Proceedings of the 55th Hawaii International Conference on System Sciences | 2022

Face It, Users Don't Care: Affinity and Trustworthiness of Imperfect Digital Humans

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

Lingyao Yuan
Iowa State University
lyuan@iastate.edu

Alan R. Dennis
Indiana University
ardennis@indiana.edu

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

Abstract

Digital humans are growing in application and popularity, both as avatars for people and as standalone artificial intelligence-controlled agents. While the technology to make a digital human look more realistic is improving, we know little about how realistic they need to be. Humans are exceptionally good at identifying imperfect digital reproductions of human faces, so it has been reasoned that the slightest imperfections in the visual design of digital humans may translate into reduced acceptance and effectiveness. The broadly held wisdom is that digital humans should be photorealistic and indistinguishable from real people. To examine this common belief we collected data on individuals' affinity and trustworthiness in photorealistic digital humans when engaged in a product bidding situation, along with a human presenter with varying degrees of video imperfections. The results reveal that participants noticed some of the video imperfections, but this did not adversely affect their willingness to pay, affinity, or trust. We found that once digital humans become close to realistic, users simply do not care about visual imperfections

1. Introduction

There has been a steady increase in the use of realistic digital characters, both avatars and artificial intelligence (AI)-controlled agents [17]. Digital humans have been widely adopted in many industries, such as fashion, entertainment, gaming, education, and corporate communications. The field of digital humans extends from digital representations of people in films and videos to fully synthetic AI-controlled call center agents, sales assistants, and digital influencers. The application of these agents is underpinned by advances in computer graphics and new technologies, such as neural rendering, an advanced form of "deep fakes" based on machine learning. When presenting a digital human, a choice of visual appearance must be made. While some applications have deliberately selected a cartoon or graphical representation, such as Apple's emojis, many have focused on a realistic simulation of the human face.

Prior research has demonstrated that more realistic digital humans are considered more trustworthy than stylized cartoon versions of humans [18]. However, not much research exists on whether the realism of digital humans must reach a level of perfection to achieve the same level of trustworthiness as real humans.

Most of the early deployment of digital humans has been in the film and entertainment industries. In these fields, the need for realism and believability is often paramount. Millions of dollars are spent in film production to create believable characters. These believable characters make impossible stories believable. For example, in the film *Gemini Man*, a digital version of actor Will Smith fights a younger clone of himself. In *Star Wars: The Rise of Skywalker*, the late actor Carrie Fisher reprised her role, even though the audience was aware she had recently died.

In these and many other cases, the illusion of the digital human must be as realistic as possible, or the story will fail to be believable, regardless of the impossible storyline of clones or intergalactic star wars. If the illusion is obviously fake, the audience frequently loses empathy for the character or plot. With the story itself being fantastical, the essence of a film is often to tell an impossible narrative as plausibly as possible, so the viewer is not 'taken out' of the viewing experience.

Understandably, other industries using digital humans as customer agents and sales representatives have adopted the movie and gaming industry standard to create digital humans. These industries also require digital humans to be as visually and perceptibly realistic as possible. However, the question whether industries other than the film industry actually benefit from the cost of creating exceptionally realistic digital humans remains unanswered.

The rapid advances in computer graphics, specialist rendering GPU hardware, and new technological innovations in neural rendering have all made it possible to cost-effectively produce near-

URI: https://hdl.handle.net/10125/79858
978-0-9981331-5-7

realistic digital humans that are interactive and scalable in deployment. Neural rendering technology produces near perfect digital humans by inferring realistic human faces from training data using deep learning techniques, such as generative adversarial networks (GANs). However, while current technology is making breakthroughs frequently, visual imperfections still exist and are detectable.

The media and entertainment (M&E) industry has the economic advantage of being able to spend vast amounts of money and time to achieve the results seen in tent-pole theatrical releases. Although non-M&E applications are likely to be produced with smaller budgets, their industrial applications often require the resulting digital human to be produced much faster and yet still provide a plausible result. For example, feature films may take hours to render a single frame, while real-time interactive applications have only milliseconds. Such an ability to create imperfect yet "good enough" digital humans at a more attainable cost for non-M&E businesses has opened new opportunities. However, before investing in such applications, companies must address the question: How good is good enough for users?

The uncanny valley theory [12] was developed before the widespread consumption of media. During that time, smartphones and other portable devices were not common for delivering high-resolution imagery at our fingertips. Technology has fostered a sustained period of increased image quality and audio fidelity, bringing users immediacy, convenience, and immediate gratification. People's understanding and appreciation of technology have evolved along with innovations in technology. However, recent socio-technical trends, such as the consumption of vast amounts of user-generated content, have not focused on image fidelity. This situation may have changed the underlying sense-making users exhibit concerning digital humans.

