



# The Uncanny Effect of Speech: The Impact of Appearance and Speaking on Impression Formation in Human–Robot Interactions

Neelu Gurung<sup>1</sup> · Janie Busby Grant<sup>2</sup> · Damith Hearth<sup>1</sup>

Accepted: 11 February 2023 / Published online: 13 April 2023  
© The Author(s) 2023

## Abstract

This study explores the impact of appearance and speech on human perceptions of faces in human-robot interactions. Three videos were generated depicting the real face of an artist and two virtual versions of the same artist, with increasing resolution and fidelity. Each video was presented with and without speech, with matching levels of fidelity to the faces (real human speech and machine-generated speech of two levels of realism). Participants viewed all six videos and rated them on measures such as convincing, trustworthy, realistic, likable, showed biological movement, reassuring, friendly, familiar, and humanness. We found that the inclusion of speech (real) had a significant positive impact on the impression formation of real human appearance. In contrast, perceptions of the virtual avatars were more negative when speech (machine-like) was incorporated. This study illustrates the differential effects that speech can have on virtual faces in interactive settings.

**Keywords** Uncanny · Uncanny effect · Uncanny valley effect · Speaking voice · Appearances · Speech

## 1 Introduction

Speech, which includes speaking voice and facial appearance, is one of the essential elements in forming a positive impression of a face. Many studies have investigated the causes associated with Freud [3] and Mori's [15] hypothesis; the “uncanny and uncanny valley” [1, 6, 8, 9, 13, 16]. However, few studies have investigated the effect of speaking voice on the positive impression formation of a face. The key contributions of our study are insights into how speaking voice and fidelity of digital faces can hinder or help positive and comfortable human-robot interactions.

This study aimed to assess whether the incorporation of a speaking voice makes a virtual robot face to be perceived

as more convincing, trustworthy, realistic, likable, showing biological movement, reassuring, friendly, familiar, and human-like. We aimed to validate the effects of the uncanny and the uncanny valley of speech. We used Stelarc's Prosthetic Head (PH), a virtual avatar, as the research platform. Stelarc is a well-known Australian performance artist. His work includes the ear on the arm (2003–2006), exoskeleton (2003), the third arm (1982), and Prosthetic head (2003). The original Prosthetic.

Head (PH V1) is a hand-crafted rendering of the artist's face animated using technologies over a decade ago as an art installation. A new version of the installation called the Prosthetic Head 2.0 (PH V2) was created recently using high-resolution images of the artist's face rendered using advanced photogrammetry techniques and animated using a state-of-the-art animation engine. Each Prosthetic Head version has an associated voice. For the baseline, we used a video recording of the artist, Stelarc speaking, which had the authentic voice of Stelarc. PH V2 and V1 speaking had machine-like voices based on two text-to-speech engines. PH V1 sounds more like a machine than PH V2. Using technically and visually different virtual faces of the same artist, with their associated voices, PH V1, PH V2, and the artist's (Stelarc) real face, this study explores the success of interaction between those faces and human participants by assessing how the faces are perceived on a range of measures

---

✉ Neelu Gurung  
neelu.gurung@canberra.edu.au

Janie Busby Grant  
janie.busbygrant@canberra.edu.au

Damith Hearth  
damith.herath@canberra.edu.au

<sup>1</sup> Faculty of Science and Technology, University of Canberra, Canberra, ACT 2617, Australia

<sup>2</sup> Faculty of Health, University of Canberra, Canberra, ACT 2617, Australia

(convincing, trustworthy, realistic, likable, showed biological movement, reassuring, friendly, familiar, and human-like).

Our study is part of a broader project to develop a robotic art installation by Stelarc called the Articulated Head 2.0 (AH 2.0), an attention-driven interactive artificial intelligent system so that people feel safe and comfortable interacting with it in public. The project has a long history, starting with the PH (2003), Articulated Head 1.0 (2008) [10], and Articulated Head 2.0 [4, 5].

