Proceedings of the 52nd Hawaii International Conference on System Sciences | 2019

Mapping Beyond the Uncanny Valley: A Delphi Study on Aiding Adoption of Realistic Digital Faces

Mike Seymour
University of Sydney
mike.seymour@sydney.edu.au

Kai Riemer
University of Sydney
kai.riemer@sydney.edu.au

Judy Kay
University of Sydney
judy.kay@sydney.edu.au

Abstract

Developers and HCI researchers have long strived to create digital agents that are more realistic. Voice-only versions are now common, but there has been a lack of visually realistic agents. A key barrier is the "Uncanny Valley", referring to aversion being triggered if agents are not quite realistic.

To gain understanding of the challenges of the Uncanny Valley in creating realistic agents, we conducted a Delphi study. For the Delphi panel, we recruited 13 leading international experts in the area of digital humans. They participated in three rounds of qualitative interviews. We aimed to transfer their knowledge from the entertainment industry to HCI researchers. Our findings include the unexpected conclusion that the panel considered the challenges of final rendering was not a key problem. Instead, modeling and rigging were highlighted, and a new dimension of interactivity was revealed as important.

Our results provide a set of research directions for those engaged in HCI-oriented information systems using realistic digital humans.

1. Introduction

Central to Human Computer Interaction (HCI) is the nature of the interaction itself. Given that the most common and often preferred form of human communication is face-to-face, it is interesting that the dominant HCI metaphor is a desktop. There has been little success in achieving an emotionally engaging interface[1] that has a realistic digital version of a human face. Yet, such an approach might change the way we interact with computers. While such faces could prove valuable, a key barrier to their acceptance is the phenomenon known as the 'Uncanny Valley'[2]. An additional barrier has been the limits of technology, which is only now achieving quite realistic implementations of faces.

Emotion plays a key role in human interaction and the face is one of the most expressive non-verbal tools for conveying emotion. In human interaction, emotion is highly efficient, imbuing verbal communication with meaning and context. Realistic faces for interaction has the potential to greatly impact some key areas of HCI.

In this area of research, there is much to learn from the film and entertainment industry. The professionals in these areas have been working for a long time to produce human simulations for feature films, television and computer games. These industries are large, highly computerized and with dedicated teams researching this area. Even in animated films, the animators tend to study and reference human actors' faces, to give their non-human animated characters emotional energy and relevance.

To capitalize on this expertise and insights from largely unpublished commercial research, a Delphi Study was undertaken with 13 of the world's leading experts in facial animation and simulation. This research explored their collective wisdom about what drives realism. It exposed new and previously unexpected opinions that run contrary to accepted doctrine, particularly the quite new idea that interactive movement can greatly *reduce* the Uncanny Valley effect. The panel raised the possibility that emotional interactions positively change the way people perceive computer avatars, robots and agents.

This positive response to interaction has not appeared in previous published work; rather, the accepted Uncanny Valley original theory states that the effect worsens with movement.

We note that our research sought to gain insight into what is required for an effective implementation of a digital human, but that we did not study the simulation of the human responses or the artificial intelligence that might power such faces.

In summary, we had expected the panel of experts to primarily discuss approaches to improve the later stages of rendering faces, to address bridging the Uncanny Valley. As a real face produces no negative effect, we expected to be focused only on what is stopping a digital face from appearing real or photo-realistic.

In summary, the result of the Delphi study is a set of insights into the complex visual hurdles that interact, as people appear to evaluate faces holistically, and "see the person" rather than the individual aspects of the facial representation. A person's acceptance of a synthetic face is then moderated by interacting with it in real time, making the complexity of creating a

URI: https://hdl.handle.net/10125/59916
ISBN: 978-0-9981331-2-6

digital human face multifaceted. This provides insights into the challenges needed to be addressed in order to avoid triggering a negative response in users.

2. Background

To create a realistic digital agent as a user interface element is highly complex. Even in high-end film production there are technical challenges to overcome in producing a realistic human face. This section provides background on three key aspects of this work. First, we introduce the core under-pinning foundation of the Uncanny Valley. Then we introduce the range of technical challenges in creating a realistic face.

2.1. Uncanny Valley

The 40-year-old Uncanny Valley[2] theory plays a key role in the research on users' reactions to avatars and agents. According to the theory users have greater affinity for agents that are more realistic. User affinity increases as the agent becomes increasingly realistic, until the agent is semi-realistic, at which point affinity drops dramatically because a partially realistic agent triggers unease in users (see Figure 1).

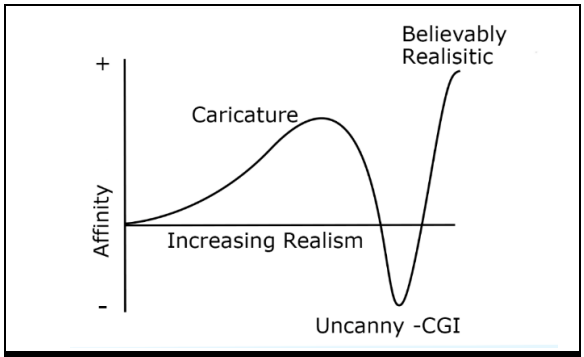

Figure 1. Uncanny valley

As realism increases, there comes a point where the valley has been crossed and the avatar's affinity increases to its highest level. It does not require the realistic agent to be imperceptibly real, just very close. Thus, "crossing the Uncanny Valley" has been identified as a significant hurdle to the use of realistic faces in HCI.

