Computers in Human Behavior Reports 14 (2024) 100430

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

Deviation from typical organic voices best explains a vocal uncanny valley

Alexander Diel [a,b,c,*], Michael Lewis [a]

[a] Cardiff University School of Psychology, Park Pl 70, Cardiff, CF10 3AS, UK
[b] University of Duisburg-Essen, Clinic for Psychosomatic Medicine and Psychotherapy, LVR University Hospital Essen, 45147, Essen, Germany
[c] Center for Translational Neuro- and Behavioral Sciences (C-TNBS), University of Duisburg-Essen, 45147, Essen, Germany

ARTICLE INFO

Keywords:
Uncanny valley
Voice processing
Pathological voice
Voice distortion
Text-to-speech
Deviation from familiarity

ABSTRACT

The uncanny valley describes the negative evaluation of near humanlike artificial entities. Previous research with synthetic and real voices failed to find an uncanny valley of voices. This may have been due to an incomplete selection of stimuli. In Experiment 1 ($n = 50$), synthetic, normal, and deviating voices (distorted and pathological) were rated on uncanniness and human likeness and categorized as human or non-human. Results showed a non-monotonic function when the uncanniness was plotted against human likeness indicative of an uncanny valley. However, the shape could be divided into two monotonic functions based on voice type (synthetic vs deviating). Categorization ambiguity could not predict voice uncanniness but moderated the effect of realism on uncanniness. Experiment 2 ($n = 35$) found that perceived organicness, animacy, and mind attribution of voices significantly moderated the effect of realism on uncanniness. Results indicate a vocal uncanny valley driven by deviations from typical human voices. While voices can fall into an uncanny valley, synthetic voices successfully escape it. Finally, the results support the account that uncanniness is caused by deviations from familiar categories, rather than categorical ambiguity or the misattribution of mind or animacy.

1. Introduction

Artificial humanlike entities with imperfect human appearance are evaluated negatively, a phenomenon called uncanny valley (MacDorman & Ishiguro, 2006; Mori, MacDorman, & Kageki, 2012). The relationship between human likeness and likability or uncanniness has often been mathematically defined as a polynomial function consisting of a gradual increase of likability with increasing human likeness and a drop into the negative at near human likeness (Diel, Weigelt, & MacDorman, 2021; Diel, Sato, Hsu, & Minato, 2023; Mathur et al., 2020; Mori et al., 2012; Mara, Appel, & Gnambs, 2022). The uncanny valley remains a pressing issue in human-machine interaction, yet the underlying cognitive mechanisms remain unclear.

1.1. The vocal uncanny valley

The uncanny valley is not only relevant in the perception of artificial agents' visual appearance, but could also play a role in the acceptance of artificial voices. Synthetic voices gain increasing human likeness with technological development (Oord et al., 2016), and their auditory qualities have been found to be essential for adequate human-technology interaction (Seaborn, Miyake, Pennefather, & Otake-Matsuura, 2021). humans process natural synthetic voices akin to human voices, for example by seeing the speakers as potent as distinct agents (Whang & Im, 2021) or by attributing personality traits to the speaker (Nass & Lee, 2001). Voice realism and naturalness influence the perception of a virtual characters (Thomas, Ferstl, McDonnell, & Ennis, 2022; Zibrek, Cabral, & McDonnell, 2021), and social robots (Niculescu, van Dijk, Nijholt, Li, & See, 2013; Schreibelmayr & Mara, 2022; Trovato et al., 2017). Given that virtual voices will find increasing presence in the future (Chang, Kim, Beom, Won, & Jeon, 2020), a greater understanding of what human processing mechanisms make a voice more or less acceptable is essential for the development of adequate and practical voices.

The uncanny valley has been observed in the context of android appearance and behaviour and their mismatch with voices (Meah & Moore, 2014; Mitchell et al., 2011). However, previous research has consistently failed to find a 'vocal uncanny valley' when isolated voice stimuli were used: likability increased with a voice's human likeness in a linear manner (Baird et al., 2018a; Baird et al., 2018b; Kimura & Yotsumoto, 2018; Kühne, Fischer, & Zhou, 2020; Romportl, 2014; Schreibelmayr & Mara, 2022). However, except for one study (Kimura &

* Corresponding author. Cardiff University School of Psychology, Park Pl 70, Cardiff, CF10 3AS, UK.
E-mail address: diela@cardiff.ac.uk (A. Diel).

https://doi.org/10.1016/j.chbr.2024.100430
Received 25 January 2024; Received in revised form 14 May 2024; Accepted 14 May 2024
Available online 17 May 2024

Yotsumoto, 2018), research investigating a vocal uncanny valley have used exclusively synthetic voices and/or fully human voice stimuli. There are four explanations on why an uncanny valley of voices may not have been found: 1) a vocal uncanny valley does not exist; 2) stimulus selection has sufficient range but lacks stimuli that fall into the valley; 3) stimulus selection does not extent into the valley and stops before the drop (Mara, Appel, & Gnambs, 2022); 4) stimulus selection begins at the valley and ends at full human likeness. An uncanny valley and the four possible explanations are depicted in Fig. 1.

These explanations urge different implications for the design of artificial voices: If an uncanny valley of voices has not yet been reached, technological development may yet lead to its emergence. If, on the other hand, today's synthetic voices already overcome an uncanny valley or if a vocal uncanny valley does not exist, then this particular issue can be disregarded for the design of artificial voices.

Multiple theories on the uncanny valley have been proposed (Diel & MacDorman, 2021; Kätsyri, Förger, Mäkäräinen, & Takala, 2015; Wang, Lilienfeld, & Rochat, 2015). Two of these theories have received some critical attention in the past years, namely deviation-based explanations (Diel & MacDorman, 2021; Kätsyri et al., 2015) and categorization-based explanations (Mathur et al., 2020). Hence, these two theories will be focused on for Experiment 1. Both theories would predict the existence of an uncanny valley in voice stimuli.

1.2. Uncanniness and deviation from typical variation

Closeness to realistic human appearance may activate human-specific schemata and processing mechanisms, and sufficiently human-like yet deviating stimuli may activate these schemata and processes which would sensitize to the stimuli's relative atypicality, leading to uncanny sensations (Chattopadhyay & MacDorman, 2016; Kätsyri et al., 2015). Sensitivity to deviations is increased for more familiar stimulus categories, as is the case for human-specific stimuli like faces, potentially due to the recruitment of more specialized processing mechanisms (Chattopadhyay & MacDorman, 2016; Diel & Lewis, 2022a; Diel & Lewis, 2022b; Diel & MacDorman, 2021; Jung, Lee, & Choi, 2022). Deviating stimuli may recruit additional processing need, decreasing aesthetic evaluation through disfluent processing (Reber, Schwarz, & Winkielman, 2004). Alternatively, stimuli belonging to more familiar categories may have more solidified or strict predictive patterns, increasing the likelihood of prediction errors (Friston & Kiebel, 2011; Saygin, Chaminade, Ishiguro, Driver, & Frith, 2012). In any case, humanlike yet deviating voices would suffer from aesthetic devaluation.

Certain disorders like vocal fold paresis, Reinke's Edema, or muscle tension dysphonia, can lead to changes in the voice. Pathological voices are more likely to be categorized as atypical (Kreiman, Auszmann, & Gerratt, 2018; Kreiman & Gerratt, 2005; Kreiman, Gerratt, & Precoda, 1992) and are evaluated more negatively across various social dimensions compared to healthy voices (Altenberg & Ferrand, 2006; Amir & Levine-Yundof, 2013; Eadie, Rajabzadeh, Isetti, Nevdahl, & Baylor, 2017; Schroeder, Rembrandt, May, & Freeman, 2020). In analogy, previous research has suggested that dysmorphic, diseased, or very unattractive faces are perceived negatively or even uncanny (Diel & MacDorman, 2021; Rosa, Villacampa, Corradi, & Ingram, 2021). Pathological voices, similarly to disfigured faces, may hence fall into an uncanny valley as highly realistic yet deviating stimuli.

1.3. Uncanniness and categorization difficulty

Stimuli difficult to categorize may fall into an uncanny valley (Mathur et al., 2020; Chattopadhyay & MacDorman, 2016; Cheetham, Pavlovic, Jordan, Suter, & Jancke, 2013; Yamada, Kawabe, & Ihaya, 2013). Categorization difficulty may decrease likability due to processing disfluency (Carr, Hofree, Sheldon, Saygin, & Winkielman, 2017; Winkielman, Schwarz, Fazendeiro, & Reber, 2003) or cognitive conflict (Weis & Wiese, 2017). As categorization theories do not depend on stimulus domain, categorical ambiguity should thus also predict the uncanniness of voices.

2. Experiment 1

The aim of the experiment is to investigate the existence of a vocal

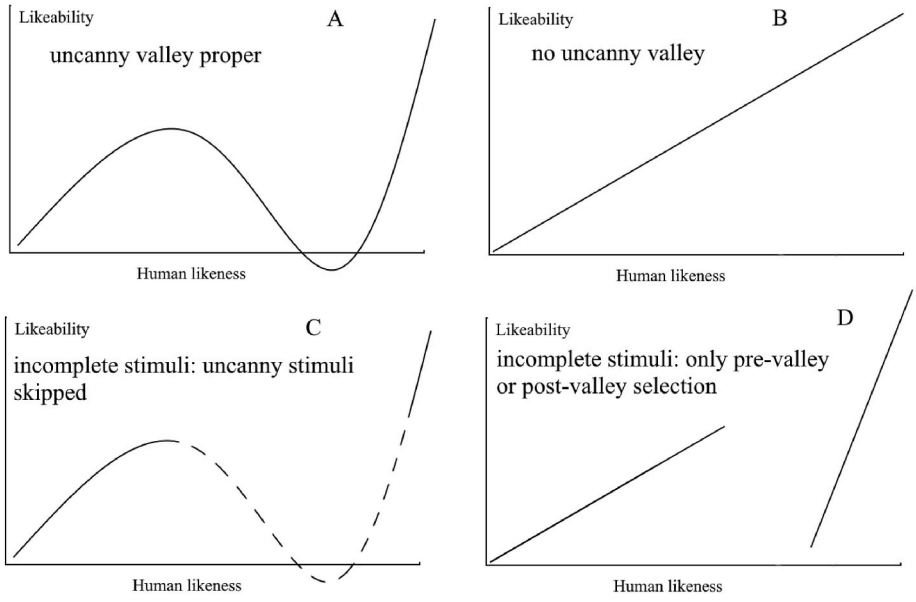

Fig. 1. A schematic representation of the uncanny valley (Fig. 1A) and possible data distributions depicting explanations on why a monotonic increase of likeability is observed instead of an uncanny valley.