In an earlier study with AH 2.0, we found that the type of embodied face (no face, static image of PH V2, animated virtual face PH V2) did not affect the participant on all measures, including the distance mode between the robot and interlocutor, total interaction time, animacy, likability, and perceived safety before and after the interaction [4]. However, we found a significant difference between perceived safety before and after interaction regardless of the presence (Stelarc real voice, PH V2 machine-like voice, PH V1 more machine-like voice) and absence of face [5]. We believe this is because the robot's behavioural system generates adequate aliveness through robot motion in response to human movements compared to randomly generated motions [2]. This research extends that knowledge further into the realms of speech and speech perception in HRI.

The rest of the paper is organized as follows:

1. Section two discusses the related literature and hypotheses.
2. Section three presents the methodology.
3. In sections four and five, we have presented the result of the study.
4. In Section six, we discuss the results.
5. Finally, section seven discusses the implications of our findings.

We further discuss the limitations of current studies and suggest recommendations for future work.

## 2 Literature Review

Sigmund Freud, in his famous essay (*Das Unheimliche*) [3] describes the uncanny as the eerie feeling one feels when seeing inanimate figures come to life, ghosts or severed limbs, or seeing your *doppelgangers* (one's double). Independently, Mori [15] observed that “in climbing toward the goal of making robots appear like a human, our affinity for them increases until we come to a valley, which I call the uncanny valley” (NB, Jasia Reichardt [9] coined “uncanny valley” as a translation of “*bukimi no tani*”). He further observed that the presence of movement tends to steepen the slopes of the uncanny valley. The rather subjective notions of the Uncanny and the Uncanny Valley has since been of

much interest and speculation within the robotics community. There is increased interest in the uncanny in Human-Robot Interactions, especially in appearances and motion. However, few researchers have examined the uncanny effect associated with speech.

For instance, Hanson et al. [6] argue that the uncanny valley theory needs to be revised. They posit a new theory, the path of engagement, which suggests that creating realistic and nearly realistic characters might not be the cause of the uncanny valley effect. Instead, they suggest that social intelligence could also contribute to the uncanny valley effect. Brenton et al. [1] have argued that Mori's hypothesis is merely conjecturing. They propose that scientifically-based additional research is required to understand human perception of factors such as behaviour and appearance. This suggests that Mori's hypothesis does not seem to apply to every aspect of creating a robot with a human-like appearance and suggests other factors may affect interactions between robots and humans.

Speech is important in impression formation. In uncanny effect research, we could hypothesize that the type of audio affects impression formation. Romport [17] studied the effect of voice (‘machine-like and natural voice’) on text-to-speech systems using the “Wizard of Oz” technique. The findings revealed that a natural voice was preferred to a machine-like voice. Another study examined the effect of a robot's gender (male and female) and voice<sup>1</sup> (male (William) and female (Sarah) in forming the uncanny valley effect [16]. They found that the robot's gender did not affect the participant's perceived sense of eeriness, but the voice influenced the impression of the robot's gender such that a female voice was preferred more than a male voice. The robot's gender and voice did not negatively impact children. This research suggests the human-like voice does not necessarily cause an uncanny valley effect, but it aids in forming an impression of a robot's gender.

It has been observed that a human voice is liked more than a ‘human-like’ voice. For instance, Kuhne et al. [11] conducted an online study to investigate how audio affects the impression formation of the speaker. They used (human, Watson,<sup>2</sup> Sophia<sup>3</sup>) big five personality traits and kept the female voices constant. The study's findings demonstrated that a human speaker was preferred more than a human-like and synthesized voice, and a human voice was preferred more than a human-like appearance but did not relate to the participant's personality. That is, irrespective of the participant's personality, the human voice was rated higher than the human-like voice in terms of being confident and trustworthy, whereas the human speaker was rated higher on human-

<sup>1</sup> <https://www.cereproc.com/>.

<sup>2</sup> <https://www.ibm.com/au-en/watson>.

