The Effect of Onset Asynchrony in Audio Visual Speech and the Uncanny Valley in Virtual Characters

Tinwell, Angela; Grimshaw, Mark; Abdel Nabi, Deborah

Published in:
International Journal of Mechanisms and Robotic Systems

DOI (link to publication from Publisher):
10.1504/IJMRS.2015.068991

Publication date:
2015

Document Version
Accepted author manuscript, peer reviewed version

The effect of onset asynchrony in audio-visual speech and the Uncanny Valley in virtual characters

Angela Tinwell*

Faculty of Arts and Media Technologies,
The University of Bolton,
Deane Road, Bolton, BL3 5AB, UK
E-mail: A.Tinwell@bolton.ac.uk
*Corresponding author

Mark Grimshaw

Aalborg University,
Institute for Communication,
Kroghstræde 6, Lokale 19,
9220 Aalborg Ø, Denmark
E-mail: grimshaw@hum.aau.dk

Deborah Abdel Nabi

Faculty of Wellbeing and Social Sciences,
The University of Bolton,
Deane Road, Bolton, BL3 5AB, UK
E-mail: A.Nabi@bolton.ac.uk

Abstract: This study investigates if the Uncanny Valley phenomenon is increased for realistic, human-like characters with an asynchrony of lip movement during speech. An experiment was conducted in which 113 participants rated, a human and a realistic, talking-head, human-like, virtual character over a range of onset asynchronies for both perceived familiarity and human-likeness. The results show that virtual characters were regarded as more uncanny (less familiar and human-like) than humans and that increasing levels of asynchrony increased perception of uncanniness. Interestingly, participants were more sensitive to the uncanny in characters when the audio stream preceded the visual stream than with asynchronous footage where the video stream preceded the audio stream. This paper considers possible psychological explanations as to why the magnitude and direction of an asynchrony of speech dictates magnitude of perceived uncanniness and the implications of this in character design.

Keywords: onset asynchrony; audio-visual speech; Uncanny Valley; virtual characters; realism; human-like; lip-sync; asynchrony of speech; lip movement; 3D environments; video games; strangeness; familiarity; audio stream; visual stream; viewer perception; digital human.

Reference to this paper should be made as follows: Tinwell, A., Grimshaw, M. and Nabi, D.A. (xxxx) 'The effect of onset asynchrony in audio-visual speech and the Uncanny Valley in virtual characters', *Int. J. Mechanisms and Robotic Systems*, Vol. X, No. Y, pp.000–000.

Biographical notes: Angela Tinwell, following her PhD, 'Viewer perception of facial expression and speech and the Uncanny Valley in human-like virtual characters', she has published studies on the Uncanny Valley in the journal *Computers in Human Behavior* and for Oxford University Press. She has presented her research on the Uncanny Valley with animators from Frame store at the London Science Museum. As part of the Digital Human League, she is working with visual effects professionals at Chaos Group aimed at overcoming the Uncanny Valley. The body of her research is summarised in her book, *The Uncanny Valley in Games and Animation* published by CRC Press in 2014

Mark Grimshaw is the Obel Professor of Music at Aalborg University, Denmark where he is the chair of the Music and Sound Knowledge Group. He has a BMus (Hons) from the University of Natal, South Africa, an MSc (Music Technology) from the University of York, UK, and a PhD on the Acoustic Ecology of the First-Person Shooter from the University of Waikato, New Zealand. Mark has published over 60 works across subjects as diverse as sound, virtuality, the Uncanny Valley, and IT systems and also writes free, open source software for virtual research environments (WIKINDX). His last two books were an anthology on computer game audio published in 2011 and the Oxford Handbook of Virtuality for Oxford University Press (2014). With co-author Tom Garner, a monograph entitled Sonic Virtuality is due in 2015 from OUP.

Deborah Abdel Nabi has a PhD in Human Brain Electrophysiology/ Pre-Attentive Visual Processing from The University of Manchester, she later evolved an interest in the topic of cyberpsychology. Her work with the University of Bolton Computer and Cyberpsychology Research Unit includes projects on the psychological factors underlying perception of the Uncanny Valley in virtual characters, impression formation and identity management in cyberspace; and e-learning, including innovative uses of emergent information technology to facilitate conversational models of teaching and learning. Her current research centres around the Bolton Affect Recognition Tri-stimulus Approach (BARTA), a new database of synthetic and human images to improve emotion recognition training in children with autistic spectrum disorders; a paper on the work has been submitted to The *Journal of Non-Verbal Communication*. She is also currently assessing the EEG correlates of uncanny characters expressing the six basic emotions.