We surveyed Amazon Mechanical Turk participants on their affinity and trust regarding multiple versions of a live-action video promoting a consumer product to research this issue. We added different visual imperfections to a professionally created video and audio recording to produce different versions. The imperfections range from subtle changes to highly visible distortions. In addition, we created a video using an M&E standard photorealistic digital human. All videos were rated for video quality, affinity, and trustworthiness of the presenter, and ultimately, how much a user was willing to pay for a product advertised by the presenter. All videos had the same script, environment, and base actors. Thus, the paper addresses two questions:

RQ1: Do visual imperfections adversely affect viewers' bidding behavior (willingness to pay), affinity and trust toward the actor in a video presentation in the online auction context?

RQ2: Are digital humans able to approach the same level of bidding behavior (willingness to pay) affinity and trustworthiness as a human presenter in online auction context?

2. Background

The area of digital humans is expanding rapidly, with commercial digital humans available from several companies (e.g., see soulmachines.com, neon.life, and digitaldomain.com). It is no longer just the domain of high-end media and entertainment projects to create high-fidelity digital humans. It is not difficult for organizations to create custom digital humans. It is even possible for individuals to create personalized digital avatars. Many companies provide a platform and technology for users to create digital humans. One example is the MetaHuman Creator, a tool developed by Epic Games (Figure 1). This tool enables users to create computer-generated digital humans with a vast range of races and appearances, achieving diversity in digital humans.

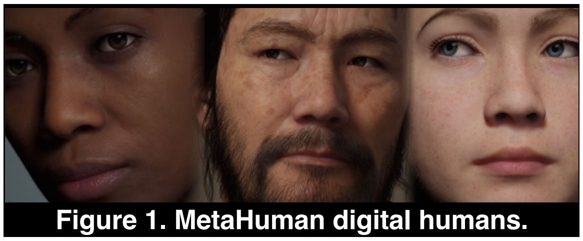

Figure 1. MetaHuman digital humans.

The $600 billion fashion industry is an early adopter of digital humans in e-commerce. Visually plausible AI-controlled digital humans assist, advise, and influence sales by providing online digital fittings to reveal how consumers look in the latest fashion styles.

Virtual influencers are another critical fashion industry driver. For instance, Lil Miquela and other digital influencers are globally successful, with millions of followers. Their success is not due to their followers believing they are real but instead that the experience and digital influencers' sentiments seem 'authentic' and 'genuine' for the digital personality presenting them.

This result suggests that the visual quality of the characters must reach a minimum level of quality to communicate emotional content adequately; however, perfect realism does not appear to be the primary driver. Industries such as health, education, finance, and corporate communications are all moving to implement digital humans and avatars. Given this trend, it is vital to examine the role image quality plays in providing a trusted customer experience.

2.1. Uncanny Valley

The uncanny valley theory argues that users have a greater affinity (or natural liking) for avatars that are more realistic and humanlike [12]. This theory explains why it is crucial to provide a high-quality digital human avatar. User affinity increases as the avatar becomes more realistic until the avatar is almost realistic, at which point affinity drops dramatically, creating a "valley." The valley occurs because a semi-realistic avatar triggers unease in users (Figure 2). As realism increases further, after crossing the valley, the avatar's affinity resumes increasing to the highest level [18].

Affinity is an indicator of how humanly realistic and favorable users perceive an avatar to be. We can easily observe the uncanny valley for affinity. However, the original theory, especially the graphical representation, was created purely based on theory without empirical evidence. Therefore, the implied right-hand side of the diagram (after crossing the valley) is simply a theoretical prediction. There is no evidence to prove how the sharp and steep gradient of affinity improves as one approaches the level of complete realism. The diagram has no empirical scale for either affinity or increasing realism.

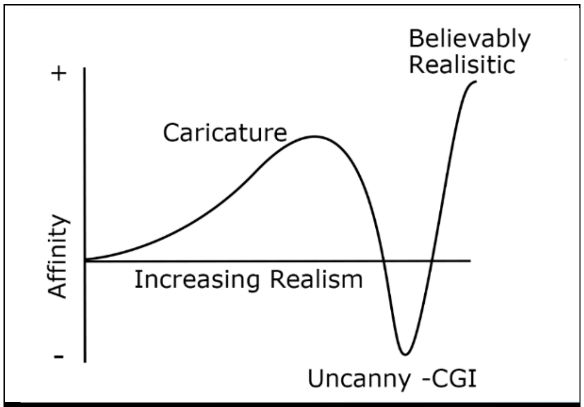

Figure 2. Uncanny Valley

Prior research has demonstrated that the uncanny valley effect exists for digital humans. Specifically, digital humans with poor fidelity invoke a low level of affinity [18]. It has also been established that a modern high-fidelity digital human can be considered more trustworthy than a cartoon version [18]. In other words, it is worth the effort to produce a more complex and realistic digital human. There is a greater affinity for a highly believable representation than for a caricature [18].

The question remains; by how much does a current high-fidelity digital human fall short in affinity, compared to a real human? Furthermore, for a digital human that has crossed the uncanny valley, but not undetectably real, is there still a significant deficit in the level of trustworthiness compared with a real person?

Piror research has also questioned the nature of any manipulation that seeks to decrease realism. Kätsyri et al. [6] states that "the uncanny valley could be understood as the naïve claim that any kind of human-likeness manipulation will lead to experienced negative affinity at close-to-realistic levels. It further suggests that the uncanny valley phenomenon could be caused by a perceptual mismatch between artificial and actual human features. It draws attention to the role of different kinds of manipulations and also suggests that a "generally humanlike character with subtle flaws in some focal features (e.g., eyes), would be likely to elicit negative affinity".