<sup>3</sup> <https://www.hansonrobotics.com/sophia/>.

like. The researchers concluded that creating the impression of a human-like voice depends on the person's perception of factors such as gender and age. The authors presented a compelling argument on the impression formation of voice, speaker, and big five personality traits. However, the mean duration of the audio was 5.8 seconds, and only the female voices were tested. Thus in our study, we addressed these limitations by using male voices and lengthening the duration of the video.

Another way the uncanny valley effect can be evoked is through asynchronous lip movement with voice. For instance, Tinwell et al. [21] studied the effect of different virtual faces ranging from most human-like to zombie, considering human-like, lip synchronization, and voice. The findings suggested that the human face was perceived as the most familiar and human-like, and virtual characters' mismatched voices and lip synchronization also created negative feelings. In another study conducted by Tinwell et al. [20], the authors explored the effect of audiovisual speech synchronization of human and virtual heads on impressions of familiarity and human-like. They found the effect of asynchronous lip motion to be more pronounced in the virtual head compared to a human head. They also found that a human character with synchronized speech was judged to be more familiar and human-like relative to a virtual character. However, in this study, all the participants were male, which reduces generalizability, only one type of virtual face was used, and the duration of the video was very short—only 4 seconds.

Additionally, McDonnell et al. [14] reviewed the impact of a male actor's rendering style, ranging from abstract to realistic, on perception measures (e.g., realistic, familiar, trustworthy, friendly, and reassuring) using lip motion and a static photo. The video included lip motion, which was presented for 6–10 seconds, in which the actor was speaking, but the audio was excluded. The study showed that the realist rendering was considered the most realistic, and the abstract style was considered the most friendly, with both being equally highly rated as familiar, trustworthy, and reassuring. The main finding of this study was that the character's lip motion affected perceived familiarity, but it did not affect assessments of reassuring, friendly, realistic, and trustworthiness. There was no significant interaction between lip motion and appearances, which might be due to the study excluding the audio. Thus in our study, we chose to include speaking with audio to investigate its impact on a positive impression of a face.

Along with non-verbal behaviour, human expectations also play a role in the uncanny valley effect. Thepsonthorn et al. [19] illustrated how the non-verbal behaviour of an NAO robot causes the uncanny valley effect. The results showed that non-verbal behaviour such as hand gestures,

body language, and speaking appears to affect the formation of a positive affinity. The results showed that non-verbal behaviour such as hand gestures, body language, and speaking appears to affect the formation of a positive affinity. They suggested that human expectations seem to affect the perception of the robot, such that the more the robot is expected to be human-like, the more it is expected to look like a human. A related study concluded that appearance influences how moral the robot is seen to be, such that the more human-like a robot looks, the more participants perceive the decision made by these robots to be less ethical. At the same time, the decision made by humans and non-uncanny robots is more ethical than human-like robots. The authors referred this observation, 'moral uncanny valley effect' [12].

Over the decades, many researchers have devoted themselves to examining how voice is perceived by people interacting with robots and virtual faces, but relatively less attention has been paid to the effect of speech in informing positive impressions [11, 12, 14, 17, 19–21]. The existing literature does not provide an answer as to whether speaking or not speaking interacts with the appearance of a robot to affect perceptions.

## 2.1 Hypotheses

In this study, we examined the effect of appearance (real face versus two rendered versions of the same face PH V1 and PH V2) and speaking (not speaking, speaking) on a range of outcome measures such as how convincing, trustworthy, realistic, likable, showed biological movement, reassuring, friendly, familiar, and human-like the faces were perceived to be. Consistent with previous literature, we hypothesized that:

**H1** An order of preference would be observed, such that speaking Stelarc videos (real person, real voice) would be rated highest in terms of whether the interaction was trustworthy, realistic, likable, showed biological movement, reassuring, friendly, familiar, and human-like, followed by PH V2 while speaking (PH V2, machine voice), PH V1 with speaking (PH V1, machine voice), the Stelarc without speaking, the PH V2 without speaking, and then the PH V1 without speaking.

