Investigating the Uncanny Valley Effect for Prosody

Ahdiyeh Alipour

Snr:2088437

Master's Thesis

Communication and Information Sciences

Specialization New Media Design

Department Communication and Cognition

School of Humanities and Digital Sciences

Tilburg University, Tilburg

Supervisor: Dr. Tom Lentz

Second Reader: Dr. Jan de Wit

July 2023

Abstract

The integration of human-like characteristics into Intelligent Virtual Assistants (IVAs) is a promising strategy to boost Human-Robot Interaction (HRI) by leveraging users' natural tendency to anthropomorphize technology. However, anthropomorphism can also elicit negative reactions due to the uncanny valley effect (UVE). The prosody of computer-generated speech plays an important role in conveying human-like qualities, but its impact on the UVE remains unclear. This study aims to identify the certain point of human likeness in the prosody of computer-generated speech that can trigger the UVE, through its effects on participants' perceptions of anthropomorphism, robomorphism, eeriness, and trust. Two experiments, involving a total of 88 participants, were conducted to investigate that certain point since it is known to be in proximity to fully human-like prosody, while its precise location remains unexplored. The level of roboticness in the prosody was manipulated as the IVAs read descriptions of museum paintings. Both experiments consistently yield similar findings. There is only evidence for a linear relationship between the level of roboticness in the prosody and participants' aforementioned perceptions. Additionally, the study reveals the mediating role of eeriness in the relationship between anthropomorphism and trust. However, the expected valley pattern of the UVE is not observed. The results highlight the need for further investigation, considering the various features and levels in the prosody. The findings of this study may also contribute to the design and implementation of IVAs, alleviating concerns about negative emotional and behavioral responses when employing human-like voices.

Keywords: Intelligent Virtual Assistants, Prosody, Human-Robot Interaction, Anthropomorphism, The Uncanny Valley Effect

Contents

Investigating the Uncanny Valley Effect for Prosody

The proliferation of virtual agents in our daily lives can be attributed to significant advancements achieved in Artificial Intelligence (AI) and machine learning technologies (Agarwal et al., 2022). Intelligent virtual agents (IVAs) or chatbots are the most widely used form of AI in Human-Robot Interaction (HRI) and are designed to simulate human conversation and speech in order to interact with users (Bansal & Khan, 2018). The recent evolution in machine learning as well as natural language processing (NLP) has improved the ability of modern IVAs to comprehend and use human language. As a result, there has been a surge in their applications across diverse fields such as education, healthcare, and marketing (Abdulla et al., 2022). This has led to a heightened awareness of the substantial value and potential benefits of IVAs. Therefore, it is becoming increasingly imperative to understand how humans interact with these machines and identify the specific features that can improve the quality of HRI (Westerman et al., 2019).

One prominent way to enrich the immersive user experience and interaction is the deliberate inclusion of human-like characteristics by implementing features such as language, visual elements, and voices in IVAs (Følstad & Brandtzaeg, 2017). According to the Computers Are Social Actors (CASA) paradigm, people tend to treat technologies as real humans, and an interface resembling a human is preferred since it requires users to make fewer adjustments when starting social interactions with it (Nass, et al., 1994). Moreover, CASA entails that the users are more likely to respond positively to IVAs that resemble human beings (Westerman et al., 2019). As an example, the use of human-like agents can foster greater levels of trust (Lee & Nass, 2010; Natarajan & Gombolay, 2020), enable more profound connections to be formed between users and IVAs (K. M. Lee et al., 2005), increase purchase intention (Han, 2021), and cause customer loyalty (Jenneboer et al., 2022). Consequently, some HRI researchers are striving to humanize IVAs by adding strong anthropomorphic cues.

The concept of anthropomorphism refers to a subconscious process in which our brains automatically assign human-like qualities to non-human objects with triggering characteristics when we encounter them, even if we consciously recognize that the object is not alive nor capable of human-like behavior (Mithen & Boyer, 1996). One of the key elements of anthropomorphism is the prosody of computer-generated speech (Nishio et al., 2012). Prosody refers to the patterns of duration, intonation (pitch), and loudness. Through prosodic features, speakers can convey emotions, emphasize certain words or phrases, and express communicative intent such as sarcasm or irony, all of which add to human speech's unique qualities (Rao et al., 2013). Accordingly, research focusing on the prosody of computer-generated speech has the potential to improve the perception of human-like qualities in the voice of IVAs, thereby fostering HRI.

Anthropomorphism holds great potential for improving user experiences but can also present challenges. While there are some studies have highlighted the potential benefit of using anthropomorphized IVAs in HRI (Złotowski et al., 2015; Roesler et al., 2021), other studies have revealed negative consequences associated with their use (Strait et al., 2015; Yogeeswaran et al., 2016). The negative findings regarding anthropomorphism and likeability may be attributed to the non-linear relationship inherent in the uncanny valley effect (UVE), wherein a salient yet imperfect resemblance to humans can evoke negative responses and feelings of eeriness (Mori, 1970/2012). In other words, simply anthropomorphizing an artificial inanimate entity does not always lead to increased likeability. Therefore, it is important to delve into the various features that may activate the UVE in IVAs in order to mitigate negative effects.

However, the findings regarding the negative effect and the non-linear relationship between anthropomorphism and the feeling of eeriness have been inconsistent. On the one hand, numerous studies have provided evidence that increasing anthropomorphism, especially in the visual aspect effect (Song & Shin, 2022; Mara et al., 2022; Strait et al., 2015) or through the incorporation of both visual and audio features (Mitchell et al., 2011; Tinwell& Grimshaw, 2009), can result in adverse consequences and trigger the UVE. These findings suggest that as the resemblance to humans is very large, the sense of eeriness becomes more pronounced, aligning with the predictions of the UVE. On the other hand, other studies have taken a different approach, focusing solely on the impact of anthropomorphism on the face (MacDorman et al., 2009) or audio features (Baird et al., 2018; Kimura & Yotsumoto, 2018; Kühne et al., 2020). Surprisingly, these studies have revealed that the most human-like appearance or speech does not necessarily elicit the strongest sense of eeriness, contradicting the expectations derived from the UVE. These findings suggest that the relationship between anthropomorphism, and the experience of eeriness is followed by a linear relationship. Considering this inconsistency, to the best of the present researcher's knowledge, a significant research gap still exists when it comes to examining the isolated impact of the prosody to identify the presence of the UVE.

The present study aims to identify a certain point in the level of human likeness in prosody that is not too high to cause the UVE and evoke an emotional response of eeriness. In addition, the primary goal of individuals when using technology is commonly seeking qualities that can strengthen their trust and reliance on it (McKnight et al., 2002). Consequently, the present study evaluates if the level of trust is influenced by anthropomorphized computer-generated speech prosody and can be mediated by the feeling of eeriness.

Considering the gap in understanding the role of prosody in audio-only contexts and its impact on the UVE, exploring the level of human likeness in prosody that can elicit feelings of eeriness can broaden our understanding of the UVE. Furthermore, this investigation can contribute to the effective design and implementation of IVAs, potentially preventing negative emotional and behavioral responses through the optimization of prosodic features. Hence, the research question is:

> ***RQ****: Is there an uncanny valley effect for prosody in computer-generated speech? If so, at which level of human likeness does this effect occur?*

Theoretical Framework

This literature review starts with the IVAs and their potential advantages**.** After that, the concept of anthropomorphism and its implication in enhancing HRI is presented. It also addresses the critical question of whether incorporating anthropomorphic features is always beneficial, highlighting the phenomenon of the UVE and explaining the underlying reasons that contribute to the feeling of eeriness. Subsequently, it discusses anthropomorphic prosody for IVAs, investigating its potential to trigger the UVE. Finally, this review concludes with the presentation of formulated hypotheses.

Intelligent Virtual Agents

The remarkable advancements in voice recognition technology and IVAs software have been exponential over the past few years, resulting in numerous applications in diverse fields, including home automation (Terzopoulos & Satratzemi, 2020); assistive tools for disabled people (Pradhan et al., 2018), education (Mekni et al., 2020), e-commerce (Chai et al., 2001), and even the provision of companionship and support for the elderly (Wolters et al., 2016). As an example, Wolters et al. (2016) indicated the valuable use of IVAs as assistive technology for individuals with dementia, emphasizing their ability to provide a continuous verbal presence, respond patiently to repetitive questions, and offer motivation and reassurance. Therefore, IVAs have the potential to promote the quality of HRI.

The quality of HRI is substantially influenced by the parameters, which can be categorized into two groups: appearance (static) and behavior (dynamic) (Von Zitzewitz et al., 2013). According to their study, the appearance category includes parameters that relate to a particular system's visual appearance and sound. These parameters describe the observable or perceptible physical characteristics, where the voice assumes the primary role as a communication channel, while the visual appearance influences the user's perception. On the other hand, their category of behavior encompasses parameters that describe the dynamic aspects of an entity such as actions, movements, and interactions. To enhance the user experience and interaction, IVAs are frequently engineered to mimic human characteristics in appearance and behavior (Følstad & Brandtzaeg, 2017).

Anthropomorphism

Anthropomorphism is the spontaneous and unconscious cognitive process whereby individuals ascribe human-like characteristics to non-human entities, even when acknowledging the non-human nature of those entities (Mithen & Boyer, 1996). Studies have indicated that anthropomorphism can occur for any non-human entity, including animals, mechanical devices, and technological apparatuses (Epley et al., 2007).

The tendency to anthropomorphize technology can be described by the Computers Are Social Actors (CASA) paradigm, where individuals commonly tend to attribute human characteristics to computers and other technological devices, thereby interacting and applying similar social norms and expectations to human-like technologies as they would to actual humans (Nass et al., 1994). Essentially, when encountering technology or other non-human entities, we tend to default to using familiar human-like traits and behaviors as a way to comprehend and interact with them in order to bridge the gap between the unfamiliar and the familiar, enabling us to engage with these entities more easily (Morana et al., 2020). Consequently, interfaces resembling humans are perceived to be more appealing and user-friendly, requiring lesser adaptation by users (Reeves & Nass, 1996).

The Modality, Agency, Interactivity, Navigation (MAIN) model developed by Sundar (2008) provides further support for the implication of human likeness in interactions with IVAs. According to this model, the activation of the human likeness heuristic is a necessary step to perceiving IVAs as credible. This is because individuals tend to perceive entities resembling the natural world as more real and credible. Therefore, IVAs that resemble humans are prone to be seen as more credible by users, thereby enhancing the overall effectiveness of their interactions (Sundar, 2008).

Aligned with the perspective of the CASA paradigm and the MAIN model, the tendency to attribute human-like qualities to inanimate objects is contingent upon features exhibiting a certain degree of humanness (Epley et al., 2007). Humanness perception is the degree to which users ascribe characteristics resembling human traits to technology (Hu et al., 2021). According to their study, the perceptions of technology as human-like are shaped by the presence and intensity of human-related attributes, such as the type of voice employed, ranging from synthetic to more human-like. In simpler terms, the more human-like features there are, the more users perceive the entity as being human-like. Since the process of human perception is a complex psychological phenomenon, verbal and nonverbal cues can influence human perception, including visual characteristics such as the face, auditory cues such as voice, and personal traits such as identity (Netter et al., 2021).

The interaction between users and IVAs indicates that incorporating human-like characteristics into different aspects of IVAs enhances their human-likeness quality, which leads to more favorable responses from users (Lee et al., 2005; Lee & Nass, 2010; Salem et al., 2013; Złotowski et al., 2015). One of the positive outcomes of anthropomorphized IVAs is their ability to foster a stronger emotional bond with users (Lee et al., 2005). For instance, their study involved a month-long interaction with Sony's robot dog AIBO, revealing that users displayed social responses by developing affection, perceiving the robot as a social being. Also, the simple manipulation of anthropomorphic qualities in AIBO 's form and behavior enhanced these natural social responses. Moreover, implementing anthropomorphic gestures or speech has been found to contribute to positive user experiences (Salem et al., 2013). In their experiment with a humanoid robot, they investigated the impact of the robot's hand and arm gestures on perceptions of human likeness, likability, and future contact intentions. The findings revealed that when the robot used gestures and verbal together during the interaction, participants exhibited a higher tendency to anthropomorphize, found more likable, and expressed greater intentions for future contact, even when the gestures were partly incongruent with speech.

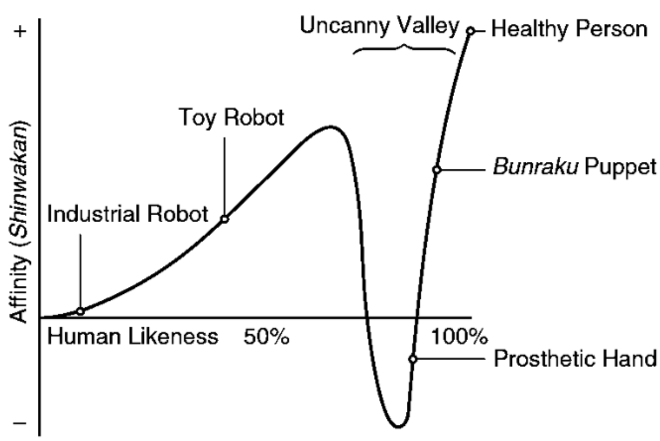
The literature on the interaction between users and IVAs suggests on the one hand anthropomorphized IVAs have the capabilities to elicit more positive responses from users. On the other hand, empirical studies on anthropomorphized IVAs have reported inconsistent results with this notion, and some studies have even found negative effects from humanizing IVAs (Dehn & Van Mulken, 2000; Strait et al., 2015; Yogeeswaran et al., 2016; Song & Shin, 2022). These studies suggest that negative effects may be explained by the UVE proposed by Mori (1970/2012).

The Uncanny Valley Effect

The UVE is the psychological phenomenon that the emotional response toward human replicas follows a non-linear pattern (Mori, 1970/2012): namely as the resemblance of a replica to human characteristics increases, individuals generally experience a growing sense of affinity which is the sense of liking or attraction, but when the replica reaches a stage where it closely resembles a human but falls slightly short, there is a sudden dip in emotional response, characterized by feelings of discomfort, eeriness, or strangeness. This dip is the UV, a valley in the graph representing the relationship between human likeness and emotional response (see Figure 1). Mori (1970/2012) suggested that as human likeness increases, the emotional response may turn negative until a certain point is reached, beyond which a further increase in human likeness leads to a positive emotional response again. Also, he stated that the UVE is typically observed when encountering humanoid objects, such as IVA or prosthetics, that possess human-like features and behaviors but do not fully replicate human appearance or behavior.

Figure 1

Mori's graph (1970/2012)



Since the introduction of the term UV, extensive studies have been conducted to elucidate its underlying existence and triggers (Hanson, 2005; MacDorman & Ishiguro, 2006; Wang et al., 2015). One possible explanation is the Evolutionary Aesthetics hypothesis, which proposes that humans have developed an elevated sensitivity toward aesthetics such as appearance characteristics due to evolutionary pressures (Hanson, 2005; MacDorman & Ishiguro, 2006). According to this hypothesis, the sensation of uncanniness arises when encountering appearances that deviate from established universal aesthetic norms, given the failure to meet the expected standards of human likeness. MacDorman and Ishiguro (2006) put forth that human preference for specific physical attributes is shaped by selection pressures associated with fitness, fertility, and health, and that the uncanny sensation experienced towards human replicas is not attributable to their lack of realism, but rather their perceived low level of attractiveness. To investigate this, Hanson (2005) conducted experiments employing morphed images and observed the presence of the UVE. The result indicated that with the enhancement of aesthetic qualities in the uncanny face, the presence of the UVE diminished or disappeared.

The Violation of Expectation hypothesis offers an alternative explanation for the UVE focusing on cognitive processes (MacDorman & Ishiguro, 2006). According to this hypothesis, human replicas generate expectations for human-like characteristics, but these expectations are not met, resulting in an uncanny feeling. Mori (2012) further suggests that the sensation of uncanniness arises from perceptual tension caused by conflicting perceptual cues or mismatches. Cross-modal and incongruent features contribute to the uncanny sensation, including discrepancies between appearance and motion (Saygin et al., 2012), disproportionate facial features such as enlarged eyes and head (Seyama & Nagayama, 2007), or mismatches between facial features and corresponding voice (Mitchell et al., 2011).

An additional assumption that contributes to the UVE is the Categorical Uncertainty hypothesis. This hypothesis explores the perception of ambiguity or uncertainty regarding the boundary of classification for an entity (Wang et al., 2015; Ramey, 2006). In simpler words, when individuals encounter new information that deviates from their established categorical expectations of human likeness, it can lead to a sense of discomfort or eeriness. Mathur and Reigling (2016), in support of this hypothesis, found that when participants were presented with a sample of 80 real-world robot faces, they took longer to categorize entities as human-like with ambiguous faces and rate them as less likable compared to those with categorically unambiguous and clearly defined features.

Building upon these insights, numerous empirical studies have been conducted to investigate different aspects of the UVE, employing various terms such as attractiveness (Ho & MacDorman, 2010), familiarity (Hanson, 2005), pleasantness (Seyama & Nagayama, 2007), and eeriness (MacDorman &Ishiguro, 2006), each with different assessment methods and materials. However, there is an imbalance in the selection criteria for experimental materials, as both the uncanny feeling and human likeness are multifaceted phenomena, necessitating the use of diverse stimuli in research (Zhang et al., 2020). In the comprehensive review of recent studies spanning over 15 studies, Zhang et al. (2020) found that visual stimuli, such as pictures or video clips focusing on the face, head, or entire body, were commonly used while none of the studies isolated the role of auditory stimuli concerning evoking the UVE. Indeed, the majority of studies have primarily focused on the visual aspect of humanizing IVAs when observing the presence of UVE. Consequently, there is a scarcity of research related to the potential negative effects of humanizing auditory-only factors, without relying on visual cues.