1 Introduction

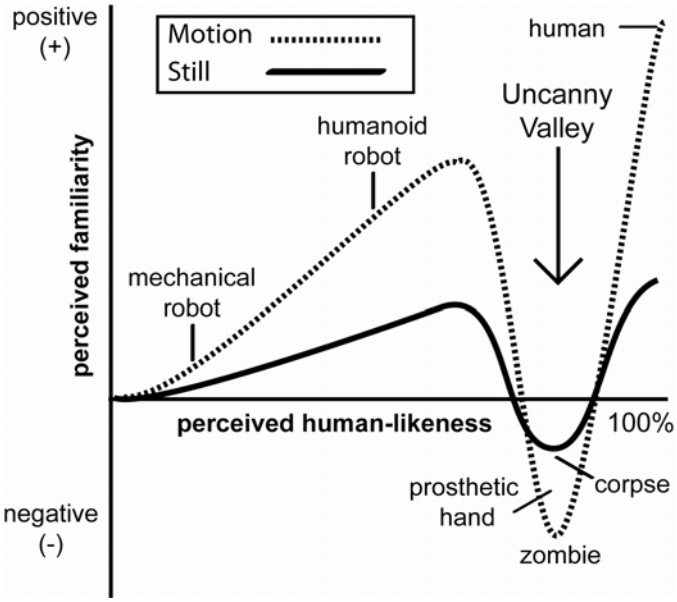
1.1 The Uncanny Valley

In 1906 the psychologist Ernst Jentsch gave consideration as to why objects may appear uncanny and frightening to the viewer. Jentsch described the uncanny as a mental state where one cannot distinguish if an object is animate or inanimate for example, when encountering objects such as human-like wax-work figures, artificial dolls and automata. Jentsch (1906) also suggested that the uncanny effect was evoked by the perceived manifestation of mental dysfunction such as witnessing an epileptic seizure where a human may have lost control of his normal bodily behaviours. He suggested that in such circumstances, the viewer may take a defensive stance against such a person as they are no longer able to predict how that person will behave, or what control they may have

over the situation. Building on this work, Freud (1919) suggested that the uncanny occurs as a realisation or revelation of the repressed, in another or one's self. More contemporary perspectives draw on ideas similar to that of Jentsch by proposing that an uncanny sensation may be elicited when an object is regarded as a threat and may cause harm (Kang, 2009). For example the feeling of anxiety or fear caused by a powerful machine is due to it being perceived as more dominant and powerful than the viewer, beyond human control. Similarly, an encounter may also be traumatic if one is confronted with a man-made object, such as a virtual character or robot that causes a disruption to one's worldview of what we understand to account for the essence of humanity (Kang, 2009).

In 1970, Mori applied the theory of the uncanny to robot design, his work culminating with the notion of 'The Uncanny Valley', which he represented as a hypothetical graph (see Figure 1). Mori had observed that as a robot's appearance became more human-like it continued to be perceived as more familiar, and that this was accompanied by a positive emotional response. This continued until the robot approached near full, authentic human-likeness. At that point, imperfections from the human norm in the robot's appearance and behaviour evoked a negative affective response from the viewer. The robot was regarded as more strange and creepy than familiar, creating a valley shaped dip in the otherwise linear relationship between perceived familiarity and human-likeness (see Figure 1). Mori placed objects such as zombies, corpses and lifelike prosthetic hands in the Uncanny Valley and predicted that the phenomenon would be exaggerated with object movement.

Figure 1 Mori's plot of perceived familiarity against human-likeness as the Uncanny Valley taken from a translation by MacDorman and Minato of Mori's 'The Uncanny Valley'



1.2 The Uncanny in realistic, human-like virtual characters

Virtual characters with a realistic, human-like appearance are commonly featured in 3D immersive environments, animation and simulations used for psychological assessment (MacDorman et al., 2010; Von Bergen, 2010). As well as for entertainment purposes (in video games and film), virtual characters are increasingly being used to present moral and ethical dilemmas in the legal, medical and recruitment professions (MacDorman et al., 2010; Von Bergen, 2010). For example, potential employees are being placed in scenarios that may include perturbed customers, prospective clients or difficult co-workers to assess their reaction to the given scenario (Von Bergen, 2010). However, with the increased drive towards realism in these virtual characters, the Uncanny Valley has also been associated with synthetic, human-like characters used in a wide variety of such applications (Brenton et al., 2005; MacDorman et al., 2009, 2010; Geller, 2008; Pollick, 2010; Tinwell, 2009; Tinwell and Grimshaw, 2009; Von Bergen, 2010). Particular concern has been raised as to how the appearance and behaviour of a virtual character may influence the decisions made by participants confronted with an ethical dilemma and the reliability and validity of psychological assessments which employ such synthetic agents (Von Bergen, 2010; MacDorman et al., 2010).

Previous studies suggest that viewers respond more positively to characters when the degree of behavioural fidelity (e.g., motor activity) matches their human-like appearance and less favourably when their actions are perceived as unnatural, with rigid or jerky movements (Bailenson et al., 2005; Ho et al., 2008; MacDorman et al., 2010; MacDorman and Ishiguro, 2006; Vinayagamoorthy et al., 2005). For example, viewers experience the uncanny more when there is a perceived lack of human-likeness in a character's speech and facial expression (Tinwell et al., 2010; Tinwell et al., 2011a, 2011b). It has also been observed that perception of lip-synchronisation error in virtual characters can increase the uncanny (Gouskos, 2006; Tinwell et al., 2010). However, empirical evidence was still required to explain how much asynchrony is necessary for uncanniness to be evoked, whether the direction of asynchrony (voice before lip movement or vice versa) interacts with magnitude to affect uncanniness experienced and, importantly, why asynchrony results in this transient negative affective state. Furthermore, the implications of this for designers had yet to be considered.