One of the difficulties in researching the original Uncanny Valley theory is that there is no clear metric for the notion of 'affinity'. It is not a dependent variable against which one can test with some independent variable(s). The word is a translation from the original Japanese and thus is itself an interpretation of the meaning of the Japanese word Shinwakan (親和感). Affinity is the currently accepted translation. Other English translations have also been used to describe the theory's vertical axis, such as: familiarity, rapport, and comfort level[3]. We therefore did not restrict our discussion only to the contemporary Western notion of 'affinity'.

Masahiro Mori' 1970s paper focused on robots; he termed this affinity drop "bukimi no tani", translated and popularized as "Uncanny Valley". However, the non-linear response shown in Figure 1 has also been shown to apply to how users judge computer graphics images (CGI) of faces[4][5] or avatars. We restrict our definition of a digital agent or avatar to the digital facial representation or facsimile of a person.

It was postulated that the Uncanny Valley effect occurs for a variety of reasons. One such reason is known as the death mask effect, whereby a face that falls in the Uncanny Valley is associated with death as the face appears not fully life-like[6].

Further theories have been proposed, including

1) that lifelike faces are simply judged more like faces, therefore are held to a higher standard[7],

2) that lifelike faces are repulsive because they challenge the idea of what is 'human'[8] and we avoid such faces as they look sick or wrong. By avoiding them we avoid possible infection or contamination[9].

The original theory further contends that movement will magnify the effect positively and negatively. According to the death mask explanation movement or animation of the face is therefore 'moving death' – or the undead moving, a common device of fictional drama horror associated with zombies or similar characters[6].

Exploration of movement is relevant for our research given its focus on applications such as film, video, gaming, and most specifically the use of faces in computer interfaces.

Figure 2. Caricature, 'Repulsive' version, and Realistic. All CGI of the same face (L-R).

Humans are hardwired to interpret faces. From birth, a child responds and learns from their parents' faces, and these interpretations are fundamental for the successful growth and functioning of humans[10]. As such, we have developed the ability to read faces far more specifically and with greater fidelity than any other object. This has left us with both the ability to see a face in a few line strokes of a cartoon or in a puffy

cloud (pareidolia) but also to identify and reject those artificial faces which are only approximately close to realistic as covered in the Uncanny Valley theory[11].

An HCI using an agent with an incomplete solution can mean not just a lack of acceptance but a visceral repulsion (figure 2). Faces, unlike other HCI artifacts, can trigger strong non-linear responses to trust, concern and repulsion.

2.2. Technical State of the Art

There are several approaches for pre-rendered and real-time realistic face synthesis in the entertainment industry. Many of these pipelines share common approaches, and the experts build on their experience in these long-established pipelines[12].

A generalized face pipeline consists of 7 stages

1. Scanning or modeling
2. Expressions or poses
3. Correspondence
4. Rigging
5. Texturing
6. Animation
7. Rendering

In broad terms, a face is created either from computer-aided *scanning* such as photogrammetry, or artist interpretation using computer modeling tools.

A set of *poses or expressions* is then made. This stage defines the range of motion. This 'expression space' defines the extent of expressions that the digital character will be able to display. Often these key poses relate to the theory of Facial Action Coding System (FACS) which break down the face's expressions into Action Units (AU)[13]. This is the standard industry practice, as validated in this research.

Correspondence is achieved between expressions so that the model may move between key expressions seamlessly. This stage connects the various separate expressions into one range of facial movements.

The *rigging* stage allows controls for moving the face to be presented for either manual or data manipulation of the face. The 'rig' allows the face to be controlled and animated.

The fifth stage of *texturing* adds realism with skin and hair detail, and the correct responses to light. The face is now complete. The last two stages *animate* the face and *render* a final output at the appropriate frame rate and resolution with appropriate lighting.

Each of these stages is open to variation, but even in the creation of non-human characters a real person is commonly used to re-target to a character face.

2.3. Delphi Studies

The Delphi method has a long and successful history in structuring group communication for forecasting the development trajectories of new technology[14]. The nature of the 'structured group communication' is to explore a topic in *rounds* and provide a summary as feedback, with individual contributions reported *anonymously* to the group. While it was originally designed to seek quantitative consensus, it is now used mostly qualitatively[15].

The Delphi approach emerged in the late 1960s as a way of getting an expert view of future developments in a specialist field. From the outset, the application areas included clarifying real or perceived human motivations and developing causal relationships in complex phenomena. Two appropriate uses of a Delphi study are:

a) a problem that does not lend itself to precise analytical techniques but benefits from selective expert judgment.
b) a situation where diverse individuals contribute to a complex problem[14].

For our purposes a Delphi study provides a rich source of interrelated 'knowledgeable insights' on how a face might be designed to provoke positive affinity. This follows the principle that, "when the problem is directed toward analysis of a number of interdependent variables in complex structures the natural choice would be to go deeper... instead of increasing the number of cases"[16].

Central to the design of the Delphi process is the notion of the 'panel', as a curated list of experts, and their anonymity. This allows for "effective and reliable utilization of a small sample from a limited number of experts in a field of study to develop reliable criteria that inform judgment and support effective decision-making"[17]. No expert, or outside party should be able to identify the comments of any one expert, but rather the comments are disclosed as having come from the panel as such.