Note. Fig. 1A: The uncanny valley function. Fig. 1B: An uncanny valley does not exist, leading to a monotonic function of likeability and human likeness. Fig. 1C: An uncanny valley exists but uncanny stimuli are not included in the experiment, resulting in a monotonic function. Fig. 1D: Stimulus range is either not sufficiently humanlike to fall into an uncanny valley (left line; see Mara et al., 2022), or stimuli are selected whose range starts within the uncanny valley (right line; see Diel et al., 2023); in both cases, monotonic functions would be observed even though a proper uncanny valley would be found if plotted with a sufficiently wide variation of human likeness.

uncanny valley using (manipulated) natural voices, synthetic voices, and pathological voices. In addition, it is investigated whether the uncanniness of voices can be explained by deviation from familiar categories or categorical ambiguity.

First, the role of specialization on the negative evaluation of deviating voice stimuli is investigated by comparing the effects of distortion on uncanniness for specialized (human) voices and less specialized (cat) voices. If specialization sensitizes the detection and negative evaluation of deviating stimuli, then distortion should increase the negative evaluation of both human and cat voices, but more strongly so for human voices. Hypothesis 1 is thus:

1. Distortion of human voices increases uncanniness more than distortion of cat voices.

Furthermore, it is investigated whether, similar to previous research, no vocal uncanny valley is found when using only natural and synthetic voices. However, a vocal uncanny valley is expected when distorted and pathological voices are included. It is hence tested whether a vocal uncanny valley exists in principle but is successfully avoided by contemporary synthetic voices. A monotonic function (i.e., an incremental decrease of uncanniness with increasing human likeness) would be evidence against an uncanny valley, while a non-monotonic function (specifically a strong increase in uncanniness within an otherwise monotonic decrease of uncanniness with increasing human likeness) would be evidence in favor of an uncanny valley (see Fig. 1). Hypotheses 2 and 3 are thus:

2. A monotonic function of human likeness explains the uncanniness of synthetic and natural voices better than a non-monotonic function.
3. A non-monotonic function akin to an uncanny valley can explain the uncanniness of synthetic, natural, distorted, and pathological voices better than a monotonic function.

Finally, it is investigated whether ambiguity in categorizing a voice as either human or non-human can best explain the uncanniness ratings. Categorization ambiguity is operationalized as 1. Categorization reaction time, and 2. Categorization uncertainty, i.e., the inconsistency of categorizations across participants. Because categorization reaction time may be impacted by other variables (e.g., stimulus type, participant attention or fatigue), effects on both reaction time and categorization uncertainty are tested. Hypothesis 4 is thus:

4. Categorization reaction time and categorization uncertainty predict uncanniness ratings of voices.

2.1. Methods

2.1.1. Participants

Power analysis revealed that $n = 50$ participants are sufficient to exceed a power of 1-$\beta = 0.8$ with a six-voice-conditions within-subject design and a standard effect size of $d = 0.5$ (Cohen, 1988). Participants were Psychology students at the Cardiff University School of Psychology, recruited via the Experimental Management System (EMS). Participants were on average 19 years old ($SD_{age} = 1.05$), 37 identified as female, 11 as male, one as other, and one preferred not to say. Participants were compensated with 4 credits equivalent to the advertised compensation of a 60 min online study.

2.1.2. Stimuli

Ten typical and 15 pathological voices were taken from the Perceptual Voice Qualities Database (PVQD; Walden, 2020). The database consists of standardized voice samples of healthy individuals and individuals with diverse voice pathologies that have been rated by voice professionals on several dimensions. Specifically, the 15 pathological voices with the highest subjective severity ratings were selected as stimuli. Specific pathologies included Reinke's Edema (x3), lesions (x3), vocal fold paralysis (x3), muscle tension dysphonia (x2), ulcerative laryngitis, adductor spasmodic dysphoria, and one unrecorded pathology. Ten distorted voice variants were created by using the STRAIGHT software, specifically by multiplying the normal voices' fundamental frequencies by 1000 to create a considerable level of distortion (Kawahara et al., 2008). Fundamental frequency manipulation would correspond to a deviation in a variable necessary to recognize and differentiate voices, akin to face structure variables in face research (Andics et al., 2010; Barsics, 2014; Latinus, McAller, & Bestelmeyer, 2013; Loffler, Yourganov, & Wilson, 2005). As face structure distortions can cause uncanniness, voice fundamental frequency distortions would be expected to create analogous effects for voices (Diel & Lewis, 2022a). In addition, 10 normal cat meowing sounds were selected from www.freesound.org, and 10 distorted cat voice variants were created with STRAIGHT by multiplying the fundamental frequency by 1000. Finally, 15 synthetic voices were selected from various sources: Four mechanical sounds were taken from www.freesound.org, five voices from IBM Watson (https://www.ibm.com/uk-en/cloud/watson-text-to-speech), three voices by Azure Microsoft TTS (https://azure.microsoft.com/en-us/services/cognitive-services/text-to-speech/#features), Microsoft Sam (https://tetyys.com/SAPI4/), one voice created by a Stephen Hawking Voice Generator (https://lingojam.com/StephenHawkingVoiceGenerator), and one generic Google TTS voice.

Fifteen pathological and synthetic voices were selected instead of 10 (as in the other conditions) because both conditions were expected to be more heterogenous and thus would need a higher stimulus number to be adequately statistically represented.

Because the fifteen pathological voices consisted of 9 female and 6 male voices, the same ratio was selected for typical voice counterparts. As distorted voices were created by manipulating the typical voices, the same ratio was present for those. For synthetic voices, six were artificial female voices, five were artificial male voices, and four were mechanical sounds. Accent of speech was not controlled.

All stimuli were cut to be around 5 s in length. For standardization, all typical, pathological, distorted, and synthetic voices (except for the mechanical sounds) were expressing the same sentences. The spoken sentences, "The blue spot is on the key again. How hard did he hit him?", were used as basic sentences in the PVQD database and were recreated for synthetic voices. More details on the voice stimuli are shown in Table A1. All stimuli are available at the OSF link below.

2.1.3. Rating task

For the rating task, participants had to rate each sound based on three rating scales: "not eerie/uncanny" to "eerie/uncanny", "strange/weird" to "not strange/weird" (reversed scale), and "not humanlike/realistic" to "humanlike/realistic", presented in that order. Scalesranged from the extremes of 0–100 and participants could choose to place the slider on any point of the scale. Scales were selected as some of the most effective items in measuring the uncanny valley according to a meta analysis (Diel et al., 2021). Although the scales "strange/weird" and "eerie/uncanny" represent two different constructs (statistical anomality and a subjective negative experience), it has been noted that the uncanny valley is marked by both (Diel et al., 2021), for example in that the subjective negative experience may emerge from a perception of strangeness due to a stimulus' atypicality or deviation (Diel & MacDorman, 2021; Kätsyri et al., 2015). Hence, measurements sensitive to both constructs are adequate at capturing the uncanny valley effect and thus scales are often used in combination (e.g., Ho & MacDorman, 2017; Kätsyri, Mäkäräinen, & Takala, 2017).

Voices were presented in a random order for each participant and were replayed for each item. Participants had an unlimited amount of time responding to the items. Because uncanniness and human likeness are here understood as subjective experiences and assessments, the terms were presented with minimal information to the participants to

gauge their own interpretations.

2.1.4. Categorization task

Categorization task followed the rating task for each stimulus. For the categorization task, all sounds except for the normal and distorted cat meow voices were used. For each presented sound, participants had to do a two-alternative forced choice task on whether the voice was humanlike or not. Participants first heard 2 s of the sound before the choice text appeared, at which point participants had the ability to decide by pressing either the left or right key on their keyboard. Participants were instructed to be as accurate and fast as possible.

2.1.5. Procedure

The whole procedure was conducted online on the platform pavlovia (pavlovia.org). After giving informed consent and filling out a demographic questionnaire asking for participants' gender and age, participants were redirected to the experiment. They first went through the rating task followed by the categorization task. For both tasks, all stimuli were together presented in a random order.

The human likeness ratings were used to operationalize the x-axis of the uncanny valley function. Meanwhile, human categorization responses (both reaction times and response inconsistencies) were used as indicators of categorical ambiguity.

2.1.6. Analysis and ethics statement

Analysis was conducted via R. Linear mixed models (LMM) were used to control for participants and stimulus as random effects, as well as analyses of variance (ANOVAs) and linear regressions. Specifically, stimulus and participant were used as random effects and random slopes in LMM analyses to control for the repeated measures design and the repeating base stimuli for the distorted and undistorted voice stimuli. For LMM analysis, the function *lmer()* using packages *lme4* and *lmerTest* was used with degree of freedom estimation based on Satterthwaite's method. Effect sizes of LMMs are reported as R^2 calculated according to Nakagawa and Schielzeth (2013) and Johnson (2014). The assumption of normality of residuals was checked using QQ-plots (Figure A5).

Data cleaning was conducted by removing all outlier (1.5*IQR) uncanniness (index) and human likeness ratings, and categorization reaction times for each stimulus on a trial level. A total of 17 trials were removed and not used in the analyses. The experiment was approved by the Cardiff University School of Psychology Ethics Committee in October 2021 (reference number: EC.21.09.14.6411G). All methods were performed in accordance with the Declaration of Helsinki and informed consent was collected from all participants.

2.1.7. Data availability

Stimuli and datasets generated and analysed during the current studies and the analysis scripts are available on OSF: https://osf.io/7xs6j.

2.2. Results

The eerie/uncanny and strange/weird items were combined into an uncanniness index with a Cronbach's alpha of $\alpha = 0.79$, indicating acceptable, almost good construct validity.

2.2.1. Voice distortion: human vs cat

A within-subject 2x2 ANOVA was conducted with distortion (normal vs distorted) and species (cat vs human) as factors of uncanniness. The analysis showed main effects of both distortion ($F(1,48) = 567.02$, $p < 0.001$, $d = 0.77$) and species ($F(1,48) = 51.84$, $p < 0.001$, $d = 0.20$), as well as an interaction between these two ($F(1,48) = 47.35$, $p < 0.001$, $d = 0.15$). The interaction is visualized in Fig. 2.

Follow-up p-adjusted post-hoc Tukey tests showed that distortion increased the uncanniness of both cat ($t(1825) = 33$, $p_{adj} < 0.001$, $d = 2.16$) and human voices ($t(1825) = 52.48$, $p_{adj} < 0.001$, $d = 3.43$).

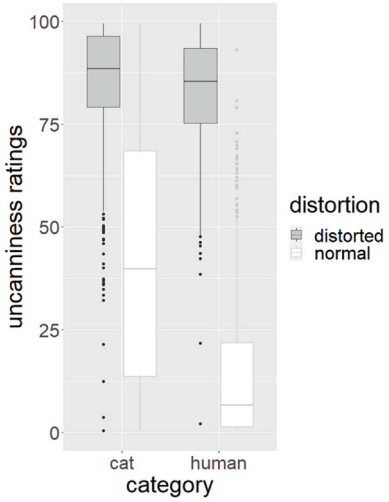

Fig. 2. Boxplots showing uncanniness ratings across species (category) and distortion condition including box plots. Error bars (wider than the boxplots) represent standard errors.

