given a score of 3.3 and 3.8 respectively as compared with low scores for VHs (avg. score 1.7).

The correlation between 'real' from questionnaires and 'uncanny' from EEG data (Fig. 6) shows the higher the Likert score for an agent, the probability of an uncanny EEG response is higher. Both measures do not provide the same result e.g., although according to the questionnaire Digital Emily was 'real'(M = 5.28, SD = 0.76 vs real human score M = 5.69, SD = 0.75) we know from EEG data that she is not perceived as 'real'(peak amplitude of real human = -4.5 vs Emily = -7.6). If we looked at only the data from the questionnaire we would conclude Emily was perceived as human as real-humans. This is clearly not true from the EEG data i.e. there is a mismatch between Emily and real-humans.

DISCUSSION

We used the N400 component of ERPs to find the uncanny neural response to cg characters. The amplitude of N400 (negative peak 400ms after stimulus) brain response is a well established measure associated with mismatched expectation i.e. you see something that does not match your expectation of what it should look like. We also know the exact electrode locations (parts of brain) where this signal comes from (from fMRI and EEG studies) [20, 25]. This makes the N400 suited for exploring the uncanny effect.

The ERP responses in our study reflect this in terms of the 'uncanniness' of the cg characters. According to the N400 component, the uncanny response is strongest when the cg character appears highly realistic (Fig. 4). An analysis of the average amplitude of the N400 component for each character shows a distinct ordering from the most uncanny agents, i.e. Emily to the least uncanny, i.e. real humans (5). Interestingly, digital Emily, which is a state-of-the-art digital human and is rated the highest in terms of realism, has the greatest amplitude at the N400 component. This supports the predictive coding hypothesis where the uncanny valley is related to expectation violations in neural computing when the brain encounters highly realistic characters. Oddly, we would have expected Digital Ira (another highly realistically rendered human) to rate highly not only in the perceptual study (Fig. 2) but also to evoke an ERP response similar to Emily's. However, Ira rates low in terms of realism (avg. score 3.4) as compared to Emily (avg. score 5.31) and consequently, the ERP N400 response to Ira evokes a lower amplitude than Emily (Fig. 5). One explanation is that in appearance Digital Ira is less human-like than Digital Emily.

Given the results from the ERP responses, we train an SVM to categorize and predict the level of 'uncanniness' of any given cg animated character. The SVM is trained on the responses to Emily which are labelled as 'uncanny' and the responses to real humans which were labelled as 'not uncanny'. Based on the responses to only these two characters the SVM is able to create an ordering of the agents based on the difference in the amplitude of the N400 which represents the anomalies between internal expectations and external stimuli i.e. 'uncanniness'. The ordering of characters created by the SVM correlate with the user rating for each agent. The more realistic a cg character is rated the more toward uncanny the neural response is. Conversely, the less like a real human a cg character is scored on realism, the lower the probability of a uncanny neural response (Fig. 6). So for example, 'comic' is scored the lowest of all characters on realism (Fig. 2, likert score of 1) and it is also classified by the SVM with a low level of 'uncanniness' (Fig. 6). Similarly, Virtual Humans are rated low in terms of realism, and the SVM also gives them a lower 'uncanniness' probability than Emily, 'HeadTech', 'Kara' and 'Ernst'. Interestingly, the SVM also classifies Ira with a low probability of 'uncanniness' which is in-line with the realism score from the perceptual study and the ERP response.

Uncanny has previously been defined with concepts such as 'familiarity' , 'eeriness' or 'relatedness' [9, 26]. We argue that this is exactly what makes the hypothesis difficult to study. In contrast, we define it based on what the EEG responses show i.e. the difference in the amplitude of the N400 ERP component. So for us, 'uncanny', as defined by EEG responses, is the mismatch between what we know is human versus what is presented on screen. EEG allows us to analyse the implicit responses as they occur during participant viewing versus relying on remembered feelings. So we can determine which components are activated on a ms scale. It is not possible to get this data from questionnaires. The likert scale provides a single answer which depends on the question asked and as stated above would not provide the same result as the ERP responses.