The process is designed so that interviews from one 'round' are collated and presented to the panelists for further discussion as part of the next round. This process of rounds also highlights the role of the Delphi designer, whose role is to conduct the interactions, balance the various communication goals and give context to each stage of the process, while maintaining the objective of the Delphi research.

3. Research Objective

Our research question is: what needs to be done to be able to create human faces that cross the Uncanny Valley and can be effective in a range of contexts?

To explore this, we designed a Delphi study with leading experts in the field of digital humans, from entertainment, games and advanced facial research. In

so doing, we sought to define a research roadmap of relevant issues and inform an HCI research agenda.

4. Research Approach

The study design is a qualitatively exploratory study of human simulation, using an abductive, reflective approach based on the iterative abductive method of Peirce[18], also called 'systematic combining'[16], [19] as it uses both inductive and deductive approaches. This makes it different from grounded theory[20], which aims to find truth "in" the data itself without a particular theory guiding the analysis[21]. We will now outline our approach.

4.1. Our Delphi Method

We chose a qualitative Delphi methodology for several reasons. Firstly, we are investigating an estimation of an emotional response. This is hard to quantify, as it involves the Uncanny Valley's notion of 'affinity'. Secondly, while we are proposing a path forward for enabling the creation of digital faces, we do not have a preconceived hypothesis to test against, as there is a shortage in literature researching a comprehensive prescriptive approach to the Uncanny Valley. Finally, an alternate approach of grounded theory was considered and rejected as it denied the perspective of the researcher as an active participant in the curation and interpretation of the study.

This research does not depend on large-scale empirical data, but on theoretical generalizations from in-depth iterative analysis of expert opinions. Through this iterative process, we gain cumulative insight into the phenomenon, and form an agenda for subsequent research.

Ensuring rigor is a primary concern in research. We therefore outline our study design and how the research was executed.

4.2. Study Design

The initial questions for the first round of the Delphi panel were selected to define the range of the discussion and solicit new and unexpected opinions on what might be fruitful research.

Each expert is sufficiently distinguished in that they alone could drive a valuable research agenda. The panelists were contacted and formally accepted participation. They were then interviewed in person (sometimes via skype), and the interviews recorded.

Each set of interviews represents one round of the study. After each interview the transcripts were captured in *NVivo* and a summary of the comments of the panel produced as a discussion document for the next round. There were three Delphi rounds in total.

4.3. Delphi Panel

Our panel members were selected based on their recognized international expertise in deploying digital humans, with everyone active in the field. The panel included academics, two former CTOs, five games industry experts and VR specialists. One indicator of the 'expert' nature of the panel is that it featured a combined 14 Visual Effects Oscars and Scientific Technical Academy Awards (Sci-Tech Oscars). None of the original panelists dropped out of the study.

The group not only represents the leading researchers in this field, but as a group, they are responsible for how major commercial research resources are allocated in this field. The list of experts is as follows, all agreed to have their names published:

Rob Bredow: Head of Industrial Light and Magic (ILM, Lucasfilm) VFX Supervisor & Producer.
Dr. Paul Debevec: USC - ICT Research Professor, now Senior Staff Engineer, Google. Sci-Tech Oscar.
Christopher Evans: Face Technical Director, Epic Games.
TJ Galda: Autodesk, Creative Senior Product Management, Innovation, Change Management, and Strategic Planning.
Ben Grossmann: Magnopus co-founder, VFX supervisor. Oscar Winner.
Christophe Hery: PIXAR, Global Tech & Research Technical Director. Multiple Sci-tech Oscar Winner.
Dr. J.P Lewis: Weta Digital & Victoria University, Assoc. Prof, now Electronic Arts. Multiple Sci-Tech Oscars.
Kim Libreri: Chief Creative Officer, Epic Games, Multiple Sci-tech Oscar winner.
Dr Iain Matthews: Principal Research Scientist, Disney Research, Hon. Prof. Now FaceBook Reality Labs. Sci-tech Oscar winner.
Stephen Rosenbaum: VFX Supervisor. Two time Oscar winner.
Dr. Mark Sagar: Founder, Soul Machines and University of Auckland. Multiple Sci-Tech Oscars.
Sebastian Sylwan: CTO at Weta Digital, now CTO Félix & Paul VR Studios.
Edson Williams: Co-founder Lola, VFX supervisor.

The panelists each have highly specialized knowledge across the broad range of face simulation technology. Their areas of expertise, while overlapping, are complementary and provide different points of view. For example, the domain expertise of Edson Williams is as a world expert in changing or

replacing faces with image compositing (2D), as compared to 3D graphics which is the domain of the other experts. This is extremely specialized and complex work, but it affords him a unique perspective. Dr. Mark Sagar was instrumental in the adoption of FACS in the 3D effects industry as a whole. Christophe Hery is a world expert in rendering and simulation but not necessarily real-time graphics, while TJ Galda is an expert in rigging, especially in real-time games, but not advanced rendering algorithms. And so forth, with all the panel and their individual strengths complementing the whole.

4.4. Open Ended Question Examples

Below is a sample of the open-ended questions that were used for the interviews. The first question aimed to establish the core topic. Follow-up questions sought both higher level opinions and detailed technical discussions. The open format of the questions allowed the discussion to take different directions based on the expert's expertise and the content of the answers they gave.

