Discovering the uncanny valley for the sound of a voice.

Dennis Jansen
STUDENT NUMBER: 2022966

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE
DEPARTMENT OF COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE
SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
TILBURG UNIVERSITY

Thesis committee:

Prof. Dr. M.M. Louwerse
Dr. A. Alishahi

Tilburg University
School of Humanities and Digital Sciences Department of Cognitive Science & Artificial Intelligence Tilburg, The Netherlands
June 2019

Preface

First of all, I would like to thank Nick for getting me onboard of this crazy ride. Also, Katherine for channelling our ideas and guiding us towards a decent research question. Furthermore, Julija for being of great help before, during and after the experiment. A special thank you to Afra for providing me with helpful feedback on the thesis itself. And lastly, Max for supervising the entire project and for keeping us focussed.

The sound of a voice that leads you into the Uncanny valley.

Dennis Jansen

With the increasing possibility to create artificial voices that are more human-like, it is important to understand how humans perceive this voice in a human-robot interaction. The current study has researched whether the sound of a voice could cause an Uncanny valley effect. This effect claims that at a certain point on the scale from a robotic entity to a full human being a sudden decrease emerges in how familiar a human-being feels in relation to an entity. The study at hand focused on the categorical uncertainty hypothesis, which proposes that the uncanny valley effect emerges due to problems with the categorisation of an entity. More specifically, the aim of this research was to find if auditory features of an entity could cause an uncanny valley effect. Participants were asked to categorise voices on a nine-level scale from robot to human whether a voice was human or not human. Results supported this hypothesis, both the answers as well as the response times to these answers indicated that an uncanny valley indeed emerges at the expected location on the scale from robot to human.

Keywords*: Uncanny valley; Auditory features; Categorical uncertainty; Robot; Humanness*

1. Introduction

Ever since the creation of robots and artificial avatars, the interaction between humans and robots has been a topic of debate amongst researchers. Hancock et al. (2011) suggest that due to the increasing technical abilities regarding the development of robots, different working fields start implementing the use robots. Sheridan (2016) states that because of this increase, research on the interaction between humans and robots, but also the design of robots is needed. Moreover, with the growing possibility to produce robots that are more and more human-like, the relationship between robots and humans is subject to change. Ruijten and Cuijpers (2017) suggest that in order to have a natural like interaction between robots and humans, the appearance of an entity has to be perceived as human-like. However, Stein and Ohler (2018) stated that developers of robots should always keep the uncanny vally (UV) theory in mind. This UV effect was first introduced by Mori in 1970. Mori (1970/2012) proposed this so-called uncanny valley (UV) theory which shows a sudden decrease, a valley, in how a perceiver's affinity towards an entity increases as an entity becomes more human-like (Figure 1).

Mori (1970/2012) proposed this monotonically increasing function starting at an industrial robot, where someone does not show much affection towards, up to a healthy human-being which has a 100% affection from the perceiver. Around the 60-80% range on the human-likeness scale a sudden decrease emerges on the affection scale. Mori illustrated this by looking at a prosthetic hand. Since a prosthetic hand has been developed such that it somewhat resembles a normal hand at first sight, a closer look shows us that it is artificial which causes an eerie feeling. This is the point where the UV appears in the theory. Moreover, Mori proposes that movement of an entity is vital in how humans perceive this entity. He states that if an industrial robot is programmed in such a way that it can move its own parts, humans feel more affection towards this robot. Contrary to this, if a prosthetic hand that is somewhere around the lowest point of the UV starts moving, the eeriness becomes stronger. Mori concludes that adding movement to his original model causes higher peaks, but also deeper lows, thus a deeper UV.

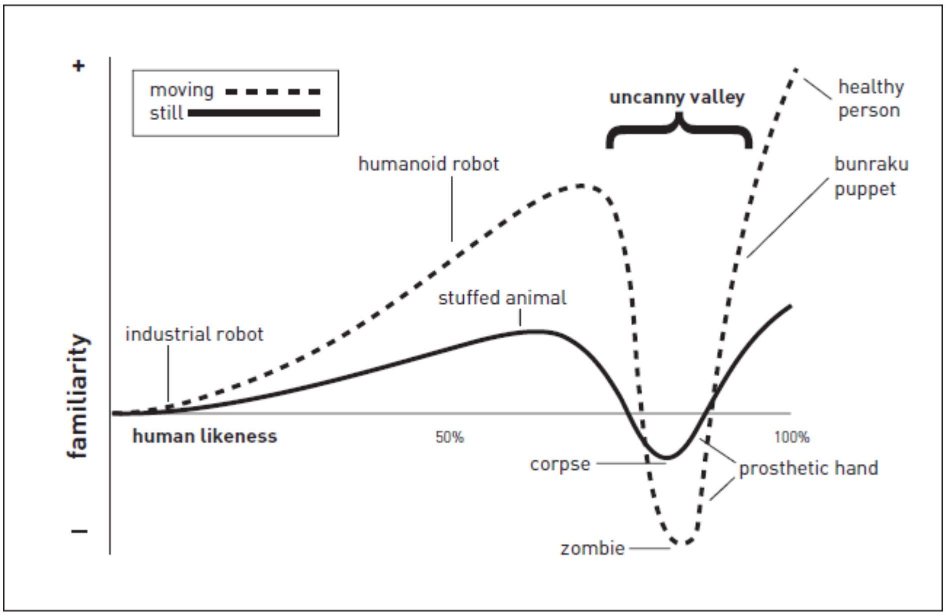

Figure 1. Both lines in the figure show a monotonically increasing function with a sudden decrease in affinity from a human towards an entity, the uncanny valley, between the 60% and 80% range on the scale of human-likeness. An increased effect is illustrated for moving entities with the dotted line. Adapted from "Avoiding the uncanny valley in virtual character design.," by V. Schwind, K. Wolf and N. Henze, 2018, *Interactions, 25*(5), 47. Copyright 2018 by "ACM, Inc.".

Seyama and Nagayama (2007) suggest that the UV theory not only applies to robot entities, but any object that resembles a human being such as dolls, masks and characters in video games. Furthermore, the researchers state that entities that are developed to interact with humans, should be built with the impact and psychological impact on humans in mind. Mori (1970/2012) proposed that developers of robotic entities should aim for a moderate level of realism in order to avoid the UV. Moreover, Seyama et al. (2007) state that Mori's (1970/2012) theory has been a widely adopted guideline for developers.

Despite the great interest and the importance of the UV theory, no consensus was found about the exact mechanisms that underlie this theory. Throughout the years, the definition of UV, ways of measuring and the cognitive processes that underlie this phenomenon have been a topic of debate amongst researchers. Wang, Lilienfeld and Rochat (2015) conducted a review study in which they stated this difficulty. They reviewed a set of hypotheses which could explain the UV effect (Table 1). The hypotheses as studied by Wang et al. (2015) will be outlined below.

