# The Uncanny Valley: A Focus on Misaligned Cues

Lianne F.S. Meah and Roger K. Moore

Department of Computer Science, University of Sheffield,
United Kingdom
{lfsmeah1,r.k.moore}@sheffield.ac.uk

**Abstract.** Increasingly, humanoid robots and androids are easing into society for a wide variety of different uses. Previous research has shown that careful design of such robots is crucial as subtle flaws in their appearance, vocals and movement can give rise to feelings of unease in those interacting with them. Recently, the Bayesian model for the uncanny has suggested that conflicting or misaligned cues at category boundaries may be the main attributing factor of this phenomenon. The results from this study imply that this is indeed the case and serve as empirical evidence for the Bayesian theory.

**Keywords:** Uncanny valley, social robotics, human-robot interaction.

## 1 Introduction

Although the phenomenon of the uncanny valley was first proposed by Mashiro Mori [7], the concept of the uncanny can be traced back as far as 1906. In his essay, psychiatrist Ernst Jentsch described the uncanny as *intellectual uncertainty* [4], and several years later it was revisited by Sigmund Freud, who described it as something which seems familiar and yet foreign simultaneously [2]. In his report, Mashiro theorized that an object that is more humanlike in appearance will seem more familiar with an observer.

For example, a robotic arm used in industry may be seen as less familiar than a humanoid robot, as it is visually far less humanoid. This is depicted in Fig. 1, where industrial robots are placed near the origin of the graph with low familiarity and low human likeness. Humanoid robots are placed just before the peak in familiarity. It might then be expected that robots that look *especially* human will continue the trend in the graph, however, they instead fall into the uncanny where their familiarity ratings are akin to those of zombies or corpses. With this drop in familiarity comes an increase in eeriness, which manifests as a feeling of unease or repulsion in observers.

Before proceeding, it is important to clarify what is meant by the terms *robot*, *humanoid robot* and *android* in this study. The term robot shall refer to a programmable machine, or automaton, that bears little to no resemblance of a human being. A humanoid robot, then, is a robot which is humanlike in some sense (it may possess a humanoid body or face) but can visibly be distinguished

M. Beetz et al. (Eds.): ICSR 2014, LNAI 8755, pp. 256–265, 2014.

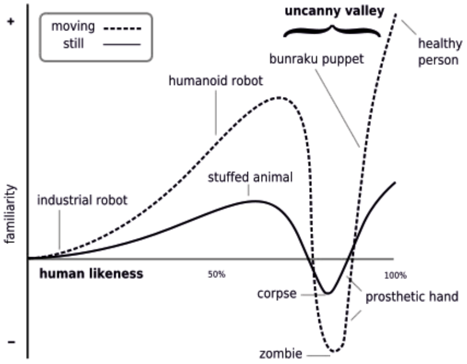

**Fig. 1.** The uncanny valley diagram. [7]

from a human being, in other words, it is easy to classify as a robot. The term android refers to a humanoid robot with an added layer of complexity; androids are designed to pass as human beings and will own more intricate assets such as artificial skin, hair and so on. They are visually almost human, to the point where they fall into the uncanny.

As the original illustration of the uncanny valley depicts familiarity against increasing visual human likeness, many studies have been carried out with a focus on the visual domain. However, the uncanny valley has also been shown to exist in the audio continuum [3]. As such, it can be suggested that a person's response to a stimulus can be altered by changing either the visuals, audio or both. Indeed, the link between a character's voice and face has already been investigated [8], and a mismatch in these features can induce the uncanny valley effect. For example, pairing a human voice with a robotic, mechanical face produces feelings of unease in observers [8], suggesting that a person or robot's voice and face play a major role in communication. In particular, the eyes are thought to provide a multitude of cues. Abnormal alterations of the eyes alone is enough to produce the uncanny effect [5], [6].

More recently, a Bayesian explanation of the uncanny has been suggested [9]. Based on the categorical perception model of Feldman, Griffiths and Morgan [1], the model of the uncanny proposes that stimuli containing conflicting cues cause 'differential perceptual distortion' which in turn induces perceptual tension. It is suggested that this tension manifests as feelings of eeriness. The key to perceptual distortion is categorization; the uncanny is predicted to manifest from observing androids as they contain multiple conflicting perceptual cues, some of which cause a greater amount of uncertainty regarding their category membership, thus giving rise to perceptual tension (also see [11]). Androids cannot easily be classified into the human or robot category; they lie within or near a category boundary (see [9] for example illustrations), and we find that the "*inability to categorize will then lead to a state of dissonance*" [10].