These initial questions were derived based on the professional expertise of the lead researcher who had conducted over a 1,000 industry interviews since 1999 on one of the industry's leading web sites.

While the questions and discussion varied, interviews had most questions in common, such as:

- How far do you think we are from being able to reliably cross the Uncanny Valley?
- Do you draw a distinction between photo-real and crossing uncanny valley?
- Do you think acceptance is influenced by race?
- Does age of the face effect its difficulty?
- Do you think the brain sees faces differently, from other objects?
- What do you think of FACS for animation?
- What is the easiest face to generate?
- How important is spectral rendering?
- What do you think we must solve to rig faces? Generally, is there any recent research that you think holds promise in the research of faces?

4.5. Round 1

Round 1 of the study explored the traditional face pipeline and sought to gauge areas of consensus and important areas of new insight for Round 2 discussion. The first-round interviews were approximately 45-60 minutes each, with a written summary of the discussion sent to the panellists for the next round.

4.6. Round 2: Surfacing Critical Topics

Round 2 mirrored the first in implementation and duration. The points from the first round were clarified and then discussed in detail in Round 2. It was noted that not all rejection of faces is due to some special neurological response; there are also just poorly attempted face simulations, "*I think the Uncanny Valley is kind of a glib way to say lots of people have done facial animation badly and everyone hates it.*"

The largest single shift from Round 1 to 2 was the focus on how real-time interaction changed the viewers'/users' response. The second-round was the most informative, as summarized in section 5.

4.7. Summation: Final round

The third round was shortest in duration. It confirmed the outcomes of the prior rounds and the characterization of the issues in the study. Five key outputs are discussed in the next section.

5. Results: Key Issues for Faces

The panel initially examined individual aspects of realistic digital faces. But rather than focusing on these multiple isolated or decontextualized aspects, what emerged was a complex interrelated view of acceptance. Overall there was agreement on the current standard industry approach, as it was outlined in section 2.2. Several general points are now noted.

5.1. General insights

The panel agreed that a face needed to be sampled to a very high degree of fidelity, much higher than might be expected given the final display resolution.

The surprising outcome of Round 1 was that while rendering is often the center of discussion about CGI faces, rendering was not seen as the critical element for improvement as increasing compute power has already greatly improved non-real-time rendering. Major improvements have been made in the specific areas of ray tracing and physically plausible materials. While final rendering was still seen as vital, rendering alone, was not seen as the area needing the greatest innovation.

By contrast, real-time rendering is computationally very costly, and thus lagging in realism. As computers get faster it was expected that real-time engines would be able to take advantage of newer physically plausible lighting and shading models that are currently more common in non-real-time applications.

These newer approaches were thought to be producing good results, especially for static shots.

Animation was explicitly stated in Round 2 to be a much bigger issue than the rendering for achieving realistic agents. It was suggested that animation needs a more scientific approach to reliably produce work that was believable. Notwithstanding recent advances in motion capture, rendering was thought to be more 'solved' than the area of animation.

A critical point was the difference between reproducing a known individual compared to a generic person. A famous person or someone personally known, was said to be much harder to reproduce.

In Round 2, many panelists highlighted that there are many aspects to human faces that people find hard to articulate, but when missing one of these, the face feels 'wrong' and unnatural. This emerged as a core reason why the Uncanny Valley is so hard to address. We may not see faces in a simple way; instead we process faces with highly developed and specific facial neurological processes. In round 2 there were points that were not agreed upon by the panel. (See table 1).

Table 1. Points Raised

Summary of Points	Expected	Disputed
Underlying muscles (5.1.1)	Yes	
Scope of the face, Hockey Mask (5.1.2)	Yes	
Linearity of blend shapes (5.2.1)	Yes	
Use of FACS as a base (5.2.1)	No	Yes
Blood flow - Blush response (5.2.2)	No	Yes
Skin Solutions (5.3.1)	Yes	
Spectral rendering (5.3.3)	Yes	Yes
Movement vs. Interaction (5.4.1)	No	
Display space (5.4.2)	No	
Context (5.5)	No	
Knowing the subject (5.5)	Yes	

The Delphi Study identified five major areas relating to the simulating of digital humans. We now describe these, organized around the main areas that emerged: Modeling and Sampling (which includes scanning and correspondence) (5.1); Rigging and animation (5.2); Rendering (5.3); Interaction & Environment (5.4) and Questioning assumptions (5.5).

5.2. Modeling and sampling

There are two major points in this specific part of the Delphi discussion. First, current approaches for creating faces did not allow for differences in individual facial muscles underneath the facial skin. All current approaches assumed an average or typical muscle structure, and this may not be valid when trying to make a digital human match an actual person.

The second point was even more far-reaching; many panelists mentioned that the human perceptive system has developed in evolutionary biological terms to process different parts of the human face via specific regions of the brain. The panel agreed that there is no single unified face recognition system in the brain responsible for the Uncanny reaction. It was hypothesized that it may not just be a poor rendition of a face that causes an Uncanny response, but a dissonance between different parts of the brain when processing the incoming face.