1.3 Audio-visual speech synchrony detection

Multisensory signals emanating from humans do not have to be exactly physically synchronous to achieve temporal coordination so that the sound and image are perceived as a singular temporal event (Conrey and Pisoni, 2006; Stein and Meredith, 1993). Based on such studies, a 'synchrony window' has been identified over which asynchronies are not readily detected by normal-hearing/seeing adults and desynchronised auditory and visual events are normally perceived as synchronous (Conrey and Pisoni, 2006; Dixon and Spitz, 1980; Grant and Greenberg, 2001; Grant et al., 2004; Lewkowicz, 1996). Despite differences in participant tasks, stimuli and statistical analysis techniques, data emerging from these studies has highlighted a number of consistent characteristics in the synchrony window (Conrey and Pisoni, 2006). Firstly, the duration of the window is over several hundred milliseconds long, extending to a range of ±400 ms; beyond this, asynchrony is more readily detectable. Secondly, it is asymmetrical: viewers are more sensitive to an asynchrony for audio-visual (AV) speech when the audio stream precedes

the visual stream than when the audio steam lags behind the visual stream; viewers detected an asynchrony when audio preceded video by only 50 ms, whereas an asynchrony of 220 ms was required for audio to lag behind video before the asynchrony was noticed (Grant et al., 2004).

Asynchronies generally reduce viewer enjoyment because, although when viewing *people* onscreen, an audience can interpret speech using just visemes that visually represent the mouth movement of each phoneme sound (similar to how those with hearing impediments can use visemes to lip-read and understand the spoken language when unable to hear sound) (Conrey and Pisoni, 2006; Macaluso et al., 2004; Mattys et al., 2000; Munhall and Vatikiotis-Bateson, 2004; Murray et al., 2005). Any notable asynchrony of that movement with the auditory input of speech can lead to interpretative conflict. This is possibly related to cognitive load issues or an over-reliance on what has been seen or what has been heard, both of which may lead the viewer to inaccurate interpretations of what was actually said (McGurk and MacDonald, 1976). In response to this and as a way to reduce potential stress or annoyance for the viewer caused by asynchrony, standards set by the television broadcasting industry require that the audio stream should not lag behind the video stream by more than 125 ms, nor precede the video stream by more than 45 ms (ITU-R, 1998).

Given the increasing involvement of virtual characters in computer applications relating to human work and social life, and the potential for AV asynchronies to exist in their speech, the present study comprised a comparison of the effects of AV speech asynchrony (that is, asynchronies of specific variable lengths and directions) and experience of the uncanny in response to a virtual character versus human.

1.4 Hypotheses

The experimental hypotheses were as follows:

H1A Effect of condition: regardless of asynchrony conditions, (sound before movement or vice versa) humans will be rated less uncanny (higher for both perceived familiarity and human-likeness) than virtual characters.

H2A Effect of asynchrony: in both conditions, (human and virtual character) perceived uncanniness will increase with increasing levels of asynchrony.

H2B Effect of asynchrony direction: in both conditions, stimuli where the audio stream follows the visual stream will be rated less uncanny (more familiar and human-like) than stimuli where the audio precedes video.

2 Method

2.1 Design

A 2 × 5 repeated measures multifactorial design was used in the present study. The independent variables were:

1 the type of character speaking in a short video clip:
   a human
   b virtual character

2 the magnitude of asynchrony between facial movement and vocal output.

Five AV asynchrony windows of various specified ranges were used: ±400 ms, ±200 ms, and 0 ms (the negative value represents asynchronies where the sound stream preceded the video stream). The dependent variables were:

1 ratings of familiarity

2 ratings of human-likeness as the characters were subjected to each of the five asynchrony conditions in the videos.

To reduce risk of cumulative effects on consequent uncaninness, the ten videos (human = 5 and virtual character = 5) were presented in a random order to participants.

2.2 Participants

The opportunity sample of participants consisted of 113 undergraduate, male university students. One hundred and eleven were within an age range between 18–30 years and two between 31–40 years. Students were selected from the subject areas of: video game art, video game design, multimedia and website development, special effects development for television and film, and sound engineering. It was expected that (especially male) students from these particular disciplines would have a similar level of exposure to virtual characters of a realistic, human-like appearance in video games and film.[1] No participants had visual or hearing impediments that would have impeded their ability to participate in this task.

2.3 Human and virtual character stimuli

Two characters were used in the study; 'Barney' an empathetic, realistic, male human-like, character from the video game *Half-Life 2* (Valve, 2007) was used for the group, virtual character and a male actor used for the group, human.