Furthermore, normal human voices were significant less uncanny than cat voices ($t(1825) = 21.328$, $p_{adj} < 0.001$, $d = 1.39$), but not in the distortion conditions ($t(1825) = 1.82$, $p_{adj} = 0.19$, $d = 0.12$). Thus, the same distortion procedure increased the uncanniness of human voices more than the uncanniness of cat voices. Hypothesis 1 is thus supported.

2.2.2. An uncanny valley of voices

An uncanny valley of voice stimuli was investigated using a linear mixed model with realism ratings as fixed effects and participants and stimuli as random effects and random slopes on uncanniness, using the following formula: uncanniness ~ humanlike + humanlike^2 + humanlike^3 + (1+humanlike|stims) + (1+humanlike|participant). For quadratic and linear models, the formula was adapted by removing cubic and quadratic terms respectively. Normality of residuals was confirmed by investigating QQ-plots (Figure A5A).

Cat sounds were excluded from the analysis to focus on humanlike and mechanical voices. Results show that a cubic term ($t(2419) = -4.17$, $p = 0.007$, $R^2_{adj} = 0.81$, 95% CI [−0.0001, −0.00004]) could explain the variance better than a linear term ($\chi^2 = 45.13$, $p < 0.001$) or a quadratic term ($\chi^2 = 17.34$, $p < 0.001$). The model is plotted in Fig. 3. As can be seen in the plot, confidence intervals in the curves' "valleys" (i.e., the boundaries of the grey areas in the local minima) do not overlap with the confidence intervals (boundaries of the grey areas) of the curves' maxima. Taken together with the significant cubic term, a non-monotonic relationship explains the relationship between uncanniness and realism across voice categories.

In a second step, distorted and pathological voices were removed and the analysis was redone. Normality of residuals was confirmed by investigating QQ-plots (Figure A5B). The results show that neither the linear, quadratic, nor cubic term was significant. The formulas were identical to the ones reported above. The function, depicted in Fig. 4, however does not reflect an uncanny valley plot: Given that at no point in the functions in Fig. 4, the confidence intervals seem to significantly decrease, but only increase with increasing realism, the function indicates a monotonic relationship between uncanniness and realism. Thus, a non-monotonic relation between uncanniness and human likeness seems to result from a combination of multiple stimulus categories. Thus, hypotheses 2 and 3 are supported.

2.2.3. Categorization difficulty as a predictor of voice uncanniness

A linear mixed model with reaction time as a fixed effect and stimuli and participants as random effects and random slopes on uncanniness showed that reaction time could not predict voice uncanniness ratings (t

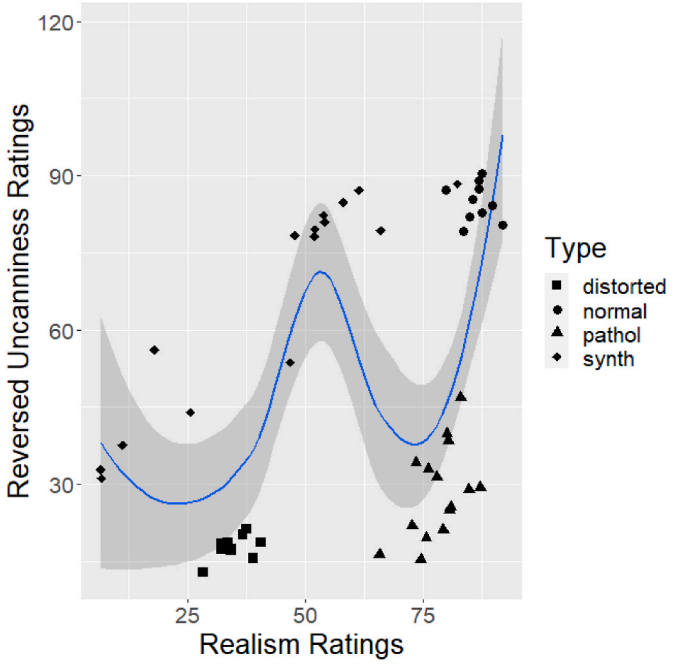

Fig. 3. Reversed uncanniness ratings plotted as a function of realism ratings across the four voice conditions. The blue line represents the best fit local regression line. Each point represents a stimulus. The grey area represents the 95% confidence interval of the running mean. "Reversed" indicates uncanniness ratings subtracted by hundred (higher score equals to lower uncanniness ratings). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

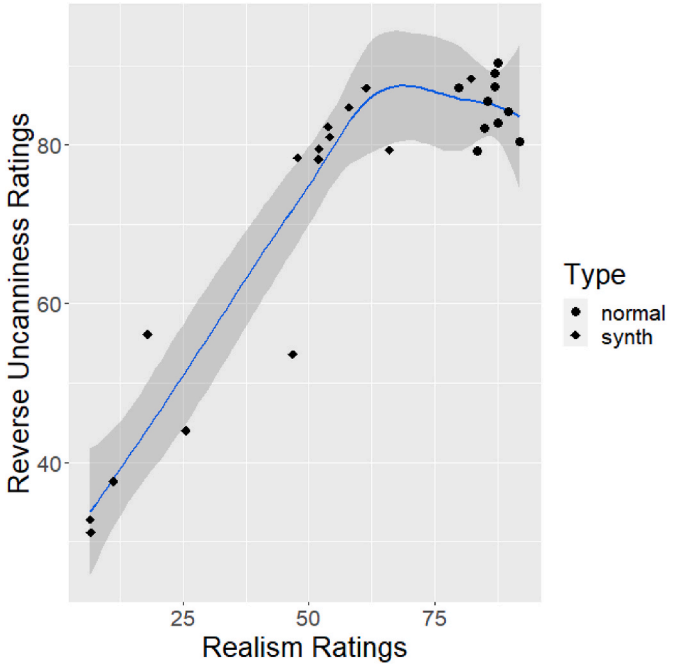

Fig. 4. Reversed uncanniness ratings across realism ratings when only normal human and synthetic voices are used. The blue line represents the best fit local regression line. Grey areas represent 95% confidence intervals of the running mean. "Reversed" indicates 100 – uncanniness ratings. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

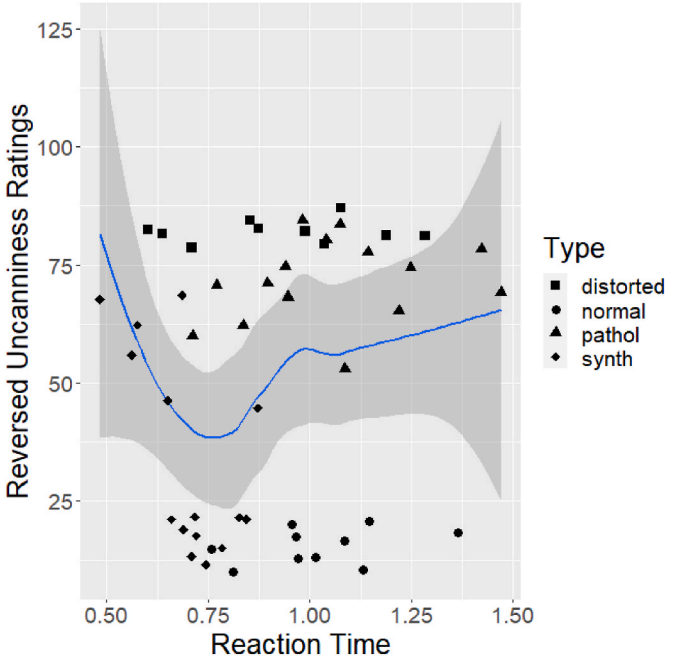

Fig. 5. Reversed uncanniness ratings plotted against categorization reaction time. No significant relation between the variables was found. Each point represents a stimulus. The blue line represents the best fit local regression line. Grey areas represent 95% confidence intervals of the running mean.
Note. The model was plotted using the following formula: uncanniness ~ rt + (rt|stims) + (rt|participant; with *rt* referring to reaction time. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

(27) = 0.7, $p = 0.489$, 95% CI [−0.63, 1.25]). Normality of residuals was confirmed via QQ-plots (Figure A5C). The data is plotted in Fig. 5.

Voice categorization data was transformed into a voice certainty variable by coding participants' non-human categorizations as 0 and human categorizations as 1. Absolute average categorizations were then subtracted by 0.5, creating a range from 0 (inconsistent categorization, equivalent to 50:50 categorization) and 0.5 (consistent categorization).

Because the transformed data was already aggregated across participants for each stimulus, a linear regression model was used to investigate the effect of categorization certainty on uncanniness. The results show that categorization certainty could not predict voice uncanniness ratings ($t(50) = 0.15$, $p = 0.88$, 95% CI [−93.2, 108.48]), and the data is visualized in Fig. 6.

Post-hoc tests were conducted to test differences between conditions: While distorted voices were more ambiguous and uncanny than synthetic (ambiguous: $t(46) = -4.553$, $p < 0.001$; uncanny: $t(46) = 9.192$, $p < 0.001$) or human voices (ambiguous: $t(46) = -3.197$, $p = 0.008$; uncanny: $t(46) = 11.59$, $p < 0.001$), pathological voices were more uncanny than synthetic ($t(46) = 8.03$, $p < 0.001$) and human voices ($t(46) = 10.69$, $p < 0.001$), while not being more ambiguous (synthetic: $t(46) = 1.29$, $p = 0.475$; human: $t(46) = -1.05$, $p = 0.621$).

As neither reaction time nor categorization ambiguity predicted uncanniness, hypothesis 4 was not supported.

2.2.4. Human categorization as a moderator of human-deviation on uncanniness

To test for a potential moderation effect, a post-hoc linear regression analysis has been conducted for the interaction between categorization response (human vs non-human) and human likeness on uncanniness. The results show main effects of response ($t(46) = 10.011$, $p < 0.001$, 95% CI [286.9, 431.3]), human likeness ($t(46) = -8.922$, $p < 0.001$, 95% CI [−2.14, 1.35]), and an interaction between these two ($t(46) = -6.163$, $p < 0.001$; $R^2_{adj} = 0.80$, 95% CI [−3.38, 1.72]). Thus, a

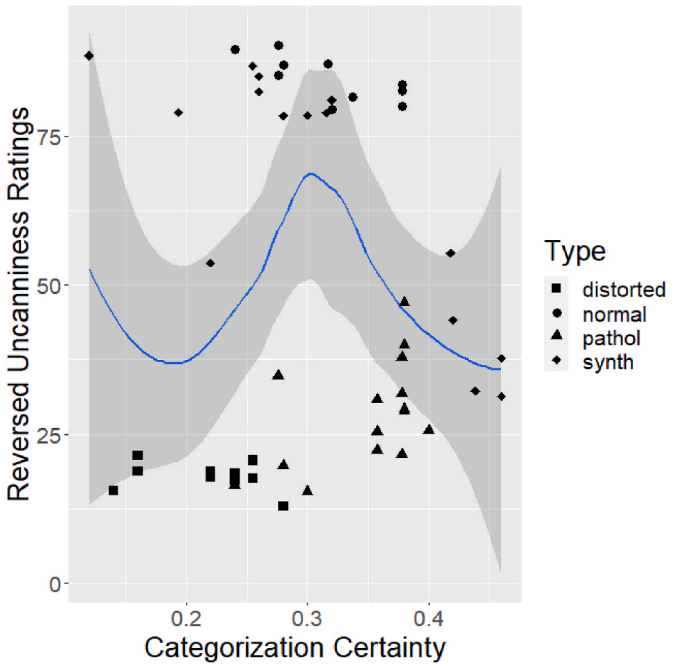

Fig. 6. Uncanniness ratings plotted against voice categorization consistency (0 = inconsistent categorization, 0.5 = consistent categorization across participants). Each point represents a stimulus. The blue line represents the best fit local regression line. Grey areas represent 95% confidence intervals of the running mean.
Note. The model was plotted using the following formula: uncanniness ~ resp; with *resp* indicating categorization certainty. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

moderated linear relationship between human likeness, uncanniness, and categorization as "human" is indicated. The model was plotted using the following formula: (uncanniness ~ resp*humanlike), with *resp* indicating categorization certainty.