5.2.1. Sampling surfaces for underlying muscles. It was suggested that the historical difficulty of producing a realistic animated human face reflects the way that surface properties can be witnessed but faces are driven by unseen facial muscles, and these cannot currently be measured or sampled when building digital humans. In the first round of the Delphi study, one panelist noted how unique human facial muscles are, compared to other primates, and how evolved human faces are as communication tools.

Building further on this point, it was noted that normal human faces are not similar in actual muscle size to each other, yet most CG models assume a similar underlying facial muscle profile. The differences between any two people, which are often significant, can be seen in autopsies, noted one panelist, who had attended real autopsies.

"Some people's muscles looked like a tiny piece of string and in other people it looks more like the thick strip that you see in the anatomy book. The individual differences were interesting. It makes sense, why should your facial muscle anatomy be consistent?"

5.2.2. Scope of the face. The panel agreed that the whole head is important when modeling and animating a digital human to produce a likeness or fully express a range of emotions. While the 'face' is often discussed in terms of a 'hockey mask', the face and most of the head and neck are key to realism and need to be accurately modeled or sampled. Building on the notion of extending beyond the hockey mask region, in discussion about movement, one panelist stressed how widely facial *animation* extends beyond just the face. This is important as much prior research had assumed the primary front of the face could be thought of as being independent for animation.

For example, a popular interpretation of a smile is that it is only apparent on the mouth. Specialists go

further and normally agree that the lower face, and the muscles at the side of the eyes are also triggered. The panel agreed it goes further and that "your hair does go up and down when you smile... the muscles in the sides of your neck... It's even down into your neckline that you have to really start worrying about on some poses" commented one panelist.

Several panelists pointed out that this is related to unexpected comments coming from people reviewing digital humans, and they attributed this to the difficulty in articulating a problem when one sees a face that seems 'wrong'.

The consensus was to make sure that any 'face' solution extends well beyond the hockey mask region that is often all that is focused on.

5.3. Rigging and animation

The two points raised were: the validity of linear blend shape combinations with the dominant reliance on FACS; and the importance of blood flow.

5.3.1. Blend shapes and FACS. There are several ways to animate a facial model, or 'rig' the face for later animation. The primary method discussed by the panel was a blend shape rig which moves between expressions for different parts of the face via a notional slider or value. This approach is often based on FACS action units or AUs. For each sub-expression, an animator or motion capture solver can 'dial in' a percentage of sub-expression (AUs).

A FACS pipeline requires actors to strike a series of poses in a separate FACS scanning session. It was stated that the validity of the 'performance' and the interrelationship of different parts of the face can be lost in the subsequent animation stage.

There was some disagreement over how far a FACS and blend shape rig approach could go in achieving realism. Some panelists stated that they were not comfortable with the level of detail and accuracy that can currently be captured and produced with a blend shape driven FACS animation solution.

The process of creating the range of motion comes from the actor producing a series of FACS poses. This set of facial expressions is of the order of 40 or so expressions. The FACS poses (and the AUs they are decomposed into) are co-opted from the non-CGI research of Psychologist Paul Ekman. This was originally developed to identify and classify human facial expressions. While FACS have been very successful (one panelist suggested every major face-pipeline has a FACS component), some other panelists raised whether it's 'fit for purpose': "*I just don't think we really understand well enough how to parameterize a face*".

An example offered was an actor who, when providing their FACS still poses, did not produce an authentic emotional response; thus, the capture FACS reference is partially incorrect. The facial response that controls a smile receives its input from both subcortical and cortical areas of the brain. This means that a person can normally not control their face to smile in a genuine way unless the smile is motivated by a genuine emotional sentiment (Panelists pointed to this as reinforcing the value of 'method acting'). Any FACS pipeline will reference this inauthentic smile if the tracked points on the face later 'get solved' to a smile. There are always effective ways of adjusting such animation iteratively by hand, but it is expensive.

Another key aspect of animation discussed by the panel, was the issue of non-linearity. This refers to the combinatorial nature of the sub-expressions or FACS and their component AUs. This was identified as a more complex issue to resolve.

Each part of an expression is called an action unit or AU. In simple terms, if we call an AU eyebrow raised 'A', and an AU smirk with the mouth 'B', then any face pipeline system around the world will allow A+B = A and B *happening at once*. The problem is that this assumes what is known as 'linear combinatorial expressions'. It assumes that the way an actor raises an eyebrow (AU: A) when not smirking is the same as how they would raise it if they were smirking. This is at the heart of why one can combine or build up expressions by adding AUs together. Since one cannot capture all the combinatorial variations of every AU with every other AU permutation, the problem is fundamental to current approaches to face capture.

One panelist commented that there is not an orthogonal set of combinations of AUs. In other words, no two AUs can just be added or combined arbitrarily in their opinion. For example, two AUs may be valid and seemingly happening on independent parts of the face but an actor could not have achieved both AUs together. The face has odd combinations which may be hard or impossible to achieve in real life. Nor could the actor get from one expression directly to another, without intermediate expressions.

"Linearity is very important, faces are incredibly non-linear within one expression, a smile is a good example. A smile will start out as sort of stretching the lips, but then after a certain point the lips are stretched tight around the teeth that they almost widen, and then you'll get the teeth showing, all are very non-linear."

FACS was heavily defended by some panelists in later rounds. For some, AUs are directly linked to facial muscles, and a core approach to successful facial animation. There was never agreement, and the panelists remained divided.