The uncanny valley theory was proposed in the 1970s. At that time, it had no empirical basis. Subsequent studies have validated the valley phenomena but have not focused on the last part of the theory regarding approaching parity in affinity as it pertains to trustworthiness. Importantly the uncanny valley theory has been co-opted in some popular circles as shorthand for being a type of visual Turing test of believability, although that was not the theory's original purpose.

The uncanny valley theory provides two important perspectives for our research. First, the digital human must be realistic enough to be clear of the off-putting valley effect of partial realism. If this is not met, then many users find interacting with digital humans to be very negative. The second point is much more implicit. As a poor or unfaithful representation is off-putting, exhibiting low affinity, it is assumed that any imperfection must result in a significant level of negative consequences such as an accompanying loss of trust. However, this assumption has not been investigated or confirmed.

The differences in affinity created by different types of imperfection may seem insignificant, but identifying the differences has considerable practical implications. For many businesses, the commercial viability of many applications pivots on this point. Humans are good at spotting facial imperfections, even small ones. We have been trained since birth to respond to faces and people's body language in general. Obvious imperfections that create mental blocks for viewers may compromise the ability of viewers to accept the content delivered in the videos. Slight real-world facial imperfections may be noticed but not dealt with mentally in a way that generates behavioral consequences. Perfect visual facial features may be desirable in an ideal world, yet imperfect, but adequate facial features are much more probable and normal in everyday life.

The same rule can be extended to digital humans. Creating indistinguishable perfect digital humans is ideal but expensive and economically challenging. Nevertheless, creating "good enough" digital humans can be more cost-effective than filming real humans. Users may notice the imperfections and are cognitively aware that the digital human is a simulation from various visual cues. However, if such awareness does not adversely reduce the digital human's effectiveness, then a wide range of new applications becomes commercially viable.

2.2. Video Distortion

In this study, we independently examined five different visual distortions. These five approaches were selected to provide a range of degrees of noticeable distortion and focus on the eyes, mouth and body separately. This focus reflects the widely understood importance of the eyes and mouth in human communication and allows us to explore the sensitivity to the lack of realism in these critical aspects of human representation. For each distortion, we first want to understand whether it is visually detectable. For each distortion, we looked to understand whether the treatment generated behavioral consequences in trustworthiness.

Each visual distortion was designed to produce a visual effect that would not appear to the subject as simply a local playback or streaming content error. Pauses, visual glitches, or loss of playback audio synchronization may be perceived as nonexperimental Internet issues and thus be disregarded. Therefore, it was essential to produce previously unseen digital distortions simulating a lack of realism.

We also included one digital human version of the same human presenter. The digital human was produced with industry-level high quality. Even with the highest level of technical fidelity, the digital human still carries noticeable imperfections from the real human agent. However, those imperfections are unique from the five purposeful distortions.

2.2.1 Nondistorted

The nondistorted video presents a standard view of a presenter in a medium close-up shot, lit professionally and addressing the camera, which is characteristic of how a presenter may be filmed in any typical setting. The lighting is even, and there are no additional foreground elements. The audio is clear and without distortion and the audio was also not altered in any of the subsequent treatments. This video is the baseline that serves as the control.

2.2.2 Stretch

The imagery was scaled horizontally (i.e., made wider) to a degree that an observer could not miss. The effect is to produce a video that is still clearly a person but grossly affects the aspect ratio of the presenter. Such a transformation applies equally to all aspects of the presenter: the head, eyes, mouth, and body. This transformation keeps the relative aspects of the presenter consistent with each part and does not highlight any individual aspects of the presenter for special visual consideration. The purpose of this video is to test the effects of a significant distortion that should be obvious to all viewers.

2.2.3 Warped

It is commonly held that the classic gaze pattern on a face is a triangle of interest from each eye to the mouth and back again. In this treatment, as with the next, these areas of high interest were isolated and degraded. The most noticeable characteristic of the mouth is its ability to accurately reflect the shape of the sounds heard when delivering dialog. The mouth is primarily affected by degrading it through speed warping techniques, as this leads to specific higher frequency lip movements being lost. The effect is not isolated to the lips, but the lips are one of the most affected parts of the face. This warped loss of fidelity compares to simply offsetting the audio timing, which would maintain the fidelity of the movement but with an audio delay. This simpler audio timing treatment was rejected as it could be perceived as an audio issue and not a visual realism artifact. By comparison, the warped approach keeps the audio synchronized, but the presenter 'speaks' incorrectly. The purpose of this distortion is to test the effects of a distortion focusing on the mouth.

2.2.4 Eyes Stabilized

An approach was taken to stabilize the eyes and recomposite them back onto the face to focus the distortion effect on just the eyes. This method provided an incorrect stare or look. Unlike the mouth, which must accurately provide a high volume of specific shapes that match the dialog, the eyes' shape, convergence, and subtle movement determine the perception of where someone is looking. By stabilizing the eyes, the perception is of an altered gaze that is unnaturally fixed on an oddly off-screen location. This effect only influences the gaze. The speech, head movement, and body motion remain unaffected. The purpose of this distortion is to test the effects of a distortion focusing on the eyes.