CONCLUSION

Findings from our study provide empirical evidence into the nature of the uncanny neural response for computer-generated faces. Our work is of particular interest for animation artists, and video game developers as a method for evaluating their animated humans. Wireless EEG headsets are now affordable (typically $<$ \$1000), readily available, wireless and gel-less [13]. Similaly, open source tool-kits with interactive graphical user interfaces such as the Matlab EEGLAB toolbox [11] and FieldTrip [35] make it easier to process and analyse EEG data. This makes EEG more accessible and lowers the barrier for use in research and application.

Our method provides a quantitative way to measure the kind of reaction any given computer generated agent might evoke in the intended audience. To our knowledge, this is the first work that classifies computer-generated characters based on the level of the evoked 'uncanniness' as measured by neural data. This initial study shows promise for using neurological measures for determining the perceived realism of virtual humans. To be able to land on the other side of the uncanny valley further experimentation is required into specific aspects of computer-generated humans.

In particular we are interested in comparing digital Emily with the real actress on which digital Emily is modelled. This was beyond the scope of our current work since, to be able to make a valid comparison we would need footage of real Emily with similar hair, camera angle etc. which we is currently not available. Also, because familiarity is a consideration we would need to conduct multiple experiments showing one set of participants only real Emily and another set showing only digital Emily, so as to discount influence of one animation over the other e.g. real Emily would influence how digital Emily is perceived. For this paper we were interested in looking at the neural responses to digital characters in isolation without influence from their human counter-parts. However, this kind of comparison is part of our future work plan. In the future we would also like to work on the changes in perception of cg humans based on changing anatomy i.e., a bigger nose or more pronounced ears etc. Such a detailed analysis would allow us to pinpoint the salient features of animated humans that contribute most to the uncanny valley phenomenon.

ACKNOWLEDGMENTS

The work was supported by the German Science Foundation DFG MA2555/15-1 "Immersive Digital Reality" and by ERC Grant No.256941 "Reality CG".

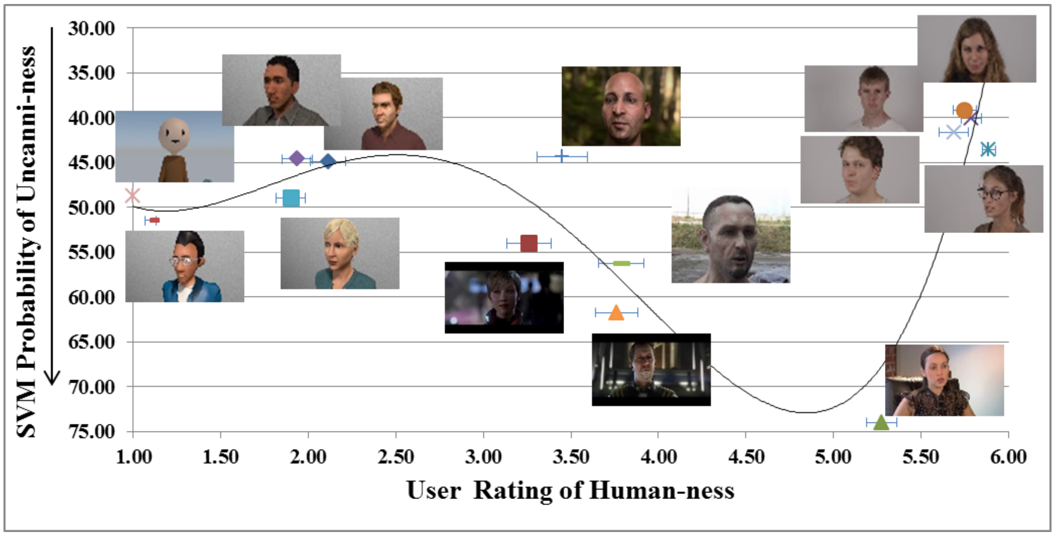
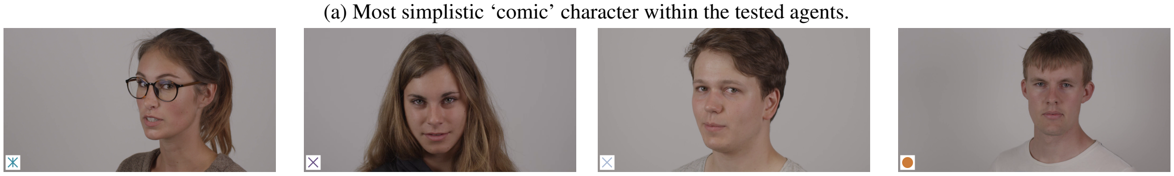