2.3. Discussion

2.3.1. Voice distortion and specialization

Voice distortion created by multiplying the fundamental frequency by 1000 increased the uncanniness of both human cat voices. The increase was stronger for human compared to cat voices. A higher degree of specialization to a voice category may sensitize uncanniness caused by deviation.

Differences in fine details between human voices carry vital information about spoken messages and characteristics and states of the speaker (Kreiman & Gerratt, 2005; Kreiman, Gerratt, Precoda, & Berke, 1992). The recognition of analogous information is less important for the perception of cat voices. Thus, the degree of specialization (and change sensitivity) in humans is lower for cat compared to human sounds. Higher uncanniness sensitivity for human compared to cat voices can thus be explained by higher specialization to typical voice patterns and sensitivity to deviations from these patterns.

2.3.2. An uncanny valley of voices

A function with only synthetic and normal human voices showed a monotonic relationship between human likeness and uncanniness akin to previous research (Baird et al., 2018; Baird, Jorgensen, et al., 2018; Kimura & Yotsumoto, 2018; Kühne et al., 2020; Romportl, 2014). However, adding voices that are either deliberately distorted or naturally deviating produces a non-monotonic function of uncanniness and human likeness. When excluding either distorted or pathological voices, the hypothetical new curve would appear similar to an *N*-shaped uncanny valley plot, and the deviating voices would lie within an uncanny valley akin to the prediction of dead bodies falling into an uncanny valley (Mori et al., 2012). This is especially the case for pathological voices, as removing the distorted voices would create an uncanny valley curve with mechanical sounds at the non-human section, TTS voices at the pre-valley section, and pathological voices within the valley part of the function, creating a "proper" uncanny valley. Such an interpretation would favor explanations of the uncanny valley related to mortality salience or disease avoidance (MacDorman & Ishiguro, 2006). Meanwhile, the observation of a "smaller" uncanny valley with distorted voices is consistent with proposals that multiple uncanny valleys exist along the human likeness axis that may emerge due to different mechanisms (Kim, de Visser, & Phillips, 2022).

Previous researchers have noted that an uncanny valley could occur at any point at a graph, allowing multiple valley-shaped functions, potentially due to a multicausal emergence of the effect (Bartneck, Kanda, Ishiguro, & Hagita, 2009; Diel & MacDorman, 2021; Hanson, 2006; Kim et al., 2022; Yam, Bigman, & Gray, 2021). An uncanny valley may not necessarily occur on just one area on the human likeness axis, and polynomial functions more complex than an N-shaped curve may occur depending on the stimuli selected, as in this study.

2.3.3. Categorization ambiguity does not predict uncanniness

Categorization ambiguity has been proposed to underlie the uncanny valley effect (Cheetham et al., 2013; Weis & Wiese, 2017; Yamada, Kawabe, & Ihaya, 2013). This study failed to find evidence for the categorization ambiguity hypothesis: Neither categorization reaction time nor categorization response consistency could predict uncanniness ratings. While distorted voices were both uncanny and difficult to categorize, pathological voices were not. Categorization ambiguity may correlate with stimulus deviations when stimuli are incremental morphs between two easily categorizable stimuli (Yamada et al., 2013) and thus may be both uncanny and difficult to categorize due to their deviation. However, certain stimuli can be uncanny despite being easy to categorize (Chattopadhyay & MacDorman, 2016; Diel & MacDorman, 2021; Mathur et al., 2020). Thus, uncanniness cannot be explained solely by categorization ambiguity.

2.3.4. A moderator of uncanniness?

The model plotted in Fig. 2 indicates a *W*-shaped relationship with "two valleys". Such a relationship may be a consequence of choosing different categories of voices which interact differently with human likeness to affect uncanniness. The effect of voice type could be moderated by a variable influencing the perception of a decrease in realism (or closeness to the human norm) on uncanniness. Hence, a third variable may underlie the observed data by moderating a linear relationship between human likeness and uncanniness.

It has been recently suggested that the uncanny valley may be explained by a moderated linear function: perceptual specialization increases the uncanniness of distortions, leading to an increase of uncanniness for deviating stimuli close with a realistic appearance (Diel & Lewis, 2024). For voices, humans may be especially sensitive to deviations in voices close to natural human voices (i.e., those categorized as human voices), increasing uncanniness if such otherwise natural human voices are deviating – as would be the case for the pathological voices used in this study.

Concordantly, a significant interaction between human categorization and human likeness was found that could explain uncanniness better than a polynomial model of human likeness. Categorization as human sensitized the effect of deviation on uncanniness. As dehumanization can decrease the uncanniness of androids (Yam et al., 2021), categorization as human may activate a stricter evaluation of stimuli based on their proximity to the human norm. As a humanization manipulation can affect the specialized processing of faces (Fincher & Tetlock, 2016; Fincher, Tetlock, & Morris, 2017), an increase of humanization (and human categorization) may also further sensitize the

detection of configural deviations and thus uncanniness.

Similarly, as mind perception increases configural processing (Deska, Lloyd, & Hugenberg, 2017), it may also increase the sensitivity to deviations and thus uncanniness when a stimulus is perceived as both having a mind and deviates from the norm of appearance. Finally, as the uncanny valley has been linked to perceptions of markers of death and disease avoidance (MacDorman & Ishiguro, 2006; Mori et al., 2012), the effect may be linked to the perception of organic appearance. Thus, a perceived high "organicness" of a voice may increase the sensitivity of uncanniness towards deviations from human likeness, potentially due to evolutionary disease avoidance mechanism.

However, human likeness and categorization choice was highly correlated in this study, decoupling human likeness (or deviation) from human categorization (or humanization) would be required, which however should be difficult given the conceptual similarity of these constructs.

3. Experiment 2

For Experiment 2, standardized voice selections of multiple voice categories (normal, distorted, and pathological human voices, synthetic voices) were rated on uncanniness, realism, organicness, animacy, and mind attribution by participants using self-assessment rating scales. Ranges of distorted voices were created by multiplying normal voice fundamental frequencies by multiples of 250. Normal and ranges of pathological human voices were selected from the PVQD (Walden, 2022). Synthetic voices were identical to Experiment 1. Methods are described in detail in section 6.

The aim of Experiment 2 is to investigate a potential third variable that may moderate a monotonic effect of human likeness on uncanniness. As the uncanny valley may be explained by a moderated linear function for which specialization increases the sensitivity to deviations (Diel & Lewis, 2024), a variable associated with perceptual specialization should increase the uncanniness caused by incremental voice distortions. Several candidates for this third variable were explored.

Pathogen avoidance: Perception of organic voice. Uncanniness may be a response to the detection of indicators of contagious disease (MacDorman & Entezari, 2015; MacDorman & Ishiguro, 2006; Moosa & Ud-Dean, 2010). Disease indicators may appear as physical anomalies or deviations co-occurring with pathology or physical disabilities (Park, Faulkner, & Schaller, 2003; Workman et al). As disease threat is only relevant for organic material, the perception of an entity being organic (vs synthetic) should then increase negative response towards norm deviation in a stimulus. A high specialization for organic-appearing stimuli would allow the detection and negative evaluation of deviating stimuli which may then protect against disease and contamination. Meanwhile, a voice recognized as inorganic should pose no disease-related threat even despite deviating from the norm. Thus, hypothesis 1 is as follows:

1. Perception of organicness moderates the relation between human likeness and uncanniness across voice categories

Mind attribution and animacy. Uncanniness may be elicited when human qualities like mind or animacy are attributed to non-human entities (Gray & Wegner, 2012; Stein & Ohler, 2017). Mind perception relates to the attribution of the ability to have subjective experiences (e.g., sense or feel), and has been associated with uncanniness in otherwise non-humanlike entities (e.g., a presumed supercomputer; Gray & Wegner, 2012). Attributions of human qualities may increase the subjective importance of a stimulus, leading to a higher level of specialized process which would make it more likely to detect (and negatively evaluate) potentially deviating information. Thus, less humanlike voices not perceived as having a mind or being animate should not elicit uncanniness, while deviating voices which appear to have a mind or to be animate should be uncanny. Thus, hypotheses 2 and 3 are as follows:

2. Attribution of mind moderates the relation between human likeness and uncanniness across voice categories
3. Perception of animacy moderates the relation between human likeness and uncanniness across voice categories

3.1. Methods

3.1.1. Participants

According to a power analysis, $n = 35$ participants are sufficient to exceed a power of $1\text{-}\beta = 0.8$ for a within-subject design with a standard effect size of $d = 0.5$ (Cohen, 1988). Participants were Psychology students at the Cardiff University School of Psychology, recruited via the Experimental Management System (EMS). Participants were on average 19.26 years old ($SD_{age} = 1.29$), 34 identified as female and one as male.

3.1.2. Stimuli

Per category (distorted, normal, pathological, synthetic), five stimuli were selected from Experiment 1. In addition, variation of distortion degree was created for distorted and pathological voices: For distorted voices, fundamental frequencies of normal (base) voices were increased by 250, 500 and 750, in addition to the present distorted voices with an increase by the value of 1000. These distortion levels were created to simulate an incremental increase of distortions starting with the normal counterparts. As the goal of the experiment is to investigate a moderated linear function of uncanniness, an incremental increase of distortion may reflect a linear function for one value of the moderator variable. For pathological voices, additional sets of five voices were selected based on the level of perceived severity ratings as reported in the PVQD (Walden, 2022). The five most severe pathological voices were selected for the severity rating maxima of 100, 75, 50, and 25, out of a range of 0–100 (i.e., for 100 the most severe voices with a maximum severity of 100 were selected; for 75, the most severe voices with a maximum severity of 75 were selected; etc.). Spoken sentences were the same as in Experiment 1. In addition, the 15 synthetic voices from Experiment 1 were used. The stimuli are summarized in Table A2. All stimuli are available in the OSF repository linked below.