5.3.2. Blood flow. The timing and nature of blood flow to the face was raised as an important issue in the first round.

Some panelists stated that blood redistribution affected skin hue and it was a failing if this was not modeled and animated. Still other panelists who work with facial blood flow maps had introduced a delay offset between pose and hemoglobin redistribution, so color changes visibly lagged poses. It was stated that such a lag would be of the order of multiple seconds. While panelists believed that the issue was important, they also questioned if such blush or flush responses are 'readable' by a viewer explicitly. It was suggested that due to our evolved way of reading facial emotions, people were affected by such color changes, but the same people would find it very hard to 'see' them separately or articulate their impact on a face.

Most panelists suggested that modeling hue shifts might be important but only a few panelists expressed a strong opinion that there should be a time delay between expression and a color change. It was suggested that more quantitative research was needed.

5.4. Rendering

Rendering is a complex issue involving the simulation of light interacting with objects. Current methods favor solving the render equation with a physically plausible unidirectional path tracing approach. This is not yet possible for most real-time applications.

The area differs greatly between real-time agents and avatars and non-real-time pre-rendered faces. While there was confidence in the technological approaches used in the entertainment industry, the limitations of rendering an interactive character using all these techniques is prohibitive. It was expected by the panel that this will be addressed over time thanks to rapid increases in compute performance. Hence a discussion of non-real-time approaches was the focus.

The panelists commented that energy conserving approaches, ray tracing and detailed subsurface scattering in the skin were all key technologies.

The areas of discussion focused on skin solutions and the recent move to spectral rendering.

5.4.1. Skin Solutions. Facial realism is heavily related to skin rendering and realism, a point universally agreed upon. Most panelists agreed upon the significance of recent advances in diffuse Sub-Surface Scattering (SSS). Only a few panelists felt that the current approaches to skin were holding back character acceptance.

The general sentiment could be characterized as agreeing that poor SSS is very noticeable, and good SSS is still hard to achieve, but current strong implementations are close to acceptable and this was no longer such a large contributor to the Uncanny Valley effect as it had been.

5.4.2. Spectral rendering. A panelist in the first round stated that spectral rendering (rendering over a wider range of light spectrum sample points than R G and B) was contributing to successful face pipelines at award winning companies such as *WETA Digital* (which has recently created an in-house spectral renderer called *Manuka*). Specular rendering requires not only the rendering to accommodate a wide gamut/greater spectral frequency sampling, but more complexity when creating the facial textures.

While the SSS is inherently going to be affected to some extent by spectral rendering (as skin diffusion is based on wavelength), panelists considered that it was primarily significant in allowing accurate rendering into a specific scene or lighting setup. Its greatest contribution in face rendering was in producing a believable face in context, so that it sat well in a live-action background. The main exception was a benefit for rendering eye caustics and modeling the way some eye light causes skin caustics.

5.5. Interaction and environment

An aspect of the original Uncanny Valley theory was that movement would magnify the effect. This secondary aspect of the Uncanny Valley Theory is rarely focused on in research. It should be noted that the original paper offered no empirical evidence to validate this theoretical claim. Until recently, due to technical complexity, highly interactive user interfaces with realistic digital faces have remained largely untested in respect to this theory.

The panel also asked if the Uncanny effect was amplified or moderated by interaction compared to recorded movement. A secondary question was raised regarding context of digital humans.

5.5.1. Movement vs. Interaction. Based on three of the panelists' observations and subsequent rounds of discussion in the study, the panel raised that emotions positively change the way people perceive avatars, agents, and even robots when these figures engage interactively. This positive user reaction is unpredicted by current accepted behavioral models. The original Uncanny Valley theory states that the effect will worsen with movement.

Importantly, this was speculated to be related to interactivity and not just movement. The amplification effect suggested in the original research was generally

agreed to by most of the panelists, but only if one considers *pre-recorded* movement. Prior research with both recorded still-images and video clips have borne out the existence of this phenomenon[22]. What the panel did not feel had been researched was movement in the form of *interactivity*.

This opposite outcome occurs when these figures are exhibiting emotional 'Affective Computing' style feedback loops, such as matching eye contact, smiles, and conversational non-verbal responses[23][24].

The panel suggested that this explains why certain computer games, with lower levels of realism than corresponding 'blockbuster' films, enjoy greater success than their more realistic film counterparts. It was perhaps why videogame 'cut scenes' in the same game draw criticism. As one panelist pointed out, while playing with the game characters in an interactive environment, the characters "*seem OK*", but when they stop and just stand in a '*waiting loop'* they seem "*less believable… less likeable*" ….

The implication is that as the video character is less believable in a non-response mode, its 'Uncanny' effect increases (there is less affinity with the loss of interactivity). In contrast to limited video game characters, several panelists cited the work of BabyX[10] where the interaction is critical, in the form of voice (audio), face tracking (vision), and manual keyboard input. In this simulation, the BabyX cognitive agent 'seems' to see, hear, watch, and react to the user and not just respond to button presses on a keyboard. In this way, BabyX is exhibiting far more user awareness than most video games and also makes eye contact with the user.

The emotional component of a cognitive agent directly interacting and responding to a user appears to trigger a different kind of perception, and this is an emotionally influenced response that is 'more forgiving' or more accepting than an impression made of a static or pre-recorded digital human.