Video recordings of a human actor were made for Condition 1, (human). The videos were filmed within a video production studio using a Panasonic (AG-HMC/50P), portable, high definition video camera. For the recordings, the human was asked to recite the line, 'the cat sat on the mat' using a neutral expression for both face and voice. The neutral expression was utilised to avoid participants attempting to interpret facial expression of emotion in the characters (as this may bias perception on character mood and intent and, thus their reaction to the character). This also facilitated participants to concentrate on mouth movement with speech.

For the virtual character (condition 2), the line of neutral speech from the human was automatically synchronised with the virtual character using the phoneme extractor tool within the software *FacePoser* (Valve, 2008). All facial expression sliders were set to zero to create a neutral expression for the virtual character. Headshots against a dark background were used as settings for all ten videos. Adobe Premiere CS3 was then used to adjust and process the audio and video for each of the ten videos to the required asynchronies (in this case ±400, ±200, and 0 ms). The videos were: Standard PAL video (4:3 interlaced), 48 kHz (16 bit) audio, 25 frames/s, and all of a duration of four seconds.

2.4 Apparatus and delivery

Participants were presented with the ten video clips via an interactive web-based questionnaire. Participants used individual computer stations in a computer lab to observe the video footage using 'Dell E207WFPc 20.1 inch Widescreen' monitors, with the auditory stimuli presented via 'Speed Link, Ares² Stereo PC Headphone' sets.

2.5 Procedure

Having watched each video clip, participants rated the character on a nine-point scale in terms of how human-like they perceived it to be from 1 (*very human-like*) to 9 (*non-human-like*) and how strange or familiar they judged it to be (1 conveying they thought it to be *very strange* and 9 indicating that they thought it *very familiar*). Nine point scales have been used previously in experiments investigating the uncanny with virtual characters for the dependent variables human-likeness and familiarity (see e.g., MacDorman 2006; Tinwell, 2009; Tinwell and Grimshaw, 2009; Tinwell et al., 2010), and were used in the present study so that results could be compared with those previous experiments. Participants were also required to rate their level of experience both playing video games and using 3D modelling software from the options, (1) none, (2) basic or (3) advanced.

3 Results

3.1 Effect of stimulus condition

The first hypothesis (H1A) proposed that, regardless of asynchrony size and direction, human videos would be rated higher for both familiarity and human-likeness than videos of the virtual character. Table 1 (and Figure 2 and Figure 3) shows the mean ratings for familiarity and human-likeness associated with each asynchrony time frame for each condition and indicates support for the hypothesis. Consistently, videos of humans were rated as more familiar and human-like than those of virtual characters.

Table 1 Mean ratings (and SD) for familiarity and human-likeness in the two conditions across all levels of asynchrony (N = 113).

Condition	*Familiarity*				*Human-likeness*			
	Human		*VC*		*Human*		*VC*	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Asynchrony(ms)								
–400	6.19	2.07	3.97	1.85	7.81	1.77	3.90	1.90
–200	6.43	2.17	4.18	1.79	7.89	1.72	4.13	1.71
0	7.00	1.89	4.59	1.68	8.35	1.16	4.55	1.75
+200	6.76	2.13	4.47	1.78	8.15	1.54	4.35	1.71
+400	6.34	2.03	4.06	1.80	7.83	1.73	3.96	1.72

Notes: Judgments were made on a nine-point scale from *very strange* (1) to *very familiar* (9) and *non-human-like* (1) to *very human-like* (9).

Figure 2 Line graph showing mean ratings of familiarity in the two conditions across all levels of asynchrony

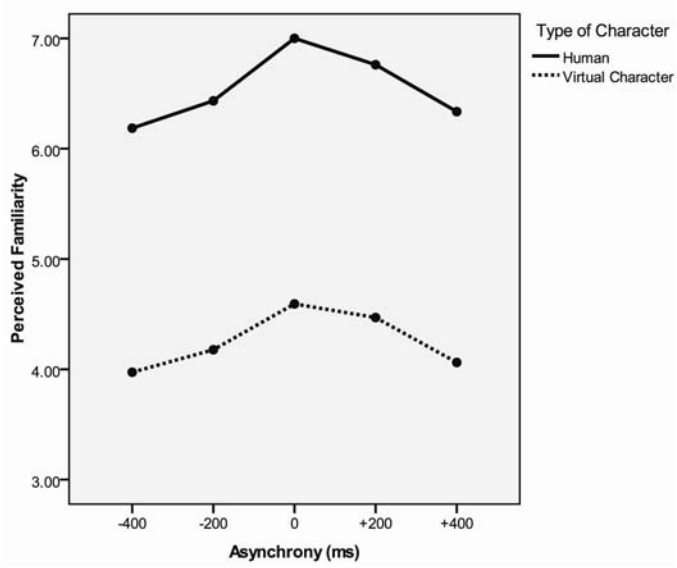


Figure 2 demonstrates that, regardless of level of asynchrony, Condition 1 (human) attracted the highest ratings of familiarity for all asynchronies. In both conditions synchronised footage (0 ms) was rated as the most familiar with increasing levels of asynchrony reducing scores for familiarity, rated in the following descending order: 0 ms, +200 ms, –200 ms, +400 ms, and –400 ms. The more asynchronous the AV vocalisations, the stranger the participants felt the characters to be.