2.2.5 Shirt Stabilized

Stabilizing the shirt leaves the face unaffected while the whole body is stabilized at one point on the presenter's shirt. This stabilization provides an unnatural overall body motion that is undetectable on a single still frame. In doing so, we focus on the degradation of the movement of the whole presenter in a subtle yet impossibly unrealistic way. It is impossible to move and talk while maintaining a perfectly centered, fixed core body position. It is both unnatural and subtle to the untrained or uninformed observer, but is not impossible to identify if the footage is viewed closely. This distortion aims to test the effects of a subtle distortion outside the facial area because digital human development often focuses on the face, and body language is also important to effective communication.

2.2.6 Blink Reduction

Reducing blinking is the subtlest distortion. The aim was to do something that would perhaps trigger the sense that something was imperfect but at a level that was expected to be the hardest to detect. If there was some unconscious response to *any* type of degradation, no matter how small, it was reasoned that this treatment would highlight the phenomenon. The eyes in this treatment do not fully blink, but the lids do partially move as if they were blinking, but the actual blinks are removed. The remaining lid twitch would be impossible to perform in reality, but the visual difference overall is extremely minor. This distortion aims to test the effects of a very minor distortion that may not be consciously detected.

2.2.7 Digital Human

A digital human is created to represent the fully digital view of the same presenter. It is crucial to place a state-of-the-art digital human in the context of the degraded presenter videos to measure whether the level of the visual quality of all or any of the treatments is of the same order as the visual 'unreality' of a fully digital human. This video is not indistinguishable from the control human presenter video. It represents a practical level of actual human replication fidelity and is a real-world working digital human, not a simulation. The video was produced with an underlying performance as close to the other treatments as possible. The purpose of this treatment is to test the effects of commercially available digital humans.

2.3. Affinity, Willingness to Pay, and Trustworthiness

To observe the behavioral consequences of the distortions, we examined three outcomes (affinity, willingness to pay, and trustworthiness) tied to digital humans and the application domain of e-commerce. Willingness to pay refers to the amount of money an individual is willing to pay for a given product [24]. It is set subjectively by individuals [24], reflecting the consumers' perceived value of the product [16]. In the e-commerce context, such as online auctions, customers' willingness to pay is influenced by many factors, such as product information, product image, product reviews, pricing strategy [21, 2], and system design. Bidding decisions are not purely rational and are prone to be influenced by uncertain elements embedded in the environment and unique user characteristics [19, 26]. For example, if using video to deliver product information, video quality naturally would affect the bidding decision.

Trust is an individual's willingness to be vulnerable to the actions of others [10]. Trustworthiness is an assessment of whether another person or thing is worthy of trust [10]. Trust can describe interpersonal relationships and a person's attitude toward avatars, virtual agents, machines, and information systems [9, 22, 1, 8]. As indicated, affinity captures how realistically viewers perceive a virtual agent or avatar and the level of positive attitude generated as a result of realism. Affinity and trustworthiness in online avatars and virtual agents are important factors that influence whether consumers visit and purchase from online retailers [4]. In this research, product information is delivered through a human or digital human presenter. Therefore, the visual quality of the presenter may influence other outcomes.

3. Method

3.1 Participants

We recruited 775 participants from Amazon Mechanical Turk following the recommendations of [20]. We recruited adults in the United States who had completed more than 1,000 HITS with a 98% success rate. We removed 49 (6%) who failed one or more of the four attention checks. We also removed nine who said they would not buy the tablet. Thus, we have a final sample of 727 participants. About 61% were male, 81% Caucasian, 9% Black, and 7% Asian. Age ranged from 21 to 83, with a mean of 38.9 years.

3.2 Task

Participants watched a video with a presenter describing a new tablet from Apple with a list price of $329 and entered the amount they would bid for the tablet. They reported their perceptions of the video quality and the affinity and trustworthiness of the presenter. The video script is provided in the appendix.

3.3 Treatments

Participants were randomly assigned to watch one of the seven videos. The seven videos were produced using state-of-the-art computer graphics and professional cinematography. The control video was a direct presentation of the original script and served as the ground truth of the correct delivery and visuals. The video was shot against a chromakey green background so that all presentations, both real and artificial, would have the same background. The five distortion video treatments were all variations of this control treatment, modified during post-production.