Anthropomorphic Prosody and the Uncanny Valley Effect

Since early machines did not possess the capability of speaking, the focus was primarily on their visual appearance in the initial phases of automation, which led to a limited discussion of the voice of historical automatons (Männistö-Funk & Sihvonen, 2018). Pettorino (2015) offered a historical analysis of early talking apparatuses, exploring two distinct approaches. The first approach, known as voice transport, involved the use of trickery to convey a hidden person's voice through a device. The second approach, artificial voice, aimed to imitate human speech to reproduce similar sounds. According to Pettorino (2015), these approaches have played a significant role in the development of talking apparatuses throughout the centuries. However, there is a lack of detailed information and studies concerning auditory functions in this context. The advanced IVAs have led to the recognition of the significance of the voice or speech characteristics as a key indicator of appearance cues in HRI which can signal the presence of human-like minds (Schroeder et al., 2016). The auditory cues can hold the same level of importance as visual cues when it comes to influencing individuals' perception of robots in conveying emotions and human-like qualities (Karle et al., 2018; Sims et al., 2009). Undoubtedly, the ability to speak has been a fundamental attribute that sets humans apart from other living and non-living entities. Speech is commonly regarded as a uniquely human trait, and its integration into artificial machines has long been seen as a means of creating a hybrid of human and machine (Männistö-Funk & Sihvonen, 2018).

Speech is a medium through which individuals convey their emotions, intentions, and moods (Von Zitzewitz et al., 2013). Prosody is partially the result of the natural flow of human speech, which is defined by variations in duration (the length of sounds and pauses), intonation (pitch variation), and energy patterns (loudness and other types of emphasis). These prosodic cues serve as the perceptual qualities of speech that humans rely on to accomplish diverse speech tasks, such as conveying emotions or intention through emphasis. For example, when expressing active emotions like anger, pitch and energy values tend to be higher, whereas these same parameters are relatively lower for passive emotions like sadness (Rao et al., 2013). Moreover, when it comes to entities with voice, James et al. (2018) have shown that listeners tend to favor voices that sound more human-like than robotic because robots have difficulty conveying more nuanced emotions.

Since prosody and its cues can distinguish voices with human-like characteristics from those that produce robotic sounds, numerous studies have investigated the effect of these features on speech acceptability in HRI (Montano et al., 2017; Niculescu et al., 2013). For instance, Montano et al. (2017) found a positive relationship between pitch and trust perception. In their study, it was found that women exhibited a preference for placing trust in men with lower-pitched voices when considering them as political leaders. However, when it comes to mating situations, women tend to favor men with higher-pitched voices. Furthermore, Niculescu et al. (2013) demonstrated a substantial impact of the voice pitch on quality aspects of user interaction, and the perception of factors such as appeal, enjoyment, and empathy. They created two robot characters: Olivia, an extroverted and humorous robot with a higher-pitched voice, and Cynthia, an introverted and composed robot with a lower-pitched voice. Their results revealed Olivia received better ratings in terms of overall appearance, voice appeal, behavior, and personality. Moreover, Zhong et al. (2022) aimed to support the idea that prosody is important for likeability. They conducted an evaluation involving two distinct robots, Pepper and Joey, and found that Pepper, with its more natural-sounding voice, was generally perceived as more likable compared to Joey, which had a monotonous, machine-like voice. The likability cannot be solely attributed to the voice of the robots, as the study also considered the visual design, which was different for each robot. Therefore, voice alone may not be responsible for the observed differences in likability. In addition, these studies have not considered the UVE.

There have been some studies exploring the role of the voice of IVAs and its potential relationship to the UVE along with visual cues. Mitchell et al. (2011) used four 14-second videos, featuring a character reading neutral phrases. The videos were categorized into two groups: matched (robot figure with synthetic voice, human figure with human voice) and mismatched (robot figure with human voice, human figure with synthetic voice). The findings showed that a discrepancy between the authenticity of a subject's facial appearance and its corresponding voice can contribute to the emergence of the UVE. Similarly, Tinwell and Grimshaw (2009) revealed that the absence of synchronization between lip movements and voice amplifies the perception of eeriness towards an entity. Interestingly, they proposed that the dim presence of human-like characteristics in the voice produces a comparable effect. Moreover, Eyssel et al. (2012) examined the manipulation of a robot's voice, considering variations in robot gender and the use of human-like or robot-like synthesized speech. Participants in the study formed judgments about a gender-neutral robot named Flobi based solely on its vocal cues. The results unveiled a notable tendency among participants to exhibit greater anthropomorphism towards robots of the same gender, particularly when these robots possessed a human voice. However, the study focused on gender and did not explore the specific level of human likeness in the robot's voice. While these findings are important regarding the role of auditory features in the UVE, they do not explicitly demonstrate that audio alone can be the sole anthropomorphic characteristic leading to the UVE, as the presence of an entity's voice and speech is always associated with the visual parameter.

The presence of the UVE can sometimes be unclear, particularly when the stimuli are presented separately. The UVE may be less typical when the focus is solely on visual or vocal stimuli. Interestingly, the visual stimuli have not always contributed to the UVE. MacDorman et al. (2009) found a linear relationship that a computer-generated face does not necessarily become the eeriest as it becomes more human-like in appearance, contrary to the UVE. A similar and more complicated story unfolds when considering audio-only stimuli.

Certain studies have investigated the vocal UVE using isolated voice stimuli, but they have not been able to observe the expected valley pattern (Baird et al., 2018; Kühne et al., 2020). Anthropomorphism on the one hand increases if a human-like voice is used, but there is no evidence of a non-linear, UV-like pattern on the other hand. Kühne et al. (2020) examined how people perceive synthesized voices in terms of eeriness and likability. They used audio clips representing synthesized (Watson, IBM), humanoid (Sophia, Hanson Robotics), and human voices. This study also investigated how individual listener characteristics influenced evaluations of synthesized voices. The results showed that human voices and speaker characteristics received higher ratings in most aspects, except for eeriness. Ratings of synthesized voices were influenced by agreeableness and neuroticism, with females generally rating them more positively. Contrary to their expectations, voices and speaker characteristics perceived as more human-like were perceived as less eerie by participants. In addition, Baird et al. (2018) aimed to assess the likeability and human likeness of synthesized voices produced through different approaches. They evaluated a corpus of 13 German male voices and 39 utterances created using five synthesis methods, with a human control group for comparison. Their results indicated that the likeability of synthesized voices consistently improved as the human likeness increased. They concluded that the concept of vocal UVE may not apply to synthesized voices.

However, these studies had different approaches to investigating the prosody. In the study by Kühne et al. (2020), they used five clips without manipulating the pitch within each category or exploring the optimal pitch for likability. Although this study found that the human voice exhibited the most natural prosody, it remains unclear to what extent the proximity to natural prosody elicits an eerie response, especially since other categories were synthesized voices. Also, the inclusion of non-native speakers might have affected their sensitivity to voice qualities. Their qualitative analysis identified intonation, sound, and emotion as important factors in making human-like voices, suggesting that manipulating these elements may yield different results in further studies. Moreover, Baird et al. (2018) employed various synthesis methods such as formant synthesis, and concatenative diphone synthesis to create different types of artificial voices within the corpus. These synthesis methods have limitations in producing rich human prosodic characteristics, such as restricted pitch variation, limited intonation patterns, and insufficient rhythmic variability, resulting in synthetic voices that may sound monotonous. Also, using voices from only one gender may introduce biased evaluations in both previous studies. Considering the different approaches of these studies as well as the uncertainty surrounding the sole influence of prosody as the main driving factor, the potential presence of the vocal UVE remains.

The possibility of the vocal UVE occurring can be similar to when solely visual stimuli are investigated. The presence of the UVE in visual stimuli has yielded mixed results, with instances where it has been observed and others where it has not. Researchers may have been looking in the wrong direction or missed the vocal UVE due to their approaches. Furthermore, considering the importance of prosody in creating human-like voices and the rapid improvement in the quality of synthesized voices, it is plausible that the vocal UVE may not have been evident in the past. However, as synthesized voices increasingly resemble human voices, they may have the potential to elicit feelings of eeriness. Considering the importance and the intricate nature of prosody, the present study aims to delve into this aspect to uncover further details about the UVE. To date, the role of prosody in isolation from visual stimuli in the perception of uncanniness in IVAs has not been explored regarding the potential contribution of prosody to the presence of UVE. The object is to determine if there is a specific threshold based on Mori's graph (2012), which will be referred to as the *'certain point'* in the present study's hypotheses, that can elicit negative feelings like eeriness.

Although the current study aims to investigate the prosody of computer-generated speech in IVAs, it is important to acknowledge the existing limitations of computer-generated speech in achieving flawless prosody. In order to ensure the highest quality of prosody at various levels, the approach taken in the present study involves manipulating human speech to preserve its naturalness. Currently, it is easier to make human speech sound less human-like or robotic, but the technology is not yet advanced enough to adjust computer-generated speech to sound more human-like.

Through manipulation of prosodic features in speech, the present study seeks to explore the relationship between the anthropomorphism of prosody in the speech of IVAs and the UVE, using the human likeness scale of Hu et al. (2021). This scale covers diverse factors that contribute to the human likeness of speech. In order to prevent any confusion, the term human likeness (of prosody) will be used to refer to the characteristic of IVAs speech (how objectively similar the speech is to human speech), while anthropomorphism pertains to the level of perception individuals have about the speech (how much they think the speech is produced by a human-like entity). This paper tests the relationship between prosody and anthropomorphism in the following hypotheses:

H_{1a}. *The perceived level of anthropomorphism of an IVA increases as its prosody becomes more human-like (less robotic).*

H_{1b}. *The perceived level of anthropomorphism of an IVA increases as its prosody becomes more human-like, but at a certain point, the perceived anthropomorphism is lower than at the other levels of the human likeness of prosody.*

The concepts of dehumanization and robomorphism are frequently employed to define the degree of lacking human likeness attributed to an IVA. Building on this understanding, Schouten et al. (2022) proposed the notion of robomorphism which pertains to the attribution of robot-like qualities to humans, and they introduced a scale to evaluate this concept within computer-mediated interactions. With the use of their scale, the current study investigates the role of robomorphism in prosody within an IVA in relation to the UVE. The following hypotheses are proposed:

***H_{2a}*.** *The perceived level of robomorphism of an IVA decreases as its prosody becomes more human-like.*

***H_{2b}*.** *The perceived level of robomorphism of an IVA decreases as its prosody becomes more human-like, but at a certain point, the perceived robomorphism is higher than at the other levels of the human likeness of prosody.*

Trust and Eeriness

Users' initial interaction with technology involves seeking attributes that enhance trust (McKnight et al., 2002). Since trust signifies the level of confidence in computer systems, it plays a crucial role in promoting users' willingness to depend on and actively engage with computer systems (Lee & Nass, 2010). Moreover, behavioral patterns and anthropomorphic attributes inherent in IVAs play an important role in predicting the degree of trust and compliance demonstrated by individuals towards the robot (Mathur & Reigling, 2016; Natarajan & Gombolay, 2020). Due to the importance of trust, the other aim of this study is to examine the level of human likeness in prosody and its impact on the development of trust.

Trust is defined as relying on the agent to provide support and assistance in circumstances, particularly where the outcome is uncertain and the individual feels exposed or at risk (Natarajan & Gombolay, 2020). Human-like trusting beliefs, closely related to trust in technology, refer to the cognitive construct encompassing individuals' perceptual evaluations of an agent's competence (the entity's ability to perform effectively), benevolence (its intentions and motivations), and integrity (its reliability and adherence to ethical principles) (McKnight et al., 2002). Incorporating human-like attributes, such as voice or visual elements, may elicit a heightened sense of familiarity and resemblance to human interactions, prompting individuals to evaluate the technology based on their preconceived notions of competence, benevolence, and integrity, grounded in their past experiences in human-to-human interactions (Lankton et al., 2015). Their study showed that the inclusion of voice in technology potentially contributes to the formation of human-like trusting beliefs, as it adds a certain degree of human likeness. In a study conducted by Følstad et al. (2018), individuals highlighted that the trust placed in chatbots was shaped by their perceived human resemblance. While previous studies have examined the relationship between trust and anthropomorphized technology (Følstad et al., 2018; Natarajan & Gombolay, 2020; Nordheim et al., 2019), the analysis of prosody features in isolation from visual appearance requires further exploration. Consequently, the following hypothesis is formulated:

***H_{3a}*.** *The level of trust in an IVA increases as its prosody becomes more human-like.*

Furthermore, the experience of eeriness resulting from the UVE can have adverse effects on trust, as it may lead to a negative emotional state that can induce biased social judgments

(Nowak, 2006). This research suggested that increasing the human likeness of an avatar if perceived as insufficiently portraying human properties, can negatively impact its credibility. Similarly, some studies have highlighted the excessive level of anthropomorphism of agents can negatively impact trust (Mathur & Reichling, 2016; Song & Shin, 2022). The occurrence of the UVE might have adverse repercussions, as it can erode the user's trust in the IVAs. For instance, the study by Song and Shin (2022) examined how the degree of chatbot resemblance to humans influences the perception of eeriness, trust, and users' behavioral intentions within the e-commerce context. Through their experiment, they found that interacting with a hyper-realistic animated avatar intensified the feeling of eeriness, which subsequently led to a decrease in trust which can significantly influence individuals' purchase intention. To date, there has been a scarcity of investigations exploring the UVE solely in the prosody, but it may result in a diminished level of trust due to the UVE:

> ***H*$_{3b}$.** *The level of trust in an IVA increases as its prosody becomes more human-like, but at a certain point, the level is lower than at the other levels of the human likeness of prosody.*

Additionally, the relation between the feeling of eeriness and human likeness in the UVE can be described by an inverted U-shaped; the UV is, of course, a peak of eeriness (Wang & Rochat, 2017). This peak occurs when there are subtle deviations from human-like features, eliciting negative reactions. Ho and MacDorman (2010) conducted extensive research on developing a scale to measure the eerie feeling from various perspectives. With the use of their scale, this study explores the relationship between the level of human likeness in speech prosody of IVA and the level of eeriness by following the hypotheses:

H_{4a}. *The perceived level of the eeriness of an IVA decreases as its prosody becomes more human-like.*

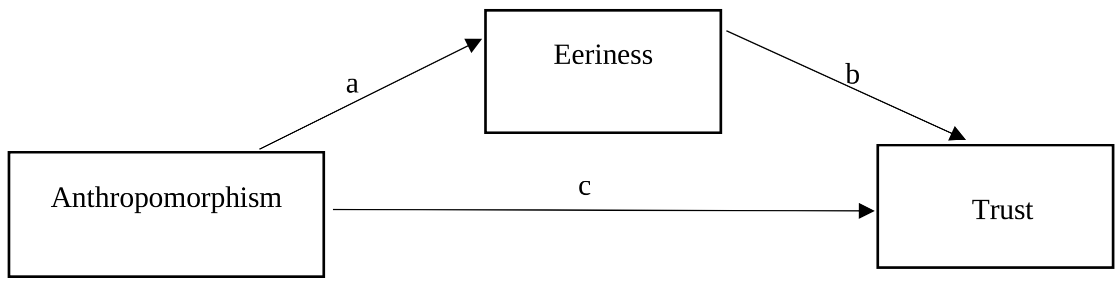
H_{4b}.*The perceived level of the eeriness of an IVA decreases as its prosody becomes more human-like, but at a certain point, the perceived eeriness is higher than at other levels of the human likeness of prosody.*

The present study's other objective is to investigate the intricate relationship between anthropomorphism, trust, and eeriness. It is hypothesized that the extent to which the prosody of IVAs is perceived as human-like (anthropomorphized) can directly influence users' level of trust. However, the role of feeling eerie comes into play as a mediator in this relationship (see Figure 2). The investigation on the mediation effect of eeriness primarily concentrates on visual stimuli within the e-commerce context and business (Grazzini et al., 2023; Shin et al., 2019). For instance, the study by Shin et al. (2019) explained that the eerie evoked by the chatbot interferes with the evaluation process of it. In their study on virtual social networking avatars, hyper-realistic avatars were found to evoke more eeriness and result in biased user perceptions compared to cartoonish avatars. This bias could negatively affect trust within the chatbot. The same effect may occur with auditory stimuli:

H_5. *Perceived eeriness plays a mediating role in the relationship between perceived anthropomorphism and trust*.

Figure 2.

Model of the effect of anthropomorphism on trust, mediated by the feeling of eeriness.



Note. Anthropomorphism has a direct effect on eeriness (a), and eeriness, in turn, impacts trust directly (b). Additionally, anthropomorphism also has a direct effect on trust (c).

General Method

Design

In this survey-based experiment, data were collected using two experiments of the same within-subjects design. The independent variable was 'Roboticness of Prosody' (RoP, i.e., a manipulated recording of a human voice) with four levels. The study's dependent variables were the perceived level of Anthropomorphism, Robomorphism, Trust, and Eeriness. Two experiments were used due to the lack of a priori knowledge regarding the certain point mentioned in the hypotheses, only that it is close to fully human-like prosody, i.e., the prosody of actual humans. The first experiment aimed to discover the point of the valley in the UVE, with the second experiment serving to find it at different levels of RoP in case Experiment 1 missed it, or to zoom in on the exact location in case Experiment 1 would show the UVE at one of its levels.