Figure 3 Line graph showing mean ratings for human-likeness in the two conditions across all levels of asynchrony

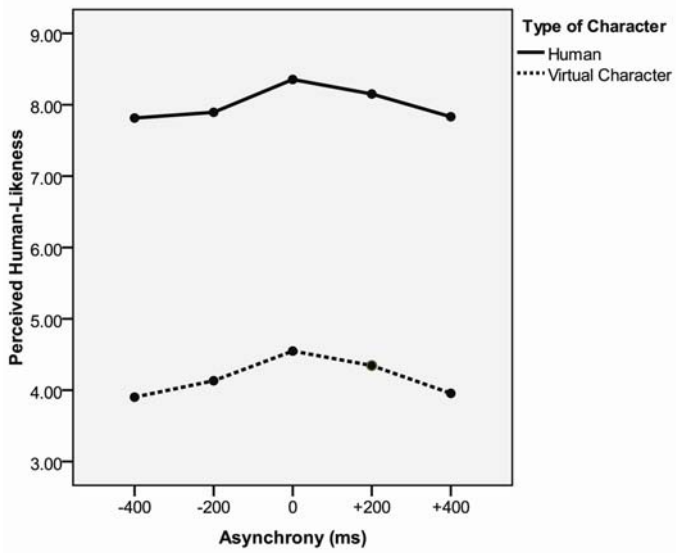


Figure 3 demonstrates that, once again, for all asynchronies, videos of the human (Condition 1) attracted the highest ratings of human-likeness, followed by the virtual character (Condition 2). As found with familiarity, synchronised footage (0 ms) received highest ratings for human-likeness with scores for human-likeness decreasing as asynchrony increased in both conditions. Scores for human-likeness were rated in the following descending order: 0 ms, +200 ms, –200 ms, +400 ms, and –400 ms.

To assess the significance of these results, two 2 x 5 repeated measures analysis of variance (ANOVA) and planned comparisons were applied to the data. The ANOVA revealed a significant main effect of character-type (human vs. virtual character) for familiarity ratings $F(1, 224) = 107.829$, $p < .001$, and for human-likeness, $F(1, 224) = 392.766$, $p < .001$, thus supporting Hypothesis H1A. There was no significant interaction effect between character type and differing levels of asynchrony for perceived familiarity, $F(3.751, 840.124) = .246$, $p = .902$, and human-likeness, $F(3.738, 837.347) = .233$, $p = .910$; indicating the effect of asynchrony on perceived uncanniness was the same for both humans and virtual characters.

3.2 Effect of asynchrony type

The second hypothesis (H2A) proposed that for both humans and virtual characters, uncanniness, indexed by decreases in ratings of familiarity and human-likeness, would increase the greater the level of asynchrony. A significant main effect of asynchrony was identified for familiarity ratings $F(3.751, 840.124) = 16.589$, $p < .001$, and for human-likeness, $F(3.738, 837.347) = 16.047$, $p < .001$ implying that perceived uncanniness was influenced by the magnitude of asynchrony.

Planned comparisons on the familiarity data showed significant differences between ratings of synchronised footage (0 ms) versus: +200 ms, $F(1, 224) = 3.600$, $p < .05$; 0 ms versus –200 ms, $F(1, 224) = 26.827$, $p < .001$; 0 ms versus +400 ms, $F(1, 224) = 34.687$, $p < .001$; and 0 ms versus –400 ms, $F(1, 224) = 43.517$, $p < .001$. Significant differences were also identified between –200 ms versus –400 ms $F(1, 224) = 4.592$, $p < .05$; and +200 ms versus +400 ms, $F(1, 224) = 12.900$, $p < .001$, offering support for Hypothesis H2A.

Similar results were observed with regards to human-likeness. Planned comparisons demonstrated significant differences between ratings of synchronised footage (0 ms) versus: +200 ms, $F(1, 224) = 6.666$, $p < .05$; 0 ms versus –200 ms, $F(1, 224) = 23.721$, $p < .001$; 0 ms versus +400 ms, $F(1, 24) = 32.413$, $p < .001$; and 0 ms versus –400 ms, $F(1, 224) = 43.206$, $p < .001$. Significant differences were also identified between –200 ms versus –400 ms $F(1, 224) = 3.018$, $p < .05$, and +200 ms versus +400 ms, $F(1, 224) = 14.242$, $p < .001$, again lending support to Hypothesis H2A. Overall, characters were regarded as more uncanny with increasing levels of asynchrony.

3.2.1 Effect of direction of asynchrony

Hypothesis H2B proposed that perceived uncanniness (evidenced by lower ratings of familiarity and human-likeness) would increase for stimuli where the audio stream played before the video. Planned comparisons showed for both human and virtual characters, there were significant differences between familiarity ratings for –200 ms versus +200 ms, $F(1, 224) = 8.145$, $p < .05$; but not for –400 ms versus +400 ms, $F(1, 224) = 1.764$, $p = .093$. For human-likeness, significant differences were found

between –200 ms versus +200 ms, $F(1, 224) = 6.020$, $p < .05$, but again, not for –400 ms versus +400 ms, $F(1, 224) = 1.76$, $p = .34$.