The final treatment, the digital human, was produced using state-of-the-art motion capture, 3D facial simulation, and 3D and neural rendering to simulate the original presenter as closely as possible. The motion capture suit was an Xsens suit connected to a computer with a high-end Nvidia GPU supporting a UE4 live-action 3D character, including facial animation. This setup was processed with a custom paGAN machine-learning face replacement. All treatments used the same audio. The seven treatments were (Figure 3) as follows:

1. **Control**. In the control version, the presenter is posed in a typical office background. This version is the most technically correct version, with no treatment or degradation. Video link: https://tinyurl.com/HICSS2022v1

2. **Stretch**. This version distorted the live-action presenter so that the aspect ratio was incorrect. This distortion gave the effect of the presentation being grossly distorted or stretched horizontally. The reference was an old-style 4:3 video being played back on a 16:9 aspect ratio monitor. Video link: https://tinyurl.com/HICSS2022v2

3. **Warped**. All but every tenth frame was discarded and replaced with a set of interpolated frames. Thus, the duration remained the same, but all micro-movements were removed. The lip synchronization was less accurate because the live action was only sampled every 10 frames. As a result, the presenter appears unnaturally smoothed or averaged in their speech patterns and visemes. Video link: https://tinyurl.com/HICSS2022v3

4. **Eyes stabilized**. The presenter's eyes were modified to be fixed and unnatural while still tracking any head movements. The effect is to make the face seem not unlike a painted mask. Video link: https://tinyurl.com/HICSS2022v4

5. **Shirt Stabilized**. The face is unaffected, but the whole body is stabilized around the shirt. The net effect is that the body motion is unnatural, but the face is unaltered directly. Video link: https://tinyurl.com/HICSS2022v5

6. **Blink reduction**. The eyelids are altered never to blink fully, which is the most subtle effect. One would expect this effect to be the most difficult to notice because the alteration is the most minor. Video link: https://tinyurl.com/HICSS2022v6

7. **Digital human**. This version uses a digital human avatar, a copy of the presenter rendered digitally. Video link: https://tinyurl.com/HICSS2022v7

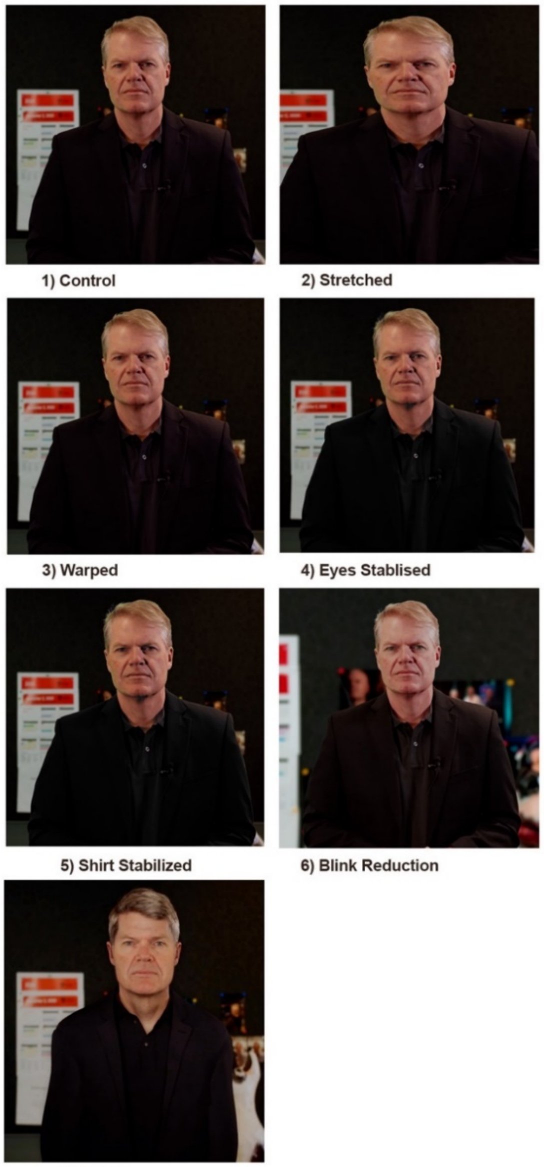
1) Control
2) Stretched
3) Warped
4) Eyes Stablised
5) Shirt Stabilized
6) Blink Reduction
7) Avatar

Figure 3. Experiment Conditions

3.4 Measures

We quantified four outcome measures. The first was the bid amount. The second was video quality (four items from [23], alpha = .84). The third and fourth were affinity (four items from [18], alpha = .93) and trust (seven items from [18], alpha = .95), respectively. The items and exploratory factor analysis loadings are listed in Table 1.

A power analysis using G*Power [5] determined that the design had a power of .92 to detect minor effects of 0.15. Next, we analyzed the data with four separate analyses of variance (ANOVAs), one for each outcome variable. Treatment had significant effects only on video quality (*F* (6,720) = 9.27, *p* = .000). Table 2 presents the treatment means, standard deviations, and statistical results.