The prosody of recordings of a human voice reading the descriptions of museum paintings was manipulated to differ at four levels. The concept of the museum tour was derived from Rekkers' study (2022), but most details and the procedure were altered. A human voice was used due to the imperfect prosody currently found in computer-generated speech and the difficulty of manually making synthetic speech more human, as it was previously discussed. In both experiments, participants were presented with four IVA tour guides, always two male and two female voices. Each tour guide described one painting at a different level of RoP. There were four lists in which each painting was connected with one level of RoP, and these connections were counterbalanced across lists. Participants were randomly assigned to one of four lists. Each participant was presented with all four paintings separately, and all four levels of RoP, hence, one painting was described with a voice at one level of RoP. For example, one participant received the first painting with a male voice (male1) at level 1, the second painting with a female voice (female1) at level 2, followed by the third painting with a different male voice (male2) at level 3, and finally the fourth painting with another female voice (female2) at level 4. The rotation of the lists and the randomization of paintings with the levels of RoP are presented in Appendix A.

Experiment 1

Participants

Participants were recruited using convenience and snowball sampling methods. The online survey was sent through various social media platforms and among the researcher's personal social circles. Participants were native and non-native English speakers. A total of 45 participants were recruited with a mean age of 27.48% (*SD*= 6.1, age range:18-40). Among the participants, 22.2% identified as male, 75.6% identified as female (n_{females} =34, n_{males}= 10), and one person preferred not to disclose their gender. Also, 73.4% of participants had an advanced level of English proficiency or were native English speakers, and 82.2% of them had a bachelor's degree or a higher level of education.

Materials

The survey was constructed using Qualtrics. The voices that participants heard in the experiment were prerecorded human voices. As a precaution against gender bias, two women and with two men were selected. They were aged 25 between 35 and all were native English speakers from the United States to avoid noise in the measurements due to accent differences. The speakers were informed of the study's objectives and instructed to read the stimulus descriptions naturally. Each speaker recorded a description of one painting.

A set of four paintings by Van Gogh were presented with descriptions in one to four levels of RoP (male and female variations). These descriptions were read aloud by those prerecorded voice-operated IVA museum tours. Four Van Gogh paintings (see Figure 3) and their descriptions were obtained and modified from the official website of the Van Gogh Museum(https://www.vangoghmuseum.nl/en/collection?q=&Artist=Vincent+van+Gogh&Type=painting). Their website included a collection of Van Gogh's paintings which are publicly available online. The descriptions were standardized to be three sentences long for each painting.

Figure 3.

Four of Van Gogh's paintings

The level of RoP varied across four levels: level 1 (0% roboticness, being not robotic at all; natural human speech), level 2 (33% roboticness), level 3 (66% roboticness), and level 4 (100% robotic). To create the levels of RoP, the prosody was manipulated per sentence in the following way. First, the natural pitch decline was calculated as the regression line of pitch and time. This line represents the natural decline of the pitch, which is common for a declarative sentence, but has no other accentuation or natural pitch movements. The pitch of the human voice recording was moved towards this line, reducing the distance of the actual pitch point to the line with 0% (no change, i.e., fully human-like prosody), 33% or 66%, or 100% (most robotic, hence highest RoP). Praat 6.2.01 (Boersma & Weenink, 2021) was used to perform the manipulation, with pitch detection settings based on gender: For male voices, the pitch detection window was set at 75 Hz, and 400 Hz, while it was 110- 500 Hz for female voices.

Procedure

Participants were given an explanation of the task and were asked to provide their consent. Their objective was to select the most suitable IVA tour speech from four different options that provided information on Van Gogh's paintings. They listened to a verbal description by the IVA and subsequently answered several questions about their experience following each painting and its distinct IVA tour guide. Participants were requested to wear headphones while completing the survey to have a better auditory experience.

After seeing each painting, participants had to press the play button to listen to the three-sentence description (see Appendix B for the full script). In order to ensure that participants listened to the description, they were not allowed to proceed to the next page for evaluation until the narration had concluded.

The participants evaluated the spoken description using a series of Likert-scale questions to measure the perceived levels of Anthropomorphism, Robomorphism, Trust, and a semantic differential scale to assess the perception of Eeriness (A description of the measurement scales can be found in the Measures section below). In addition to this, participants answered demographic questions about their age, gender, level of English proficiency, and highest level of education. Afterward, they responded to an open-ended question about one of the paintings to ensure they had listened to all stimuli. However, 33% of the participants (15 individuals) provided an incorrect response to the question. Some of the participants commented that they were primarily focused on evaluating the speech itself rather than considering the content and, though they attentively listened, failed to provide the correct answer. Also, the results of these 15 participants were examined separately, and their responses did not notably deviate from the other participants (see Appendix C for the result). Consequently, this question was not used to exclude participants. Finally, the participants were given a debriefing and thanked for their time and participation.

Measures

After each painting, a questionnaire was measured to assess several dependent variables, namely anthropomorphism, robomorphism, trust, and eeriness. The subsequent sections provide further information about each construct. All variables had acceptable reliability values (see Table 1). The details of each measurement scale are provided in Appendix D.

Anthropomorphism

The scale developed by Hu et al. (2021) was used to assess the degree of humanness present in computer-generated speech. It consisted of five items to measure the naturalness of pronunciation and the degree of similarity to human-like language expression and speech. Participants were asked to rate these items on a 5-point Likert scale (1 = *strongly disagree* to 5 =

strongly agree). The average score of the five items was calculated to provide an overall measure of participants' responses.

Robomorphism

Schouten et al. (2022) introduced a scale to assess the degree of perceived robomorphism. The present study adapted the scale and one item was excluded, as it did not apply. The scale comprised four items, each rated on a 5-point Likert scale. The scale's response options ranged from 1, representing strong disagreement, to 5, indicating strong agreement. The respondents' score was determined by calculating the average of the four items.

Trust

Participants' trust in the IVA was measured using a 5-point Likert-type scale (1=*completely disagree*; 5=*completely agree*), which consisted of five items. This was adopted from a previous study conducted by Nordheim (2018). The respondents' score was obtained by averaging the four items.

Eeriness

The IVA's tour speech was evaluated for its eerie feeling using three 5-point semantic differential scale items adapted from a study by Ho and MacDorman (2010). The original scale consists of eight items, of which three items were selected for the present study due to their effectiveness in studying eerie feelings, as demonstrated in Song and Shin's research (2022). The participants' score was measured by taking the average of the three items.

Table 1.

Cronbach's Alpha Values for Dependent Variables

Dependent Variables	Experiment 1
Anthropomorphism	.78
Robomorphism	.74
Trust	.73
Eeriness	.76

Results

Correlation of the Dependent Variables

Table 4 shows correlations between the four dependent variables. Bootstrapping was used for certain levels at which the data were not normally distributed (see Table 2). As the correlations observed between the dependent variables were significant (see Table 3), the variables were possibly measuring similar underlying concepts (see Appendix E for the complete report). Hence, significance values for groups of hypotheses that only differ in their dependent variable should be interpreted together and with caution to avoid spurious effects.

Table 2

z-scores of Skewness and Kurtosis for Dependent Variables at all levels of RoP

Variables	*M.*	*SD.*	*z*-scores skewness	*z*-scores kurtosis
Anthropomorphism level 1 of RoP	3.9	0.6	0.6	-1.1
Anthropomorphism level 2 of RoP	3.7	1.0	**-2.82**	0.62

Anthropomorphism level 3 of RoP	3.2	1.1	-0.42	-1.52
Anthropomorphism level 4 of RoP	2.3	1.0	**2.00**	-0.29
Robomorphism level 1 of RoP	2.0	0.9	1.6	-1.00
Robomorphism level 2 of RoP	2.2	1.0	**3.42**	1.28
Robomorphism level 3 of RoP	2.8	1.2	1.25	-1.80
Robomorphism level 4 of RoP	3.7	1.1	-1.88	-1.08
Trust level 1 of RoP	3.7	0.6	0.77	-0.76
Trust level2 of RoP	3.6	0.7	-0.34	-0.84
Trust level 3 of RoP	3.3	0.8	0.54	-1.04
Trust level 4 of RoP	2.7	0.8	0.57	0.26
Eeriness level 1 of RoP	2.3	0.8	0.57	-0.96
Eeriness level 2 of RoP	2.4	0.8	**2.4**	1.95
Eeriness level 3 of RoP	2.8	0.9	1.11	-0.20
Eeriness level 4 of RoP	3.4	0.9	1.4	-0.98

Table 3

Pearson's Correlation Values for Dependent Variables at all levels of RoP

	Robomorphism level 1 of RoP	Robomorphism level 2 of RoP	Robomorphism level 3 of RoP	Robomorphism level 4 of RoP
Anthropomorphism level 1 of RoP	-.75			
Anthropomorphism level 2 of RoP		-.85 [-0.94, -0.73]		
Anthropomorphism level 3 of RoP			-.94	
Anthropomorphism level 4 of RoP				-.90 [-0.95, -0.83]
	Trust level 1 of RoP	Trust level 2 of RoP	Trust level 3 of RoP	Trust level 4 of RoP
Anthropomorphism level 1 of RoP	.60			
Anthropomorphism level 2 of RoP		.33 [-0.02, 0.63]		
Anthropomorphism level 3 of RoP			.40	
Anthropomorphism level 4 of RoP				.52 [0.28, 0.74]
	Eeriness level 1 of RoP	Eeriness level 2 of RoP	Eeriness level 3 of RoP	Eeriness level 4 of RoP
Anthropomorphism level 1 of RoP	-.41			
Anthropomorphism level 2 of RoP		-.60 [-0.80, -0.32]		
Anthropomorphism level 3 of RoP			-.60	
Anthropomorphism level 4 of RoP				-.55 [-0.74, -0.32]

	Trust level 1 of RoP	Trust level 2 of RoP	Trust level 3 of RoP	Trust level 4 of RoP
Robomorphism level 1 of RoP	-.52			
Robomorphism level 2 of RoP		-.36 [-0.59, -0.08]		
Robomorphism level 3 of RoP			-.38	
Robomorphism level 4 of RoP				-.46
	Eeriness level 1 of RoP	Eeriness level 2 of RoP	Eeriness level 3 of RoP	Eeriness level 4 of RoP
Robomorphism level 1 of RoP	.54			
Robomorphism level 2 of RoP		.64 [0.43, 0.80]		
Robomorphism level 3 of RoP			.56	
Robomorphism level 4 of RoP				.52
	Trust level 1 of RoP	Trust level 2 of RoP	Trust level 3 of RoP	Trust level 4 of RoP
Eeriness level 1 of RoP	-.53			
Eeriness level 2 of RoP		-.35 [-0.56, -0.12]		
Eeriness level 3 of RoP			-.60	
Eeriness level 4 of RoP				-.49

Note. All of the correlations were significant ($p < .05$).

Analyses

Linear regression analyses were conducted to test $H_{1\text{-}4a}$, while repeated-measures ANOVAs were conducted for $H_{1\text{-}4b}$, both using SPSS. Finally, a mediation analysis was performed to investigate H_5. As the UVE was expected to occur just before natural human prosody, the second level of RoP is used as the certain point in the hypotheses. To ensure the effect was not missed, level 3 was also compared to the other levels, considering the possibility that level 2 may be too close to natural human prosody.

The normality assumption was violated for the dependent variables, Anthropomorphism, Robomorphism, and Eeriness, (see *z*-scores of skewness and kurtosis in Table 4).

Table 4

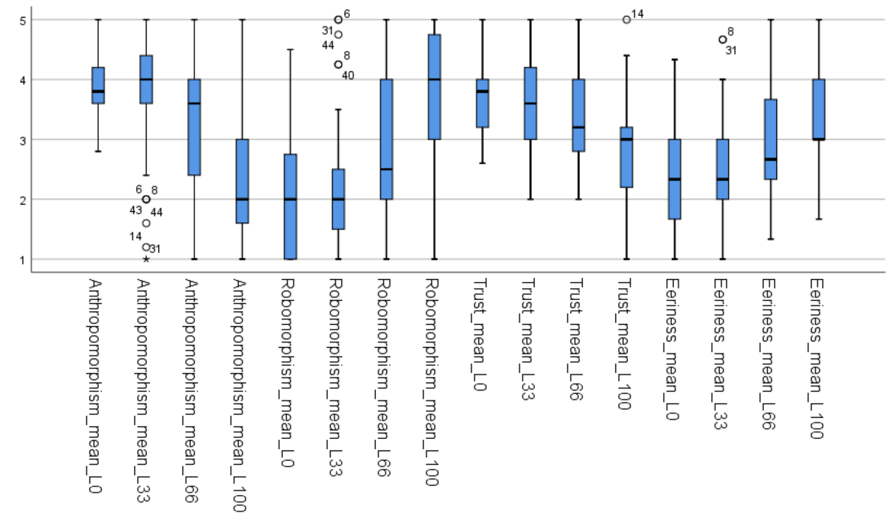
z-scores of Skewness and Kurtosis for all Dependent Variables

	M.	*SD.*	*z*-scores skewness	*z*-scores kurtosis
Anthropomorphism	3.3	1.1	-4.25	-4.80
Robomorphism	2.7	1.3	3.69	-6.27
Trust	3.4	0.8	-1.17	-1.92
Eeriness	2.8	0.9	5.17	0.02

A total of 14 outliers were identified including six for Anthropomorphism, five for Robomorphism, and two for Eeriness at level 2; as well as a single outlier for Trust at level 4 (see Figure 4). The outliers warrant further discussion since most of them occurred at the identified certain point at level 2. Moreover, certain participants (e.g., participants 8 and 31) consistently appeared as outliers for Anthropomorphism, Robomorphism, and even Eeriness at level 2. Given the outliers observed in the data did not necessarily indicate measurement errors, they were not excluded from the analysis.

Figure 4

Outliers at Each Level of RoP



Linear Regression

Four linear regression analyses were performed to test $H_{1\text{-}4a}$, one for each dependent variable with RoP as the independent variable (M = 2.5, SD = 1.1). All the coefficients were based on a scale ranging from one to four of RoP. The results of all four regression analyses were significant.

The result of the analysis with Anthropomorphism by RoP showed that the residuals were normally distributed ($z_{skeweness}$ = -0.05, $z_{kurtosis}$ = -1.45), and as a visual inspection showed the assumption of homoscedasticity was met. RoP significantly predicts Anthropomorphism (see the result in Table 5), indicating that for one step (33%) on the RoP scale, the Anthropomorphism answer changed by -0.53 on the Likert scale (see Figure 5). The model explained 27% of the variance in the number of data shared. In conclusion, the data supported H_{1a}.

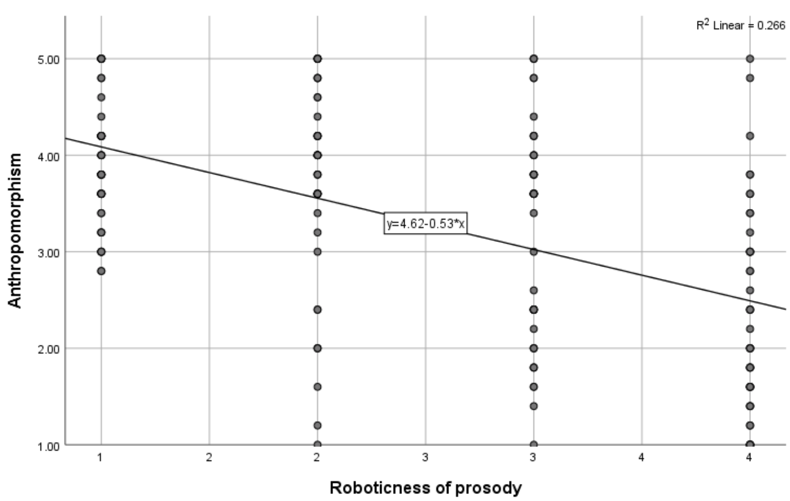
Table 5

Regression Result of Anthropomorphism by RoP

	b	*SE*	β	*t*	*F*	R^2	*DF*	*p*
Anthropomorphism by RoP	-0.53	.07	-0.52	-8.03	64.52	.27	178	< .000

Figure 5

Simple Scatterplot of Anthropomorphism by RoP



The regression of Robomorphism by RoP had residuals that were not normally distributed ($z_{\text{skeweness}}$ = 0.70, z_{kurtosis} = -2.17). Bootstrapping was performed. The assumption of homoscedasticity was met according to visual inspection. The regression analysis indicated that

RoP significantly predicted Robomorphism (see the result in Table 6), indicating that for one step (33%) on the RoP scale, the Robomorphism answer changed by 0.57 on the Likert scale (see Figure 6). The model explained 24% of the variance in the number of data shared. Since the bootstrapped 95% confidence interval did not include zero, H_{2a} was supported.

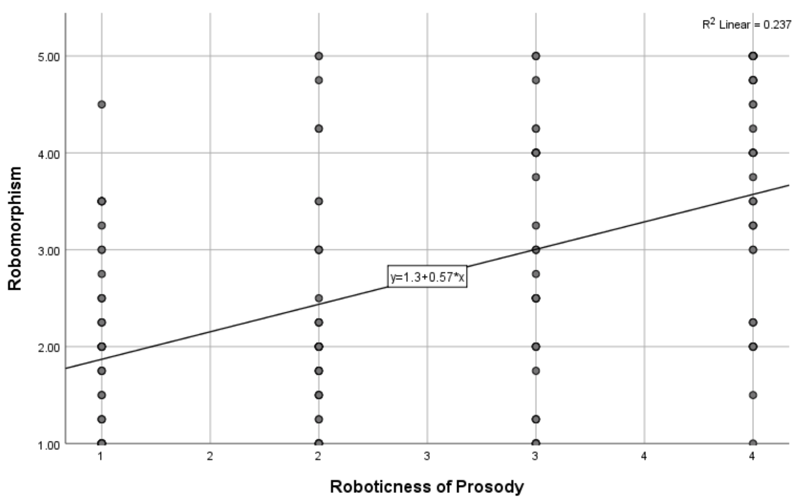
Table 6

Regression Result of Robomorphism by RoP

	b	*SE*	β	*t*	*F*	R^2	*DF*	*p*	95%CI	
									LL	UL
Robomorphism by RoP	0.57	.07	.50	7.44	55.33	.24	178	< .000	0.42	0.70

Figure 6

Simple Scatterplot of Robomorphism by RoP



The analysis results using Trust by RoP revealed that the residuals were normally distributed ($z_{skeweness}$ = 0.83, $z_{kurtosis}$ = -1.08), and as visual inspection showed the assumption of homoscedasticity was met. The regression analysis indicated that RoP significantly predicted Trust (see the result in Table 7), suggesting that for one step (33%) on the RoP scale, the Trust answer changed by -0.31 on the Likert scale (see Figure 7). The model accounted for 17% of the variance in the number of data shared. Hence, H_{3a} was supported.