A significant difference was found between –400 ms and +200 ms for both familiarity $F(1, 224) = 22.894$, $p < .001$, and human-likeness $F(1, 224) = 19.329$, $p < .001$, but no significant difference was observed between –200 ms and +400 ms for both familiarity $F(1, 224) = 1.130$, $p = .14$, and human-likeness, $F(1, 224) = 2.338$, $p = .06$, offering partial support for Hypothesis H2B.

4 Discussion

This study investigated the effect of an asynchrony between sound and lip movement on perceived uncanniness in realistic, human-like, virtual characters. As was hypothesised, virtual characters were rated the least familiar and human-like (i.e., most uncanny) when compared to humans, and the magnitude of this perceived uncanniness increased with magnitude of asynchrony. Significant differences in perceived familiarity and human-likeness were found to exist between synchronised footage (0 ms) and asynchronies of ±200 ms and ±400 ms for both the human and virtual character. Significant differences were also found between –200 ms versus –400 ms, and +200 ms versus +400 ms. Interestingly, the size of this increased uncanniness varied depending on the direction of asynchrony. Stimuli with an asynchrony of –200 ms and –400 ms, where the audio stream preceded the visual stream, were rated as significantly more uncanny than some timing offsets where the audio stream instead followed the visual stream.

As commonplace in psychological studies, only male undergraduate students were used as participants in the present study; a fully stratified sample of the population was not employed. Even though these significant findings in our results infer that these effects are representative of the population and that there is no empirical evidence to suggest that factors such as gender, age, or level of education, would have an effect on performance in this particular task it might, nevertheless, be prudent in future studies to include a wider demographic. For example, to include females, non-gamers, those with different educational backgrounds, and those of different age groups as participants in order to test that generalisation. We now discuss how our results may be explained by considering the typically proposed psychological substrates of the uncanny, findings on the characteristics of the synchrony window in humans, and the current modelling limitations of automated lip-syncing tools.

The uncanny sensation evoked by virtual characters with an AV asynchrony may have similar origins to the uncomfortable feeling commonly experienced when encountering automata. But where does this response originate? The answer to this may lie in the typical human response to android and robotic characters. Writers of science fiction and horror frequently depict tales of violent conflicts between man and machine, which exploits the underlying sense of unease humans can experience when confronted with a self-operating (seemingly sentient) machine, such as an android or robot (Kang, 2009). Whilst we may feel awe at the strength, efficiency and productivity of a powerful man-made machine, for example a locomotive train, objects such as android robots and high fidelity, human-like characters may challenge one's schema of reality. Kang (2009, p.49) stated that a worldview is established based on one's experience and the classification of objects into particular categories 'such as day/night, human/animal, living/dead, man/woman, adult/child, safe/dangerous etc.' When confronted with an

object that may not fit easily within one of our predefined categories, (i.e., a nearly but not quite human character) such an anomaly can cause fear and concern that one's predisposed schema of the world has been questioned. As such, any man-made objects that mimic life as we know it (such as realistic, human-like, virtual characters) are generally perceived as less trustworthy, possibly motivated by a potentially malevolent force that may exist behind that object. With an asynchrony of speech, one may equate the non-sensical 'flapping jaw' presented by the synthetic character to that often demonstrated by a mechanical robot, triggering the same catalogue of doubts, suspicions, fears as a robot and serving to increase a sense of the unfamiliarity and unpredictability in terms of behaviour; we fear that we may no longer be in control of the object, due to the evident malfunction in that character's actions.

In relation to the results of the current study, those characters with synchronised speech were regarded as significantly less uncanny than those in experimental conditions with an asynchrony of speech. Based on these findings, one might speculate that whilst the viewer may have been somewhat in awe of the virtual character, once they detected an asynchrony of speech (impairing the ability to interpret what was being said), the character was then regarded as less familiar and predictable, evoking an intensified feeling of mistrust and discomfort.

This sense of unpredictability has, of course, also previously been associated with the uncanny with regard to human behaviour in those who present sudden, unexpected transitory states (such as the convulsions of epileptic seizures). In these conditions, where they may not appear to be in control of their movement, they are often regarded as uncanny (Jentsch, 1906). A similar effect may be evoked when there is an unexpected asynchrony of speech in a character with a realistic, human-like appearance. This may raise the possibility of mental dysfunction in that character, thus a potential threat to the viewer.

The results of the current study imply that due to a human's increased sensitivity to asynchrony where sound precedes the visual stream (Conrey and Pisoni, 2006; Grant et al., 2004), ratings for perceived familiarity and human-likeness were significantly lower for stimuli where audio preceded the video than vice versa. However, this effect does not extend to larger asynchronies. When –400 ms was compared to +400 ms the comparatively more deleterious effect (in terms of uncanniness) of sound preceding lip movement was reduced to non-significant levels. It appears that when AV footage is asynchronous to the extent of 400 ms, the direction of asynchrony is irrelevant with regards to perceived uncanniness.