Firstly, Wang et al. (2015) reported the pathogen avoidance hypothesis as proposed by MacDorman and Ishiguro (2006). This hypothesis claims that the emotion of disgust underlies the UV phenomenon. More specifically, MacDorman, Green, Ho and Koch (2009) proposed that due to the imperfections of a (humanoid) robot, people perceive the robots as carriers of diseases and illness. Since more human-like robots are perceived as more human, their lacks and deficits are perceived more intense and thus cause more eeriness. Wang et al. (2015) suggest that prior research has tried to focus on the pathogen avoidance hypothesis, but that the hypothesis has never been tested directly, meaning that no empirical evidence in favour of this hypothesis was found.

Secondly, the cause of the UV effect could also lie in the fact that a robot reminds humans of their own mortality. Since robots are in the basis dead objects that suddenly come to live, it triggers a defence-system in humans that copes with the fear for mortality (Ho, MacDorman, & Pramono, 2008). Although researchers have tried to test this mortality salience hypothesis, the results could not provide conclusive evidence that robots elicit this fear for mortality. Ho et al. (2008) proposed that both the fear of death and the feeling of disgust as proposed in the pathogen avoidance theory are predictors for the UV effect. They furthermore claim that it is not one single phenomenon that underlies the effect, but that it could be an interaction that causes this.

Contrary to earlier findings (MacDorman et al. 2009; MacDorman et al. 2006; Ho et al. 2009), Hanson (2005) suggested that not the realism of a robot causes the UV effect, but that it is caused by its low attractiveness. The evolutionary aesthetics hypothesis proposes that different physical attributes that indicate the fitness of a person, its fertility and health are important in how people perceive someone's attractiveness. Hanson tested this hypothesis by using morphed faces. As soon as the aesthetics of the face increased, the UV seemed to disappear. Albeit promising, after simultaneously manipulating both the aesthetics and the realism of the face, Hanson could not prove that the two were not associated.

Whereas the pathogen avoidance hypothesis, the mortality salience hypothesis and the evolutionary aesthetics hypothesis focussed on the appearance of a robotic entity, Wang et al. (2015) described three more hypotheses that focusses on the cognitive processes within the perceiver.

Firstly, Wang et al (2015) discussed the violation of expectation hypothesis. Just as Mori (1970/2012) proposed when he first introduced the UV theory, this hypothesis describes that the realization of something being artificial, though it seems human-like, causes the UV. Mitchell et al. (2011) found that a mismatch between the realism of a voice and face can cause an eerie feeling, hence the UV effect. Moreover, Brenton, Gillies, Ballin, and Chatting (2005) found that a mismatch between the appearance of an entity and its behaviour can cause a sense of eeriness. A similar result was found by MacDorman et al. (2009). They found that a mismatch in the physical proportions of a face can cause an UV effect. Although the violation of expectation hypothesis and the research conducted concerning this hypothesis seems promising, Wang et al (2015) suggest that this only partially explains the emerging of the UV effect. According to the researchers it could be that the violation of expectations explains the UV effect, but it could also be triggered by other mechanisms.

One of these underlying mechanisms could be found in the categorical uncertainty hypothesis. Other than the violation of expectation hypothesis, which describes that the eerie feeling associated with the UV is caused by expectations about an entity that are violated, the categorical uncertainty hypothesis describes that uncanniness comes to play when categorical boundaries are unclear (Ramey, 2006). Mathur and Reigling (2016) suggested that the UV occurs at the point where people take longer to categorize an entity. They based this claim on the research done by Yamada, Kawabe and Ihaya (2012). Their research showed that a peak in response times occurred at the point where people had difficulties categorising a morphed face. Moreover, the researchers found a negative correlation between the response times and the likeability of an entity. Contrary to the findings by Yamada et al. (2012) and the replication study by Mathur et al. (2016), Cheetham, Wu, Pauli and Jancke (2015) did not find such results. In their study using electrophysiological measures the researchers did not find a negative affect for the categorically ambiguous avatars. Moreover, they found a negative affect at the more robotic side of their morph range. Furthermore, Kätsyri, Förger, Mäkäräinen, and Takala (2015) suggest that this correlation found by Yamada et al. (2012) might have occurred as a confounding effect of the morphed faces itself. Prior research could not provide conclusive evidence for supporting the categorical uncertainty hypothesis

Although the violation of expectation hypothesis as well as the categorical uncertainty hypothesis try to explain how conflicts in cognitive processes cause the UV to emerge, none of these hypotheses explain the underlying believes about robots and humans and how a violation of these can cause the UV effect. In order to answer these questions Wang et al. (2015) state the mind perception hypothesis. This hypothesis tries to explain that the difference between robots and humans not only lies in their physical appearances, but more so in how humans perceive them in more detail. Gray and Wegner (2012) propose that two main features on which humans differ from non-humans are agency and experience. Agency describes the possibility to make plans and to do things. Experience on the other hand is based on the ability to feel and sense. These are two concepts that robots are not very well capable of according to the researchers. This made them propose that the UV does not come from the fact that a robot does not appear human enough, but that humans associate the human-like appearance with certain key capabilities of human which a robot does not have. In other words, they proposed that since robots do look very similar to humans it is expected that the robots have the same abilities as humans do. The researchers found support for their claim that the lack of agency and experience causes the UV to emerge, but Wang et al. (2015) question these findings because they suggest that attributing these abilities to artificial entities by humans does not seem to provoke negative feelings in various domains.

To conclude the given hypotheses, it can be said that there have been two main approaches in explaining the UV effect. Firstly, the pathogen avoidance hypothesis, the mortality salience hypothesis and the evolutionary aesthetics hypothesis tried to explain the UV effect based on the physical attributes of an entity. Secondly, the violation of expectation hypothesis, the categorical uncertainty hypothesis and the mind perception hypothesis looked at the cognitive processes that could underlie the UV effect. The claims made in the last three hypotheses suggest that not only the physical appearances of an entity can cause the UV effect, but that it could well be caused by other aspects of an entity.

Table 1
An overview of the hypotheses regarding UV as stated by Wang et al. (2015).

Hypothesis	Summary
Pathogen Avoidance	Due to the imperfections of a (humanoid) robot, people perceive the robots as carriers of diseases and illness.
Mortality Salience	A robot reminds humans of their own mortality.
Evolutionary Aesthetics	Different physical attributes that indicates the fitness of a person, its fertility and health are important in how people perceive someone’s attractiveness
Violation of Expectation	The realization of something being artificial, though it seems human-like causes an UV effect.
Categorical Uncertainty	The feeling of uncanniness emerges when categorical boundaries are unclear.
Mind perception	The difference between robots and humans not only lies in their physical appearances, but more so in how humans perceive them.