3.1.3. Procedure: rating task

The experiment consisted only of a rating task conducted online on the platform pavlovia (pavlovia.org). The rating task was identical to the one in Experiment 1, except participants rated each voice based on the items eerie, strange, and humanlike only, in addition to its perceived animacy ("not animate" to "animate", mind attribution ("has not mind" to "has a mind"), and organicness "not organic" to "organic") from 0 to 100. The additional rating scales were presented the same way as the previous ones described in Experiment 1. Stimuli were presented together in a random order.

3.1.4. Analysis, ethics statement, and data availability

Analysis was conducted via R. Linear mixed models were used to control for participants, as well as analyses of variance (ANOVAs) and linear regressions. Data cleaning was conducted by removing all outlier (1.5*IQR) uncanniness (index) and human likeness ratings, and categorization reaction times for each stimulus across all subjects on a trial level. A total of 13 trials were removed and not used in the analyses. For LMMs, stimulus and participant were used as random effects and random slopes in LMM analyses to control for the repeated measures design and the repeating base stimuli for the distorted and undistorted voice stimuli. For LMM analysis, the function *lmer()* using the packages *lme4* and *lmerTest* was used with degree of freedom estimation based on Satterthwaite's method. Effect sizes of LMMs are reported as R^2 calculated according to Nakagawa and Schielzeth (2013) and Johnson (2014). The assumption of normality of residuals was checked using QQ-plots (Figure A5). All methods were performed in accordance with the Declaration of Helsinki and informed consent was collected from all

participants.

3.1.5. Data availability

Stimuli and datasets generated and analysed during the current studies and the analysis scripts are available on OSF: https://osf.io/7xs6j.

3.2. Results

Eerie and strange items were combined into an uncanniness index with a Cronbach's alpha of $\alpha = 0.8$, indicating good consistency.

3.2.1. Moderating effects

Linear mixed models with human likeness and either animacy (formula: (uncanny ~ humanlike * animate + (animate|stims) + (1+animate animate|participant)), mind attribution (formula: (uncanny ~ humanlike * mind + (mind |stims) + (mind |participant)), or organicness (formula: (uncanny ~ humanlike * organic + (organic|stims) + (organic|participant)) as fixed effects stimuli and participants as random effects showed that the interaction between human likeness and animacy ($t(612) = -2.18$, $p = 0.03$, $R^2_{adj} = 0.65$, 95% CI [−0.003, −0.0001]), mind attribution ($t(421) = 2.17$, $p = 0.03$, $R^2_{adj} = 0.62$, 95% CI [0.0001, 0.002]), or organicness ($t(656) = -2.96$, $p = 0.003$, $R^2_{adj} = 0.63$, 95% CI [−0.004, −0.0007]) each significantly predicted uncanniness. Normality of residuals was confirmed by investigating QQ-plots (Figure A5C-A5F).

To test whether a moderated function can explain uncanniness better than a quadratic function of human likeness, the linear moderator models were tested against a quadratic human likeness function. A quadratic human likeness model (formula: uncanny ~ humanlike + humanlike^2 + (1+humanlike^2|stims) + (1+humanlike^2|participant)) was able to predict uncanniness ($t(109) = -2.96$, $p = 0.004$, $R^2_{adj} = 0.84$, 95 % CI [−0.006, −0.002]). Model comparisons showed that the moderating models with animacy ($\chi^2 = 764$, $p < 0.001$), mind attribution ($\chi^2 = 737$, $p < 0.001$), and organicness ($\chi^2 = 834$, $p < 0.001$) fitted the data significantly better than the quadratic human likeness model. Thus, a moderated linear function of human likeness could explain the results better than a quadratic function of human likeness.

Relations between uncanniness and the other variables are depicted in Figures A1 to A4.

3.2.2. Differences between voice types

P-adjusted Tukey tests on differences between voice categories showed that distorted voices were more uncanny than normal ($t(56) = 6.789$, $p_{adj} < 0.001$) and synthetic voices ($t(56) = 7.097$, $p_{adj} < 0.001$). However, while distorted voices were perceived as less animate ($t(55) = -9.825$, $p_{adj} < 0.001$) and as having less mind ($t(55) = -9.725$, $p_{adj} < 0.001$) compared to normal voices, they did not differ from synthetic voices.

3.3. Discussion

3.3.1. "Uncanny valley" as a moderated linear function

A third variable of organicness, animacy, or mind attribution moderates a linear relationship between human likeness and uncanniness. A moderating function may appear as an increase of the slope with increasing organicness: While distinctively artificial voices can deviate from the human norm without suffering from uncanniness, deviations in organic-sounding voices may quickly become unnerving, for example due to the threat of contamination from infected organic entities.

These results are consistent with the prediction that the uncanny valley can be explained by a moderated linear function (Diel & Lewis, 2024): Specifically, organic voices may increase the level of specialized processing due to their proximity to natural human voices, which would also increase the sensitivity to deviations and their negative evaluation. Such an effect may reflect an evolutionarily adaptive response to protect against contamination, as especially organic and humanlike stimuli may carry diseases, which would be indicated by deviating appearance. A higher sensitivity to such deviations would thus work as a defensive mechanism.

However, all tested predictors were highly intercorrelated, and correlated highly with human likeness. Thus, it is not clear whether organicness itself is the third variable, or whether the third variable can be better described by a different construct.

3.3.2. Animacy and mind perception

Previous research aimed to explain the uncanny valley phenomenon through the attribution of humanlike characteristics like animacy or mind onto visibly artificial or inanimate stimuli[48]. However, the present results suggest that voice uncanniness also occurs for deviating voices clearly perceived as animate or having a mind (i.e., pathological voices). Meanwhile, artificially distorted voices perceived as inanimate or lacking mind were still uncanny. These results cannot be explained by misattribution of human qualities onto artificial entities.

4. General discussion

4.1. Uncanny valley of voices

In two experiments, non-monotonic relationships between uncanniness and human likeness for voices were observed (Fig. 2), although the function differs from a typical uncanny valley function. The cognitive processing underlying the uncanny valley effect may be analogous across visual and auditory domains. Distinct face and voice variants elicit stronger activity in neural substrates specific to these categories (Andics et al., 2010; Latinus, McAleer, Bestelmeyer, & Belin, 2013; Loffler et al., 2005), which may indicate increased processing need. Increased processing need may in turn decrease the aesthetic appeal of a stimulus (Winkielman, Schwarz, Fazendeiro, & Reber, 2003). Alternatively, a higher specialization with a face or voice category may sensitize to errors or deviations, leading to prediction error signals (Friston & Kiebel, 2011; Saygin et al., 2012).

The present results are the first show that uncanny valley effects can be replicated in a non-visual (in this case auditory) modality, and that similar perceptual mechanisms (the negative evaluation of deviating information) may underlie visual and auditory uncanny valley effects. Furthermore, and in accordance with previous research, the results highlight that synthetic (TTS) voices successfully avoid an uncanny valley of voices.

4.2. Synthetic voices and the uncanny valley

Synthetic voices were allocated around an uncanny valley of voices (Fig. 2), and when only observing synthetic and human voices, they form a monotonic function without a valley (Fig. 3). Some modern TTS synthetisation may have the potential to successfully replicate human voices. Specifically, participants consistently rated one of the Watson voices to be about as humanlike as typical human voices (however, the same voice was ambiguously categorized with a 53% human categorization rate). Thus, the results indicate some synthetic voices may even overcome an uncanny valley in terms of human likeness ratings.

It may be easier to replicate a synthetic voice than a synthetic face without errors: Synthetic voice replication can rely on recorded natural voices while synthetic faces must be artificially reconstructed. Alternatively, as human identity discrimination ability is more sensitive to faces than to voices (Barsics, 2014), visual human processing may also be more sensitive to deviations compared to auditory human processing, making errors in design more apparent and appalling.

Our results replicate previous findings on an absence of an uncanny valley when only natural and synthetic (i.e., no distorted) voices are used (e.g. Kühne et al., 2020; Schreibelmayr & Mara, 2022). Some synthetic voices even managed to overcome the vocal uncanny valley in

the current study. These results cast a favourable light onto the development of synthetic voices for human-computer interaction: For a voice to be uncanny, it ought to be noticeably distorted, which appears to be easier to avoid for the development of synthetic voices compared to artificial human faces. Especially with increasingly sophisticated tools to create synthetic voices using artificial intelligence (AI; Amershi et al., 2019; Chang et al., 2020), future synthetic voices are likely suitable to avoid falling into an uncanny valley.

It remains an open question whether higher level processing aspects of synthetic voices may still trigger eerie perceptions, such as adequately adapting prosody and affect according to the social interaction. A synthetic voice may appear eerie if its affective intonation does not fit the social setting (e.g., being cheerful in a sad context). Research also shows a dislike of synthetic voices if voice emotion and content emotion do not match (Nass, Foehr, Brave, & Somoza, 2001). Correctly recognizing the interaction partner's affect and adequately responding to it them is an essential ability of artificial social agents (Picard, 2000), and failing to do so may elicit uncanny responses in otherwise acceptable synthetic voices.

4.3. Theories on the uncanny valley

The present results conflict with two existing theories on the uncanny valley: That uncanniness is caused by either 1) categorical ambiguity or categorization difficulty, or 2) by misattribution of human qualities onto nonhuman entities. While distorted voices in Experiment 1 were both uncanny and categorically ambiguous, pathological voices were uncanny despite being clearly categorized as human (see Figs. 4 and 5). In Experiment 2, distorted voices were uncanny despite having less mind or animacy attributed to them than normal voices, and with no differences compared to synthetic voices. Furthermore, pathological voices were uncanny in both experiments, contrasting the misattribution theory's prediction that uncanniness is caused by non-human entities.

The present data can be better explained by a deviation-from-familiarity account (Diel & Lewis, 2022a; 2022b): both distorted and pathological voices may be uncanny because they deviate from the typical pattern of human voices. However, there are still differences between distorted and pathological voices in that pathological voices tended to be categorized consistently while distorted voiced did not, indicating that the voices differed in their level or type of distortion. Categorical ambiguity can correlate with stimulus uncanniness as categorically ambiguous stimuli (Yamada et al., 2013) also deviate from typical appearance. Similarly, mind attribution can enhance configural processing of faces (Deska, Almaraz, & Hugenberg, 2017), which in turn may sensitize the negative evaluation of deviations (Diel & Lewis, 2022b). Thus, mind attribution may increase uncanniness by sensitizing to deviations (Müller, Gao, Nijssen, & Damen, 2021; Yam et al., 2021; Yin, Wang, Guo, & Shao, 2021). The interaction between attribution of human qualities, degree of configural processing, and uncanniness sensitivity can be explored in future research.

4.4. A moderated monotonic function of uncanniness

Rather than being a non-monotonic, valley-shaped function, the uncanny valley may consist of two or more monotonic functions with different slopes (e.g., one for an increase of likability from synthetic to full human variants, and one for a decrease of uncanniness from deviating or abnormal to typical humanlike variants). To test this, both experiments have investigated a moderated linear function of uncanniness.