The question then arises as to why negative asynchronies elicit greater perceived uncanniness and especially so in virtual characters. The answer possibly lies in the human ability to capitalise on other, *visual*, cues to process speech. When watching humans speak, if we are unable to hear any sound, viewers can use the technique of lip-reading in an attempt to interpret what is being said (Macaluso et al., 2004; Mattys et al., 2000). However, for speech generated for virtual characters in real time or via text input methods, mouth articulation is restricted to a preset class of mouth shapes (visemes) to represent each phoneme sound. The set of visemes available for virtual characters is more limited compared to that used by humans. Hence, the viewer may have a greater difficulty in lip-reading virtual characters post-sound, further reducing speech processing efficacy and accuracy. Viewers may not only misinterpret a particular sound, but may invent a new sound to compensate for the unintelligible dialogue, [as evident with the

McGurk effect, (McGurk and MacDonald 1976)], hence further exaggerating the uncanny in virtual characters.

So how does lead to an increased sense of uncanniness? The possibility exists that under such conditions, viewers may be less able to cognitively process simultaneity for the auditory and visual inputs and the viewer is reminded that the character is simply a man-made object (not human, thus unpredictable and, possibly, to be feared). The information normally gained from observation of mouth movement is occluded to the extent that comprehension of what the character is attempting to communicate is impaired; the character appears non-sensical and unfathomable. A perception that the voice is disembodied from the character evokes the conceptual conflict that an unnatural or mechanical process is behind the actions of that character, thus provoking the uncanny. Whilst this effect may work to the advantage of characters designed to intentionally frighten the viewer, such as zombies within the horror game genre, a viewer risks being repulsed by or rejecting an empathetic character that displays such odd and unnatural behaviour. With the intention to introduce virtual characters to replace the role of humans in simulations for educational purposes or psychological assessments, it is important that a participant's performance is not impeded by any factors that evoke an experience of the uncanny as this may reduce the believability and trustworthiness of that character (Von Bergen, 2010; MacDorman et al., 2010). A tighter window of acceptable asynchrony in virtual characters, smaller than that already defined for the television broadcasting industry, and ensuring that if asynchrony exists at all, that it does not take the form of speech preceding lip movement, may help to reduce the risk that lip-sync error has a detrimental effect on the viewer who interacts with virtual characters.

References

Bailenson, J.N., Swinth, K.R., Hoyt, C.L., Persky, S., Dimov, A. and Blascovich, J. (2005) 'The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments', *Presence: Teleoperators and Virtual Environments*, Vol. 14, No. 5, pp.379–393.

Brenton, H., Gillies M., Ballin D. and Chatting, D. (2005) 'The Uncanny Valley: does it exist?', paper presented at the *HCI 2005, Animated Characters Interaction Workshop*, Napier University: Edinburgh.

Conrey, E. and Pisoni, D. (2006) 'Auditory-visual speech perception and synchrony detection for speech and nonspeech signals', *The Journal of the Acoustical Society of America*, Vol. 119, No. 6, pp.4065–4073.

Dixon, N.F. and Spitz, L. (1980) 'The detection of auditory visual desynchrony', *Perception*, Vol. 9, No. 6, pp.719–721.

Freud, S. (1919) 'The uncanny', in Strachey, J. and Freud, A. (Eds.): *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 17, pp.217–256, Hogarth Press, London.

Geller, T. (2008) 'Overcoming the Uncanny Valley', *IEEE Computer Graphics and Applications*, Vol. 28, No. 4, pp.11–17.

Gouskos, C. (2006) *The Depths of the Uncanny Valley* [online] http://uk.gamespot.com/features/6153667/index.html (accessed 16 April 2011).

Grant, K.W. and Greenberg, S. (2001) 'Speech intelligibility derived from asynchronous processing of auditory-visual information', paper presented at the *ISCA International Conference on Auditory-Visual Speech Processing*, Scheelsminde, Denmark.

Grant, W., Wassenhove, V. and Poeppel, D. (2004) 'Detection of auditory (cross-spectral) and auditory–visual (cross-modal) synchrony', *Speech Communication*, Vol. 44, Nos. 1–4, pp.43–53.

Ho, C-C., MacDorman, K. and Pramono, Z.A.D. (2008) 'Human emotion and the Uncanny Valley. A GLM, MDS, and ISOMAP analysis of robot video ratings', *Proceedings of the Third ACM/IEEE International Conference on Human-Robot Interaction*, 11–14 March, pp.169–176, Amsterdam.

ITU Recommendations for Broadcast Television Production (1998) *BT. 1359: Relative Timing of Sound and Vision for Broadcasting (Question ITU-R 35/11)* [online] http://www.itu.int/rec/R-REC-BT.1359/en (accessed 18 April 2011).

Jentsch, E. (1906) 'On the psychology of the uncanny', (R. Sellars, Trans.), *Angelaki*, Vol. 2, No. 1, pp.7–16.

Kang, M. (2009) 'The ambivalent power of the robot', *Antennae, The Journal of Nature in Visual Culture*, Vol. 1, No. 9, pp.47–58.