**H2** Further, we hypothesized that there would be a difference such that all speaking versions would be rated higher than the non-speaking versions on all measures for each appearance type (real Stelarc, PH V2, and PH V1).



Fig. 1 Articulated head 2.0

## 3 Methodology

### 3.1 Participants

The sample comprised 49 students and staff from the University of Canberra (Australia) and the Aalborg University (Denmark). There were 28 females ( $M=44.5$  years,  $SD=12.7$ ), 20 males ( $M=40.4$  years,  $SD=12.4$ ), and one participant who did not report gender. The participants' ages ranged from 23 to 70 years. The Human Research Ethics Committee of the University of Canberra approved the study (20204462). All the participants were voluntarily recruited through university communication channels such as UC chat and the faculty forum. The survey was conducted from 6 May 2021 until 22 May 2021 (Fig. 1).

### 3.2 Procedures

The participants were informed about the online survey through email. The participants were advised that they would be required to watch 10 seconds of each video, a total of 60 seconds comprising 6 videos. The videos were presented to the participants in random order to reduce order bias and cumulative bias [20]. After each video, the participants completed the questionnaire. The Jamovi platform was used to analyze the data.

### 3.3 Design

*Independent Variables (Within Subjects):* 1. Speaking (absence, presence (Stelarc real voice, PH V2 machine-like voice, PH V1 more machine-like voice) (Stelarc real voice, machine-like Stelarc voice, machine-like voice). 2. Appearance (human-likeness) (Stelarc, PH V2, PH V1).

*Dependent Variables (DVs):* convincing, trustworthy, realistic, likable, showed biological movement, reassuring, friendly, familiar, and human-like.

*Variables that were controlled in all conditions:* 1. Exposure time (10 seconds per face). 2. Expression (neutral). 3. Head motion (matched across conditions). 4. Speaking: each head said the same phrase, “Hello, my name is Stelarc, and I am a performance artist.”

*Conditions:* 1. Stelarc speaking (Fig. 2a). 2. Stelarc non-speaking (Fig. 2b). 3. PH V2 speaking (Fig. 2c). 4. PH V2 non-speaking (Fig. 2d). 5. PH V1 speaking (Fig. 2e). 6. PH V1 non-speaking (Fig. 2f).

## 3.4 Materials

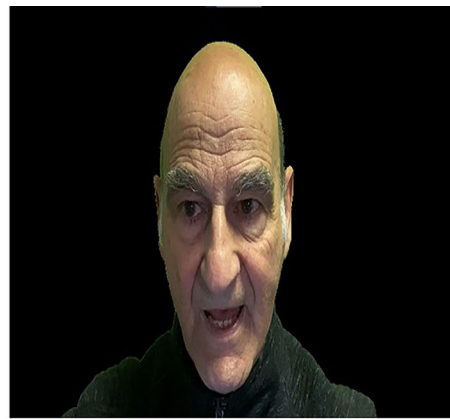
The participants completed elements of the Ho and Mac Dorman Questionnaire [7], Racheal McDonnell Questionnaire [14], and Valentin Schwind Questionnaire [18]. The questionnaire for this study included nine items to—eight items from the above questionnaires. These were questions about whether the interaction was convincing, trustworthy, realistic, likable, showed biological movement, reassuring, friendly, familiar, and human-like. We added one item to the questionnaire asking how convincing the interaction was. The biological movement, realistic, and human-like items were from the Ho and Mac Dorman questionnaire [7], the trustworthy, reassuring, friendly, and familiar items were from the Racheal McDonnell Questionnaire [14], and the likable item from Valentin Schwind Questionnaire [18]. These nine items were selected as we considered these were fundamental elements that were likely to influence the interaction such as the uncanny (familiar) and uncanny valley (human-like, likable). An example of a sample question is: “Based on the video watched, please rate your impression of the face on the following scales:(1=not at all, 7=very much).” The Qualtrics platform was used to conduct the survey.