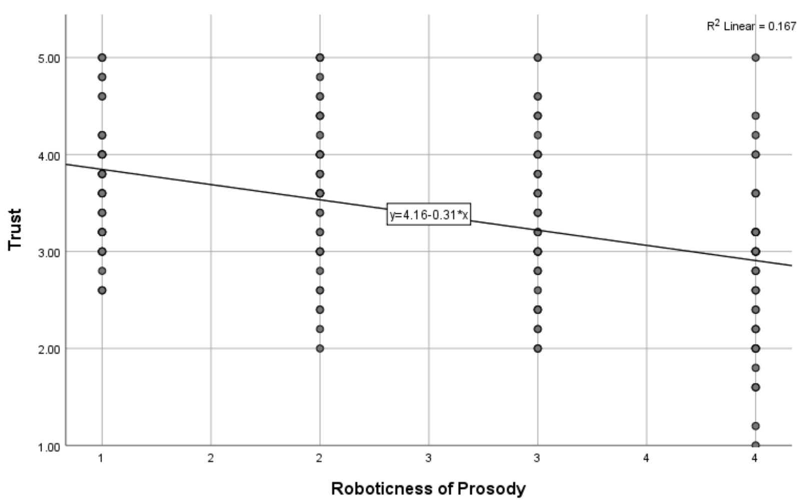
Table 7

Regression Result of Trust by RoP

	b	*SE*	β	*t*	*F*	R^2	*DF*	*p*
Trust by RoP	-0.31	.05	-0.41	-6.00	35.74	.17	178	< .000

Figure 7

Simple Scatterplot of Trust by RoP



Finally, the analysis results using Eeriness by RoP indicated that the residuals were not normally distributed ($z_{skeweness} = 2.36$, $z_{kurtosis} = -0.80$). Bootstrapping was conducted. The assumption of homoscedasticity was met according to a visual inspection. The regression analysis revealed that RoP significantly predicted Eeriness (see the result in Table 8), suggesting that for one step (33%) on the RoP scale, the Eeriness answer changed by 0.37 on the Likert scale (see Figure 8). The model accounted for 19% of the variance in the number of data shared. Since the bootstrapped 95% confidence interval did not cross zero, H_{4a} was supported.

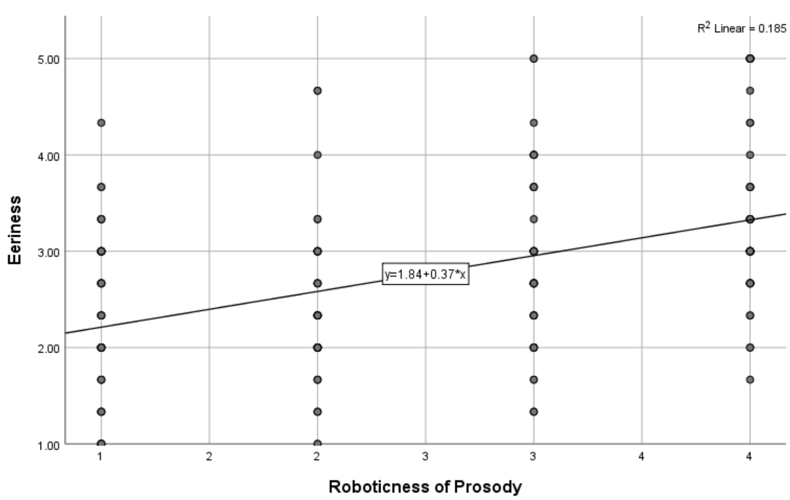
Table 8

Regression Result of Eeriness by RoP

	b	*SE*	β	*t*	*F*	R^2	*DF*	*p*	95%CI	
									LL	UL
Eeriness by RoP	0.37	.04	.43	6.35	40.32	.19	178	< .000	0.25	0.45

Figure 8

Simple Scatterplot of Eeriness by RoP



All of $H_{1\text{-}4a}$ were supported by the data. Expanding on the previous cautionary note regarding the interpretation of these correlated dependent variables, they essentially measured the same underlying construct. Hence, it is plausible that the significant results could be attributed to a latent variable exerting simultaneous influence on all four variables. In any case, the correlation among the dependent variables corresponded to the same effect four times, eliminating the possibility of a spurious effect on one of the four DVs.

Repeated-Measures Analyses of Variance

Repeated-measures ANOVAs were conducted to test $H_{1\text{-}4b}$. The normality assumption was violated, but unfortunately, bootstrapping is not implemented in SPSS's repeated-measures ANOVA. Therefore, checking the confidence intervals for each variable to determine any overlaps was a solution. Overlapping intervals suggestes a lack of differences. In such cases, it is important to interpret the results with more caution. Additionally, the assumption of sphericity was checked by Mauchly's Test. The *p*-values for Mauchly's test were all above .05, indicating no significant violation (see Table 9). There was a significant effect of RoP on all four dependent variables (see the result in Table 10).

Table 9

Result of Mauchly's Test of Sphericity

Within Subjects Effect	Measure	Mauchly's *W*	Approx. χ^2	*df*	*P*
Level	Anthropomorphism	0.97	1.37	5	.928
	Robomorphism	0.90	4.43	5	.490
	Trust	0.84	7.44	5	.190
	Eeriness	0.79	9.78	5	.082

Table 10

Analyses of Variance Using Repeated Measures ANOVA Assuming Sphericity

	M^2	F (3, 44)	p	η^2
Anthropomorphism	23.04	25.62	<.000	.37
Robomorphism	26.44	22.56	<.000	.34
Trust	8.35	16.35	<.000	.27
Eeriness	11.27	14.05	<.000	.24

Then we conducted a deviation contrast between the dependent variables, with level 2(33%) and level 3 (66%) each contrasted against the mean of the other three levels. The contrast comparing level 2 with the mean of others was found significant for all dependent variables. Although statistically significant differences were found, it was not in a predicted direction. There was no significant difference between level 3 and the mean of others (see the result in Table 11).

Table 11

Result of the Deviation Contrast

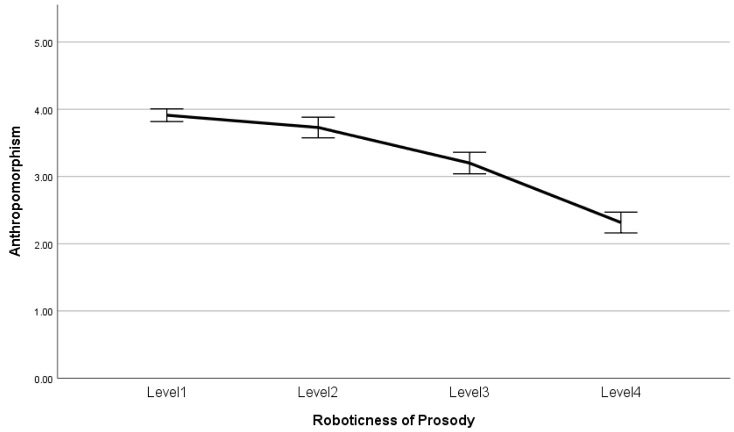
Dependent variables	Level of RoP	M^2	F (3, 44)	p	η^2
Anthropomorphism	level 2 vs. Mean	8.71	12.03	**<.001**	.21
	level 3 vs. Mean	0.36	0.48	.490	.01
Robomorphism	level 2 vs. Mean	11.19	10.95	**<.002**	.20

	level 3 vs. Mean	0.51	0.49	.487	.01
Trust	level 2 vs. Mean	3.90	8.96	**<.004**	.17
	level 3 vs. Mean	0.07	0.17	.680	.00
Eeriness	level 2 vs. Mean	5.63	10.38	**<.002**	.19
	level 3 vs. Mean	0.57	.93	.341	.02

Since the expected valley pattern was not observed, it suggested that there was no significant effect to be concerned about the violation of normality assumptions (While there was no need to refer to the confidence intervals, they are available in Appendix F for anyone seeking consultation). The mean values for each variable across the four levels appeared to follow a linear change, which means hypotheses H_{1b}, H_{2b}, H_{3b}, and H_{4b} were not supported; level 2 is never higher (or lower) than the surrounding levels (see Figures 9.1-9.4). There was no evidence for the presence of the UVE. See also the estimated marginal means in Table 12.

Figure 9.1

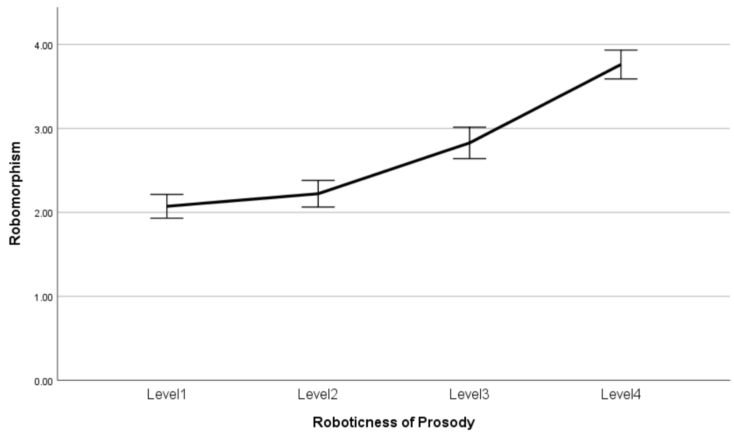
Anthropomorphism Across Levels of RoP



Note. Error Bars represent 95% confidence intervals.

Figure 9.2

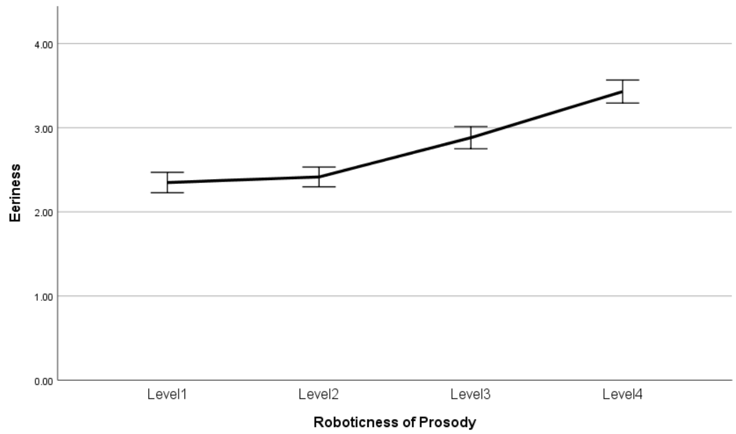
Robomorphism Across Levels of RoP



Note. Error Bars represent 95% confidence intervals.

Figure 9.3

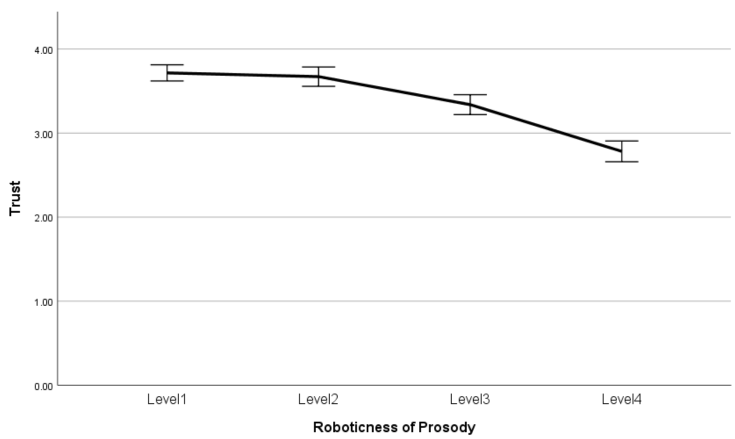
Eeriness Across Levels of RoP



Note. Error Bars represent 95% confidence intervals.

Figure 9.4

Trust Across Levels of RoP



Note. Error Bars represent 95% confidence intervals.

Table 12

Estimated Marginal Means

	Level of RoP	*M*	*SD*	95%CI	
				LL	*UL*
Anthropomorphism	1	3.9	0.1	3.72	4.11
	2	3.7	0.1	3.41	4.05
	3	3.2	0.1	2.87	3.53
	4	2.3	0.1	1.10	2.63
Robomorphism	1	2.1	0.1	1.78	2.36
	2	2.2	0.1	1.90	2.55
	3	2.8	0.2	2.44	3.21
	4	3.8	0.2	3.41	4.11
Trust	1	3.7	0.1	3.52	3.91
	2	3.6	0.1	3.43	3.91
	3	3.3	0.1	3.10	3.58
	4	2.8	0.1	2.53	3.04
Eeriness	1	2.3	0.1	2.10	2.60
	2	2.4	0.1	2.17	2.66
	3	2.8	0.1	2.61	3.15
	4	3.4	0.1	3.15	3.71

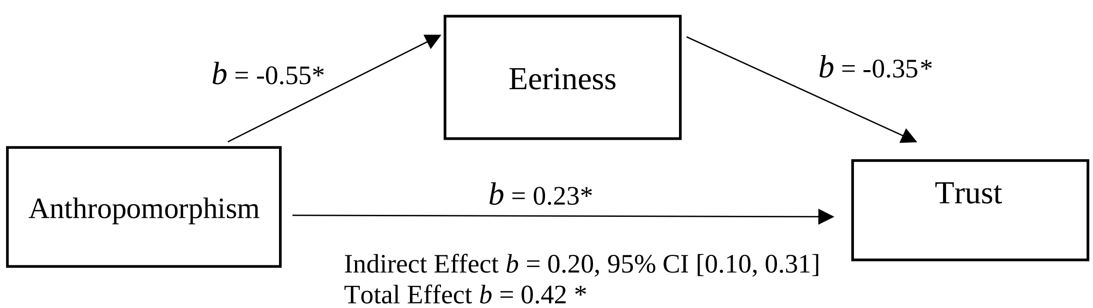
Mediation Analysis

A mediation analysis was conducted to explore the connection between Anthropomorphism and Trust, mediated by Eeriness. The analysis employed the PROCESS macro for SPSS (Model 4) by Hayes (2023), incorporating bootstrapping (with 5000 samples.

Anthropomorphism had a significant effect on Eeriness, b = -0.55, t (178) = -11.60, p <.000, accounting for 43% of the variability in Eeriness R^2 =0.43, F (1, 178) =134.46. Moreover, there was a significant effect of Eeriness on Trust b = -0.35, t (177) =-5.10, p <.000 explaining 41% of the variability in trust, R^2 =0.41, F (2, 177) = 61.10. Also, Anthropomorphism had a significant effect on Trust b = 0.23, p < .000, explaining 41% of the variability in Trust R^2 =0.41, F (2,177) =61.10. There was a significant indirect effect of Anthropomorphism on Trust b = 0.20, SE = 0.05, 95% CI [0.10, 0.31], supporting H_5. The total effect of Anthropomorphism and Eeriness on Trust was 0.42, p<.000. See Figure 10 for visualization.

Figure 10

Meditation Analysis Summary



Note. * p<.05.

Discussion

The first experiment aimed to identify the certain point in prosody that leads to the UVE and elicits eeriness as level 2 of the manipulated prosody. Additionally, the study explored whether the level of trust is influenced by anthropomorphized computer-generated speech prosody and whether eeriness affects trust resulting from human likeness perception.

Experiment 1 showed a linear relationship between RoP and the four dependent variables; Anthropomorphism, Robomorphism, Eeriness, and Trust, as predicted in $H_{1\text{-}4a}$. While the level of roboticness in the prosody increased, participants perceived higher levels of robomorphism and eeriness but lower levels of anthropomorphism and trust. However, the data did not align with the expected valley pattern of the UVE as hypothesized in $H_{1\text{-}4b}$. There was a significant mediation role of Eeriness between Anthropomorphism and Trust. The supported hypothesis H_5 suggests that as the perceived human likeness decreased, feelings of eeriness increased, leading to decreased levels of trust.

Experiment 1 also had 14 outliers. Interestingly, these outliers were consistent for both Anthropomorphism and Robomorphism, as well as Eeriness, which means they were the same participants at the designated level 2 of RoP. This suggests that these particular participants differ from others in their perceptions at level 2 of the human likeness of prosody. These participants may be more sensitive to imperfections in the human likeness of prosody, and consequently, the UVE may occur for them.

If there is the UVE for prosody, the failure to find it in Experiment 1 may be due to two different reasons. The first reason may be that the location of the UVE is between levels 1 and 2 (0-33% roboticness). This reason can be explored by testing additional levels between 0% and 33% of RoP. The second possible reason is that the effect occurs at a higher level and that the increase in eeriness from level 1 to 4 is the start of the valley and that level 4 might be close to a peak of eeriness. If so, a further increase in roboticness beyond level 4 might uncover the true peak.

Although both possibilities seemed worthy of investigation, considering the constraint of time, the decision was made to continue with the second one. This involves stronger manipulations of roboticness, with a focus on removing more human-like aspects of the prosody.