Figure 6: Probabilities of an uncanny neural response plotted against the user responses from the perceptual study with a fourth order polynomial fitted against the data points. The results show a correlation between the neural responses as classified by the SVM and how realistic each agent is. The more human-like a cg character, the more the neural response tends towards the 'uncanny'. The results resemble Mori's uncanny valley hypothesis as seen in Fig. 1

(a) Most simplistic 'comic' character within the tested agents.



(b) Real agents recorded in our lab.

(c) Virtual human (VHT) agents recorded from the Institute of Creative Technologies Virtual Human Toolkit [18].

(d) Virtual human agents with varying degrees of realism and higher 'uncanny' responses ('Kara' [12], 'Ernst' [37] and 'HeadTech' [22])

(e) Highly realistic state-of-the art rendered humans Emily [2] and Ira [1]

Figure 7: Exemplars of animated agents with varying degrees of human-ness used in EEG study to explore and predict neural responses to different levels of realism in animated humans.

REFERENCES

1. Oleg Alexander, Graham Fyffe, Jay Busch, Xueming Yu, Ryosuke Ichikari, Andrew Jones, Paul Debevec, Jorge Jimenez, Etienne Danvoye, Bernardo Antionazzi, and others. 2013. Digital ira: Creating a real-time photoreal digital actor. In *ACM SIGGRAPH 2013 Posters*. ACM, 1.
2. Oleg Alexander, Mike Rogers, William Lambeth, Jen-Yuan Chiang, Wan-Chun Ma, Chuan-Chang Wang, and Paul Debevec. 2010. The digital emily project: Achieving a photorealistic digital actor. *IEEE Computer Graphics and Applications* 30, 4 (2010), 20–31.
3. Shlomo Bentin, Truett Allison, Aina Puce, Erik Perez, and Gregory McCarthy. 1996. Electrophysiological studies of face perception in humans. *Journal of cognitive neuroscience* 8, 6 (1996), 551–565.
4. George Borshukov and John P Lewis. 2005. Realistic human face rendering for the matrix reloaded. In *ACM Siggraph 2005 Courses*. ACM, 13.
5. Kai Bötzel, Stefan Schulze, and Stefan RG Stodieck. 1995. Scalp topography and analysis of intracranial sources of face-evoked potentials. *Experimental brain research* 104, 1 (1995), 135–143.
6. Tyler J Burleigh, Jordan R Schoenherr, and Guy L Lacroix. 2013. Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in Human Behavior* 29, 3 (2013), 759–771.
7. Maxime Cauchoix, Gladys Barragan-Jason, Thomas Serre, and Emmanuel J Barbeau. 2014. The neural dynamics of face detection in the wild revealed by MVPA. *The Journal of Neuroscience* 34, 3 (2014), 846–854.
8. Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2, 3 (2011), 27.
9. Marcus Cheetham, Ivana Pavlovic, Nicola Jordan, Pascal Suter, and Lutz Jancke. 2013. Category processing and the human likeness dimension of the uncanny valley hypothesis: eye-tracking data. *Frontiers in psychology* 4 (2013), 108.