To date, the Bayesian model of the uncanny has not been documented in an empirical study. We investigated to what extent contradictory or misaligned cues contribute towards feelings of eeriness in both a unimodal and multimodal setting. The model suggests that an increase in uncertainty between cues results in an increase in perceptual tension, thus it follows that a decrease in uncertainty will reduce perceptual tension. In the unimodal setting, we examined the role of an android's eyes and how the removal of conflicting cues from them might alter an observer's response. In the multimodal setting, an experiment performed originally by Mitchell et al [8] was replicated and extended to include a wider range of visual and auditory stimuli, with a particular focus on the degree of conflicting cues they might contain.

## 2 Materials and Methods

We performed two experiments, one with a focus on unimodal cues and the other focusing on multimodal cues. In both experiments, volunteers were asked to watch several videos and then provide feedback both qualitatively and quantitatively by filling out a questionnaire. Upon watching a video, a participant was required to give ratings for four different attributes of the subject in the video: humanness, eeriness, familiarity and appeal (it should be noted that only the eeriness attribute will be discussed in the results). The ratings were on a Likert scale between 1 and 5. A listening booth was provided by the University Speech and Hearing Lab, where participants could sit at a desk within a quiet environment with the videos being displayed on a computer monitor. Footage of three androids and one humanoid robot were obtained for use in both experiments: the *Geminoid DK*, *'Jules'*, the *Repliee Q* and the *iCub*, respectively. In addition, for the second experiment, a video of a human male was recorded using an HD camcorder. See Fig. 2 for all the visual stimuli.

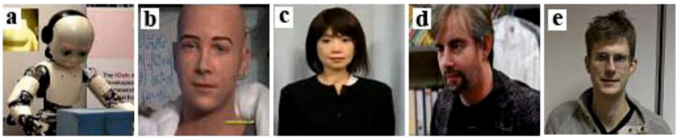

**Fig. 2.** All visual stimuli used in the experiments, composed of *a*: one humanoid robot, *b-d*: three androids and *e*: one human. Subjects *a*, *b*, *c* and *d* were used in the first experiment, subjects *a*, *d* and *e* were used in the second. Additionally, audio was recorded from *e* for the second experiment. Images are a single frame taken from each video.

### 2.1 Experiment One

The primary goal of the first experiment was to investigate the impact of unimodal cues in the visual domain, as such the videos were not combined with any auditory cues. We investigated the impact of an android's eyes and hypothesized that the removal of misaligned cues from them would significantly decrease the eeriness felt in an observer. We also investigated the impact of a humanoid robot's eyes and predicted that, since robots typically do not fall into the uncanny (although this is dependent on design), removal of cues from the eyes would not have the same effect.

To carry out this study, three videos of different androids and one video of a humanoid robot were shown to participants. The original videos were edited only to control the length of time that each video ran for and also to mute the audio. In addition, four other videos were created where the cues from the eyes were blocked by a rectangular black box, which was placed just above the lower lid and beneath the eyebrows, thus covering the eyes. In the final video reel, the videos were paired such that a 'covered' video followed after its 'uncovered' counterpart and vice versa. To summarize, there were eight videos in total, four pairs of 'covered' and 'uncovered' clips.

### 2.2 Experiment Two

In order to confirm that a mismatch in voice and face induces the uncanny effect, in the second experiment the focus changed from unimodal to multimodal cues and audio was combined with the visual stimuli. For this experiment, we extended a recent study on the uncanny [8]. In the original study, Mitchell et al combined the face and voice of a human with the face of a robot and a synthetic voice in order to create 'matched' and 'mismatched' stimuli. They theorized that matched stimuli (aligned cues) would be significantly less eerie than mismatched stimuli (misaligned cues). For example, it was shown that participants are comfortable in viewing a video of a human face combined with a human voice, but not so comfortable if the human face was paired with a synthetic voice. We extended this experiment to include the visuals of an android and dual-pitched audio, both of which should be regarded as particularly eerie by observers as they are both almost human in their respective domains, thus near category boundaries.