Experiment 2

Participants

Participants were recruited in the same way as for experiment 1. A total of 50 participants initially took part. However, seven participants did not complete the survey and were subsequently excluded. The final dataset consisted of 43 participants, with a mean age of 28.81% (*SD*=8.9, age range:18-60). Among these participants, an equal distribution of 48.8% identified as male/ female, and one person was non-binary. In terms of English proficiency, only 28% had an intermediate or upper intermediate level, while the remaining majority had a higher level or were native speakers. Also, merely 9% possessed an education level lower than a bachelor's degree.

Materials

The materials were constructed in the same way as for Experiment 1 except for the manipulation. The change was now that the pitch line to which the natural pitch was reduced was not the linear regression but the mean, and the manipulation therefore now also removed the natural decline of pitch over time.

Procedures

The open-ended question about one of the paintings was removed, otherwise, the procedure was the same as in Experiment 1.

Measures

The questionnaire used to evaluate the four dependent variables for each painting was the same, and all variables showed acceptable reliability values (see Table 13).

Table 13.

Cronbach's Alpha Values for Dependent Variables

Dependent Variables	Experiment 2
Anthropomorphism	.79
Robomorphism	.80
Trust	.71
Eeriness	.74

Results

Correlation of the Dependent Variables

The correlations between the four dependent variables were again tested, with bootstrapping for the levels with non-normal data distribution (see Table 14). According to the result (see Table 15), the correlations observed between the dependent variables were significant, suggesting that these variables might measure similar underlying concepts (see Appendix G for the full report). Hence, significance values for groups of hypotheses that only differed in their dependent variable should be again as Experiment 1 interpreted together and with caution to avoid spurious effects.

Table 14

z-scores of Skewness and Kurtosis for Dependent Variables at all levels of RoP

	M.	*SD.*	*z*-scores Skewness	*z*-scores Kurtosis
Anthropomorphism level 1 of RoP	3.6	0.7	-0.94	-0.80
Anthropomorphism level 2 of RoP	3.6	0.8	-1.77	0.58
Anthropomorphism level 3 of RoP	3.0	0.9	-0.19	-1.27
Anthropomorphism level 4 of RoP	1.6	0.6	**2.80**	0.34
Robomorphism level 1 of RoP	2.2	0.9	**2.69**	1.07
Robomorphism level 2 of RoP	2.2	0.9	1.32	-1.25
Robomorphism level 3 of RoP	2.9	0.9	-0.75	-1.24
Robomorphism level 4 of RoP	4.4	0.6	**-3.19**	0.69
Trust level 1 of RoP	3.4	0.6	-1.94	0.53
Trust level 2 of RoP	3.5	0.6	-1.59	-0.16
Trust level 3 of RoP	3.1	0.7	-0.08	-1.20

Trust level 4 of RoP	2.6	0.9	0.75	-0.28
Eeriness level 1 of RoP	2.2	0.7	-0.40	-1.34
Eeriness level 2 of RoP	2.5	0.8	0.82	-0.16
Eeriness level 3 of RoP	2.7	0.8	0.14	-1.51
Eeriness level 4 of RoP	3.6	0.9	-0.28	-1.24

Table 15

Pearson's Correlation Values for Dependent Variables at all levels of RoP

	Robomorphism level 1 of RoP	Robomorphism level 2 of RoP	Robomorphism level 3 of RoP	Robomorphism level 4 of RoP
Anthropomorphism level 1 of RoP	-.72 [-0.84, -0.55]			
Anthropomorphism level 2 of RoP		-.80		
Anthropomorphism level 3 of RoP			-.81	
Anthropomorphism level 4 of RoP				-.70 [-0.88, -0.36]
	Trust level 1 of RoP	Trust level 2 of RoP	Trust level 3 of RoP	Trust level 4 of RoP
Anthropomorphism level 1 of RoP	.54			
Anthropomorphism level 2 of RoP		.54		

Anthropomorphism level 3 of RoP			.68	
Anthropomorphism level 4 of RoP				.41 [0.20,0.64]
	Eeriness level 1 of RoP	Eeriness level 2 of RoP	Eeriness level 3 of RoP	Eeriness level 4 of RoP
Anthropomorphism level 1 of RoP	-.38			
Anthropomorphism level 2 of RoP		-.40		
Anthropomorphism level 3 of RoP			-.67	
Anthropomorphism level 4 of RoP				-.43 [-0.65, -0.16]
	Trust level 1 of RoP	Trust level 2 of RoP	Trust level 3 of RoP	Trust level 4 of RoP
Robomorphism level 1 of RoP	-.60			
Robomorphism level 2 of RoP		-.53		
Robomorphism level 3 of RoP			-.64	
Robomorphism level 4 of RoP				-.32
	Eeriness level 1 of RoP	Eeriness level 2 of RoP	Eeriness level 3 of RoP	Eeriness level 4 of RoP
Robomorphism level 1 of RoP	.30			
Robomorphism level 2 of RoP		.54		
Robomorphism level 3 of RoP			.74	
Robomorphism level 4 of RoP				.47 [0.23,0.70]

	Trust level 1 of RoP	Trust level 2 of RoP	Trust level 3 of RoP	Trust level 4 of RoP
Eeriness level 1 of RoP	-.54			
Eeriness level 2 of RoP		-.53		
Eeriness level 3 of RoP			-.72	
Eeriness level 4 of RoP				-.55

Note. All correlations were Significant ($p < .05$).

Analyses

To test $H_{1\text{-}4a}$, linear regression analyses were performed, while $H_{1\text{-}4b}$ were examined using repeated-measures ANOVAs both in SPSS. Finally, a mediation analysis was carried out for H_5.

The assumption of normality was violated for the dependent variables, Anthropomorphism, Robomorphism, and Trust (see *z*-scores of skewness and kurtosis in Table 16).

Table 16

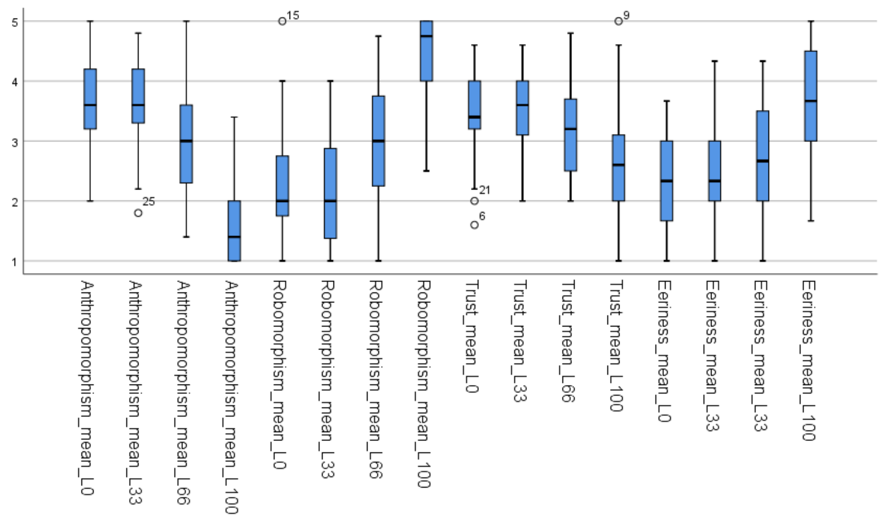
z-scores of Skewness and Kurtosis for all Dependent Variables

	M.	*SD.*	*z*-scores skewness	*z*-scores kurtosis
Anthropomorphism	3.0	1.2	-1.23	-2.92
Robomorphism	3.0	1.3	-0.80	-3.30
Trust	3.2	1.0	-2.58	-0.54
Eeriness	3.0	1.0	1.68	-1.20

A total of five outliers were detected in the dataset, with only one outlier observed specifically for level 2 (see Figure 11). As the presence of outliers in the data did not necessarily imply measurement errors, they were not excluded from the analysis.

Figure 11

Outliers at Each Level of RoP



Linear Regression

Four linear regression analyses were conducted to examine $H_{1\text{-}4a}$, one for each dependent variable with RoP as the independent variable (M = 2.5, SD = 1.1). All the coefficients were based on a scale of four RoP. The results of all four regression analyses were significant.

The result of the analysis with Anthropomorphism by RoP indicated that the residuals were normally distributed ($z_{\text{skeweness}}$ = 0.70, z_{kurtosis} = -1.72), and according to a visual inspection, the assumption of homoscedasticity was met. RoP significantly predicted

Anthropomorphism (see the result in Table 17), suggesting that for one step (33%) on the RoP scale, the Anthropomorphism answer changed by -0.67 on the Likert scale (see Figure 12). The model accounted for 43% of the variance in the number of data shared. Hence, the data supported H_{1a}.

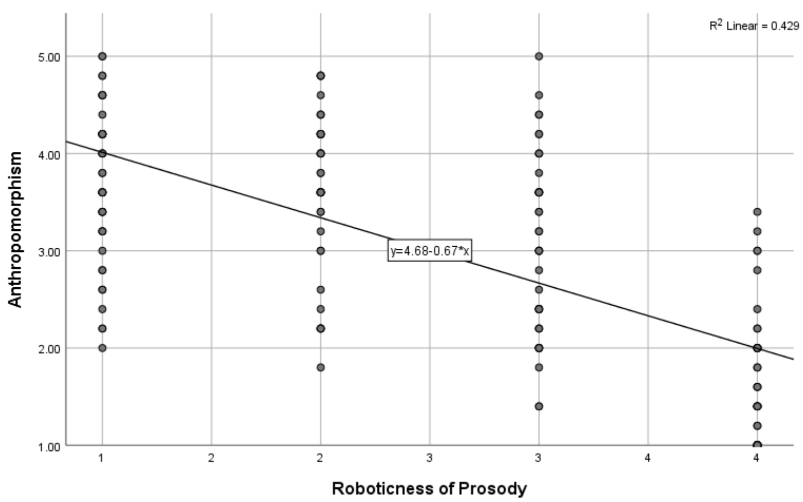
Table 17

Regression Result of Anthropomorphism by RoP

	b	*SE*	β	*t*	*F*	R^2	*DF*	*p*
Anthropomorphism by RoP	-0.67	.06	-0.65	-11.30	127.60	.43	170	< .000

Figure 12

Simple Scatterplot of Anthropomorphism by RoP



The regression of Robomorphism by RoP had residuals normally distributed ($z_{skeweness}$ = -0.14, $z_{kurtosis}$ = -0.35), and the assumption of homoscedasticity was met according to a visual inspection. The regression analysis showed that RoP significantly predicted Eeriness (see the result in Table 18), suggesting that for one step (33%) on the RoP scale, the Robomorphism answer changed by 0.74 on the Likert scale (see Figure 13). The model accounted for 43% of the variance in the number of data shared. H_{2a} was supported.

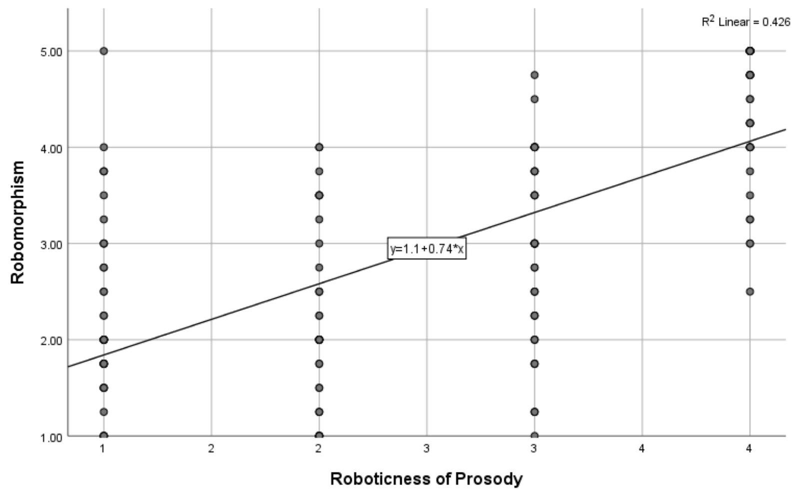
Table 18

Regression Result of Robomorphism by RoP

	b	*SE*	β	*t*	*F*	R^2	*DF*	*p*
Robomorphism by RoP	0.74	.07	.65	11.24	126.30	.43	170	< .000

Figure 13

Simple Scatterplot of Robomorphism by RoP



The analysis results using Trust by RoP revealed that the residuals were normally distributed ($z_{skeweness}$ = -0.84, $z_{kurtosis}$ = -0.03), and the assumption of homoscedasticity was met as visual inspection showed. The regression analysis showed that RoP significantly predicted Trust (see the result in Table 19), indicating that for one step (33%) on the RoP scale, the Trust answer changed by -0.27 on the Likert scale (see Figure 14). The model explained 13% of the variance in the number of data shared. The data supported H_{3a}.

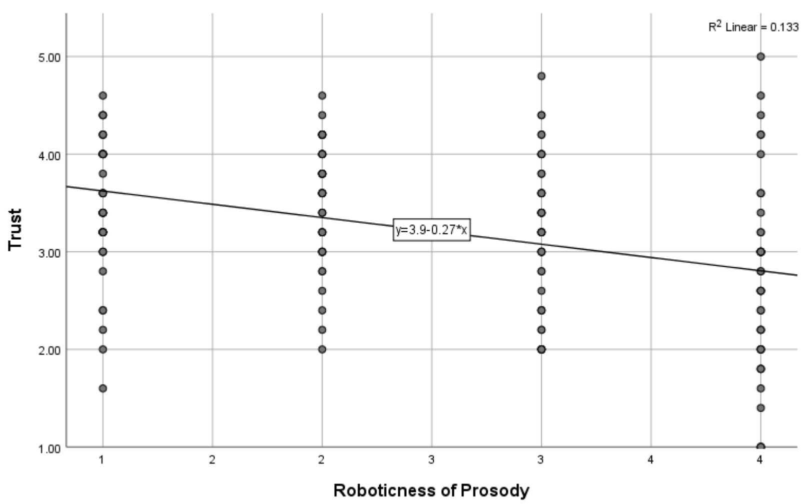
Table 19

Regression Result of Trust by RoP

	b	*SE*	β	*t*	*F*	R^2	*DF*	*p*
Trust by RoP	-0.27	.05	-0.37	-5.10	26.05	.13	170	< .000

Figure 14

Simple Scatterplot of Trust by RoP



Finally, the analysis results using Eeriness by RoP showed that the residuals were not normally distributed ($z_{skeweness}$ = 0.13, $z_{kurtosis}$ = -2.00). Bootstrapping was performed. The assumption of homoscedasticity was met according to a visual inspection. The regression analysis revealed that RoP significantly predicted Eeriness (see the result in Table 20), indicating that for one step (33%) on the RoP scale, the Eeriness answer changed by 0.43 on the Likert scale (see Figure 15). The model explained 23% of the variance in the number of data shared. Since the bootstrapped 95% confidence interval did not include zero, the data supported H_{4a}.

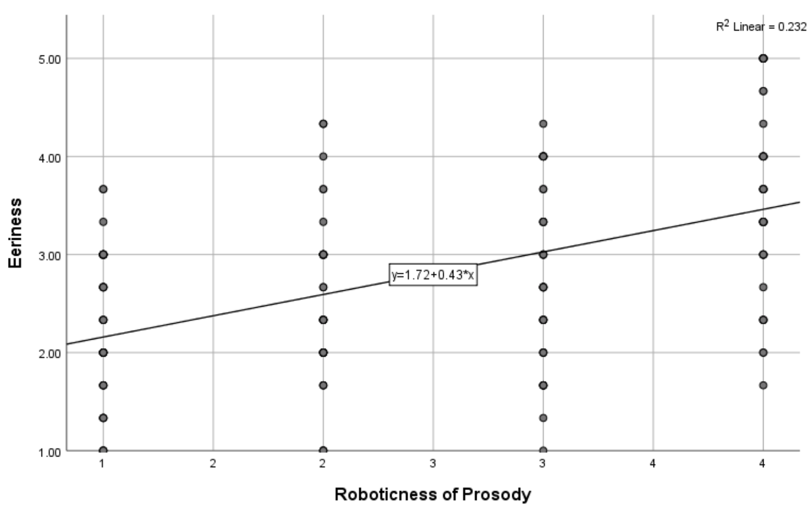
Table 20

Regression Result of Eeriness by RoP

	b	*SE*	β	*t*	*F*	R^2	*DF*	*p*	95%CI	
									LL	UL
Eeriness by RoP	0.43	.06	.48	7.16	51.34	.23	170	< .000	0.31	0.55

Figure 15

Simple Scatterplot of Eeriness by RoP



All four hypotheses ($H_{1\text{-}4a}$) were supported. Building upon the earlier reminder to interpret the correlations between these dependent variables with caution, these variables essentially captured the same underlying concept. Therefore, it is plausible that these findings may be explained by a latent variable that impacted all four variables simultaneously. In any case, the correlation between the dependent variables should correspond to the same effect four times, removing the possibility of a spurious effect on one of the four DVs.

Repeated-Measures Analysis of Variance

Repeated-measures ANOVAs were performed to test $H_{1\text{-}4b}$. The normality assumption was violated, but unfortunately, bootstrapping is not implemented in SPSS's repeated-measures ANOVA. Therefore, checking the confidence intervals for each variable to assess potential overlaps was considered as a solution. The presence of overlapping intervals indicates a lack of difference, and in such cases, it is necessary to interpret the results with caution. Also, the assumption of sphericity was assessed using Mauchly's Test. Since the *p*-values for Mauchly's test were all above .05, suggesting no significant violation (see Table 21). There was a significant effect of RoP on all four dependent variables (see the result in Table 22).