Experiment 1 found that a moderated linear function could predict uncanniness, and Experiment 2 found that it could explain uncanniness better than a non-linear function of human likeness. Although the specific moderating variables differed between experiments, both "human" categorization and perceived organicness increased the effect of deviation on uncanniness. However, both variables also highly correlated with human likeness.

The results are consistent with previous research on a moderated linear function underlying the uncanny valley (Diel & Lewis, 2024): A higher level of perceptual specialization (as would be the case for more humanlike and organic voices) may increase the sensitivity to deviations due to a deeper processing level, increasing the negative evaluation of such deviations.

The investigated moderator variables are evolutionarily sensible: Disease avoidance may underlie the uncanny valley effect (MacDorman & Entezari, 2015), and markers of infectious disease are expressed as changes from typical (human) appearance or behaviour (Park et al., 2003). Given that the threat of infection is present only in organic entities, avoidance of deviating organic or human entities should be effective for minimizing risk of infection. Meanwhile deviating yet clearly inorganic entities pose no threat of infection.

Alternatively, the increased uncanniness for less humanlike stimuli in organic entities or those categorized as hu man may be due to a higher level of perceptual experience with naturally humanlike stimuli: Perceptual expertise with a stimulus category increases the uncanniness of deviating exemplars (Diel & Lewis, 2022b).

4.5. Limitations and future directions

Interpretations of test results on a moderated linear function of the uncanny valley are limited due to the intercorrelation between the predictors. As multicollinearity cannot be excluded, the exact relationship between the predictor variables and uncanniness remains unclear. Future research may aim to tackle this problem using decorrelated predictors.

As the experiment was conducted online, no control of the participants' devices and sound systems was present. Potential differences in the quality of sound systems may have confounded the results. However, as participants were treated as a random effect in the analysis, participant-level confounding effects were controlled for.

As can be seen in the Figures for both experiments, stimuli tended to be clustered into groups representing the voice categories, as expected given the categorical nature of the stimuli. However, the lack of continuation complicates interpretations of the spaces between the stimuli. Future research can aim to select a broader range of stimuli, especially stimuli varying along vocal dimensions.

Voice distortions were created using equidistant multiplicator steps. However, alternative incremental distortions, e.g., on a logarithmic scale, may more adequately represent the relevance of changes in fundamental frequency for the auditory system. Future research may test incremental distortions of voice fundamental frequency using logarithmic steps.

Cat sounds were distinct from the other human voice categories. Their relative distinctiveness may have made them more uncanny, confounding the uncanny ratings in the process. Nevertheless, distortion effects on uncanniness were observed. In addition, undistorted cat sounds were already more uncanny than undistorted human voices, which may have decreased the degree of uncanniness that the cat voices could increase to through distortion. Future research may implement subtler distortions to investigate the effect of sound familiarity on distortion.

Figs. 2–5 indicate that synthetic voices were clustered into two groups – one close to normal human voices, the other at low levels of human likeness and with higher uncanniness. Thus, synthetic voices seem to appear in a wide range of human likeness levels, potentially due to differences in quality. Future research may aim to investigate which exact properties of synthetic voices influence their human likeness and uncanniness ratings including a wider selection of synthetic voices.

Specialization is experience-dependent. If uncanniness is caused by deviations from specialized categories, then manipulation of specialization should increase the uncanniness of a deviating stimuli given the same degree of deviation. Individuals who are experts in bird songs, for example, should expectedly be more sensitive to deviations in familiar

bird sounds, and are expected to find these deviations more uncanny, compared to a novice population. This could be investigated in future research.

The use of linguistic content in the stimuli adds additional dimensions which could have influenced the results. For the difference between distortion effects on human and cat voices, a reduced intelligibility of the human voices but not cat voices due to distortion may have been a reason for the increased uncanniness for distorted human voices. Similarly, as distorted and pathological voices could be less intelligible, the additional processing need for these voices could have been a cause of uncanniness.

5. Conclusion

Contrary to previous research, this work affirms the notion that near humanlike voices can appear uncanny. Modern synthetic voices successfully escape a vocal uncanny valley. Multiple theories on the uncanny valley have been tested, favouring deviation-based and disease avoidance accounts over categorical ambiguity or the perception of animacy. Furthermore, the results indicate that uncanniness of voices is best explained by a moderator of human categorization or perception of organicness on the effect of human likeness on uncanniness.

CRediT authorship contribution statement

Alexander Diel: Writing – review & editing, Writing – original draft, Visualization, Validation, Investigation, Formal analysis, Data curation, Conceptualization. **Michael Lewis:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data is available here: https://osf.io/7xs6j/

Appendix

Table A1
Detailed information on voice stimuli used in Experiment 1. Distorted voices are not listed as their values were identical to their typical voice counterparts. For pathological voices, PCQD database reference codes are added (Walden, 2020).

Voice type	Stimulus, PVQD reference code	Gender	Severity rating and diagnosis (pathological); Speaker/source (synthetic)	Duration (sec)
typical	1	Female		4
	2	Female		4
	3	Male		5
	4	Male		4
	5	Male		4
	6	Female		4
	7	Male		4
	8	Female		4
	9	Male		3
	10	Female		4
	11	Female		4
	12	Female		4
	13	Male		4
	14	Female		4
	15	Female		4
pathological	1 (PT008)	Female	98.67; Reinke's Edema	5
	2 (PT019)	Male	98.5; lesions	9
	3 (PT015)	Female	97.5; lesions	5
	4 (PT004)	Male	95.5; ulcerative laryngitis	4
	5 (PT001)	Male	89.17; Reinke's Edema	5
	6 (PT118)	Female	88.83; unilateral vocal fold paresis	5
	7 (PT058)	Female	88.17; atrophy, MTD	4
	8 (PT050)	Male	87.33; lesions	4
	9 (LA5003)	Female	86.33; NA	8
	10 (NYU1022)	Male	86; vocal fold paresis	5
	11 (PT126)	Female	85.5; unilateral vocal fold paresis	4
	12 (PT032)	Female	85.33; MTD	7
	13 (PT046)	Female	83.83; unilateral vocal fold paresis	3
	14 (PT136)	Male	81.5; unilateral vocal fold paresis	5
	15 (PT065)	Female	78.17; Reinke's Edema	4
synthetic	1		eSpeak (Stephen Hawking voice generator)	4
	2	Male	Google	3
	3	NA	Mechanical sounds	5
	4	NA	Mechanical sounds	5
	5	Male	Microsoft Azure	3
	6	Female	Microsoft Azure	5
	7	Female	Microsoft Azure	5
	8	Male	Microsoft Sam	4
	9	NA	R2D2 sounds	5
	10	NA	R2D2 sounds	5

(*continued on next page*)

Table A1 (*continued*)

Voice type	Stimulus, PVQD reference code	Gender	Severity rating and diagnosis (pathological); Speaker/source (synthetic)	Duration (sec)
	11	Female	Watson	3
	12	Female	Watson	3
	13	Male	Watson	3
	14	Female	Watson	3
	15	Male	Watson	3

Table A2
Summary of voice stimuli used in Experiment 2.

Voice type	Stimulus, PVQD reference code	gender	Severity rating and diagnosis (pathological)	Duration (sec)
typical	1	Female		4
	2	Female		4
	3	Male		5
	4	Male		4
	5	Female		4
pathological	1 (PT026)	Female	23.83; muscle tension dysphonia	4
	2 (NYU1016)	Female	48.83; vocal fold paresis	4
	3 (NYU1025)	Male	74.33; vocal fold paresis	4
	4 (PT008)	Female	98.67; Reinke's Edema	4
	5 (PT135)	Female	23.67; muscle tension dysphonia, atrophy	4
	6 (PT053)	Female	48.17; muscle tension dysphonia	4
	7 (PT011)	Female	74; lesions	4
	8 (PT019)	Male	98.5; lesions	5
	9 (PT016)	Female	22.67; paradoxical vocal fold movement	4
	10 (PT130)	Female	47.17; adductor spasmic dysphonia	4
	11 (BL02)	Female	73.83; NA	5
	12 (PT015)	Female	97.5; lesions	5
	13 (LA7007)	Female	23.33; NA	4
	14 (PT030)	Female	46.17; leucoplakia	4
	15 (PT097)	Female	73.5; muscle tension dysphonia	5
	16 (PT004)	Male	95.5; ulcerative laryngitis	4
	17 (SJ5006)	Male	22.25; NA	4
	18 (PT047)	Male	46.17; unilateral vocal fold paresis	4
	19 (PT054)	Male	73.17; unilateral vocal fold paresis	7
	20 (PT001)	Male	89.17; Reinke's Edema	5

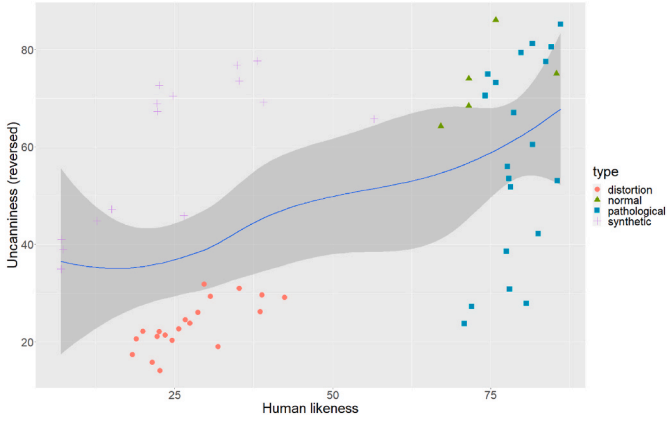

Fig. A1. Uncanniness plotted against human likeness across voice types. The blue line represents the best fitting weighted average, and the grey shaded area represents the 95% confidence range.

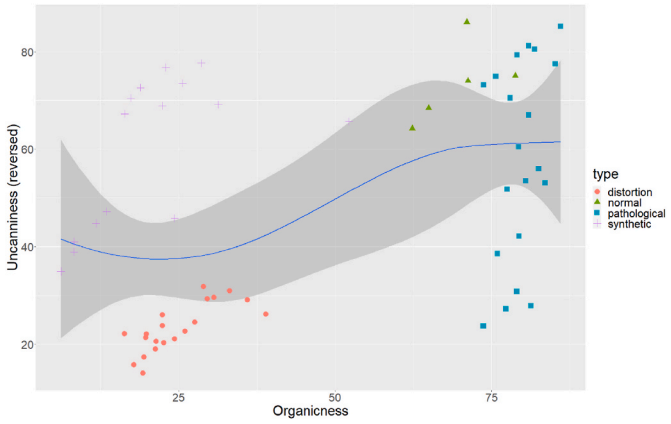

Fig. A2. Uncanniness plotted against organicness across voice types. The blue line represents the best fitting weighted average, and the grey shaded area represents the 95% confidence range.