Although research regarding the different hypotheses has been conducted, none of the hypotheses mentioned above could be fully confirmed nor refuted. This leaves the question what causes the UV effect and whether it even exists open for debate. The aim of the current research is to contribute to this specific field of research. Despite the hypotheses giving some directions on what to dive into with this research, some key aspects in the UV theory remain uncertain.

Since humans gain information about an entity in a split second it is important to look at hypotheses that address the UV effect in an early stage (Rezlescu et al., 2015). The categorical uncertainty hypothesis seems to do just so. The basis for this hypothesis lies in the categorical perception theory. Goldstone and Hendrickson (2009) suggest that categorical perception happens early on in the processing of stimuli. Moreover, the researchers suggest that the more distinguishable the stimuli, the easier it is categorised by humans. Jentsch (1906/1997) proposed that difficulties in categorising entities as being animate or not can cause feelings of unease. Yamada et al. (2012) furthermore suggest that the more fluent a stimulus can be processed, the more positive it is perceived. The categorical uncertainty hypothesis suggests that at a point where categorical boundaries are unclear, the affection humans feel towards an entity is low (Yamada et al., 2012; Ramey, 2006). Moreover, in a replicating study done by Mathur et al. (2016), the researchers indeed found that the findings of Yamada et al. (2012) could be supported.

Although promising, some limitations in the design in the study of Yamada et al. (2012) could be argued. Mori (1970/2012) proposed that with the increasing human-likeness of an entity a perceivers affection towards that entity increases, up until the point where the UV effect emerges. Yamada et al. (2012) researched the crossover between two entities on a five-step scale, but this was never between a human face and a robotic face. It could be argued that in order to test the categorical uncertainty hypothesis, which could explain the UV effect, researchers should focus on the crossover between humans and robots. Yamada et al. (2012) conducted three experiments regarding the categorical uncertainty hypothesis. Firstly, the researchers looked at the difference between humans and dolls, secondly the difference between cartoon dogs versus stuffed dogs, real dogs versus stuffed dogs and real dogs versus cartoon dogs. Lastly, they searched for this effect in a crossover between a male human face and a female human face. Although they did find some evidence for the categorical uncertainty hypothesis, none of the three experiments did entails Mori's initial idea of the UV effect. Moreover, it could be that the five-step scale that Yamada et al. (2012) used in their experiments could be too small, since Mori (1970/2012) proposed that in his model it is the slight and subtle changes that cause the UV effect.

Furthermore, in his model, Mori (1970/2012) proposed that not only the physical appearance of an entity contributes to the perception of humans, but other aspects could be of importance as well. Mori elaborates extensively on the perception of movement of an entity. Prior research done in order to test the hypotheses that might underlie the UV effect mainly focused on the physical appearances of an entity, more specifically, the faces. However, Nie, Pak, Marin and Sundar (2012) found that not only the physical appearance of a prosthetic hand, or the movement, but also the touch can cause an UV effect. The effect of holding a hand and its warmth does increase the trustworthiness of the robot, but a mismatch between the expectation of a real human touch and a robot increases the feeling of eeriness. With this finding, the researchers tap into the original idea of Mori (1970/2012) that not only the physical appearance is a cause of the UV effect.

Empirical evidence suggests that also voices can cause the UV effect comes from a study done by Karle et al. (2018). They suggest that not only facial expressions are important in receiving emotional information, but also the voice modulations. They state that both facial and vocal cues are important in successful social interaction between humans. Moreover, Baird et al. (2018) suggest that the more human-like an entities' voice, the more it is perceived as likeable. However, in their study Baird et al. (2018) found that an UV effect, as has been previously found for visual features, does not appear. It should be noted, that Baird et al. (2018) used thirteen voices of which one voice was a human control and that they did not use a gradient scale such as proposed by Mori (1970/2012). Participants were asked to rate the likeability of the voice on a 5-point Likert scale ranging from 1 (Not at all) up to 5 (Extremely). Although Baird et al. (2018) did use no other than (artificial) human voices it is debatable that the researchers did not modify those voices on a gradient scale, the results from their research could not be seen as fully refuting the UV hypothesis.

Contrary to the findings of Baird et al. (2018), Mitchell, et al. (2011) found that the voice of an entity is important in the interaction between robots and humans. The researchers looked at the interaction between a face and a voice and found that a mismatch between the realism of an entity's face and voice could indeed cause an UV effect to emerge. Although promising, Mitchell et al. (2011) do not provide conclusive results whether auditory features alone (e.g. a voice) do cause an UV effect, since it is only in combination with the presence of a face.

Additionally, Tinwell and Grimshaw (2009) found similar results stating that the lack of lip and voice synchronization causes an increased eeriness towards an entity. They also found that exaggerating the mouth articulation while producing sounds has the same effect. Interestingly, they also suggest that a lack of human-likeness of the voice shows a similar effect. Despite it being promising for the role of auditory features in the UV effect, it does not explicitly show that the audio alone could be a cause for the UV effect since the voice of an entity is always accompanied by a face.

Importantly, Latinus and Belin (2011) state that a voice is a carrier of information about an entity that humans can use to identify or obtain characteristics from an entity. Humans are even capable of doing this without understanding what is said exactly or even understanding the language that is used. Furthermore, the voice can provide information about the speaker's emotional state (Vlasenko, Prylipko, Böck & Wendemuth, 2014). Schroeder and Epley (2015) do not only suggest that the sound of a voice provides us with information about the identity of a person and the emotional state of the speaker, but also about the competences and intellect. Louwerse, Graesser, Lu and Mitchell (2005) suggest that a (mis)match between the voice and a visual of an avatar could highly affect how a person perceives this avatar. According to the researchers it highly influences the likeability of the avatar, but moreover they found effects in the comprehension scores of the avatars, suggesting that the voice plays a crucial role in the interaction between entities.

To summarize the prior research conducted regarding the UV effect and the role of audio in the cause of this effect it can be concluded that despite the great interest in the UV effect and its role in the human-robot interaction debate, a couple of questions remain unanswered. Most of the researchers interested in the UV effect have focused on either the visual features of an entity alone or in combination with other senses such as movement, the tactile senses or audio. The research at hand concerning the auditory features of an entity in relation to the UV effect seem to be promising, yet not conclusive. It shows however that the importance of auditory features, especially the voice, should not be forgotten. To the best of our knowledge the auditory features such as the voice of an entity has not gained much attention in the past. The aim of the research at hand is to find an answer to the question whether the auditory features alone could cause a similar effect as has been proposed by prior research. Besides that, keeping the original theory by Mori (1970/2012) in mind, it will focus on finding the effect on a more gradual scale from robot to human.