Table 21

Result of Mauchly's Test of Sphericity

Within Subjects Effect	Measure	Mauchly's *W*	Approx. χ^2	*df*	*P*
Level	Anthropomorphism	0.92	3.55	5	.616
	Robomorphism	0.84	7.13	5	.211
	Trust	0.92	3.55	5	.615
	Eeriness	0.83	7.33	5	.197

Table 22

Analysis of Variance Using Repeated Measures ANOVA Assuming Sphericity

	M^2	F (3, 42)	p	η^2
Anthropomorphism	39.12	56.40	<.000	.57
Robomorphism	47.38	58.24	<.000	.58
Trust	6.44	12.70	<.000	.23
Eeriness	15.23	19.30	<.000	.31

Subsequently, a deviation contrast was conducted between the dependent variables, examining the contrasts between level 2 (33%) and level 3 (66%) against the mean of the other three levels. The contrast comparing level 2 with the mean of others yielded significant results for all dependent variables. Although level 2 was significantly different, it was not in a predicted pattern. There was no significant difference observed between level 3 and the mean of others (see the result in Table 23).

Table 23

Result of the Deviation Contract

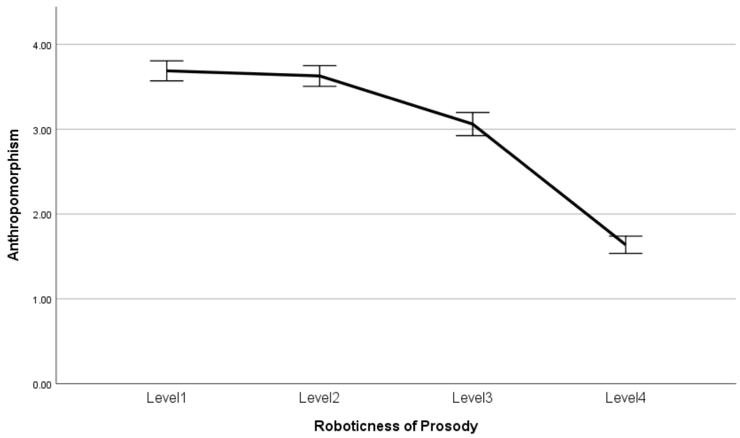
Dependent variables	Level of RoP	M^2	F (1, 42)	p	η^2
Anthropomorphism	level 2 vs. Mean	16.77	33.30	**<.000**	.44
	level 3 vs. Mean	0.14	0.23	.631	.00

Robomorphism	level 2 vs. Mean	24.47	35.90	**<.000**	.46
	level 3 vs. Mean	9.08	.00	.991	.00
Trust	level 2 vs. Mean	4.05	12.25	**<.001**	.23
	level 3 vs. Mean	0.05	0.12	.731	.00
Eeriness	level 2 vs. Mean	3.44	5.10	**<.019**	.12
	level 3 vs. Mean	0.05	0.10	.751	.00

However, the expected valley-like pattern was not noticed again, there was no need for undue concern regarding the violation of normality assumptions to check the confidence intervals (Although it was not necessary to review the confidence intervals, they are provided in Appendix H for those may require consultation). The mean values for each variable across the four levels of RoP appeared to follow a linear change, which means hypotheses H_{1b}, H_{2b}, H_{3b}, and H_{4b} were not supported; level 2 is never higher (or lower) than the surrounding levels (see Figures 16.1-16.4). There was no evidence for the presence of the UVE. See also the estimated marginal means in Table 24.

Figure 16.1

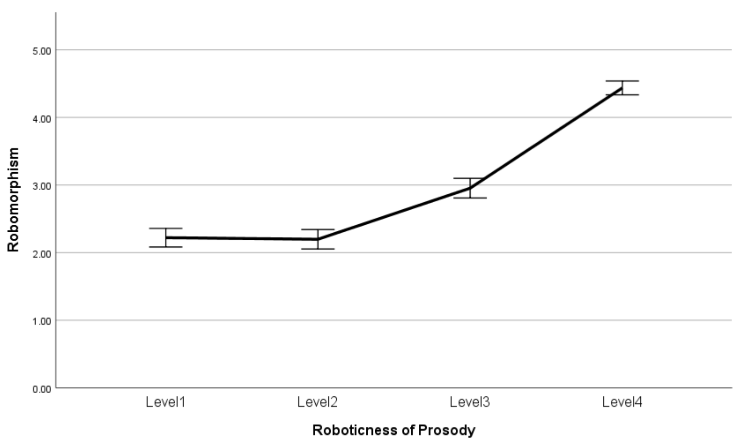
Anthropomorphism Across Levels of RoP



Note. Error Bars show 95% confidence intervals.

Figure 16.2

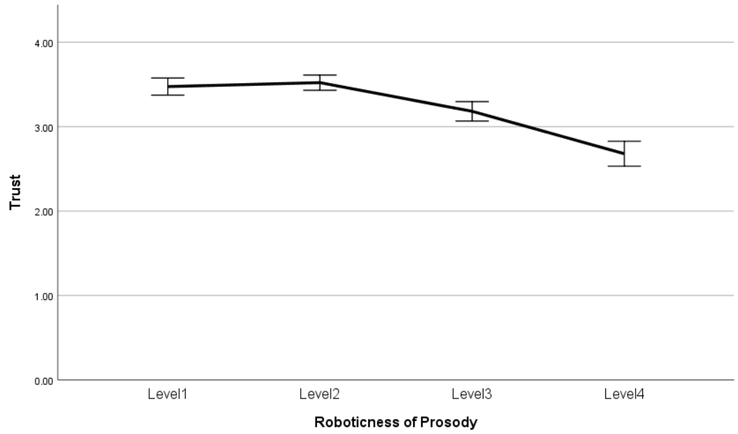
Robomorphism Across Levels of RoP



Note. Error Bars show 95% confidence intervals.

Figure 16.3

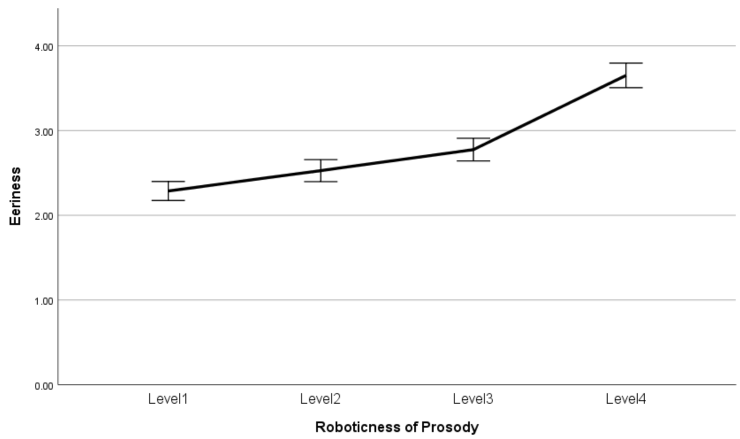
Trust Across Levels of RoP



Note. Error Bars show 95% confidence intervals.

Figure 16.4

Eeriness Across Levels of RoP



Note. Error Bars show 95% confidence intervals.

Table 24

Estimated Marginal Means

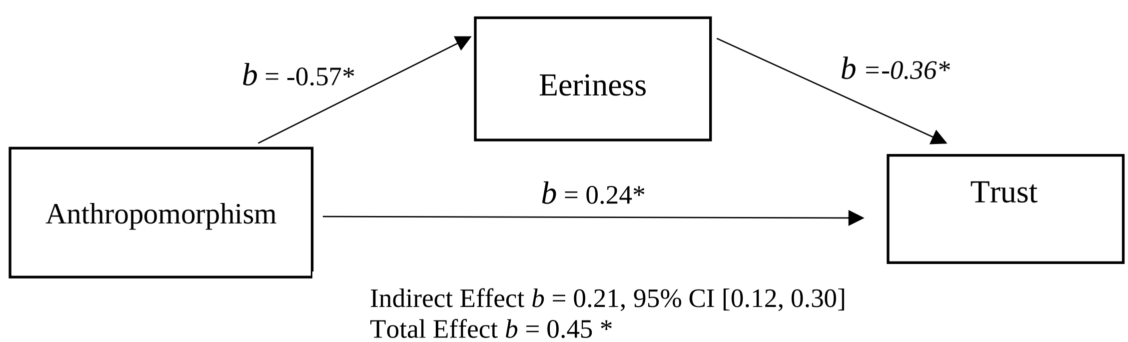
	Level of RoP	*M*	*SD*	95%CI	
				LL	*UL*
Anthropomorphism	1	3.7	0.1	3.44	3.93
	2	3.6	0.1	3.38	3.88
	3	3.1	0.1	2.80	3.34
	4	1.6	0.1	1.43	1.85
Robomorphism	1	2.2	0.1	1.94	2.50
	2	2.2	0.1	1.90	2.50
	3	2.9	0.1	2.65	3.25
	4	4.4	0.1	4.23	4.65
Trust	1	3.5	0.1	3.26	3.68
	2	3.5	0.1	3.33	3.71
	3	3.2	0.1	2.95	3.42
	4	2.7	0.1	2.37	3
Eeriness	1	2.3	0.1	2.10	2.52
	2	2.5	0.1	2.26	2.80
	3	2.8	0.1	2.50	3.05
	4	3.6	0.1	3.35	3.95

Mediation Analysis

A mediation analysis was performed to examine the relationship between Anthropomorphism and Trust, with the mediating role of Eeriness. The PROCESS macro for SPSS (Model 4) by Hayes (2023) was used for the analysis, with bootstrapping (with 5000 samples). Anthropomorphism had a significant effect on Eeriness *b* = -0.57, *t* (170) = -11.10, *p* < .000, explaining 42% of the variability in Eeriness R^2 = 0.42, *F* (1, 170) = 122.80. Additionally, a significant effect of Eeriness on Trust was observed *b* = -0.36, *t* (169) = -6.12, *p* < .000, accounting for 50% of the variability in Trust R^2 = 0.50, *F* (2, 169) = 83.53. Furthermore, Anthropomorphism had a significant effect on Trust *b* = 0.24, *t* (169) =4.171 *p* < .000, explaining 50% of the variability in Trust R^2 = 0.50, *F* (2, 169) = 83.53. The was a significant indirect effect of *b* = 0.21, *SE* = 0.04, 95% CI [0.12, 0.30], supporting $H_{5.}$ The total effect of Anthropomorphism and Eeriness on Trust was 0.45, *p* < .000. See Figure 17 for visualization.

Figure 17

Meditation Analysis Summary



Note. * *p*< .05.

Discussion

The aim of the second experiment was to continue to further explore the certain point in prosody that contributes to the UVE. The second experiment used stronger manipulations of prosody toward roboticness. The results of Experiment 2 closely align with those of Experiment 1, further supporting hypotheses $H_{1\text{-}4a}$. The linear regression analysis revealed a significant relationship between the RoP and all four dependent variables. However, contrary to hypotheses $H_{1\text{-}4b}$, the data did not reveal the anticipated valley pattern of the UVE, thus failing to provide support for those hypotheses. Notably, the findings once again supported the mediating role of eeriness between anthropomorphism and trust, supporting H_5.

General Discussion

The present study's primary objective was to investigate the potential presence of the UVE in the prosody of computer-generated speech. This was done by assessing participants' perceptions of anthropomorphism, robomorphism, and eeriness while manipulating the level of roboticness of prosody. The manipulation had a noticeable effect, as participants clearly provided different responses to the different levels of manipulation. The current study comprised two experiments, both yielding consistent findings. Notably, the similar results from both experiments provide robust evidence for the observed relationships between the variables. In both experiments, the results showed a linear relationship between each dependent variable and the RoP, thereby providing support for hypotheses $H_{1\text{-}4a}$. However, the expected valley pattern was not evident as hypothesized for $H_{1\text{-}4b}$. In conformity with H_5, experiencing eeriness had a mediating effect on the relationship between anthropomorphism and trust.

The linear relationship between human likeness and eeriness, contrary to the expected UVE, is consistent with MacDorman et al. (2009) who discovered that as a computer-generated face becomes increasingly human-like, it does not result in the highest level of eeriness. Similarly, the findings support the notion of more anthropomorphized technology leading to greater trust, in accordance with previous studies (Følstad et al., 2018; Natarajan & Gombolay, 2020). Moreover, the results of anthropomorphism on trust, mediated by the feeling of eeriness, are in line with the findings of Shin et al. (2019). The absence of the UVE also aligns with previous studies that have found no evidence of a vocal UVE when analyzing isolated voice stimuli (Baird et al., 2018; Kühne et al., 2020).

The failure to observe a valley pattern in both present experiments may be attributed to several possibilities. Firstly, as discussed in the results of Experiment 1, the decision was made to proceed with a stronger manipulation for Experiment 2 as one of two options due to time constraints. The unexplored option, namely the UVE potentially occurring between levels 1 and 2 (0-33% roboticness), may turn out to be correct, which means the UVE may have been concealed within this range. To investigate this further, additional levels between 0% and 33% RoP should be tested. The larger standard deviation of the dependent variables at level 2 of the RoP, as well as the presence of outliers that are participant-specific suggests that focusing on this range may reveal the presence of the UVE.

Some participants identified as outliers were consistently the same individuals in perceiving lower anthropomorphism, higher robomorphism, and experiencing more eeriness at level 2 of RoP (the level closest to humanlike prosody). The outliers for Trust were observed, at different levels and were not the same participants. The outliers as the same participants may possess a heightened sensitivity to imperfections in the human likeness of prosody, potentially leading to a more noticeable experience of the UVE. Individual differences in perception could be a key contributing factor to the manifestation of the UVE. For instance, according to the Evolutionary Aesthetics hypothesis, the experience of uncanniness occurs when appearances fail to meet the expected standards of human likeness, deviating from universal aesthetic norms

(MacDorman & Ishiguro, 2006). Thus, individual preference may play a prominent role in shaping the uncanny feeling, wherein individuals may have varying definitions for attractiveness in prosody. The impact of individual perception is also noteworthy in the alternative hypothesis about the emergence of the UVE, the Categorical Uncertainty hypothesis. This hypothesis posits that encountering information contradicting our established expectations of human likeness can lead to a sense of unease, stemming from perception of uncertainty regarding classification boundaries (Wang et al., 2015). Furthermore, according to the Violation of Expectation hypothesis, the sensation of uncanniness arises when human replicas fail to meet the expectations set for human-like characteristics (MacDorman & Ishiguro, 2006). Therefore, classification or expectation may be interpreted and defined differently depending on the individual.

Given the inherent variability of expectations and classification for human likeness among individuals, Wang et al. (2015) have also proposed exploring individuals' specific expectations regarding human replicas and studying the cognitive factors associated with the violation of these expectations may provide a deeper understanding of the UVE. Hence, the UVE is not a universally experienced phenomenon, but rather a subjective experience that varies among individuals based on their unique expectations and preferences. Additionally, Kühne et al. (2020) highlighted the significance of individual characteristics that impact the evaluation of synthesized voices. Thus, the presence of outliers in this study also can serve as evidence that certain individuals perceive human likeness differently, leading to the triggering of feelings of eeriness. However, in order to provide more conclusive evidence, it is important to conduct further research with larger sample sizes and explore alternative manipulation approaches and methods. Future researchers should also consider qualitative analysis to gain a deeper understanding of individuals' perceptions and expectations.

Finally, the last possibility of failure in observing the UVE stems from the separation of visual and vocal stimuli, leaving the UVE out of this type of approach. It is possible that when both visual and vocal stimuli are presented simultaneously, the UVE can be observed, as previous studies examining the integration of visual and vocal cues have suggested, such as mismatched facial appearance and voice (Mitchell et al., 2011), and unsynchronized lip movements and voice (Tinwell & Grimshaw, 2009). On the other hand, the studies conducted with only vocal stimuli did not yield any evidence supporting the existence of the UVE (Baird et al., 2018; Kühne et al., 2020). This could be due to relying solely on verbal without incorporating visual imagery since the combination of both stimuli simultaneously processing can facilitate more robust and interconnected mental representations of information (Paivio, 1986).

Allan Paivio (1986) proposed the Dual Coding Theory which suggests that information is processed and represented through two distinct and intertwined cognitive processing systems involved in human cognition. There are two coding systems, namely the verbal system and the visual system, which reinforce each other to enhance mental imagery. Mental imagery is the cognitive process of simulating perceptual information in our mind, even in the absence of sensory input. It involves retrieving stored representations from memory, enabling us to re-experience a modified version of the initial stimulus or generate novel combinations of stimuli (Pearson et al., 2015). Evoking better mental imagery, may allow for the formation of vivid images of our world, and enable a better image of the robot and its characteristics according to our established expectations. Therefore, by separating visual and vocal stimuli, the potential to engage both coding systems may be diminished, making it challenging to create mental imagery and observe the UVE.

While participants found voices that were more humanlike to be less eerie, it is important to bear in mind that comparing only four human voice samples does not sufficiently address the generalizability of these results or fully account for the potential existence of a vocal UVE. Therefore, caution is warranted in drawing definitive conclusions. Future studies should employ a comprehensive array of speech samples ranging from natural human voices to synthesized voices as well as manipulating other characteristics of prosody. Moreover, further investigations should evaluate voices in real situations and interactive settings. This includes considering factors such as the quality, quantity, and duration of interaction, as well as manipulating these variables, which can potentially influence the perceived level of eeriness. In addition, conducting functional tests, where participants engage in specific tasks or conversations with an IVA, can provide a more objective and accurate measure of the eeriness of the voice instead of subjective ratings.