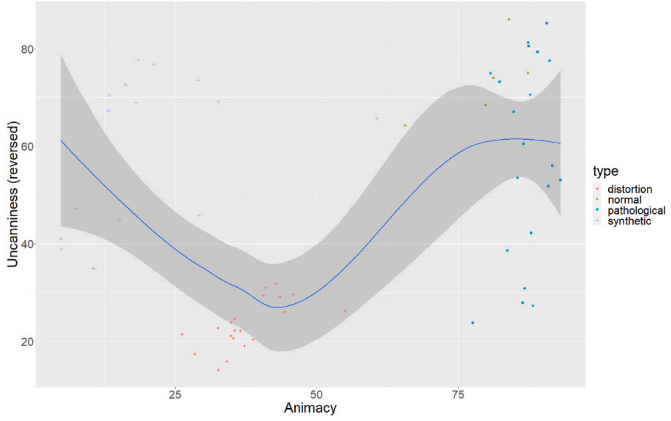

Fig. A3. Uncanniness plotted against animacy across voice types. The blue line represents the best fitting weighted average, and the grey shaded area represents the 95% confidence range.

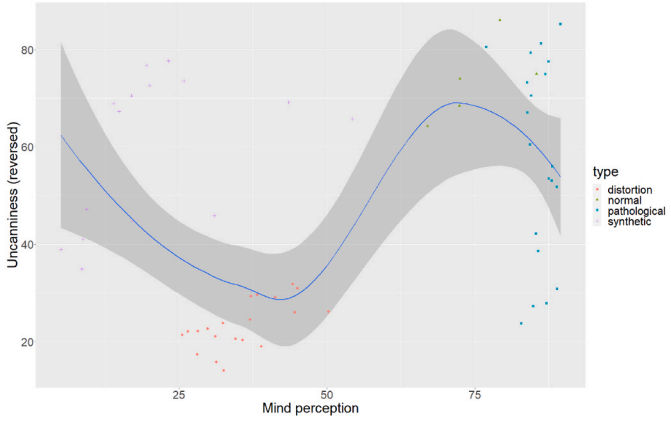

Fig. A4. Uncanniness plotted against mind attribution across voice types. The blue line represents the best fitting wei ghted average, and the grey shaded area represents the 95% confidence range.

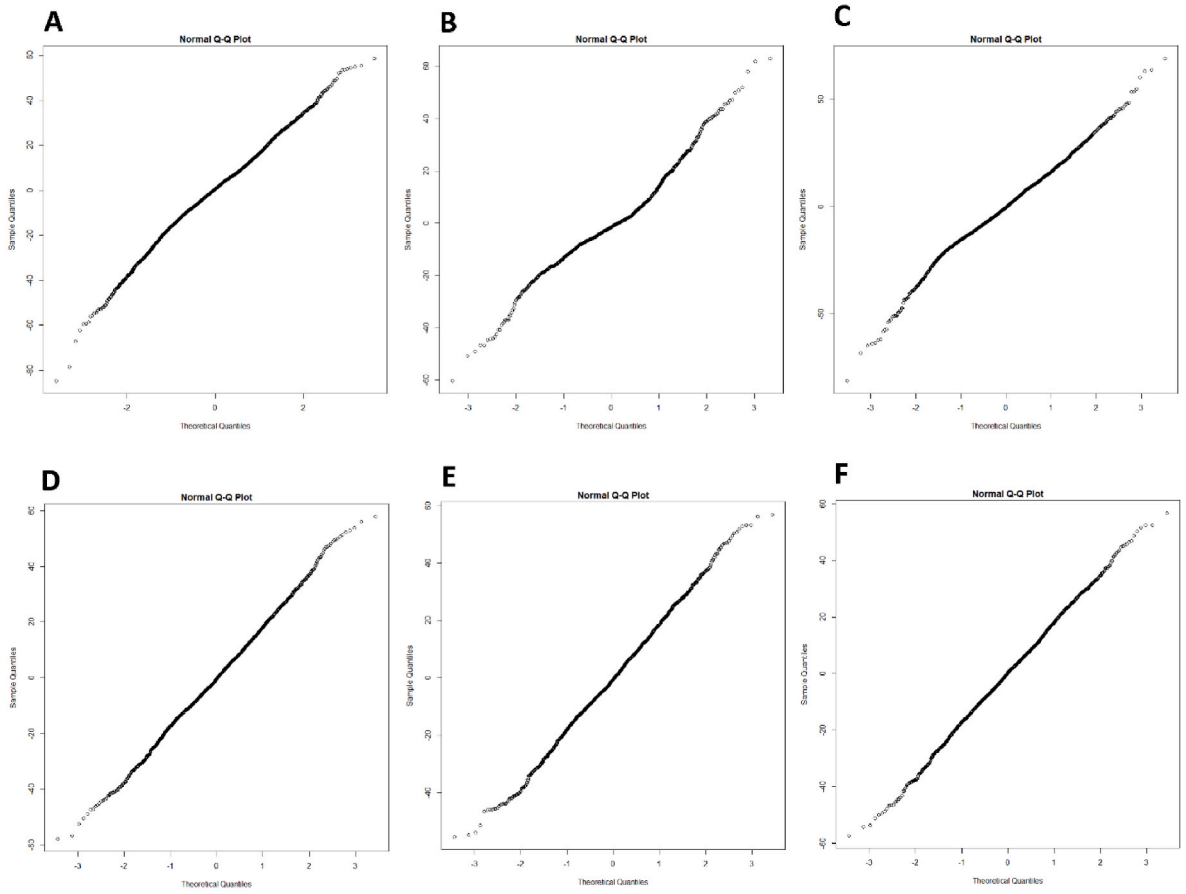

Fig. A5. QQ-Plots of the linear mixed models' residuals: a cubic function of human likeness for all human voices (A); a cubic function of human likeness excluding distorted and pathological human voices (B); the effect of reaction time on uncanniness (C); animacy as a moderator (D); mind perception as a moderator (E); organicness as a moderator (F).

References

Altenberg, E. P., & Ferrand, C. T. (2006). Fundamental frequency in monolingual English, bilingual English/Russian, and bilingual English/Cantonese young adult women. *Journal of Voice: Official Journal of the Voice Foundation, 20*(1), 89–96. https://doi.org/10.1016/j.jvoice.2005.01.005

Amershi, S., Weld, D. S., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., et al. (2019). Guidelines for human-AI interaction. In *Proceedings of the 2019 CHI conference on human factors in computing systems*.

Amir, O., & Levine-Yundof, R. (2013). Listeners' attitude toward people with dysphonia. *Journal of Voice: Official Journal of the Voice Foundation, 27*(4). https://doi.org/10.1016/j.jvoice.2013.01.015

Andics, A., McQueen, J. M., Petersson, K. M., Gál, V., Rudas, G., & Vidnyánszky, Z. (2010). Neural mechanisms for voice recognition. *NeuroImage, 52*(4), 1528–1540. https://doi.org/10.1016/j.neuroimage.2010.05.048

Baird, A. E., Jorgensen, S. H., Schuller, B., Cummins, N., Hantke, S., & Parada-Cabaliero, E. (2018). The perception of vocal traits in synthesized voices: Age, gender, and human-likeness. *Journal of the Audio Engineering Society, 66*(4), 277–285. https://doi.org/10.17743/jaes.2018.0023

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. *Interspeech*, 2863–2867. https://doi.org/10.21437/Interspeech.2018-1093

Barsics, C. (2014). Person recognition is easier from faces than from voices. *Psychologica Belgica, 54*(3), 244–254. https://doi.org/10.5334/pb.ap

Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2009). My robotic doppelganger - a critical look at the uncanny valley theory. In *Proceedings of the 18th IEEE international symposium on robot and human interactive communication, RO-man2009* (pp. 269–276). Toyama.

Carr, E. W., Hofree, G., Sheldon, K., Saygin, A. P., & Winkielman, P. (2017). Is that a human? Categorization (dis)fluency drives evaluations of agents ambiguous on human-likeness. Journal of experimental psychology. *Human Perception and Performance, 43*(4), 651–666. https://doi.org/10.1037/xhp0000304

Chang, M., Kim, T.-W., Beom, J., Won, S., & Jeon, D. (2020). AI therapist realizing expert verbal cues for effective robot-assisted gait training. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28*(12), 2805–2815. https://doi.org/10.1109/TNSRE.2020.3038175

Chattopadhyay, D., & MacDorman, K. F. (2016). Familiar faces rendered strange: Why inconsistent realism drives characters into the uncanny valley. *Journal of Vision, 16*(11), 7, 10.1167%2F16.11.7.

Cheetham, M., Pavlovic, I., Jordan, N., Suter, P., & Jancke, L. (2013). Category processing and the human likeness dimension of the uncanny valley hypothesis: Eye-tracking data. *Frontiers in Psychology, 4*, 108. https://doi.org/10.3389/fpsyg.2013.00108

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Deska, J. C., Almaraz, S. M., & Hugenberg, K. (2017). Of mannequins and men: Ascriptions of mind in faces are bounded by perceptual and processing similarities to human faces. *Social Psychological and Personality Science, 8*(2), 183–190. https://doi.org/10.1177/1948550616671404

Diel, A., & Lewis, M. (2022a). Familiarity, orientation, and realism increase face uncanniness by sensitizing to facial distortions. *Journal of Vision, 22*(4), 14. https://doi.org/10.1167/jov.22.4.14

Diel, A., & Lewis, M. (2022b). The deviation-from-familiarity effect: Expertise increases uncanniness of deviating exemplars. *PLoS One, 17*(9), Article e0273861. https://doi.org/10.1371/journal.pone.0273861

Diel, A., & Lewis, M. (2024). Rethinking the uncanny valley as a moderated linear function: Perceptual specialization increases the uncanniness of facial distortions. *Computers in Human Behavior, 108254*. https://doi.org/10.1016/j.chb.2024.108254

Diel, A., & MacDorman, K. F. (2021). Creepy cats and strange high houses: Support for configural processing in testing predictions of nine uncanny valley theories. *Journal of Vision, 21*(4). https://doi.org/10.1167/jov.21.4.1. Article 1.

Diel, A., Sato, W., Hsu, C. T., & Minato, T. (2023). The inversion effect on the cubic humanness-uncanniness relation in humanlike agents. *Frontiers in Psychology, 14*, Article 1222279. https://doi.org/10.3389/fpsyg.2023.1222279

Diel, A., Weigelt, S., & MacDorman, K. F. (2021). A meta-analysis of the uncanny valley's independent and dependent variables. *ACM Transactions on Human-Robot Interaction, 11*(1), 1–33. https://doi.org/10.1145/3470742

Eadie, T. L., Rajabzadeh, R., Isetti, D. D., Nevdahl, M. T., & Baylor, C. R. (2017). The effect of information and severity on perception of speakers with adductor spasmodic dysphonia. *American Journal of Speech-Language Pathology, 26*(2), 327–341. https://doi.org/10.1044/2016_AJSLP-15-0191

Fincher, K. M., & Tetlock, P. E. (2016). Perceptual dehumanization of faces is activated by norm violations and facilitates norm enforcement. *Journal of Experimental Psychology: General, 145*(2), 131–146. https://doi.org/10.1037/xge0000132

Fincher, K. M., Tetlock, P. E., & Morris, M. W. (2017). Interfacing with faces: Perceptual humanization and dehumanization. *Current Directions in Psychological Science, 26*(3), 288–293. https://doi.org/10.1177/0963721417705390

Friston, K. J., & Kiebel, S. (2011). Predictive coding: A free-energy formulation. *Predictions in the Brain*, 231–246. https://www.10.1093/acprof:oso/9780195395518.003.0076.