2. Study

In order to answer the question whether auditory features of an entity can cause an UV effect, participants were asked to categorise different voices as 'human' or 'not human'. The stimuli set consisted of 10 different voices, half of them were male voices, the other half were female voices and were gradually modified from completely human to completely robotic. Participants were asked to respond as quickly as possible by pressing one of two buttons. The current study resembles the experiment conducted by Yamada et al. (2012). The main differences between the two studies lie in the fact that audio was used instead of images. Apart from that, this study focusses on the gradual increase from robot to human, whereas Yamada et al. (2012) used different visual categories. Moreover, where the prior study used only five steps to morph between the two categories, in this experiment nine steps were taken.

2.1 Participants

A total of 104 students (54.81% female) at Tilburg University (*M* = 22.56 years old, *SD* = 4.19) participated in this study. In exchange for their participation, the students were all awarded 0.5 course credit concerning their human-subject pool. All participants were recruited via the human-subject pool. The vast majority of the participants originated from the Netherlands (*n* = 70) with the rest of the participants coming from different countries worldwide. Prior to the experiment participants were asked if they had problems with either their visual or auditive abilities. Three persons stated not having a normal or corrected to normal vision and one person stated not having a normal or corrected to normal hearing. Also, participants reported whether they were either left-handed, right-handed or both-handed. The majority of participants reported being right-handed (77%).

2.2 Setup

In order to control for problems during the experiment such as technical problems, distraction and internet speed, the experiment was held during six sessions of half an hour each at the campus of the Tilburg University. The room, which is normally used as a classroom, had twenty computers available to use for the experiment. All computers were connected by an ethernet cable to the university's local network. Since the experiment was conducted via Qualtrics software (Qualtrics Labs, Inc., 2019), an online tool for surveys and experiments, and the stimuli (e.g., audio files) had to be buffered, a stable connection was required.

2.3 Stimuli

The audio files used were all taken from the LibriSpeech ASR corpus (Panayotov, Chen, Povey & Khudanpur, 2015). This corpus consists of over 1000 hours of English speech which is derived from audio books. All files were documented in detail and were of the same quality with a sampling rate of 16kHz. Ten voices (50% female) were then semi-randomly selected. For each voice, one sentence was selected. All sentences had to have the same format: *pronoun + verb + determiner + (adjective) + noun.* This was done to control for sentence length and the content of the sentence. An example of a sentence used could be: "*I own a blue car.*".

The ten selected sentences were then modified using Apple's Logic Pro X software. This is an audio recording and editing software that is used by both professional and amateur musicians and sound engineers. By using the built-in Vocal Transformer tool, all voices were than manipulated to create the robotic sounding voice. A widely used method to do this is by tweaking the pitch and the formant of the audio file. This is often used by game developers, film producers and musicians (Mayor, Bonada & Janer, 2011). The pitch and the formants are the key building blocks of a voice (Pisanski & Rendall, 2011). By adjusting the pitch of a signal, the pitch itself will either go up or down, but it does not affect the timbre of the voice. This can be done by adjusting the formant of a signal. Firstly, the formant was increased up to a point where the voice was still understandable, this resulted in the formant setting at +8. Then the pitch was increased the same way, up to +2. With the Vocal Transformer tool, it is possible to make a cross-over between the original signal, keeping the output level constant. In Table 2, an overview of the percentages of the audio signal that were made up by the Vocal Transformer tool signal. All files were then exported as wav-files, to make sure none of the audio quality was lost.

Table 2

An overview of the nine different levels of voice modification ranging from completely robotic to completely human. The right column shows the percentage of the signal that was made up by the signal from the Vocal Transformer tool.

Humanness	Percentage Vocal Transformer
1 (Completely robotic)	100%
2	87.5%
3	75%
4	62.5%
5	50%
6	37.5%
7	25%
8	12.5%
9 (Completely Human)	0%

All ten voices were manipulated the same way and on nine levels, resulting in a set of 90 stimuli. Apart from the stimuli used in the experiment a different sentence was used in the practice trial. This was only presented on both the 'Completely robotic' level and the 'Completely human' level.

2.4 Design

To categorise whether a voice was either 'human' or 'not human', participants had to press one of two keys on a keyboard. If a participant perceived a voice as 'human', the participant had to press the "F" key and the "J" key if a participant thought it was 'not human'. Since the response times of the participants per trial were measured, the participants were encouraged to respond as quickly as possible, without waiting for the sentence to end. All trials were forced response, so it was not possible to skip a trial.

To control for confounds such as stimuli order, consecutive modification levels or repeating faces back-to-back, the stimuli were semi-randomly presented to the participants. Eighteen blocks consisting of five different stimuli were created. Each pair of two consecutive blocks existed of the ten different voices. The five voices within a block were then randomly presented to the participants. Since this experiment was part of another study focussing on the visual aspects of the UV effect, participants were randomly assigned to the auditory part of the experiment first or the visual experiment first. 51.92% of the participants (n = 54) were assigned to the 'audio-first' condition, all other participants were assigned to the 'visual-first' condition.

Participants were asked, in order to participate in the experiment, to bring their own headphones or earphones. A couple of spare headphones were provided by the experimenters in the case of an emergency. The output volume of the computers was set to 10% of the maximum output volume.

2.5 Procedure

Upon entry, participants were asked to take a seat behind the computer and were asked to plug in their headphones. All participants were then informed to be silent and have no contact with other participants during the experiment. They were also instructed to remain seated as soon as they had finished the experiment. As soon as everyone was finished, participants were debriefed, thanked for their participation and asked to leave the room.

Once the experiment had started, participants were first presented an information letter about the experiment and its background. Secondly, the participants were asked to sign the physical copy of the informed consent letter before continuing with the experiment. Participants were then asked some demographic questions such as age, country of origin, dominant hand and whether they had any visual or auditive disabilities. These questions were then followed by more detailed instructions about the task at hand and four practice trials (two audio trial, two visual trials). After this, the participants were first presented a fixation-cross (three seconds), followed by 90 stimuli, either visual or audio. As soon as the participant pressed the "J" key or the "F" key the next stimulus was presented and the audio would automatically start playing as soon as it was fully buffered. After 90 trials the participants were instructed that they had finished the first part of the experiment, which was then followed by the next 90 stimuli. After completing the two sets of stimuli, participants were asked if they had encountered any problems during the experiment.