To create dual-pitch voices we recorded audio from a human male (aged in his late thirties) and ran it through a dual-pitch voice changer, developed in Pure Data. This particular method of voice changing gives the impression that two voices are being spoken at once, one of which differs in pitch, and serves as a way of constructing a robotic-sounding voice without disruptions in sentence flow, as is often heard in other text-to-speech voices.

We theorized that the android, despite being visually very close to human, would still be judged as a robot, and that the dual-pitch voices, although derived from and close to the original human audio, would still be judged as robotic. As such, for the android visuals, the synthetic and dual-pitch voices were hypothesized to be the matching audio, with the human voice acting as the mismatching audio. The same was hypothesized for the robot. For the human visuals, we predicted that the human voice would serve as the matching audio, with the synthetic and dual-pitch voices acting as mismatching audio. Additionally, for the human visuals, we hypothesized that a dual-pitch mismatch would be significantly eerier than the synthetic mismatch, as the dual-pitch audio clips are closer to the original audio recorded from the human, and would thus cause a greater amount of perceptual tension.

In the experiment, footage of one android and one humanoid robot were used. Additionally, both video and audio of a human male were recorded speaking the neutral phrase *'a goal is a dream with a deadline'*. The original human audio was run through the voice changer and shifted by three different frequencies, 50Hz, 150Hz and 250Hz, in order to create three different dual-pitch stimuli. Furthermore, a text-to-speech (TTS) synthesizer was used to create a synthetic voice which spoke the same phrase. Upon completion of gathering the required audio, the voices were then overlaid onto the videos. Full lip syncing was not possible, except for the human visuals as they were recorded at the same time as the audio. As there were three different visuals (humanoid robot, android, human) and five different voice conditions (human voice, dual-pitch 50Hz, dual-pitch 150Hz, dual-pitch 250Hz and synthetic), there were fifteen clips overall for the second experiment.

Table 1 gives a summary of the stimuli used for second experiment. In the interest of clarity, hereafter the stimuli will be referred to using *visual-audio* notation, where visual refers to one of the visual categories (human, android, robot) and audio refers to one of the auditory categories (human, 50Hz dual-pitch, 150Hz dual-pitch, 250Hz dual-pitch, synthetic). For example, *Robot-Synthetic* refers to the robot face combined with the synthetic, text-to-speech voice, *Android*-50*Hz* refers to the android face combined with the dual-pitch voice that has been shifted by 50Hz, and so on.

Upon completion of all 23 videos, they were all combined into one single reel which was presented to the participant. Before the beginning of each experiment, a black screen would be presented with text in white, reading as 'Experiment One' or 'Experiment Two'. The respective stimuli would then follow, in a randomized order unknown to the participant. Each participant thus took part in both experiments and completed experiment one first.

**Table 1.** Summary of stimuli for the second experiment

| | human | android | robot |
|---|---|---|---|
| human voice | match | mismatch | mismatch |
| 50Hz shift | mismatch | match | match |
| 150Hz shift | mismatch | match | match |
| 250Hz shift | mismatch | match | match |
| TTS (synthetic) | mismatch | match | match |

## 3 Results

The study was conducted over three weeks in March. For both experiments there were 40 volunteers of varying disciplines within the University of Sheffield, 14 female and 26 male, with a mean age of 25.8. Data analysis was conducted using the matched-pairs t-test.

### 3.1 Experiment One: Unimodal Cues

The full result set for the eeriness ratings is shown in Fig. 3. The pairs of stimuli have been plotted in terms of their humanness and eeriness; the far left denoting lower humanness ratings. As expected, the humanoid robot was rated lowest in terms of humanness whilst the Geminoid DK android was given the highest ratings.

We found that blocking the eyes of the Geminoid DK, thereby decreasing perceptual tension, did indeed have a positive effect on an observer and significantly decreased its average eeriness rating. We theorize that it is because the Geminoid DK is the most humanlike of the androids that the blocking had the most impact; in the middle range of the humanness scale, the ratings for eeriness were not significantly impacted by blocking the eyes. However, on the far left of the humanness scale, covering the eyes of the humanoid robot resulted in an enhanced *negative* response from viewers and significantly increased its eeriness rating. It could be suggested, then, that the less human a robot visually appears to be, the less the covering of the eyes will impact an observer's responses in a positive way.