5.5.2. Display environments. There was agreement amongst panelists that CG people, displayed with people in real environments, is the hardest situation to make acceptable. Extending from this issue, one panelist raised the associated point that the resolution and format of the face's presentation was a complex problem, more complex than one might first imagine.

They pointed out that *"for most of the late 2000s we were watching 4:3 programs stretched on to 16:9 TV sets… they weren't saying 'I can't recognize Jennifer Aniston in Friends reruns' - that wasn't a huge problem*".

While proportions of the face relative to itself have always been assumed to be key to successful identity, an overall disproportional scale does not make the face fall into the Uncanny Valley. In this case, our visual facial perception system *"is an amazingly robust system, and it still defies a certain amount of explanation as to how we are so good at identifying faces"*. People do not find a squashed or stretched face Uncanny when watching old shows with large resolution changes.

Building on this, a panelist pointed to people who have had either weight loss or gain. In such situations, the proportions of the face do change, but we still recognize the person. Facial hair and haircuts were mentioned as they can make someone respond "*I almost didn't recognize you*!", but in most cases one *does* recognize the person but are struck with a 'sense' that something is different.

A suggested explanation was that people have different parts of their brain processing different parts of a face. This was suggested to be primarily biological and neurological and not a learnt response.

5.6. Questioning assumptions

One outcome that contradicted accepted doctrine was that the metric of affinity is not universal but specific to the individual. The panel strongly suggested the response was an individual one, built around a range of factors, from ethnic familiarity, personal history, and familiarity with the subject.

One panelist pointed out that context is important. While one may focus on the face or head as the primary driver of acceptance, the environment that this face is presented in is also very important. A face must meet the bar of the 'world' they inhabit, especially if they are shown with other real people. The metric of the Uncanny research is not 'indistinguishable real' but simply 'affinity'. Therefore, placing the face/head in a game or VR space where sometimes the environments look stylized may help acceptance of faces that are not photo-real.

Approximately half the panelists thought an older person would be easier to achieve, with a subset of these thinking darker skin would be easier as well. *"Darker skin actually is dominated more by specular reflection than subsurface scattering"*. The same panelist raised the issue that different ethnic groups may also influence successful eye simulations, adding that *"Asian eyes might have different challenges to render than Western eyes"*.

But these points were not universally agreed upon, and some pointed to it being a subjective opinion based on one's own ethnic background. They suggested there is not an absolute affinity – but a relative affinity based on one's own individuality.

This discussion led to the suggestion to research the Uncanny Valley from the point of view of actual

people who have altered their appearance (plastic surgery, Botox etc.) and are thus moving towards the Uncanny Valley from the real-world side of the equation. One panelist questioned what alterations of their appearance could trigger a lack of affinity? "*A really interesting thing if you could get an Uncanny Valley effect from a real-life person who's had plastic surgery ..., without going into absurd cases, there's a lack of natural motion in especially foreheads, (that means that) they just don't seem to be able to emote.*"

This approach might give a window on affinity sensitivity. Following this, it was suggested to research other professionals with related non-digital skills such as makeup artists; one panelist had had great success "*interview[ing] makeup artists to find out what can they get away with, [and] what they can't*".

Finally, one panelist suggested that the whole area of interactive face acceptance may be approached from the position of some form of big data or deep learning analysis once sufficient digital faces exist. "*You might need a massive database, with lots of reference material and then you can basically decompose (analyze).... I think somebody has to do a massive, joint academic research project where they've got loads of universities processing human 4D facial data".*

6. Implications and Conclusions

We now provide a summary, review limitations and an outlook on future research.

Limitations: One of the great strengths of this study was the depth and experience of the experts, but it was limited in gender and racial diversity. This is a reflection of the imbalance in the entertainment industry and especially the technical creative sector[25]. Future work should seek to address these minority positions explicitly.

Outcomes: There are several major outcomes of the Delphi study that contradicted the accepted theory and suggest future research.

First, it has always been assumed that animation or movement would magnify the Uncanny Valley response. The Delphi panel stated that this may be true in traditional animation environments, but not in *interactive* HCI.

Second, the key to this difference is thought to be emotion. It appears we interpret the interactions as emotional responses, which either override our logical facial cognitive processing or distract us from it. When we engage emotionally, we are 'swept up in the moment'. Affective computing research has aimed to provide stronger communication and more effective interaction using emotions[24]. The difference between agent movement vs. interactivity may be the difference between someone wondering what the agent might do, compared to wondering what 'they' may be thinking, as a path to predicting behavior. This difference imbues the agent with more 'humanity'. The user reacts to tight visual non-verbal loops such as eye and head acknowledgments, and both posing and emotional matching displays to emotionally engage and thus relax realism thresholds that can otherwise be unsettling.

Third, we are proposing the opposite of the Uncanny Valley phenomenon occurs when *interactively* communicating with an agent using affective computing, and high-end graphical face rendering. While the Uncanny Valley model predicts less acceptance with movement, we have reason to believe that an '*emotional flooding of the Valley*' will result in greater success.

Future directions: Our results suggest that the Uncanny Valley should be explored from the point of view of digitally altering real people to see if there can be deduced an inflection point that makes the person seem 'uncanny'.

In terms of more technical points, there was a need for more research into the FACS pipeline and its use in mapping expressions to animation. Along with doing further research into blood flow and its sub-conscious effects.