The observed correlations among the four dependent variables in this study indicate the presence of an underlying variable that could explain these relationships. One plausible assumption is that the perceived intelligence of the technology influences anthropomorphism, robomorphism, trust, and eeriness. When individuals perceive technology as intelligent, they are more likely to attribute human-like or robot-like qualities to it (Bartneck et al., 2007; Haas & Moussawi, 2020). The perception of anthropomorphism fosters perceived higher intelligence, while also influencing trust in the technology due to the belief in its capabilities and decision-making (Christoforakos et al., 2021; Moussawi & Benbunan-Fich, 2020). However, it is important to consider the contextual factor. For instance, in scenarios involving creative tasks, like the present study, a closer resemblance to human voices may lead to greater trust and reduced eeriness. Nevertheless, in different contexts, such as mathematical or complex problem-solving scenarios where computers are perceived as outperforming humans, the results may differ. Therefore, further investigations involving more experimental manipulations and diverse contexts are necessary to determine if these correlations are specific to certain scenarios.

The present study's findings can have implications for designers of IVAs in various contexts to enhance the HRI experience. Since this study did not find evidence that manipulated voices fell into the UV, IVAs designers may not need to worry about eeriness when developing or using highly human-like voices. Furthermore, having more human-like voices can enhance trust, making it beneficial for businesses, as customers tend to feel more reassured when interacting with IVAs that closely resemble human voices. Given the individual differences in voice perception that raises doubt about the universality of the UVE, it is important to consider this when developing an artificial voice for IVAs. Therefore, designers must take into account the target audience and the specific context in which the IVA would be used. It is important to note that creating a universally appealing synthesized voice for all individuals is challenging. Further investigation is needed to understand the interplay between age, personality, and voice evaluation.

Conclusion

In the present study, participants' trust is enhanced by higher levels of human-like prosody in voices, although the level of eeriness mediates this relationship. The participants' perception of human likeness increases as the prosody exhibits a greater resemblance to human speech, rather than sounding robotic. Contrary to expectations, participants always rate more humanlike voices as less eerie. The consistent findings of both experiments cast doubt on the vocal UVE. There are some possible explanations for the absence of it: 1) It is possible that a vocal uncanny valley simply does not exist unless visual and verbal stimuli are represented simultaneously; 2) The selection of voice stimuli for research may have a wide enough range, but individual difference and expectation could have prevented experiencing the typical UVE at one consistent level of manipulation; 3) The selection of voice stimuli either may not reach the valley and stop before reaching the point of negative response, resembling a drop, or conversely, start within the valley and progressively advance towards achieving a full human-like resemblance. The result of the current study calls for more exploration and experimentation to advance the understanding of voice perception, the UVE, and the design of computer-generated speech. To address the original research question regarding the existence of the UVE for solely vocal stimuli better, more points should be explored between the exact human voice and near deviation from that in robust experiments with varied stimuli and real-world scenarios. It is important to explore different settings as the effect may not be strong in the current context and could vary across individuals. Paying attention to individual differences is crucial, as there might be a subset of people who experience the vocal UVE but are overshadowed by the general mean of experiments. Given the rapid growth of voice-based systems like Alexa and Siri, collaborative studies involving psychologists are needed to delve into the details. Hence, further findings would contribute to the development and implementation of more engaging and comfortable IVAs for all users.

References

Abdulla, H., Eltahir, A. M., Alwahaishi, S., Saghair, K., Platos, J., & Snasel, V. (2022, 19-22 July 2022). Chatbots Development Using Natural Language Processing: A Review. *26th International Conference on Circuits, Systems, Communications and Computers (CSCC), Crete, Greece, 2022*. 122-128, https://doi.org/10.1109/CSCC55931.2022.00030

Agarwal, S., Agarwal, B., & Gupta, R. (2022). Chatbots and virtual assistants: a bibliometric analysis. *Library Hi Tech, 40*(4), 1013-1030. https://doi.org/10.1108/LHT-09-2021-0330

Baird, A., Parada-Cabaleiro, E., Hantke, S., Burkhardt, F., Cummins, N., Schuller, B. (2018) The Perception and Analysis of the Likeability and Human Likeness of Synthesized Speech. *Proc. Interspeech 2018*, 2863-2867, https://doi.org/10.21437/Interspeech.2018-1093

Bansal, H., & Khan, R. (2018). A Review Paper on Human Computer Interaction. *International Journal of Advanced Research in Computer Science and Software Engineering*, *8*, 53. https://doi.org/10.23956/ijarcsse.v8i4.630

Bartneck, C., Kanda, T., Mubin, O., & Mahmud, A. A. (2007). The perception of animacy and intelligence based on a robot’s embodiment. *The 7th IEEE-RAS International Conference on Humanoid Robots*, Pittsburgh, PA, USA, 2007, 300-305. https://doi.org/10.1109/ichr.2007.4813884

Bicha, Z.J.G. (2023). *Speech Errors in a Computer-Generated Voice: How Do Speech Errors Influence the Perceived Human-likeness of a Voice.* [Master's thesis, Tilburg University]. Tilburg University Library. https://www.worldcat.org/title/1382361116

Boersma, P.& Weenink, D, (2023) *Praat: doing phonetics by computer* (Version 6.2.01) [Computer program] retrieved 17 December 2021 from https://www.praat.org/

Chai, J., Lin, J., Zadrozny, W., Ye, Y., Stys-Budzikowska, M., Horvath, V., Kambhatla, N., & Wolf, C. (2001). The Role of a Natural Language Conversational Interface in Online Sales: A Case Study. *International Journal of Speech Technology*, *4*(3), 285-295. https://doi.org/10.1023/A:1011316909641

Cherif, E., & Lemoine, J. (2019). Anthropomorphic virtual assistants and the reactions of Internet users: An experiment on the assistant's voice. *Recherche Et Applications En Marketing*, *34*(1), 28–47. https://doi.org/10.1177/2051570719829432

Christoforakos, L., Gallucci, A., Surmava-Große, T., Ullrich, D., & Diefenbach, S. (2021). Can Robots Earn Our Trust the Same Way Humans Do? A Systematic Exploration of Competence, Warmth, and Anthropomorphism as Determinants of Trust Development in HRI. *Frontiers in robotics and AI, 8*. https://doi.org/10.3389/frobt.2021.640444

de Leest, B. M. L. (2022). *Focus intonation in computer-generated speech: how does focus intonation influence the perceived human-likeness of computer-generated voice*. [Master's thesis, Tilburg University]. Tilburg University Library. https://www.worldcat.org/title/1362441866

Dehn, D. M., & Van Mulken, S. (2000). The impact of animated interface agents: a review of empirical research. *International Journal of Human-computer Studies*, *52*(1), 1–22. https://doi.org/10.1006/ijhc.1999.0325

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864–886. https://doi.org/10.1037/0033-295x.114.4.864

Eyssel, F. A., Kuchenbrandt, D., Bobinger, S., De Ruiter, L., & Hegel, F. (2012). *"If you sound like me, you must be more human."* https://doi.org/10.1145/2157689.2157717

Følstad, A., & Brandtzaeg, P. (2017). Chatbots and the new world of HCI. *Interactions*, *24*(4), 38–42. https://doi.org/10.1145/3085558

Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018). What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study. In *Lecture Notes in Computer Science* (pp. 194–208). Springer Science+Business Media. https://doi.org/10.1007/978-3-030-01437-7_16

Grazzini, L., Viglia, G. and Nunan, D. (2023). Dashed expectations in service experiences. Effects of robots human-likeness on customers' responses, *European Journal of Marketing, 57 (4)*, 957-986. https://doi.org/10.1108/EJM-03-2021-0220

Haas, C., & Moussawi, S. (2020). Are Anthropomorphic Intelligent Agents More Intelligent? *AMCIS 2020 Proceedings.*3*.* https://aisel.aisnet.org/amcis2020/ai_semantic_for_intelligent_info_systems/ai_semantic_for_intelligent_info_systems/3

Han, M. C. (2021). The Impact of Anthropomorphism on Consumers' Purchase Decision in Chatbot Commerce. *Journal of Internet Commerce*, 20(1), 46-65. https://doi.org/10.1080/15332861.2020.1863022

Hanson, D. (2005). Expanding the aesthetic possibilities for humanoid robots. In *IEEE-RAS international conference on humanoid robots,* Tsukuba, Japan, 24-31. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=7d92e672ea4736657a6e74eb10e41a51a31b2bd7

Ho, C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, *26*(6), 1508–1518. https://doi.org/10.1016/j.chb.2010.05.015

Hu, P., Lu, Y., & Gong, Y. (2021). Dual humanness and trust in conversational AI: A person-centered approach. *Computers in Human Behavior, 119*, 106727. https://doi.org/https://doi.org/10.1016/j.chb.2021.106727

James, J., Watson, C., & MacDonald, B. A. (2018). *Artificial Empathy in Social Robots: An analysis of Emotions in Speech*. https://doi.org/10.1109/roman.2018.8525652

Jenneboer, L., Herrando, C., & Constantinides, E. (2022). The Impact of Chatbots on Customer Loyalty: A Systematic Literature Review. *Journal of Theoretical and Applied Electronic Commerce Research*, 17(1), 212-229. https://doi.org/10.3390/jtaer17010011

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. https://doi.org/10.1093/scan/nsy001

Kimura, M., & Yotsumoto, Y. (2018). Auditory traits of "own voice." *PLOS ONE*, *13*(6), e0199443. https://doi.org/10.1371/journal.pone.0199443

Kühne, K., Fischer, M. R., & Zhou, Y. (2020). The Human Takes It All: Humanlike Synthesized Voices Are Perceived as Less Eerie and More Likable. Evidence From a Subjective Ratings Study. *Frontiers in Neurorobotics, 14*.https://doi.org/10.3389/fnbot.2020.593732

Lankton, N. K., McKnight, D. H., & Tripp, J. (2015). Technology, Humanness, and Trust: Rethinking Trust in Technology. *Journal of the Association for Information Systems*, *16*(10), 880–918. https://doi.org/10.17705/1jais.00411

Lee, J.-E., & Nass, C. (2010). Trust in Computers: The Computers-Are-Social-Actors (CASA) Paradigm and Trustworthiness Perception in Human-Computer Communication. Trust and Technology in a Ubiquitous Modern Environment: *Theoretical and Methodological Perspectives*, 1-15. https://doi.org/10.4018/978-1-61520-901-9.ch001

Lee, K. M., Park, N., & Song, H. (2005). Can a Robot Be Perceived as a Developing Creature? *Human Communication Research, 31*(4), 538-563. https://doi.org/https://doi.org/10.1111/j.1468-2958.2005.tb00882.x

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies, 7*(3), 297-337. https://doi.org/https://doi.org/10.1075/is.7.3.03mac

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*(3), 695–710. https://doi.org/10.1016/j.chb.2008.12.026

Männistö-Funk, T., & Sihvonen, T. (2018). Voices from the Uncanny Valley. *Digital Culture & Society*, *4*(1), 45–64. https://doi.org/10.14361/dcs-2018-0105

Mara, M., Appel, M., & Gnambs, T. (2022). Human-Like Robots and the Uncanny Valley. *Zeitschrift für Psychologie*, 230(1), 33-46. https://doi.org/10.1027/2151-2604/a000486

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. https://doi.org/10.1016/j.cognition.2015.09.008

McKnight, D. H., Choudhury, V., & Kacmar, C. J. (2002). Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. *Information Systems Research*, *13*(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81

Mekni, M., Baani, Z., & Sulieman, D. (2020). A Smart Virtual Assistant for Students. In *Applications Intelligent Systems*. https://doi.org/10.1145/3378184.3378199

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. https://doi.org/10.1068/i0415

Mithen, S., & Boyer, P. (1996). Anthropomorphism and the Evolution of Cognition. *The Journal of the Royal Anthropological Institute*, 2(4), 717–721. http://www.jstor.org/stable/3034305

Montano, K. J., Tigue, C. C., Isenstein, S. G. E., Barclay, P., & Feinberg, D. R. (2017). Men's voice pitch influences women's trusting behavior. *Evolution and Human Behavior*, *38*(3), 293–297. https://doi.org/10.1016/j.evolhumbehav.2016.10.010

Morana, S., Gnewuch, U., Jung, D., & Granig, C. (2020). The Effect of Anthropomorphism on Investment Decision-Making with Robo-Advisor Chatbots. In *European Conference on Information Systems*. https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1062&context=ecis2020_rp

Mori, M. (1970). The uncanny valley. Energy, 7(4), 33–35.

Mori, M., MacDorman, K., & Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine*, *19*, 98-100. https://doi.org/10.1109/MRA.2012.2192811

Moussawi, S., & Benbunan-Fich, R. (2020). The effect of voice and humour on users' perceptions of personal intelligent agents. *Behaviour & Information Technology*, *40*(15), 1603–1626. https://doi.org/10.1080/0144929x.2020.1772368

Moussawi, S., & Koufaris, M. (2019). Perceived Intelligence and Perceived Anthropomorphism of Personal Intelligent Agents: Scale Development and Validation. *Hawaii International Conference on System Sciences*. https://doi.org/10.24251/hicss.2019.015

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors, *Conference Companion on Human Factors in Computing Systems*, Boston, Massachusetts, USA. https://doi.org/10.1145/259963.260288

Natarajan, M., & Gombolay, M. (2020). Effects of Anthropomorphism and Accountability on Trust in Human Robot Interaction *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, Cambridge, United Kingdom. https://doi.org/10.1145/3319502.3374839

Netter, S. B., Zhao, X., Barasch, A., & Schroeder, J. (2021). "Hello! How May I Helo You?": How (Corrected) Errors Humanize a Communicator. *Social Science Research Network*. https://doi.org/10.2139/ssrn.3901961

Niculescu, A. I., Van Dijk, B., Nijholt, A., Li, H., & See, S. L. (2013). Making Social Robots More Attractive: The Effects of Voice Pitch, Humor and Empathy. *International Journal of Social Robotics*, *5*(2), 171–191. https://doi.org/10.1007/s12369-012-0171-x

Nishio, S., Ogawa, K., Kanakogi, Y., Itakura, S., & Ishiguro, H. (2012). Do robot appearance and speech affect people's attitude? Evaluation through the Ultimatum Game. *IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, Paris, France, 2012, 814-814. https://doi.org/10.1109/ROMAN.2012.6343851

Nordheim, C. B., Følstad, A., & Bjørkli, C. A. (2019). An Initial Model of Trust in Chatbots for Customer Service—Findings from a Questionnaire Study. *Interacting with Computers*, 31(3), 317-335. https://doi.org/10.1093/iwc/iwz022

Nowak, K. L. (2006). The Influence of Anthropomorphism and Agency on Social Judgment in Virtual Environments. *Journal of Computer-Mediated Communication*, *9*(2), 00. https://doi.org/10.1111/j.1083-6101.2004.tb00284.x

Paivio, A. (1986). *Mental representations: a dual coding approach*. Oxford University Press; Clarendon Press.

Pearson, J., Naselaris, T., Holmes, E. A., & Kosslyn, S. M. (2015). Mental Imagery: Functional Mechanisms and Clinical Applications. *Trends in cognitive sciences*, 19(10), 590–602. https://doi.org/10.1016/j.tics.2015.08.003

Pettorino, Massimo (2015): "The History of Talking Heads: The Trick and the Research." *In: First International Workshop on the History of Speech Communication Research*, Dresden, Germany, September 4–5, 2015. ISCA Archive, https://pdfs.semanticscholar.org/ce62/d996d50d23d4138f6d99410105d18612ac71.pdf

Pradhan, A., Mehta, K., & Findlater, L. (2018). "Accessibility Came by Accident": Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal QC, Canada. https://doi.org/10.1145/3173574.3174033

Ramey, C. H. (2006). An inventory of reported characteristics 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. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=c719574caf93b1169eef6bac239a7592ddf2d48b

Rao, K. S., Koolagudi, S. G., & Vempada, R. R. (2013). Emotion recognition from speech using global and local prosodic features. *International Journal of Speech Technology*, *16*(2), 143-160. https://doi.org/10.1007/s10772-012-9172-2

Reeves, B., & Nass, C. (1996). The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Pla. *Bibliovault OAI Repository, the University of Chicago Press*.