2.6 Statistical analysis

The data gathered during the experiment consisted of the answers given by the participants and the response times. The response times were corrected by distracting the buffer times. These buffer times were recorded through a Javascript in Qualtrics. No participants were excluded from analysis, due to irregularities, but response times above 2.5 standard deviations from the mean per participant per level of humanness and below 150 ms were excluded (1.86%). Whelan (2008) argues that response times below 100 ms could be results of fast guessing, therefore a cut-off point between 100 ms and 200 ms should be used. All data was analysed using Python 3.6 and RStudio 1.1.463. In order to check for inter-rater reliability, Krippendorff's Alpha was calculated using the mean rating per level of humanness per participant for the answers given. The answers given were analysed by fitting a Mixed Effects Logistic Regression (MELR) model on the data to test if the level of humanness is a predictor for the answer given. Furthermore, the response times were analysed by performing t-tests for every level of humanness compared to a randomly shuffled baseline. By creating a random distribution from the original distribution of response times, a baseline was developed that reflects the distribution of response times as if stimuli were not categorised according to their level of humanness. This baseline reflects response times to the stimuli without dependencies between their level of humanness. According to prior research, this allows us to control for falsely interpreting random increases in response times as behavioural patterns. (Richardson & Dale, 2005; Louwerse, Dale, Bard & Jeuniaux 2012). Additionally, a Linear Mixed Effects (LME) model was fit on the data of the response times to test the differences between levels of humanness and to control for individual differences between participants and stimuli.

3. Results

In order to test the hypothesis stating that auditory features can cause an UV effect both the answers given, corresponding to the decision whether a stimulus would be either human or not human, and the response times per stimulus were analysed separately. Results from the study seem to provide support for the hypothesis. An UV effect, caused by the auditory features of an entity, emerged at the proposed spot on the scale from robot to human (in between the 60% - 80% human-like range, Mori, 1970/2012).

3.1 Answers

Firstly, the inter-rater reliability was tested using the data obtained from the answers given by the participants. To test the reliability of the data, Krippendorff's Alpha was calculated using the mean rating per level of humanness per participant. According to Krippendorff (2004, p. 241) an alpha score of one represents full agreement, a score of zero represents full disagreement. Furthermore, an alpha score of .800 or higher should be considered as reliable, but a score of .667 can be seen as acceptable. The results for the data obtained in this study lie in this range ($\alpha = 0.72$) meaning it can be considered reliable.

Further analyses regarding answers given was done by looking at the ratio between the number of responses given as being human per level of humanness divided by the number of responses given as not human. This resulted in the plot as can be found in Figure 2. As can be seen in this figure for the first four levels the ratios are close to zero indicating that few participants answered with 'human' compared to the participants that answered with 'not-human'. The slope then starts to increase slightly up to the seventh level of humanness after which the slope starts to increase more strongly. From the figure it becomes clear that the biggest disagreement emerges at the seventh level of humanness (ratio closest to one = 0.95) after which it rapidly increases in favour of the number of responses given as a stimulus being human.

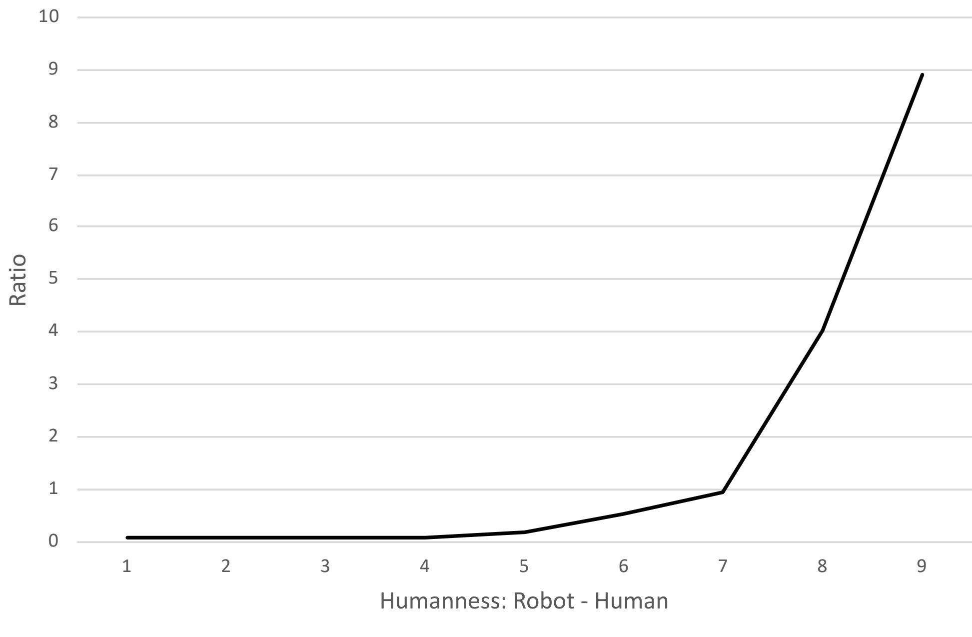

Figure 2. The ratio of answers given either being human or not human on the nine different levels of humanness.

Furthermore, to test whether the level of humanness of a stimulus can predict the answer that will be given, a Mixed Effects Logistic Regression (MELR) model was fitted on the data. The results from the MELR model can be found in Table 3. The calculated odds ratio indicate that the model can predict the answers given at the sixth and seventh level of humanness with least certainty. Furthermore, the logit scaled estimated coefficients were significantly different from each other starting at the fourth level of humanness, as is suggested from the confidence intervals. Since the MELR model can predict the sixth and seventh level of humanness with the least certainty, it shows that a deviating pattern emerges at those two levels.

Table 3

The results of the Mixed Effects Logistic Regression (MELR) model fitted on the data obtained from the answers given by the participants. The table shows the logit-scaled fixed effects estimates for each level of humanness as well as the lower and upper level 95% confidence intervals and the odd's ratios. The type of voice was used as a random factor and the level of humanness as fixed factor.

Level of humanness	Estimate	Lower level	Upper level	Odds ratio
Intercept	-2.85	-3.17	-2.53	1.18
2	0.18	-0.18	0.54	1.20
3	0.07	-0.30	0.44	1.21
4	0.35	0.01	0.70	1.19
5	1.04	0.73	1.36	1.17
6	2.20	1.91	2.49	1.16
7	2.80	2.51	3.09	1.16
8	4.27	3.96	4.58	1.17
9	5.07	4.74	5.41	1.19

3.2 Response times

Analysis of the response times shows a peak at the seventh level of humanness. Figure 3 illustrates a plot of the response times and a baseline consisting of all the response times randomly shuffled. By comparing the response times of the original distribution to the shuffled distribution, it was tested whether the recorded response times significantly differ from the randomly shuffled response times. The results of computing t-tests for all the nine different levels of humanness between this randomly shuffled baseline and the original response times can be found in Table 4. Results show that at the first, second and third as well as the sixth and seventh level of humanness the response times significantly differ from the randomly shuffled baseline. At the lowest levels of humanness participants had significantly lower response times compared to the randomly shuffled baseline. At the fifth, eight and ninth level of humanness, the reaction times did not differ significantly. Interestingly, at the sixth and seventh level of humanness, the response times were significantly higher.