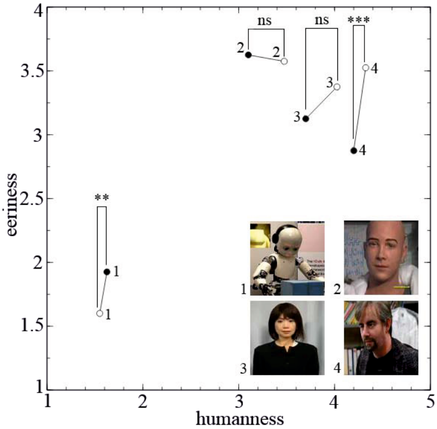

**Fig. 3.** Mean eeriness ratings from the first experiment. *Filled circles* denote videos where eyes were covered. *Open circles* denote videos were eyes were shown. Average eeriness ratings (from left to right): 1.600, 1.925, 3.625, 3.575, 3.125, 3.375, 2.875 and 3.575. ** denotes $p \leqslant 0.01$, *** denotes $p \leqslant 0.001$.

### 3.2 Experiment One Discussion

These results agree with the Bayesian model; the eyes of an android contain conflicting cues which give rise to uncertainty and perceptual tension. Covering the eyes, thereby removing the conflicting cues, decreases perceptual tension and thus decreases the eeriness felt in viewers. The impact of cue removal depends on where the subject sits on the humanness scale, or rather, how close to the category boundary it is. The model predicts that removal of cues from an object rated lower in humanness (a humanoid robot) should instead increase eeriness, which is indeed what has been found here.

### 3.3 Experiment Two: Multimodal Cues

There were two aims of this experiment. The first was to test whether a dual-pitch voice, combined with mismatching visual stimuli would be regarded as significantly eerier than a synthetic voice mismatch. The second was to repeat and extend a recent study on the uncanny, with the additional android footage (the Geminoid DK) and the dual-pitch voices to bring more dimensions to the experiment and test the Bayesian model in a multimodal setting.

The average eeriness ratings for this experiment are given in Fig. 4a and Fig. 4b. The lowest rating of eeriness was given to the Human-Human stimulus and the highest was given to the Android-50Hz stimulus. Generally, stimuli using the android face were given the highest eeriness ratings in each voice condition. Additionally, stimuli using the 50Hz dual-pitch voice were also given the highest eeriness ratings in each visual condition.

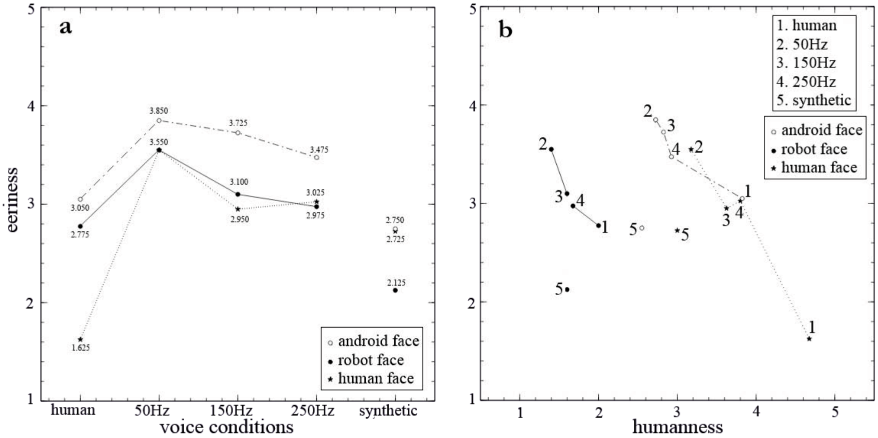

**Fig. 4.** Average eeriness ratings plotted against **a**: voice conditions, and **b**: average humanness ratings

### 3.4 Matched and Mismatched Comparisons

For the human visuals, all mismatched stimuli (Human-50Hz, Human-150Hz, Human-250Hz, Human-Synthetic) were significantly eerier than the matched stimulus (Human-Human). Since the Human-Human stimulus is a 'matched' combination and it thus follows that it received the lowest ratings of eeriness. Furthermore, we can conclude that a dual-pitch voice is indeed not judged as human.