4. Results

Table 1. Items and Factor Loadings

Item		Factor		
		1	**2**	**3**
AF1	I have affinity with the sales agent	0.446	0.786	0.077
AF2	I feel a closeness to the sales agent	0.406	0.814	0.070
AF3	I feel a likeness with the sales agent	0.485	0.759	0.084
AF4	I feel a rapport with the sales agent	0.464	0.767	0.067
T1	Overall, the sales agent is trustworthy	0.836	0.313	0.179
T2	I trust the sales agent	0.818	0.366	0.102
T3	I can rely on the sales agent	0.818	0.376	0.083
T4	The sales agent can be trusted with selling the product	0.815	0.278	0.175
T5	I have confidence in the sales agent	0.832	0.310	0.142
T6	I feel confident about the sales agent’s skills	0.777	0.283	0.188
T7	The sales agent had integrity	0.788	0.246	0.134
VQ1	The video has good quality	0.111	0.153	0.882
VQ2	I can view the video clearly	0.219	-0.277	0.699
VQ3	There is no issue with the quality of the video	0.101	0.052	0.826
VQ4	The video is high quality	0.110	0.264	0.834

Table 2. Descriptive Statistics and Results

	Bid ***Mean (SD)***	**Video Quality** ***Mean (SD)***	**Affinity** ***Mean (SD)***	**Trust** ***Mean (SD)***
Control	244.84 (86.89)	6.26^{a} (0.74)	4.36 (1.37)	5.06 (1.29)
Blink Reduction	253.67 (72.37)	6.22^{a} (0.93)	4.33 (1.36)	5.08 (1.17)
Eyes Stabilized	249.92 (72.07)	5.55^{c} (1.31)	4.19 (1.41)	4.83 (1.42)
Shirt Stabilized	245.58 (83.26)	6.23^{a} (0.76)	4.23 (1.36)	4.89 (1.26)
Warped	246.44 (81.30)	5.65^{c} (1.21)	4.14 (1.19)	4.85 (1.10)
Stretched	256.58 (67.70)	5.74^{b,c} (1.11)	4.27 (1.28)	4.85 (1.12)
Digital Human	248.62 (75.24)	6.02^{a,b} (0.93)	3.93 (1.45)	4.81 (1.19)
F	0.33	9.27	1.20	0.87
p	.920	.000	.305	.516
***R*²**	.00	.07	.01	.01

Note: The letters a, b, and c for video quality refer to the groupings from the post-hoc REGW test. Treatments with different letters are significantly different from each other.

We conducted a post-hoc analysis on video quality using a REGW *F*-test and found the control, blink reduction, and shirt stabilized treatments have the highest video quality and are not statistically significantly different from each other. The warped and eye stabilized treatments had the lowest video quality and were no different from each other. The stretched and digital human treatments were in the middle, with the digital human treatment having higher quality than the lowest group and was no different from the highest group. The stretched treatment had lower quality than the highest group and was no different from the lowest group.

We were particularly interested in the digital human avatar, which was not significantly different from the undistorted control treatment for any measure. The effect sizes of the difference in video quality were $d = .23$, affinity was $d = .32$, trust was $d = .21$, and the bid amount was $d = .05$, all of which Cohen [3] called "small."

We conducted post-hoc ANOVAs to understand the extent to which perceptions of video quality influenced the bid amount, affinity, and trustworthiness. Perceptions of video quality had no significant effect on the bid amount ($F\,(1{,}725) = 2.13, p = .144$) but significantly influenced affinity ($F\,(1{,}725) = 57.52, p = .000$) and trustworthiness ($F\;(1{,}725) = 112.76, p = .000$), with an R^2 that Cohen called "small" (.07 and .11, respectively).

In summary, participants recognized the imperfections in the video, but the effects were minor. This recognition had a negligible effect on affinity and trustworthiness for some participants, but the amount they were willing to spend was completely unaffected by the distortions. Cohen says a small difference is so slight that it is not noticeable to a casual observer (i.e., *not* "large enough to be visible to the naked eye") (p. 26). Thus, we conclude that users generally do not care about the imperfections in the videos.

5. Discussion

Participants perceived noticeable differences in video quality among the various treatments. However, there were no significant differences in the bid amount, affinity, or trustworthiness. Our participants noticed differences in video quality, but more importantly, these differences did not affect their behavior and perceptions. Simply put, the participants did not care about the video imperfections.

As we theorized, the warped and eyes stabilized versions had the lowest video quality, consistent with the broadly accepted view that people focus much of their attention on someone's eyes and mouth. Both of these scored lower than the stretched or actual digital human avatar, which implies that when the eyes or mouth are out of alignment with the rest of the person, cognitive dissonance is more profound than when the whole head, including the eyes and mouth, is degraded equally.

In the case of the stretched version, all of the head and face were stretched equally. Similarly, for the digital human avatar, the presenter was entirely synthetic, which implies that consistency is vital to the treatment of the presenter. Moreover, it implies that improvements to the digital presenter should be considered a holistic quality issue, rather than a need to focus on any single major aspect of the presenter.

The significant result is that even with this range of visual quality and fidelity of representation, once a reasonable level of representation is achieved, there is no imperative to achieve a level of reproduction that is indistinguishable from reality to create an effective digital human avatar.

The findings of this study cannot be taken without the consideration of its context. Our study shares the same limitation as any studies that use Amazon Mechanical Turk panels for data collection. Subjects on this platform are biased towards Caucasian male. We cannot rule out effects of gender, race, sexual orientation, and other demographic factors. We hope our research would inspire future research on this topic.