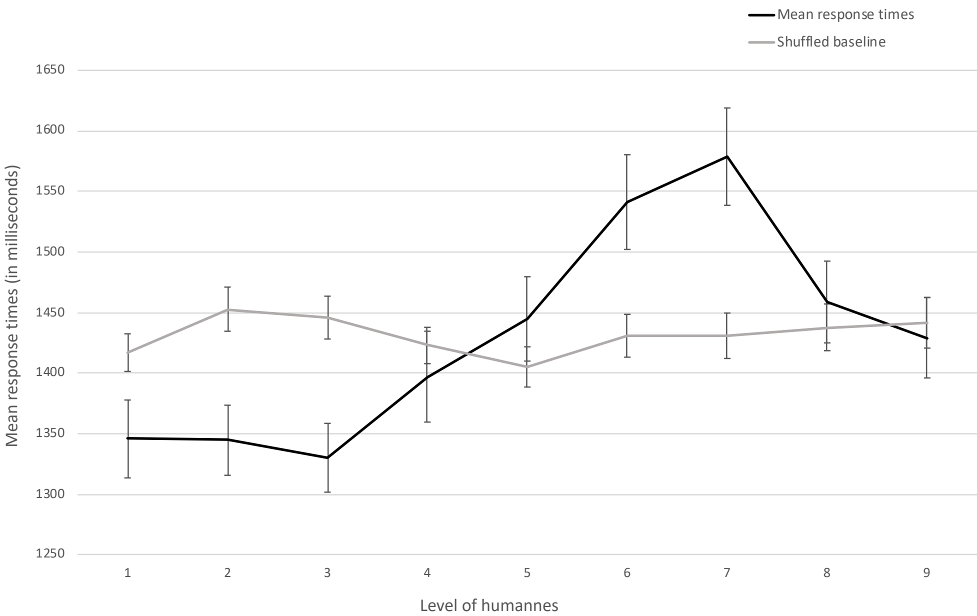

Figure 3. The response times per level of humanness. The black line represents the actual response times, the grey line is a randomly shuffled baseline. The errors bars indicate the standard error of the mean.

Table 4

The results of the t-test performed between the mean response times per level of humanness and the randomly shuffled baseline.

Humanness	*t*	*p*
1 (Completely robotic)	-2.01	0.047
2	-3.15	0.002
3	-3.48	0.001
4	-0.65	0.518
5	1.02	0.308
6	2.56	0.011
7	3.35	0.001
8	0.55	0.584
9 (Completely human)	-0.33	0.743

Furthermore, to test whether levels of humanness differ significantly from each other, multiple Linear Mixed Effects (LME) models were fitted on the data. The results of these LME models can be found in Table 5. Results indicate that the highest positive regression coefficient is found between the fifth and the sixth level of humanness. Moreover, the biggest negative regression coefficient was found between the seventh and eighth level of humanness. Both these coefficients as well as the coefficients of the third and fourth level of humanness and the fourth and fifth were significant. Results from the response times analysis seem to be incompliance with the results obtain from the answers given.

Table 5

Linear Mixed Effects (LME) models were fitted on the response times of two level of humanness each and voice type as random factor. Estimated regression coefficients, as well as the standard error (SE), p-value and the z-scores were reported.

Model	Levels of humanness	Estimate	*z*-score	SE	*p*
1	1-2	-1.693	-0.90	19.05	.929
2	2-3	-14.93	-0.76	19.70	.449
3	3-4	63.69	3.01	21.17	.003
4	4-5	51.076	2.32	21.98	.02
5	5-6	96.59	4.47	21.60	<.001
6	6-7	37.10	1.56	23.79	.119
7	7-8	-123.14	-5.22	23.61	<.001
8	8-9	-29.0	-1.38	21.06	.168

4. Discussion

Interacting with robots can cause feelings of eeriness and unpleasantness. This feeling is often referred to as the Uncanny valley (UV) effect, which was first proposed by Mori in 1970. He stated that at a certain point, between the 60% to 80% range, on a scale from a robotic entity up to a fully human entity a dip in the familiarity that a person feels towards that entity emerges. The existence, explanation and ways of measuring this effect has been topic of debate over the past centuries. Different hypotheses trying to explain why and how this effect emerges have been stated and tested in the past. The main focus in this research is on the so-called categorical uncertainty hypothesis. This hypothesis proposes that the difficulty of categorising an entity causes the UV effect. In order to test this hypothesis, a forced-response task was held amongst 104 participants. Participants were asked to categorise a stimulus as being either human or not human. The ninety stimuli were created by modulating a voice from human to robot on a gradual scale in nine steps. Both the responses as well as the response times from participants were recorded and used for analysis.

In prior studies, researchers have tried to explain the UV effect based on the visual features of an entity. Despite the fact that Mori (1970/2012) proposed that not only visual features can be the cause of the UV effect, other features of an entity have gained less attention. Auditory features such as the voice of an entity have gained small attention from researchers in the past. This is even more remarkable since it is proposed that the voice is a carrier of information about an entity (Latinus et al., 2011). The little research about the auditory features as a cause for the UV effect that is available is highly controversial and not conclusive. (Karle et al., 2018; Baird et al., 2018; Mitchell et al., 2011; Tidwell & Grimshaw, 2009).

The first contribution to the existing literature from this study might lie in the fact that the auditory features of an entity will be researched in isolation as a possible cause for the UV effect. Secondly, despite the attention the categorical uncertainty hypothesis has gained in the past, to the best of knowledge of the authors this effect has not been tested on a gradual scale from a robot entity to a human entity. (Yamada et al., 2012; Ramey, 2006; Mathur et al., 2016). This research aims to find the UV effect on a more gradual scale as proposed by Mori (1970/2012).

Both the results from the analysis of the answers given as well as the results from the analysis of the response times show that around the sixth and seventh level of humanness something happens in the perception of the participants. At his point on the scale, the discrepancy amongst the participants whether a certain stimulus is human or not human is at its peak. Before and after this point the majority of the participants seem to agree more on whether it is human or not. Although these findings do suggest that categorical uncertainty emerges amongst the different participants, it does not fully imply the UV effect at this point. The categorical uncertainty hypothesis (Wang et al., 2015) suggests that the UV effect emerges because of the difficulty a person has to categorise an entity. The analysis of the answers given alone cannot provide the information whether this emerges. It does show that no consensus was found amongst the participants, but it does not show how certain a participant made this choice. The response times, according to prior research (Yamada et al., 2012; Mathur & Reigling, 2016), can provide us with these answers. The response times show a sudden increase at the sixth and seventh level of humanness. This, in compliance with the results from the answers given, strengthens the idea that an UV exists at this point. Moreover, when compared to a randomly shuffled baseline the response times at the point on the scale do significantly differ, implying that this peak in response times does emerge. The results from the LME models show that these two levels of humanness significantly differ from the levels right before and after, indicating that the peak observed is not due to chance.