For the robot visuals, the synthetic and dual-pitch voices were theorized to be the matching auditory stimuli. However, this was not the case. The Robot-Human combination (mismatch) was significantly eerier than the Robot-Synthetic (match) combination, which was expected. However, the dual-pitch voices were also given significantly higher eeriness ratings than the Robot-Synthetic stimulus, suggesting that the dual-pitch voices are also mismatching stimuli.

We predicted that the android would be judged as a robot; thus the mismatching auditory stimulus for this visual category was proposed to be the human voice, and the matching stimuli were proposed to be the dual-pitch and synthetic voices. Generally, participants gave higher eeriness ratings for the Android-Human stimulus than the Android-Synthetic stimulus, suggesting that it was indeed a mismatched video. Additionally, the Android-Synthetic combination generated the lowest eeriness ratings for the android visuals. Here, it can be suggested that the android was being perceived as as robot. However, this implies there to be a significant increase in eeriness from the Android-Human to the Android-Synthetic stimuli. Statistically however, there was no difference between the two voice conditions. Furthermore, the dual-pitch combinations (Android-50Hz, Android-150Hz, Android-250Hz) were also seen as significantly eerier than the Android-Synthetic stimulus, suggesting that they were also mismatching audio. The full results are given in Fig. 5.

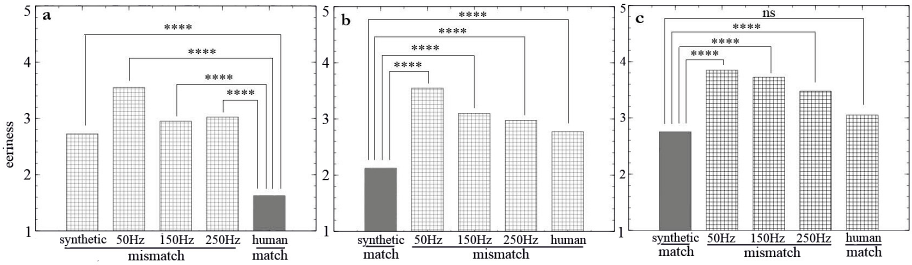

**Fig. 5.** Average eeriness ratings **a**: for the human, **b**: for the humanoid robot, **c**: for the android. Although initially thought to be matching stimuli, the dual-pitch voices for both the robot and android visuals are instead mismatching stimuli. The average eeriness rating for the Android-Human stimulus is statistically the same as the Android-Synthetic stimulus, highlighting confusion about the android's category membership. **** denotes $p \leqslant 0.0001$.

### 3.5 Experiment Two Discussion

This experiment serves as further evidence to support the Bayesian model of the uncanny valley. Here, we have shown that eeriness can be induced by mismatching stimuli, using a variety of different combinations. For a mechanical, 'obvious' humanoid robot that is far away from a category boundary, an 'obvious' synthetic voice is most suited to it. On the other end of the humanness scale, a human face, which is also far away from a category boundary, is best matched with a human voice.

For the android visuals, however, conclusions are a little more difficult to draw. The Android-Human stimulus received a higher rating of eeriness than the Android-Synthetic combination, though not significantly so. However, the Android-Human stimulus received significantly higher ratings of familiarity and appeal (data not shown), which contradicts what should happen in the presence of increased eeriness. It is also theorized that there is confusion about where the android sits in terms of categorical definition, thus why there is no statistical difference between the Android-Human and Android-Synthetic stimuli. Possibly, a dual-nature is being perceived due to there being misaligned cues at category membership.

It was already predicted by the uncanny valley model for visuals of a mechanical humanoid robot, such as the iCub, to be perceived as less eerie than the android, so it follows that generally, the videos of the humanoid robot are rated as less eerie than the videos of the android. However, the introduction of audio implements another layer of complexity to the problem. Multimodal cues are indeed influencing participant judgment, as the eeriness of a certain visual was also dependent on the voice it was combined with. Fig. 4b shows that the Android-Human combination is less eerie than the Robot-50Hz combination, and that the Human-50Hz combination is regarded as eerier than the Android-Human combination. In these cases the audio alone has reversed the uncanny effect, such that a human or humanoid robot is regarded as stranger than an android.