5.1. Good Enough Is Good Enough

Based on prior research on the uncanny valley, we expected lower quality renditions of the video character to influence user affinity and decision-making negatively. Our study reveals that this is not the case. It appears that 'good enough' is indeed good enough and that certain blemishes in quality, while detectable, do not seem to matter to other attitudes. We suggest that this is because our use of technology and media consumption has changed in the past decade, as we have become used to lower video quality. In other words, while high-fidelity video might have mattered some years ago, high-fidelity video is no longer a critical driving force in consumer attitudes and behavior.

A consideration of the developments in the music industry [14] is helpful. Audio has lower bandwidth requirements, so it precedes video developments by five to ten years. In the predigital era, innovation in the music industry was driven by the paradigm of 'perfecting sound forever' [13, 11]. The industry was organized around record production and consumption by the notion of 'high fidelity,' with consumers investing considerable resources into creating the perfect sound setups. The move from vinyl to compact disc was consistent with this paradigm. The invention and diffusion of mp3 [7, 25] triggered a paradigm shift away from record ownership and perfect sound toward music sharing and anywhere, anytime accessibility [15]. Music became mobile and is now primarily consumed in lower sound quality via headphones in often noisy environments. High-fidelity music quality is no longer a driving factor for the majority of music consumers.

We suggest that video developments and the growth of mobile devices have similarly moved most users away from the high quality of the big television screen, which is still an important but not the predominant platform on which video content is consumed. With the popularity of user-generated content on such platforms as YouTube or TikTok, the sharing of clips on social media, and video consumption on mobile devices, users are now more used to and tolerant of lower video quality. We suggest that this is the reason for user acceptance of the lower quality renditions in this study. This acceptance poses opportunities for the designers of digital humans, as users are more forgiving of less-than-perfect renditions.

5.2. Design Considerations

Our findings provide encouragement for industry practitioners, as they indicate that users accept 'good enough' digital human renditions. The concrete design implications are that a consistent and high-enough-quality digital human can be deployed as a trustworthy advocate. The audience is willing to accept the digital human as a proxy, even though they also perceive artifacts that indicate the presenter as not visually perfect.

These results have important implications for the validity of the growing industry of avatar and digital human reproduction. The assumed negative consequences of a less-than-perfect reproduction are not statistically significant. Industry practitioners should be encouraged to focus on other aspects that may be more productive, such as the traditional issues of matching the presenter to the task and the audience's preferences. The results also validate a range of applications for using digital humans, as the audience can perceive the avatars are not real, but this does not adversely the avatars' effectiveness.

5.3. Implications for Future Research

Given these findings, future research should explore four major areas. First, our research points to the importance of video quality pertaining to the eyes and mouth. Humans often focus on this triangle, perhaps because they are central to both spoken and nonverbal communication. Thus, future research should focus on facial aspects in driving video quality.

Second, our participants were able to detect certain distortions, but not others. More importantly, even when they recognized distortions, those distortions had no meaningful effect on important e-commerce outcomes. We need further research in other areas where digital humans are deployed to determine whether these effects also apply there (e.g., customer service and personal assistants). We believe that the patterns in other areas are likely to be the same—that users notice visual defects, but they do not care about them. However, more research is needed.

Third, we controlled for sound quality in all videos. Sound quality is another critical aspect of communication, so this suggests that sound quality may be important. Do sound quality reproduction and synthesis match visual reproduction and simulation findings in that distortions are noticed but do not influence other attitudes and behavior? We need more research to better understand how sound distortion is noticeable and its effects.

Finally, what other factors aside from the fidelity of reproduction influence the affinity and acceptance of digital humans? For example, does visual fidelity imply a level of cognitive fidelity? Does this statistically affect the results? In other words, if this cognitive expectation is not matched, if the digital humans fall short in our understanding, does the mismatch produce something akin to a cognitive valley or cognitive cliff? Furthermore, how much does the digital human have to back away from an acceptable perceived cognitive ability to fall off a cognitive cliff?

6. Conclusions

Our research set out to answer the practical question of whether digital human can replace real human actors in video presentations. Digital humans are not yet visually perfect but quickly approaching the same level of realism, as visual design techniques are continuously improving and new technologies emerge.

We manipulated in five different ways (stretch, warped, eye stabilized, shirt stabilized, and blink reduction) a video sales presentation featuring a real human actor and created a state-of-the-art digital human of the same human actor. We compared the effects of imperfect humans and digital humans on three behavioral outcomes, bidding behavior (willingness to pay), affinity, and trustworthiness. Our results show that individuals are more prone to notice the design details in the triangle area of eyes and mouth. Any imperfection in this area lead to an individual awareness of video imperfection. However, even though people may consciously recognize the imperfections in the video quality, this realization does not materially adversely affect the bid amount, affinity, and trustworthiness.

More interestingly, the same findings applied to both real human presenter as well as the digital human. We conclude that digital humans can match real human presenters with no material differences in bid amount, affinity, and trustworthiness. The findings of this research have implications for companies wishing to adopt digital humans in their sales presentations and other capacities. We also point out implications for research in the area of digital human design.

7. References

[1] I. Benbasat and W. Wang, "Trust in and adoption of online recommendation agents", Journal of the association for information systems, 6 (2005), pp. 4.