This raises the question if this peak in the response times and the disagreement between participants whether a stimulus is either human or not human can be explained by the UV theory or that other mechanisms underlie this. One could possibly argue that the effect found is not due to the UV theory, but that it is just another example of how categorisation tasks work and the difficulties that emerges at certain points. Edelman (1995) proposes that the categorisation of objects is slower at a point where inter-object similarity is high, meaning that category A and category B show great overlap making it thus difficult to categorise them. This is a similar effect that can be found in the stimuli used for this experiment. However, such an effect would then be at the fifth level of humanness since that is the point where the stimuli is 50% human and 50% robot, meaning that the inter-object similarity is at its highest point. This was not found in this study, but contrary to this, results indicate that the difficulty emerged around the sixth and seventh level of humanness, where the stimuli is 67.5% human and 32.5% robotic (sixth level) and 75% human 25% robotic (seventh level) respectively.

It can also be argued that the stimuli used in this study are not two perfect different categories, but that a stimulus is always more human than robotic. This could be a well-founded argument since the basis of the stimuli lies in a pure human voice which has then been modified to a certain point. Only the sound of the voice had been modified keeping all other aspects, that define the characteristics of the stimuli, constant such as syntax, semantics and tempo. However, if this were the case, then it would be expected that the sudden increase in the number of participants that rated a voice as being human would occur much earlier on the scale of humanness, at least before the fifth level where the stimuli are 50/50. The results obtained from this study indicate that that this phenomenon not occurred before or at this breaking point, but later on the scale suggesting that the results obtained from the data could indeed suggest the existence of the UV effect.

Regarding the methods used in this study, it could be that the results as found in the analysis might lie in the experimental paradigm itself. Participants were asked to rate whether a stimulus was either human or not human, meaning that the focus of the participants could probably lie more on the human part of the stimuli rather than the distinction between human and robot. The attentional biases theory could explain the shifting from the peak of response times and the sudden increase in the rating of something being human over not human away from the middle level of humanness. This theory states that a person’s attention and perception is affected by specific thoughts occurring at the time of making a decision (e.g., Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg & van IJzendoorn 2007; Kyrios & Iob,1998). Since participants were primed to make a choice between human or not human it suggests that, based on the attentional bias theory, participants would be more focused on human features of the stimuli than the robotic modifications, suggesting that they would categorise a stimulus as being human faster. If this attentional bias would indeed play a role in this study, it would be expected that the increase in responses rating a stimulus as being human would occur before the fifth level of humanness. Results show that this is not the case and that it appears, contrary to this theory, on the more human side of the fifth level of humanness.

Another limitation in both the chosen paradigm as well as the underlying idea of the categorical uncertainty hypothesis lies within the true definition of the UV theory. Since participants were primed to respond as quickly as possible, decisions about whether a presented voice was either human or not were made within a very short amount of time. The categorical uncertainty hypothesis describes that a sudden increase in response times could point towards the existence of an UV effect. However, other hypotheses should also be considered. In his first proposal of the UV theory, Mori (1970/2012) states that the UV emerges as soon as a person realizes that an entity is more artificial than was expected. For example, the prosthetic hand he talks about in his paper looks human-like at first, but a closer look shows that it is artificial, thus causing the UV effect. This has been elaborated on more by researchers considering the violation of expectation hypothesis (Mitchell et al., 2011; MacDorman et al. 2009). Our study does not take into account this idea, since participants were primed to make a decision in a split second. This could mean a couple of things. Firstly, it could be that the UV effect is not influenced by time at all and that it could emerge at different points within the time of the interaction with an entity. Secondly, it could be that the categorical uncertainty is just a categorization problem that is not related to the UV theory, but this seems unlikely since Yamada et al. (2012) not only found an increase of response times, but also a decrease in likeability of the entity at the same spot. Lastly, it could be that the violation of expectation is only partly supported. Mitchell et al. (2011) found that a mismatch between an entity and its voice causes a feeling of eeriness, but they did not take time as a factor into account. Participants were exposed to videos of fourteen seconds each, showing either the visual and the voice of an entity in a congruent or incongruent condition. This video was then looped until a participant finished rating the entity on eight seven-point Likert scales. Future research could focus on time as a factor in the UV theory debate and whether this influences how a person perceives an entity.

Furthermore, future research could focus on other features of an entity concerning the UV effect. The visual aspects of an entity have gained much attention in the past. This research contributes by investigating if auditory features can cause the UV as well. Results are promising, which could be seen as a ground for future research that focusses on other aspects such as movement, tactile senses, smell etcetera. The interaction between multiple features could be concerned as well. Prior research has focussed on the mismatch between features (e.g. the visual appearance of an entity and a voice), but little research has been conducted concerning the interaction between different features. Questions such as which of multiple features is leading in causing the UV effect or if some features could minimize the effect could have major implications in how humans interact with robot.

Results from this study provide support the categorical uncertainty hypothesis. A peak in the response times show that the uncanny valley appears towards the human side of a gradual scale from robot to human. Furthermore, it is promising that the results obtained in the study are based on audio stimuli rather than visual alone or a combination of the two, indicating that an auditory feature of an entity could indeed cause an UV effect. Claims made in this study should however be treated with cause. The exact definition, experimental paradigms and even the existence of the UV remain topic of debate. This research is a contribution to the specific field of research concerning the UV but cannot be seen as conclusive. Results do, however, suggest that auditory features of an entity can cause an UV effect. Future research could focus on further establishing a proper manner on how to measure the UV and discover if more features of an entity can cause the UV effect and which of those features are most important.

References

Baird, A., Parada-Cabaleiro, E., Hantke, S., Cummins, N., Schuller, B., Burkhardt, F., & 19th Annual Conference of the International Speech Communication INTERSPEECH 2018. The perception and analysis of the likeability and human-likeness of synthesized speech. *Proceedings of the Annual Conference of the International Speech Communication Association, Interspeech,2018-september*, 2863- 2867. doi:10.21437/Interspeech.2018-1093

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin, 133*(1), 1–24. doi:10.1037/0033-2909.133.1.1

Brenton, H., Gillies, M., Ballin, D., & Chatting, D. (2005). The uncanny valley: Does it exist? *Proceedings of the 19th British HCI Group Annual Conference.* 1–4.