We suggest that while there is an interrelated set of issues that affect realism, that there are several previously unrecognized aspects, which can mitigate negative reactions. This has important implications to the research into faces used in new forms of HCI.

7. References

[1] D. A. Norman, *Emotional design: why we love (or hate) everyday things*. Basic Books, 2004.

[2] T. U. Valley, M. Mori, and T. Minato, "The Uncanny Valley," *Energy*, vol. 7, no. 4, pp. 1–2, 1970.

[3] C.-C. Ho and K. F. MacDorman, "Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices," *Comput. Human Behav.*, vol. 26, no. 6, pp. 1508–1518, Nov. 2010.

[4] R. McDonnell and M. Breidt, "Face Reality : Investigating the Uncanny Valley for virtual faces," in *ACM SIGGRAPH ASIA 2010 Sketches*, 2010, p. 41.

[5] F. E. Pollick, "In Search of the Uncanny Valley," *Lect. Notes Inst. Comput. Sci. Soc. Telecommun. Eng.*, vol. 40, pp. 69–78, 2010.

[6] K. F. MacDorman, P. Srinivas, and H. Patel, "The uncanny valley does not interfere with level 1 visual perspective taking," *Comput. Human Behav.*, vol. 29, no. 4, pp. 1671–1685, Jul. 2013.

[7] K. F. MacDorman and H. Ishiguro, "The uncanny advantage of using androids in cognitive and social science research," *Interact. Stud.*, 2006.
[8] F. Ferrari, M. P. Paladino, and J. Jetten, "Blurring Human–Machine Distinctions: Anthropomorphic Appearance in Social Robots as a Threat to Human Distinctiveness," *Int. J. Soc. Robot.*, 2016.
[9] M. M. Moosa and S. M. M. Ud-Dean, "Danger avoidance: An evolutionary explanation of Uncanny Valley," *Biol. Theory*, vol. 5, no. 1, pp. 12–14, 2010.
[10] M. Sagar, M. Seymour, and A. Henderson, "Creating connection with autonomous facial animation," *Commun. ACM*, vol. 59, no. 12, pp. 82–91, 2016.
[11] M. Meng, T. Cherian, G. Singal, and P. Sinha, "Lateralization of face processing in the human brain," *Proc. R. Soc. B Biol. Sci.*, vol. 279, no. 1735, pp. 2052–2061, 2012.
[12] J. D. Foley, A. Van Dam, S. K. Feiner, and J. F. Hughes, *Computer Graphics: Principles and Practice in C*. 1995.
[13] P. Ekman, "Biological And Cultural Contributions To Body And Facial Movment," in *Anthropology of the body*, vol. 15, J. Blacking, Ed. Academic Press, London, 1977, pp. 40–80.
[14] H. Linstone and M. Turoff, *The Delphi Method, Techniques and Applications*, 2nd ed. Addision-Wesley Publishing Company, 1975.
[15] G. Skulmoski, F. T. Hartman, and J. Krahn, "The Delphi Method for Graduate Research," *J. Inf. Technol. Educ.*, vol. 6, 2007.
[16] A. Dubois and L.-E. Gadde, "Systematic combining: an abductive approach to case research," *Journal of Business Research*, vol. 55, no. 7. Elsevier Inc, NEW YORK, pp. 553–560, 2002.
[17] R. B. Akins, H. Tolson, and B. R. Cole, "Stability of response characteristics of a Delphi panel: application of bootstrap data expansion.," *BMC Med. Res. Methodol.*, vol. 5, p. 37, 2005.
[18] W. P. Aguayo, "Peirce's Theory of Abduction: Logic, Methodology, and Instinct," *Ideas Y Volores*, vol. 60, no. 145. Univ NAC Colombia, pp. 33–53, 2011.
[19] A. Dubois and L. E. Gadde, "'Systematic combining'-A decade later," *JBR*, vol. 67, no. 6. Elsevier Science, New York, pp. 1277–1284, 2014.
[20] K. G. Corley, "What Grounded Theory Is: Engaging a Phenomenon from the Perspective of Those Living it," *Organ. Res. Methods*, vol. 18, no. 4, pp. 600–605, 2015.
[21] E. Hafermalz and K. Riemer, "The Work of Belonging through Technology in Remote Work: A Case Study in Telenursing," *Twenty-Fourth Eur. Conf. Inf. Syst.*, 2016.
[22] K. F. MacDorman, R. D. Green, C.-C. Ho, and C. T. Koch, "Too real for comfort? Uncanny responses to computer generated faces," *Comput. Human Behav.*, vol. 25, no. 3, pp. 695–710, May 2009.
[23] T. W. Bickmore and R. W. Picard, "Establishing and maintaining long-term human-computer relationships," *ACM Trans. Comput. Interact.*, vol. 12, no. 2, pp. 293–327, 2005.
[24] R. W. Picard, *Affective computing*. MIT Press, 1997.
[25] B. Lang, "Hollywood Gender Gap: Women Comprise 7% of Directors.," *Variety*, 2015. [Online]. Available: https://variety.com/2015/film/news/women-hollywood-inequality-directors-behind-the-camera-1201626691/. [Accessed: 13-Jun-2018].

A special thanks to the expert panelists for their generous time and considered opinions.