10. Marcus Cheetham, Pascal Suter, and Lutz Jäncke. 2011. The human likeness dimension of the 'uncanny valley hypothesis': behavioral and functional MRI findings. *Frontiers in human neuroscience* 5 (2011), 126.
11. Arnaud Delorme and Scott Makeig. 2004. 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.
12. Quantic Dream. 2016. Detroit : Become Human. https://www.youtube.com/watch?v=LL2gW7hmXu4. (2016). Retrieved: 2016-01-15.
13. EMOTIV. 2011. EMOTIV EPOC+ 14 Channel Mobile EEG. https://www.emotiv.com/product/emotiv-epoc-14-channel-mobile-eeg/. (2011). Retrieved: 2016-12-20.
14. Shaojing Fan, Rangding Wang, Tian-Tsong Ng, Cheston Y-C Tan, Jonathan S Herberg, and Bryan L Koenig. 2014. Human Perception of Visual Realism for Photo and Computer-Generated Face Images. *ACM Transactions on Applied Perception (TAP)* 11, 2 (2014), 7.
15. David Friedman, Yael M Cycowicz, and Helen Gaeta. 2001. The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews* 25, 4 (2001), 355–373.
16. Tom Geller. 2008. Overcoming the uncanny valley. *IEEE computer graphics and applications* 4 (2008), 11–17.
17. Nathalie George, Julie Evans, Nicole Fiori, Jules Davidoff, and Bernard Renault. 1996. Brain events related to normal and moderately scrambled faces. *Cognitive Brain Research* 4, 2 (1996), 65–76.
18. Arno Hartholt, David Traum, Stacy C. Marsella, Ari Shapiro, Giota Stratou, Anton Leuski, Louis-Philippe Morency, and Jonathan Gratch. 2013. All Together Now: Introducing the Virtual Human Toolkit. In *13th International Conference on Intelligent Virtual Agents*. Edinburgh, UK.
19. Chin-Chang Ho, Karl F MacDorman, and ZA Dwi Pramono. 2008. Human emotion and the uncanny valley: a GLM, MDS, and Isomap analysis of robot video ratings. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*. ACM, 169–176.
20. Phillip J Holcomb and Helen J Neville. 1991. The electrophysiology of spoken sentence processing. *Psychobiology* 19 (1991), 286–300.
21. Roxane J Itier and Margot J Taylor. 2004. N170 or N1? Spatiotemporal differences between object and face processing using ERPs. *Cerebral cortex* 14, 2 (2004), 132–142.
22. Janimation, Inc. 2010-2017. HeadTech. http://janimation.com/#&target=HeadTech8_JaniWeb_720p_1500. (2010-2017).
23. D Aled Jeffreys. 1996. Evoked potential studies of face and object processing. *Visual Cognition* 3, 1 (1996), 1–38.
24. Kangsoo Kim, Arjun Nagendran, Jeremy Bailenson, and Greg Welch. 2015. Expectancy Violations Related to a Virtual Human's Joint Gaze Behavior in Real-Virtual Human Interactions. *28th Annual Conference on Computer Animation and Social Agents* (2015).
25. Marta Kutas and Kara D Federmeier. 2011. Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual review of psychology* 62 (2011), 621.