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

Hanson, D. (2006). Exploring the aesthetic range for humanoid robots. In *Proceedings of the ICCS/CogSci-2006 long symposium: Toward social mechanisms of android science*.

Ho, C.-C., & MacDorman, K. F. (2017). Measuring the uncanny valley effect: Refinements to indices for perceived humanness, attractiveness, and eeriness. *International Journal of Social Robotics, 9*(1), 129–139. https://doi.org/10.1007/s12369-016-0380-9

Johnson, P. C. (2014). Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution, 5*(9), 944–946. https://doi.org/10.1111/2041-210X.12225

Jung, N., Lee, M., & Choi, H. (2022). The uncanny valley effect for celebrity faces and celebrity-based avatars. *Science of Emotion and Sensibility, 25*(1), 91–102. https://doi.org/10.14695/KJSOS.2022.25.1.91

Kätsyri, J., Förger, K., Mäkäräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: Support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology, 6*, 390. https://doi.org/10.3389/fpsyg.2015.00390

Kätsyri, J., Mäkäräinen, M., & Takala, T. (2017). Testing the 'uncanny valley' hypothesis in semirealistic computer-animated film characters: An empirical evaluation of natural film stimuli. *International Journal of Human-Computer Studies, 97*, 149–161.

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008). Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In *2008 IEEE international conference on acoustics, speech and signal processing* (pp. 3933–3936). https://doi.org/10.1109/ICASSP.2008.4518514

Kim, B., de Visser, E., & Phillips, E. (2022). Two uncanny valleys: Re-Evaluating the uncanny valley across the full spectrum of real-world human-like robots. *Computers in Human Behavior, 135*, Article 107340. https://doi.org/10.1016/j.chb.2022.107340

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

Kreiman, J. E., Auszmann, A., & Gerratt, B. R. (2018). What does it mean for a voice to be "normal?". *Journal of the Acoustical Society of America, 143*, 1820–1820.

Kreiman, J., & Gerratt, B. R. (2005). Perception of aperiodicity in pathological voice. *Journal of the Acoustical Society of America, 117*, 2201–2211. https://doi.org/10.1121/1.1858351

Kreiman, J., Gerratt, B. R., Precoda, K., & Berke, G. S. (1992). Individual differences in voice quality perception. *Journal of Speech & Hearing Research, 35*(3), 512–520. https://doi.org/10.1044/jshr.3503.512

Kühne, K., Fischer, M. H., & 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*, Article 593732. https://doi.org/10.3389/fnbot.2020.593732

Latinus, M., McAleer, P., Bestelmeyer, P. E., & Belin, P. (2013). Norm-based coding of voice identity in human auditory cortex. *Current Biology: Cailiao Baohu, 23*(12), 1075–1080. https://doi.org/10.1016/j.cub.2013.04.055

Löffler, G., Yourganov, G., Wilkinson, F., & Wilson, H. R. (2005). fMRI evidence for the neural representation of faces. *Nature Neuroscience, 8*(10), 1386–1390. https://doi.org/10.1038/nn1538

MacDorman, K. F., & Entezari, S. O. (2015). Individual differences predict sensitivity to the uncanny valley. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems, 16*(2), 141–172. https://doi.org/10.1075/is.16.2.01mac

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

Mara, M., Appel, M., & Gnambs, T. (2022). Human-like robots and the uncanny valley: A meta-analysis of user responses based on the godspeed scales. *Zeitschrift für Psychologie, 230*(1), 33–46. https://doi.org/10.1027/2151-2604/a000486

Mathur, M. B., Reichling, D. B., Lunardini, F., Geminiani, A., Antonietti, A., Ruijten, P. A. M., et al. (2020). Uncanny but not confusing: Multisite study of perceptual category confusion in the uncanny valley. *Computers in Human Behavior, 103*, 21–30. https://doi.org/10.1016/j.chb.2019.08.029

Meah, L. F. S., & Moore, R. K. (2014). The uncanny valley: A focus on misaligned cues. In M. Beetz, B. Johnston, & M.-A. Williams (Eds.), *Social robotics: 6th international conference* (Vol. 8755, pp. 256–265). Cham, Switzerland: Springer. Lecture Notes in Computer Science.

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

Moosa, M. M., & Ud-Dean, S. M. (2010). Danger avoidance: An evolutionary explanation of uncanny valley. *Biological Theory, 5*, 12–14. https://doi.org/10.1162/BIOT_A_00016

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics and Automation Magazine, 19*, 98–100. https://doi.org/10.1109/MRA.2012.2192811

Müller, B. C. N., Gao, X., Nijssen, S. R. R., & Damen, T. G. E. (2021). I, robot: How human appearance and mind attribution relate to the perceived danger of robots. *International Journal of Social Robotics, 13*, 691–701. https://doi.org/10.1007/s12369-020-00663-8

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*.

Nass, C., Foehr, U. G., Brave, S., & Somoza, M. (2001). *The effects of emotion of voice in synthesized and recorded speech.* Retrieved from.

Nass, C., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied, 7*(3), 171–181. https://doi.org/10.1037/1076-898X.7.3.171

Niculescu, A., 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

Oord, A.V., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A.W., & Kavukcuoglu, K. (2016). WaveNet: A Generative Model for Raw Audio. ArXiv, abs/1609.03499.

Park, J. H., Faulkner, J., & Schaller, M. (2003). Evolved disease-avoidance processes and contemporary anti-social behavior: Prejudicial attitudes and avoidance of people with physical disabilities. *Journal of Nonverbal Behavior, 27*(2), 65–87. https://doi.org/10.1023/A:1023910408854

Picard, R. W. (2000). *Affective computing*. The MIT Press. https://doi.org/10.7551/mitpress/1140.001.0001

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review, 8*(4), 364–382. https://doi.org/10.1207/s15327957pspr0804_3

Romportl, J. (2014). Speech synthesis and uncanny valley. In P. Sojka, A. Horák, I. Kopeček, & K. Pala (Eds.), *Text, speech and dialogue. TSD 2014. Lecture notes in computer science* (Vol. 8655). Cham: Springer. https://doi.org/10.1007/978-3-319-10816-2_72.

Rosa, A. O.-L., Villacampa, J., Corradi, G., & Ingram, G. P. D. (2021). The creepy, the bad and the ugly: Exploring perceptions of moral character and social desirability in uncanny faces. In *Current psychology: A journal for diverse perspectives on diverse psychological issues*. Advance online publication. https://doi.org/10.1007/s12144-021-01452-w.

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (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

Schreibelmayr, S., & Mara, M. (2022). Robot voices in daily life: Vocal human-likeness and application context as determinants of user acceptance. *Frontiers in Psychology, 13*, Article 787499. https://doi.org/10.3389/fpsyg.2022.787499

Schroeder, S. R., Rembrandt, H. N., May, S., & Freeman, M. R. (2020). Does having a voice disorder hurt credibility? *Journal of Communication Disorders, 87*, Article 106035. https://doi.org/10.1016/j.jcomdis.2020.106035

Seaborn, K., Miyake, N. P., Pennefather, P. S., & Otake-Matsuura, M. (2021). *Voice in human–agent interaction* (Vol. 54, pp. 1–43). ACM Computing Surveys (CSUR).

Stein, J.-P., & Ohler, P. (2017). Venturing into the uncanny valley of mind—the influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition, 160*, 43–50. https://doi.org/10.1016/j.cognition.2016.12.010

Thomas, S., Ferstl, Y., McDonnell, R., & Ennis, C. (2022). Investigating how speech and animation realism influence the perceived personality of virtual characters and agents. In *IEEE conference on virtual reality and 3D user interfaces (VR), christchurch, New Zealand* (pp. 11–20). https://doi.org/10.1109/VR51125.2022.00018

Trovato, G., Ramos, J., Azevedo, H., Moroni, A., Magossi, S., Simmons, R., et al. (2017). A receptionist robot for Brazilian people: Study on interaction involving illiterates. Paladyn. *Journal of Behavioral Robotics, 8*(1), 1–17. https://doi.org/10.1515/pjbr-2017-0001

Walden, P. R. (2022). Perceptual voice qualities database (PVQD): Database characteristics. *Journal of Voice: Official Journal of the Voice Foundation, 36*(6), 875.e15–875.e23. https://doi.org/10.1016/j.jvoice.2020.10.001

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

Weis, P. P., & Wiese, E. (2017). Cognitive conflict as possible origin of the uncanny valley. *Proceedings of the Human Factors and Ergonomics Society - Annual Meeting, 61*(1), 1599–1603. https://doi.org/10.1177/1541931213601763

Whang, C., & Im, H. (2021). "I like Your Suggestion!" the role of humanlikeness and parasocial relationship on the website versus voice shopper's perception of recommendations. *Psychology and Marketing, 38*, 581–595. https://doi.org/10.1002/mar.21437

Winkielman, P., Schwarz, N., Fazendeiro, T. A., & Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Musch, & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 189–217). Lawrence Erlbaum Associates Publishers.

Workman, C. I.. Humphries, S., Hartung, F., Aguirre, G. K., Kable, J. W., & Chatterjee, A. Morality is in the eye of the beholder: The neurocognitive basis of the "anomalous-is-bad" steterotype. Annals of the New York Academy of Sciences, 1494(1), 3–17. https://doi.org/10.1111/nyas.14575..

Yam, K. C., Bigman, Y., & Gray, K. (2021). Reducing the uncanny valley by dehumanizing humanoid robots. *Computers in Human Behavior, 125*, Article 106945. https://doi.org/10.1016/j.chb.2021.106945

Yamada, Y., Kawabe, T., & Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon. *Japanese Psychological Research, 55*(1), 20–32. https://doi.org/10.1111/j.1468-5884.2012.00538.x

Yin, J., Wang, S., Guo, W., & Shao, M. (2021). More than appearance: The uncanny valley effect changes with a robot's mental capacity. In *Current psychology: A journal for diverse perspectives on diverse psychological issues*. Advance online publication. https://doi.org/10.1007/s12144-021-02298-y.

Zibrek, K., Cabral, C., & McDonnell, R. (2021). Does synthetic voice alter social response to a photorealistic character in virtual reality?. In *Proceedings of the 14th ACM SIGGRAPH conference on motion, interaction and games (MIG '21)* (Vol. 11, pp. 1–6) New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3487983.3488296.