[2] X. Chen, A. Ghate and A. Tripathi, "Dynamic lot-sizing in sequential online retail auctions", European Journal of Operational Research, 215 (2011), pp. 257-267.

[3] J. Cohen, Statistical power analysis for the behavioral sciences, Academic press, 2013.

[4] R. Etemad-Sajadi, "The impact of online real-time interactivity on patronage intention: The use of avatars", Computers in Human Behavior, 61 (2016), pp. 227-232.

[5] F. Faul, E. Erdfelder, A. Lang and A. Buchner, "A flexible statistical power analysis program for the social, behavioral and biomedical sciences", Behavior Research Methods 39.2 (2007): 175-191.

[6] J. Kätsyri, K. Förger, M. Mäkäräinen and T. Takala, "A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness", Frontiers in Psychology, 6 (2015).

[7] B. Kernfeld, Pop song piracy: Disobedient music distribution since 1929, University of Chicago Press, 2011.

[8] S. Y. Komiak and I. Benbasat, "The effects of personalization and familiarity on trust and adoption of recommendation agents", MIS quarterly (2006), pp. 941-960.

[9] P. B. Lowry, A. Vance, G. Moody, B. Beckman and A. Read, "Explaining and predicting the impact of branding alliances and web site quality on initial consumer trust of e-commerce web sites", Journal of Management Information Systems, 24 (2008), pp. 199-224.

[10] R. C. Mayer, J. H. Davis and F. D. Schoorman, "An integrative model of organizational trust", Academy of management review, 20 (1995), pp. 709-734.

[11] G. Milner, Perfecting sound forever, Le Castor Astral éditeur, 2017.

[12] M. Mori, K. F. MacDorman and N. Kageki, "The uncanny valley [from the field]", IEEE Robotics & Automation Magazine, 19 (2012), pp. 98-100.

[13] E. Reynolds, "Perfecting Sound Forever: An Aural History of Recorded Music", Fourth Genre: Explorations in Nonfiction, 13 (2011), pp. 169-172.

[14] K. Riemer and R. B. Johnston, "Disruption as worldview change: A Kuhnian analysis of the digital music revolution", Journal of Information Technology, 34 (2019), pp. 350-370.

[15] K. Riemer and R. B. Johnston, "Wither Interpretivism? Re-interpreting interpretation to fit a world of ubiquitous ICT", (2019).

[16] S. Rosen, "Hedonic prices and implicit markets: product differentiation in pure competition", The journal of political economy (1974), pp. 34-55.

[17] M. Seymour, K. Riemer and J. Kay, "Actors, avatars and agents: potentials and implications of natural face technology for the creation of realistic visual presence", Journal of the Association for Information Systems, 19 (2018), pp. 4.

[18] M. Seymour, L. I. Yuan, A. Dennis and K. Riemer, "Have We Crossed the Uncanny Valley? Understanding Affinity, Trustworthiness, and Preference for Realistic Digital Humans in Immersive Environments", Journal of the Association for Information Systems, 22 (2021), pp. 9.

[19] H. A. Simon, "A behavioral model of rational choice", The quarterly journal of economics, 69 (1955), pp. 99-118.

[20] Z. R. Steelman, B. I. Hammer and M. Limayem, "Data collection in the digital age", Mis Quarterly, 38 (2014), pp. 355-378.

[21] A. K. Tripathi, S. K. Nair and G. G. Karuga, "Optimal lot sizing policies for sequential online auctions", Knowledge and Data Engineering, IEEE Transactions on, 21 (2009), pp. 554-567.

[22] A. Vance, C. Elie-Dit-Cosaque and D. W. Straub, "Examining trust in information technology artifacts: the effects of system quality and culture", Journal of management information systems, 24 (2008), pp.73-100.

[23] J. D. Wells, J. S. Valacich and T. J. Hess, "What signal are you sending? How website quality influences perceptions of product quality and purchase intentions", MIS quarterly (2011), pp. 373-396.

[24] K. Wertenbroch and B. Skiera, "Measuring consumers' willingness to pay at the point of purchase", Journal of marketing research, 39 (2002), pp. 228-241.

[25] S. Witt, How music got free: The end of an industry, the turn of the century, and the patient zero of piracy, Penguin, 2015.

[26] L. Yuan and A. R. Dennis, "Acting like humans? Anthropomorphism and consumer's willingness to pay in electronic commerce", Journal of Management Information Systems, 36 (2019), pp. 450-477.

Appendix: Video Script

"The Apple 10.2" Retina Display iPad.
It has the latest and most powerful bionic A12 chip.
With the Apple Pencil, iPad OS14 and the new Handwriting function you can now hand-write in any text field and the iPad will convert your handwriting to text.
Mark up PDFs or screengrabs - or just turn hand scribbles in meetings, into searchable notes.
But you can also use it for fun – it can pair with an Xbox or PS4 controller. And with the 8MP camera you can shoot edit and finish HD Video right on the iPad and upload it immediately… – it even has Stereo speakers.
Which is great as it comes with one year's free Apple TV included."