The dips in eeriness in Fig. 4b are hypothesized to be caused by stimuli that can be easily classified, for example, a mechanical humanoid robot paired with a synthetic voice which sits within the non-human category. On the far right of the graph, the Human-Human combination is also well defined in category. The peaks in eeriness may then be explained as the result of misaligned cues. For example, the android visuals combined with dual-pitch voices, that sound almost human, are stimuli that may be regarded as eerie in both the visual and auditory domain. Thus there is an increase in perceptual tension; the face and voice combined give rise to an enhanced peak in eeriness.

## 4 Conclusions

In this study, two experiments were conducted to investigate the impact of conflicting cues from visual and auditory stimuli. We have shown that removal of unimodal, misaligned cues from the eyes of an android can significantly decrease the eeriness felt in observers and that the impact of cue removal is dependent on where the android sits in terms of humanness, or rather, how far it is from a category boundary. Thus, the eyes of an android have a great impact on observers in human-robot interaction. Humanoid robots, with low ratings of humanness, are seen as eerier when cues from the eyes are removed as they are further away from a category boundary.

We have also replicated and extended an experiment that tests the influence of multimodal cues. Our results agree that the uncanny does indeed exist within the auditory continuum and that visual stimuli regarded as non-eerie can fall into the uncanny with the introduction of audio. Our results also agree that mismatching voices and faces will induce the uncanny. Additionally, we have shown that a dual-pitch voice, derived from a human voice, is regarded as significantly eerier than a text-to-speech synthetic voice when combined with visual stimuli. It is hypothesized that the dual-pitch voices sit near to a category boundary and thus give rise to a greater amount of perceptual tension. Although developed to sound robotic, the dual-pitch voices do not match with robotic faces. Furthermore, we have demonstrated that an android sits near to a categorical boundary which in turn gives rise to perceptual uncertainty about its identity. This results in both the Android-Human and Android-Synthetic combinations being regarded as the same in terms of eeriness ratings.

In both experiments, the results agree with the Bayesian explanation of the uncanny valley and suggest that perceptual distortion, caused by misaligned cues, gives rise to perceptual tension which is felt as unease or eeriness in observers. This study thus serves as empirical evidence for the Bayesian model.

**Acknowledgments.** The authors would like to thank the Speech and Hearing Lab for providing a booth for use in the experiments.

## References

1. Feldman, N.H., Griffiths, T.L., Morgan, J.L.: The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. Psychological Review 116(4), 752 (2009)
2. Freud, S.: The uncanny (j. strachey, trans.). the standard edition of the complete psychological works of sigmund freud, vol. 17 (1919)
3. Grimshaw, M.N.: The audio uncanny valley: Sound, fear and the horror game. Audio Mostly, 21–26 (2009)
4. Jentsch, E.: On the psychology of the uncanny (1906) 1. Angelaki: Journal of the Theoretical Humanities 2(1), 7–16 (1997)
5. Looser, C.E., Wheatley, T.: The tipping point of animacy. How, when, and where we perceive life in a face. Psychological Science 21(12), 1854–1862 (2010)
6. MacDorman, K.F., Green, R.D., Ho, C.C., Koch, C.T.: Too real for comfort? Uncanny responses to computer generated faces. Computers in Human Behavior 25(3), 695–710 (2009)
7. Mashiro, M.: Bukimi no tani (the uncanny valley). Energy 7, 22–35 (1970)
8. Mitchell, W.J., Szerszen Sr, K.A., Lu, A.S., Schermerhorn, P.W., Scheutz, M., MacDorman, K.F.: A mismatch in the human realism of face and voice produces an uncanny valley. i-Perception 2(1), 10 (2011)
9. Moore, R.K.: A Bayesian explanation of the uncanny valley effect and related psychological phenomena. Scientific reports (2012)
10. Pollick, F.E.: In search of the uncanny valley. In: Daras, P., Ibarra, O.M. (eds.) UCMedia 2009. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 40, pp. 69–78. Springer, Heidelberg (2010)
11. Ramey, C.H.: The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots. In: Proceedings of Views of the Uncanny Valley Workshop: IEEE-RAS International Conference on Humanoid Robots, pp. 8–13 (2005)