26. Karl F MacDorman. 2006. Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. In *ICCS/CogSci-2006 long symposium: Toward social mechanisms of android science*. 26–29.
27. Karl F MacDorman, Robert D Green, Chin-Chang Ho, and Clinton T Koch. 2009. Too real for comfort? Uncanny responses to computer generated faces. *Computers in human behavior* 25, 3 (2009), 695–710.
28. Karl F MacDorman and Hiroshi Ishiguro. 2006. The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies* 7, 3 (2006), 297–337.
29. Rachel McDonnell and Martin Breidt. 2010. Face reality: investigating the uncanny valley for virtual faces. In *ACM SIGGRAPH ASIA 2010 Sketches*. ACM, 41.
30. Masahiro Mori, Karl F MacDorman, and Norri Kageki. 2012. The uncanny valley. *Robotics & Automation Magazine, IEEE* 19, 2 (2012), 98–100.
31. Margaret C Moulson, Benjamin Balas, Charles Nelson, and Pawan Sinha. 2011. EEG correlates of categorical and graded face perception. *Neuropsychologia* 49, 14 (2011), 3847–3853.
32. Maryam Mustafa and Marcus Magnor. 2016. EEG Based Analysis of the Perception of Computer-Generated Faces. In *Proceedings of the 13th European Conference on Visual Media Production (CVMP 2016)*. ACM, New York, NY, USA, Article 4, 10 pages.
33. Arti Nigam, James E Hoffman, and Robert F Simons. 1992. N400 to semantically anomalous pictures and words. *Journal of cognitive neuroscience* 4, 1 (1992), 15–22.
34. Alison Oddey and Christine A White. 2009. *Modes of spectating*. Intellect Books.
35. Robert Oostenveld, Pascal Fries, Eric Maris, and Jan-Mathijs Schoffelen. 2010. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience* 2011 (2010).
36. John C. Platt. 1999. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. In *ADVANCES IN LARGE MARGIN CLASSIFIERS*. MIT Press, 61–74.
37. Cloud Imperium Games Quantic Dream and Behaviour Interactive IllFonic Foundry 42. 2016. Star Citizen. https://robertsspaceindustries.com/squadron42. (2016). Retrieved: 2015-11-15.
38. Paul Rademacher, Jed Lengyel, Edward Cutrell, and Turner Whitted. 2001. Measuring the perception of visual realism in images. In *Rendering Techniques 2001*. Springer, 235–247.
39. Astrid M Rosenthal-von der Pütten and Nicole C Krämer. 2014. How design characteristics of robots determine evaluation and uncanny valley related responses. *Computers in Human Behavior* 36 (2014), 422–439.
40. Guillaume A Rousselet, Marc J-M Macé, and Michèle Fabre-Thorpe. 2004. Animal and human faces in natural scenes: How specific to human faces is the N170 ERP component? *Journal of Vision* 4, 1 (2004), 2–2.
41. Ayse Pinar Saygin, Thierry Chaminade, Hiroshi Ishiguro, Jon Driver, and Chris Frith. 2011. The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social cognitive and affective neuroscience* (2011), nsr025.
42. Ayse Pinar Saygin, Thierry Chaminade, Hiroshi Ishiguro, Jon Driver, and Chris Frith. 2012. The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social cognitive and affective neuroscience* 7, 4 (2012), 413–422.
43. Valentin Schwind and Solveigh Jäger. 2016. The Uncanny Valley and the Importance of Eye Contact. *i-com* 15, 1 (2016), 93–104.
44. Jun'ichiro Seyama and Ruth S Nagayama. 2007. The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence* 16, 4 (2007), 337–351.
45. Megan Strait, Lara Vujovic, Victoria Floerke, Matthias Scheutz, and Heather Urry. 2015. Too Much Humanness for Human-Robot Interaction: Exposure to Highly Humanlike Robots Elicits Aversive Responding in Observers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3593–3602.
46. Angela Tinwell, Mark Grimshaw, Debbie Abdel Nabi, and Andrew Williams. 2011. Facial expression of emotion and perception of the Uncanny Valley in virtual characters. *Computers in Human Behavior* 27, 2 (2011), 741–749.
47. Angela Tinwell, Mark Grimshaw, and Andrew Williams. 2010. Uncanny behaviour in survival horror games. *Journal of Gaming & Virtual Worlds* 2, 1 (2010), 3–25.
48. Burcu A Urgen, Alvin X Li, Chris Berka, Marta Kutas, Hiroshi Ishiguro, and Ayse P Saygin. 2015. Predictive coding and the Uncanny Valley hypothesis: Evidence from electrical brain activity. *Cognition: A Bridge between Robotics and Interaction* (2015), 15–21.
49. Tim Weyrich, Wojciech Matusik, Hanspeter Pfister, Bernd Bickel, Craig Donner, Chien Tu, Janet McAndless, Jinho Lee, Addy Ngan, Henrik Wann Jensen, and others. 2006. Analysis of human faces using a measurement-based skin reflectance model. In *ACM Transactions on Graphics (TOG)*, Vol. 25. ACM, 1013–1024.
50. Eduard Zell, Carlos Aliaga, Adrian Jarabo, Katja Zibrek, Diego Gutierrez, Rachel McDonnell, and Mario Botsch. 2015. To stylize or not to stylize?: the effect of shape and material stylization on the perception of computer-generated faces. *ACM Transactions on Graphics (TOG)* 34, 6 (2015), 184.