Cheetham, M., Wu, L., Pauli, P., & Jancke, L. (2015). Arousal, valence, and the uncanny valley: Psychophysiological and self-report findings. *Frontiers in Psychology, 6*. doi:10.3389/fpsyg.2015.00981

Edelman, S. (1995). Representation, similarity, and the chorus of prototypes. *Minds and machines*, *5*(1), 45-68. doi: 10.1007/BF00974189

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, *125*(1), 125–130. doi:10.1016/j.cognition.2012.06.007

Goldstone, R.L., & Hendrickson, A.T. (2010). Categorical perception. *Wiley interdisciplinary reviews. Cognitive science, 1*(1), 69-78. doi:10.1002/wcs.26

Hanson, D. (2005). Expanding the aesthetic possibilities for humanoid robots. *IEEE-RAS international conference on humanoid robots*, Tsukuba, Japan

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human factors*, *53*(5), 517-527. doi:10.1177/0018720811417254

Ho, C., MacDorman, K.F., & Pramono, Z.A. (2008). Human emotion and the uncanny valley: A GLM, MDS, and Isomap analysis of robot video ratings. *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 169-176.

Jentsch, E. (1997). On the psychology of the uncanny. *Angelaki, 2*(1), 7–16. doi: 10.1080/09697259708571910 (Original work published in 1906).

Karle, K. N., Ethofer, T., Jacob, H., Brück, C., Erb, M., Lotze, M., Nizielski, S., Schütz, A., Wildgruber, D., Kreifelts, B. (2018). Neurobiological correlates of emotional intelligence in voice and face perception networks. *Social Cognitive and Affective Neuroscience, 13*(2), 233–244. doi:10.1093/scan/nsy001

Kätsyri, J., Förger, K., Mäkäräinen, M., &Takala, T. (2015). 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*. doi: 10.3389/fpsyg.2015.00390

Krippendorff, K. (2004). Content analysis – An introduction to its methodology (2nd ed.). Thousand Oaks: Sage Publications, Inc.

Kyrios, M., & Iob, M. A. (1998). Automatic and strategic processing in obsessive–compulsive disorder: Attentional bias, cognitive avoidance or more complex phenomena? *Journal of Anxiety Disorders, 12*, 271–292

Latinus, M., & Belin, P. (2011). Human voice perception. *Current Biology, 21*(4), R143–R145 .doi:10.1016/j.cub.2010.12.033

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive science*, *36*(8), 1404-1426.

Louwerse, M. M., Graesser, A. C., Lu, S., & Mitchell, H. H. (2005). Social cues in animated conversational agents. *Applied Cognitive Psychology, 19*(6), 693 704. doi:10.1002/acp.1117

MacDorman, K. F., Green, R. D., Ho, C.-C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior*, *25*, 695–710. http://dx.doi.org/10.1016/ j.chb.2008.12.026

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems 7*(3), 297–337.doi:10.1075/is.7.3.03mac

Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the Uncanny valley. *Cognition*, *146*, 22–32.

Mayor, O., Bonada, J., & Janer, J. (2011). Audio transformation technologies applied to video games. In *Audio Engineering Society Conference: 41st International Conference: Audio for Games*. Audio Engineering Society.

Mitchell, W. J., Szerszen, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., & MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception, 2*(1), 10–*12.*doi:10.1068/i0415

Mori M (2012) The uncanny valley (K.F. MacDorman, N. Kageki, Trans.). *IEEE Robotics and Automation*, *19*(2), 98–100 doi: 10.1109/MRA.2012.2192811 (Original work published in 1970)

Nie, J., Pak, M., Marin, A. L., & Sundar, S. S. (2012). Can you hold my hand? *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '12.* doi:10.1145/2157689.2157755

Panayotov, V., Chen, G., Povey, D., & Khudanpur, S. (2015). *Librispeech:* An ASR corpus based on public domain audio books. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* doi:10.1109/icassp.2015.7178964

Pisanski, K., & Rendall, D. (2011). The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness. *The Journal of the Acoustical Society of America, 129*(4), 2201-12.

Qualtrics Labs, Inc. (2019). Copyright © [2019] Qualtrics [Survey software]. Provo, UT, USA. https://www.qualtrics.com

Ramey, C. H. (2006). An inventory of reported characteritics for home computers, robots, and human beings: Applications for android science and the uncanny valley. *Proceedings of the ICCS/CogSci-2006 Long Symposium 'Toward Social Mechanisms of Android Science'*, Vancouver, Canada.

Rezlescu, C., Penton, T., Walsh, V., Tsujimura, H., Scott, S. K., & Banissy, M. J. (2015). Dominant voices and attractive faces: The contribution of visual and auditory information to integrated person impressions. *Journal of Nonverbal Behavior, 39*(4), 355-370. doi: 10.1007/s10919-015-0214-8

Richardson, D. C., & Dale, R. (2005). Looking to understand: the coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science, 29*(6), 1045–1060. doi:10.1207/s15516709cog0000_29

Ruijten, P. A. M., & Cuijpers, R. H. (2017). Dynamic perceptions of human-likeness while interacting with a social robot. In *HRI 2017 - Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, 6-9 March 2017, Vienna, Austria* (pp. 273-274). New York: Association for Computing Machinery, Inc. DOI: 10.1145/3029798.3038361

Schroeder, J., & Epley, N. (2015). The Sound of Intellect: Speech Reveals a Thoughtful Mind, Increasing a Job Candidate's Appeal. *Psychological Science*, *26*(6), 877–891. doi: 10.1177/0956797615572906

Seyama, J., & Nagayama, R.S. (2007). The Uncanny valley: Effect of Realism on the Impression of Artificial Human Faces. *PRESENCE: Teleoperators and Virtual Environments, 16*, 337-351. doi:10.1162/pres.16.4.337

Sheridan, T. B. (2016). Human–Robot Interaction: Status and Challenges. Human Factors, 58(4), 525–532. doi: 10.1177/0018720816644364

Stein, J., & Ohler, P. (2018). Uncanny…but convincing? inconsistency between a virtual agent's facial proportions and vocal realism reduces its credibility and attractiveness, but not its persuasive success. *Interacting with Computers, 30*(6), 480-491. doi:10.1093/iwc/iwy023

Schwind, V., Wolf, K., & Henze, N. (2018). Avoiding the uncanny valley in virtual character design. *Interactions*, *25*(5), 45-49. doi: 10.1145/3236673

Tinwell, A., & Grimshaw, M. (2009). Bridging the uncanny. Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era*. MindTrek '09.* doi:10.1145/1621841.1621855

Vlasenko, B., Prylipko, D., Böck, R., & Wendemuth, A. (2014). Modelling phonetic pattern variability in favor of the creation of robust emotion classifiers for real-life applications. *Computer Speech & Language, 28*, 483-500.

Wang, S., Lilienfeld, S., & Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology, 19*(4), 393-407. doi:10.1037/gpr0000056

Whelan, R. (2008). Effective Analysis of Reaction Time Data. *The Psychological Record, 58*(3*),* 475–482. doi:10.1007/bf03395630

Yamda, Y., Kawabe, T., & Ihaya, K. (2012). Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon. *Japanese Psychological Research, 55*(1), 20–32. doi:10.1111/j.1468-5884.2012.00538.x