Lewkowicz, D.J. (1996) 'Perception of auditory-visual temporal synchrony in human infants', *Journal of Experimental Psychology*. *Human Perception and Performance*, Vol. 22, No. 5, pp.1094–1106.

Macaluso, E., George, N., Dolan, R., Spence, C. and Driver, J. (2004) 'Spatial and temporal factors during processing of audiovisual speech: a PET study', *NeuroImage*, Vol. 21, No. 2, pp.725–732.

MacDorman, K. (2006) 'Subjective ratings of robot video clips for human-likeness, familiarity, and eeriness: an exploration of the Uncanny Valley', *Proceedings of the ICCS/CogSci-2006 Long Symposium: toward Social Mechanisms of Android Science*, pp.26–29, Vancouver, Canada.

MacDorman, K.F. and Ishiguro, H. (2006) 'The uncanny advantage of using androids in cognitive and social science research', *Interaction Studies*, Vol. 7, No. 3, pp.297–337.

MacDorman, K.F., Coram, J.A., Ho, C-C. and Patel, H. (2010) 'Gender differences in the impact of presentational factors in human character animation on decisions in ethical dilemmas', *Presence: Teleoperators and Virtual Environments*, Vol. 19, No. 3, pp.213–229.

MacDorman, K.F., Green, R.D., Ho, C-C. and Koch, C. (2009) 'Too real for comfort: uncanny responses to computer generated faces', *Computers in Human Behavior*, Vol. 25, No. 3, pp.695–710.

Mattys, S., Bernstein, L.E., Edward, T. and Auer, J. (2000) 'When lipreading words is as accurate as listening', paper presented at *the139th ASA Meeting*, Atlanta, Georgia.

McGurk, H. and MacDonald, J. (1976) 'Hearing lips and seeing voices', *Nature*, Vol. 264, No. 5588, pp.746–748.

Mori, M. (1970) 'The Uncanny Valley', *Energy*, Vol. 7,No. 4, pp.33–35 (MacDorman, K.F. and Minato, T., Trans (2005).

Munhall, K. and Vatikiotis-Bateson, E. (2004) 'Spatial and temporal constraints on audiovisual speech perception', in Calvert, G., Spence, C. and Stein, B.E. (Eds.): *The Handbook of Multisensory Processes*, pp.177–188, MIT Press, Cambridge, MA.

Murray, M.M., Molholm, S., Michel, C.M., Heslenteld, D.J., Ritter, W. and Javitt, D.C., et al. (2005) 'Grabbing your ear: rapid auditorysomatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment', *Cerebral Cortex*, Vol. 15, No. 7, pp.963–974.

Pollick, F.E. (2010) 'In search of the Uncanny Valley', in Daras, P. and Mayora, O. (Eds.): *UCMedia 2009*, LNICST, Vol. 40, pp.69–78.

Stein, B. and Meredith, M.A. (1993) *The Merging of the Senses*, MIT Press, Cambridge, MA.

Tinwell, A. (2009) 'The uncanny as usability obstacle', in Ozok, A.A. and Zaphiris, P. (Eds.): *Proceedings of the HCI International 2009: Online Communities and Social Computing Workshop*, pp.622–631.

Tinwell, A. and Grimshaw, M. (2009) 'Bridging the uncanny: an impossible traverse?', *Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era*, September/October, ACM, Tampere, Finland.

Tinwell, A., Grimshaw, M. and Williams, A. (2010) 'Uncanny behaviour in survival horror games', *Journal of Gaming and Virtual Worlds*, Vol. 2, No. 1, pp.3–25.

Tinwell, A., Grimshaw, M. and Williams, A. (2011a) 'Uncanny speech', in Grimshaw, M. (Ed.): *Game Sound Technology and Player Interaction: Concepts and Developments*, pp.213–234, IGI Global, Hershey, PA.

Tinwell, A., Grimshaw, M., Abdel-Nabi, D. and Williams, A. (2011b) 'Facial expression of emotion and perception of the Uncanny Valley in virtual characters', *Journal of Computers in Human Behavior*, Vol. 27, No. 2, pp.741–49.

Valve (2007) *Half-Life 2* [Video Game], Valve Corporation, Washington.

Valve (2008) *Faceposer* [Facial Animation Software as part of Source SDK Video Game Engine], Valve Corporation, Washington.

Vinayagamoorthy, V., Steed, A. and Slater, M. (2005) 'Building characters: Lessons drawn from virtual environments', *Proceedings of the CogSci-2005 Workshop, 'Toward Social Mechanisms of Android Science'*, pp.119–126.

Von Bergen, J.M. (2010) *Queasy about Avatars and Hiring Employees* [online] http://www.philly.com/philly/blogs/jobs/Queasy_about_avatars_and_hiring_employees_.html (accessed 10 April 2011).

Notes

1 88.50% of participants had an advanced level of playing video games, with 11.50% a basic level of experience. For experience of using 3D modelling software: 47.79% had a basic level, 46.08% no experience, and 6.20% an advanced level.