Rekkers, R. (2022). *"Can I trust your voice?": the effect of doubtful intonation patterns on trust in voice-based chatbots*. [Master's thesis, Tilburg University]. https://doi.org/10.13140/RG.2.2.10333.08161

Roesler, E., Manzey, D., & Onnasch, L. (2021). A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction. *Science robotics*, 6(58), eabj5425. https://doi.org/10.1126/scirobotics.abj5425

Salem, M., Eyssel, F. A., Rohlfing, K. J., Kopp, S., & Joublin, F. (2013). To Err is Human(-like): Effects of Robot Gesture on Perceived Anthropomorphism and Likability. *International Journal of Social Robotics*, *5*(3), 313–323. https://doi.org/10.1007/s12369-013-0196-9

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. D. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience*, *7*(4), 413–422. https://doi.org/10.1093/scan/nsr025

Schouten, A. P., Portegies, T. C., Withuis, I., Willemsen, L. M., & Mazerant-Dubois, K. (2022). Robomorphism: Examining the effects of telepresence robots on between student cooperation. *Computers in Human Behavior*, 126, 106980. https://doi.org/10.1016/j.chb.2021.106980

Schroeder, J., & Epley, N. (2016). Mistaking minds and machines: How speech affects dehumanization and anthropomorphism. *Journal of Experimental Psychology: General*, *145*(11), 1427–1437. https://doi.org/10.1037/xge0000214

Seyama, J., & Nagayama, R. S. (2007). The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces. *Presence: Teleoperators & Virtual Environments*, *16*(4), 337–351. https://doi.org/10.1162/pres.16.4.337

Shin, M., Kim, S. J., & Biocca, F. (2019). The uncanny valley: No need for any further judgments when an avatar looks eerie. *Computers in Human Behavior*, *94*, 100-109. https://doi.org/https://doi.org/10.1016/j.chb.2019.01.016

Sims, V. K., Chin, M. G., Lum, H. C., Upham-Ellis, L., Ballion, T., & Lagattuta, N. C. (2009). Robots' Auditory Cues are Subject to Anthropomorphism. *Proceedings of the Human Factors and Ergonomics Society Annual Meetin*g, 53(18), 1418–1421. https://doi.org/10.1177/154193120905301853

Song, S. W., & Shin, M. (2022). Uncanny Valley Effects on Chatbot Trust, Purchase Intention, and Adoption Intention in the Context of E-Commerce: The Moderating Role of Avatar Familiarity. *International Journal of Human–Computer Interaction*, 1-16. https://doi.org/10.1080/10447318.2022.2121038

Strait, M., Vujovic, L., Floerke, V., Scheutz, M., & Urry, H. (2015). Too Much Humanness for Human-Robot Interaction: Exposure to Highly Humanlike Robots Elicits Aversive Responding in Observers *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, Seoul, Republic of Korea. https://doi.org/10.1145/2702123.2702415

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibilit*y, *The MIT Press*, 2008. 73–100. https://doi.org/10.1162/dmal.9780262562324.073

Terzopoulos, G., & Satratzemi, M. (2020). Voice Assistants and Smart Speakers in Everyday Life and in Education. *Informatics in Education*, 473–490. https://doi.org/10.15388/infedu.2020.21

Tinwell, A., & Grimshaw, M. (2009). Bridging the uncanny: an impossible traverse? *Proceedings of the 13th International MindTrek Conference*: *Everyday Life in the Ubiquitous Era*, Tampere, Finland. https://doi.org/10.1145/1621841.1621855

Von Zitzewitz, J., Boesch, P. M., Wolf, P., & Riener, R. (2013). Quantifying the Human Likeness of a Humanoid Robot. *International Journal of Social Robotics*, *5*(2), 263–276. https://doi.org/10.1007/s12369-012-0177-4

Wang, S., & Rochat, P. (2017). Human Perception of Animacy in Light of the Uncanny Valley Phenomenon. *Perception*, *46*(12), 1386–1411. https://doi.org/10.1177/0301006617722742

Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The Uncanny Valley: Existence and Explanations. *Review of General Psychology*, 19(4), 393–407. https://doi.org/10.1037/gpr0000056

Westerman, D., Cross, A. C., & Lindmark, P. G. (2019). I Believe in a Thing Called Bot: Perceptions of the Humanness of "Chatbots". *Communication Studies*, 70(3), 295-312. https://doi.org/10.1080/10510974.2018.1557233

Wolters, M., Kelly, F., & Kilgour, J. (2016). Designing a spoken dialogue interface to an intelligent cognitive assistant for people with dementia. *Health Informatics Journal*, *22*(4), 854–866. https://doi.org/10.1177/1460458215593329

Yogeeswaran, K., Złotowski, J., Livingstone, M., Bartneck, C., Sumioka, H., & Ishiguro, H. (2016). The interactive effects of robot anthropomorphism and robot ability on perceived threat and support for robotics research. J. *Human -Robot Interact*., 5(2), 29–47. https://doi.org/10.5898/JHRI.5.2.Yogeeswaran

Zhang, J., Li, S., Zhang, J., Du, F., Qi, Y., & Liu, X. (2020). A Literature Review of the Research on the Uncanny Valley. In *Lecture Notes in Computer Science* (pp. 255–268). Springer Science+Business Media. https://doi.org/10.1007/978-3-030-49788-0_19

Zhong, V. J., Murset, N., Jager, J., & Schmiedel, T. (2022). Exploring Variables That Affect Robot Likeability. *17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Sapporo, Japan, 2022, 1140-1145, https://doi.org/10.1109/hri53351.2022.9889602

Złotowski, J., Proudfoot, D., Yogeeswaran, K. et al. Anthropomorphism: Opportunities and Challenges in Human–Robot Interaction. *Int J of Soc Robotics* 7, 347–360 (2015). https://doi.org/10.1007/s12369-014-0267-6

Appendix A

Randomized Assignments Lists

The presented table illustrates the random assignment of participants to one of four lists. Within each list, participants were exposed to all four paintings, which were randomly rotated with varying degrees of RoP. The levels of RoP consisted of level 1 (0% roboticness, being not robotic at all; natural human speech), level 2 (33% roboticness), level 3 (66% roboticness), and level 4 (100% robotic).

	 Self-Portrait as a Painter	 Sunflowers	 The Bedroom	 Almond Blossom
Lists:	RoP:	RoP:	RoP:	RoP:
1	Level 1	Level 2	Level 3	Level 4
2	Level 2	Level 3	Level 4	Level 1
3	Level 3	Level 4	Level 1	Level 2
4	Level 4	Level 1	Level 2	Level 3

Appendix B

Texts used by the tour guides

Self-Portrait as a Painter

Van Gogh presented himself in this self-portrait as a painter, holding a palette and paintbrushes behind his easel. He showed that he was a modern artist by using a new painting style, with bright, almost unblended colors. Self-Portrait as a Painter was the last work Van Gogh produced in Paris; the city had exhausted him both mentally and physically.

Sunflowers

The sunflower paintings had a special significance for Van Gogh: they communicated 'gratitude'. He had already painted a new version when his friend, the painter Paul Gauguin, came to live with him for a while in the Yellow House and Gauguin later asked for one as a gift, which Vincent was reluctant to give him. He later produced two loose copies, however, one of which is now in the Van Gogh Museum.

The Bedroom

He prepared his bedroom in the Yellow House himself with simple furniture and with his own work on the wall. The rules of perspective seem not to have been accurately applied throughout the painting, but this was a deliberate choice. Vincent told Theo in a letter that he had deliberately 'flattened' the interior and left out the shadows so that his picture would resemble a Japanese print.

Almond Blossom

Almond trees flower early in the spring, making them a symbol of new life as Van Gogh's favorite subjects. The painting was a gift for his brother Theo and sister-in-law Jo, who had just had a baby son, Vincent Willem. Vincent Willem went on to found the Van Gogh Museum.

Appendix C

Analysis of the Estimated Marginal Means

The responses of the 15 participants who answered the open-up question incorrectly were separately examined to ascertain that their answers did not noticeably differ from the remaining participants.

Estimated Marginal Means

	Level of RoP	*M*	*SD*	95%CI	
				LL	UL
Anthropomorphism	1	4	0.2	3.64	4.36
	2	3.7	0.3	3.17	4.36
	3	3.1	0.3	2.47	3.74
	4	2.3	0.3	1.66	2.92
Robomorphism	1	1.1	0.2	1.46	2.40
	2	2.1	0.3	1.53	2.76
	3	2.9	0.3	2.12	3.64
	4	3.9	0.3	3.20	4.53
Trust	1	3.8	0.1	3.47	4.04
	2	3.6	0.2	3.05	4.14
	3	3.4	0.2	3.09	3.81
	4	2.8	0.2	2.23	3.28
Eeriness	1	2.7	0.2	2.31	3.11
	2	2.5	0.2	2.02	2.91
	3	2.7	0.2	2.29	3.08
	4	3.9	0.1	2.71	3.46

Appendix D

Measurement scales

Perceived Anthropomorphism

Measurement scale from Hu et al. (2021), adapted:

Please rate your impression of the IVA tour guide on these scales:

A-1: The IVA tour guide's pronunciation is natural.

A-2: The IVA tour guide has a human-like voice.

A-3: The language expression of the IVA tour guide sounds like that of a machine.

A-4: I cannot feel the distance between the voice of the IVA tour guide and that of a human being.

A-5: I characterize the IVA tour guide's speaking aspect as human.

Perceived Robomorphism

Measurement scale from Schouten et al. (2022), adapted:

Please rate your impression of the IVA tour guide on these scales:

R-1: I feel as though the IVA tour guide was a real person

R-2: I feel as though the IVA tour guide was a robot

R-3: I perceive the IVA tour guide more as a machine than a person

R-4: I sometimes perceive the IVA tour guide as a machine

Trust

Measurement scale from Nordheim (2018), adapted:

Please rate your impression of the IVA tour guide on these scales:

T-1: I feel like this IVA tour guide is trustworthy

T-2: I do not think that this IVA tour guide will act in a way that is advantageous to me

T-3: I am suspicious of this IVA tour guide

T-4: This IVA tour guide appears deceptive

T-5: I trust this IVA of tour

Perceived Eeriness

Measurement scale from Ho and MacDorman (2010), adapted:

Please choose your impression of the IVA tour guide on these scales:

Low	High
E-1: Reassuring	Eerie
E-2: Numbing	Freaky
E-3: Ordinary	Supernatural

Appendix E

Comprehensive Report on the Correlation of Dependent Variables

The data revealed that there was a strong negative correlation between Anthropomorphism and Robomorphism at levels 1 of RoP r (45) = -.75, p< .000, level 2 of RoP r (45) =-.85, p< .000, 95% CI [-0.94, -0.73], level 3 of RoP r (45) =-.94, p< .000, and level 4 of RoP r (45) =-.90, p< .000, 95% CI [-0.95, -0.83]. Also, Anthropomorphism showed a moderate positive correlation with Trust at level 1 of RoP r (45) = .60, p< .000, and level 4 of RoP r (45) = .52, p< .000, 95% CI [0.28, 0.74], also a weak positive correlation at level 2 of RoP r (45) = .33, p< .027, 95% CI [-0.02, 0.63] and level 3 of RoP r (45) =.40, p< .008. However, Anthropomorphism demonstrated a moderately negative correlation with Eeriness across all levels of RoP respectively r (45) = -.41, p< .005, r (45) = -.60, p< .000, 95% CI [-0.80, -0.32], r (45) = -.61, p< .000), r (45) = -.55, p< .000, 95% CI [-0.74, -0.32]. In terms of the relationship between Robomorphism and Trust, a moderate negative correlation was found at level 1 of RoP r (45) = -.52, p< .000 and level 4 of RoP r (45) = -.46, p< .002, also a weak negative correlation at level 2 of RoP r (45) = -.36, p< .016, 95% CI [-0.59, -0.08] and level 3 of RoP r (45) = -.38, p< .010. Contrasty, Robomorphism displayed a moderate positive correlation with Eeriness at level 1 of RoP r (45) = .54, p< .000, level 2 of RoP r (45) = .64, p< .000, 95% CI [0.43, 0.80], level 3 of RoP r (45) =.56, p< .000, and level 4 of RoP r (45) = .52, p< .000. Finally, Trust was found to have a moderate negative correlation with Eeriness at level 1 of RoP r (45) =-.53, p< .000, level 3 of RoP r (45) =-.60, p< .000, and level 4 of RoP r (45) =-.49, p< .000, while a weak negative correlation at level 2 of RoP r (45) =-.35, p< .017, 95% CI [-0.56, -0.12].

Appendix F

Bootstraps of All Four Variables for the Different Prosody Levels

Table F1.

Bootstrapped confidence intervals for Anthropomorphism across all levels of RoP

Variables 95%CI *LL* *UL*	Anthropomorphism level 1 of RoP	Anthropomorphism level 2 of RoP	Anthropomorphism level 3 of RoP	Anthropomorphism level 4 of RoP
	[3.71, 4.10]	[3.43, 4.01]	[2.90, 3.52]	[2.01, 2.64]
Anthropomorphism level 1 of RoP		**Overlap**	No overlap	No overlap
Anthropomorphism level 2 of RoP			No overlap	No overlap
Anthropomorphism level 3 of RoP				No overlap

Table F2.

Bootstrapped confidence intervals for Robomorphism across all levels of RoP

Variables 95%CI *LL* *UL*	Robomorphism level 1 of RoP	Robomorphism level 2 of RoP	Robomorphism level 3 of RoP	Robomorphism level 4 of RoP
	[1.80, 2.40]	[1.91, 2.55]	[2.47, 3.17]	[3.40, 4.12]
Robomorphism level 1 of RoP		**Overlap**	No overlap	No overlap
Robomorphism level 2 of RoP			**Overlap**	No overlap
Robomorphism level 3 of RoP				No overlap

Table F3.

Bootstrapped confidence intervals for Trust across all levels of RoP

Variables 95%CI	Trust level 1 of RoP	Trust level 2 of RoP	Trust level 3 of RoP	Trust level 4 of RoP
LL *UL*	[3.52, 3.92]	[3.44, 3.91]	[3.10, 3.56]	[2.58, 3.04]
Trust level 1 of RoP		**Overlap**	No overlap	No overlap
Trust level 2 of RoP			**Overlap**	No overlap
Trust level 3 of RoP				No overlap

Table F4.

Bootstrapped confidence intervals for Eeriness across all levels of RoP

Variables 95%CI	Eeriness level 1 of RoP	Eeriness level 2 of RoP	Eeriness level 3 of RoP	Eeriness level 4 of RoP
LL *UL*	[2.10, 2.60]	[2.20, 2.64]	[2.62, 3.15]	[3.15, 3.73]
Eeriness level 1 of RoP		**Overlap**	No overlap	No overlap
Eeriness level 2 of RoP			**Overlap**	No overlap
Eeriness Level 3 of RoP				No overlap

Appendix G

Comprehensive Report on the Correlation of Dependent Variables

The findings showed a strong negative correlation between Anthropomorphism and Robomorphism at level 1 of RoP r (43) = -.72, p< .000, 95% CI [-0.84, -0.55], level 2 of RoP r (43) = -.80, p<.000, level 3 of RoP r (43) = -.81, p< .000, and level 4 of RoP r (43) = -.70, p< .000, 95% CI [-0.88, -0.36]. Additionally, Anthropomorphism revealed a moderate positive correlation with Trust at level 1 of RoP r (43) = .54, p< .000, level 2 of RoP r (43) = .54, p< .000, level 3 of RoP r (43) = .68, p< .000, and level 4 of RoP r (43) = .41, p< .007, 95% CI [0.20,0.64]. However, Anthropomorphism displayed a moderately negative correlation with Eeriness at levels 1 to 4 of RoP respectively r (43) = -.38, p< .013, r (43) = -.40, p< .008, r (43) = -.67, p< .000, and r (43) = -.43, p< .004, 95% CI [-0.65, -0.16]. Regarding the relationship between Robomorphism and Trust, a moderate negative correlation was observed at level 1 of RoP r (43) = -.60, p< .000, 95% CI [-0.80, -0.35], level 2 of RoP r (43) = -.53, p< .000, and level 3 of RoP r (43) = -.64, p< .000, while a weak negative correlation was found at level 4 of RoP r (43) = -.32, p< .037, 95% CI [-0.55, -0.11]. Furthermore, Robomorphism showed a weak positive correlation with Eeriness at level 1 of RoP r (43) = .30, p< .049, 95% CI [0.03,0.56], and a moderate positive correlation at level 2 of RoP r (43) = .54, p< .000 and level 4 of RoP r (43) = .47, p< .001, 95% CI [0.23,0.70], but a strong positive correlation at level 3 of RoP r (43) = .74, p< .000. Finally, there was a negative moderate correlation between Eeriness and Trust at level 1 of RoP r (43) = -.54, p< .000, level 2 of RoP r (43) = -.53, p< .000, level 4 of RoP r (43) = -.55, p< .000. However, there was a strong negative correlation at level 3 of RoP r (43) = -.72, p< .000.

Appendix H

Bootstraps of All Four Variables for the Different Prosody Levels

Table H1.

Bootstrapped confidence intervals for Anthropomorphism across all levels of RoP

Variables 95%CI *LL* *UL*	Anthropomorphism level 1 of RoP [3.46, 3.91]	Anthropomorphism level 2 of RoP [3.37, 3.87]	Anthropomorphism level 3 of RoP [2.78, 3.32]	Anthropomorphism level 4 of RoP [1.44, 1.85]
Anthropomorphism level 1 of RoP		**Overlap**	No overlap	No overlap
Anthropomorphism level 2 of RoP			No overlap	No overlap
Anthropomorphism Level 3 of RoP				No overlap

Table H2.

Bootstrapped confidence intervals for Robomorphism across all levels of RoP

Variables 95%CI *LL* *UL*	Robomorphism level 1 of RoP [1.96, 2.51]	Robomorphism level 2 of RoP [1.91, 2.50]	Robomorphism level 3 of RoP [2.67, 3.27]	Robomorphism level 4 of RoP [4.21, 4.64]
Robomorphism level 1 of RoP		**Overlap**	No overlap	No overlap
Robomorphism level 2 of RoP			No overlap	No overlap
Robomorphism level 3 of RoP				No overlap

Table H3.

Bootstrapped confidence intervals for Trust across all levels of RoP

Variables 95%CI *LL* *UL*	Trust level 1 of RoP [3.25, 3.66]	Trust level2 of RoP [3.34, 3.70]	Trust level3 of RoP [2.96, 3.40]	Trust level 4 of RoP [2.40, 2.98]
Trust level 1 of RoP		No overlap	**Overlap**	No overlap
Trust level 2 of RoP			No overlap	No overlap
Trust level 3 of RoP				No overlap

Table H4.

Bootstrapped confidence intervals for Eeriness across all levels of RoP

Variables 95%CI *LL* *UL*	Eeriness level 1 of RoP [2.06, 2.50]	Eeriness level 2 of RoP [2.30, 2.80]	Eeriness level 3 of RoP [2.51, 3.05]	Eeriness level 4 of RoP [3.36, 3.92]
Eeriness level 1 of RoP		**Overlap**	No overlap	No overlap
Eeriness level 2 of RoP			**Overlap**	No overlap
Eeriness level 3 of RoP				No overlap