## 4 Results I—Ranking of the Heads

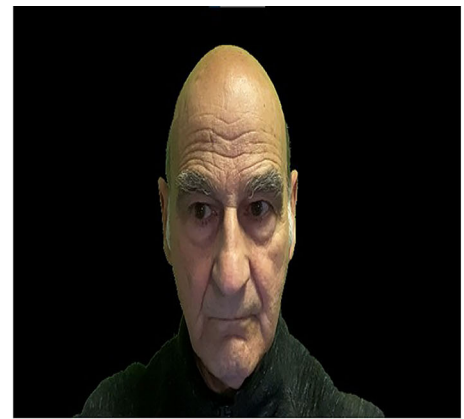
An order of preference would be observed, such that speaking Stelarc videos (real person, real voice) would be rated highest in terms of whether the interaction was trustworthy, realistic, likable, showed biological movement, reassuring, friendly, familiar, and human-like, followed by PH V2 while speaking (PH V2, machine voice), PH V1 with speaking (PH V1, machine voice), the Stelarc without speaking, the PH V2 without speaking, and then the PH V1 without speaking.

To test the hypothesis, we conducted a one-way repeated measures analysis of variance ANOVA ( $\alpha=.05$ ) directly comparing all conditions, and if significant, follow-up paired samples t-tests ( $\alpha=.01$ ) identified significant differences. The descriptive statistics are presented in Table 1.

Fig. 2 Six conditions



(a) Stelarc speaking



(b) Stelarc not speaking



(c) PH V2 speaking



(d) PH V2 not speaking



(e) PH V1 speaking



(f) PH V1 not speaking

#### 4.1 Convincing

The result showed a significant difference in the measure of ‘convincing’ across the six conditions,  $F(5,240) = 68.9, p < .001$ . We followed up the significant difference with paired t- tests and found that Stelarc speaking conditions were rated

significantly higher than all other conditions,  $p < .01$  on ‘convincing.’ The next highest was Stelarc in the non-speaking condition, which was significantly higher than the remaining four conditions  $p < .01$ . It was followed by PH V2 without speaking, which was rated significantly higher than the remaining three conditions  $p < .01$ . There was no significant

**Table 1** Descriptive statistics for Stelarc, PH V2, PH V1 with speaking and without speaking

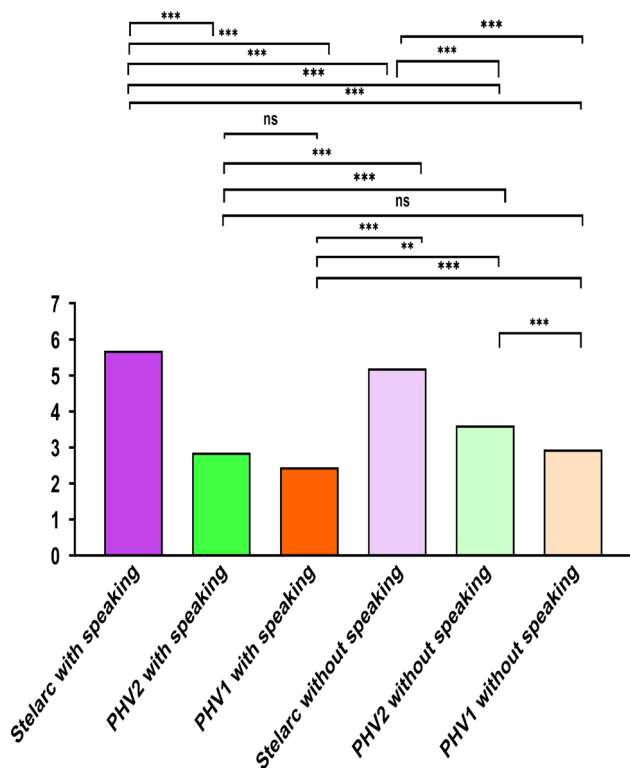
	Speaking	Appearance	Mean	Standard deviation
Convincing	With speaking	Stelarc	<b>5.69</b>	1.37
Trustworthy			<b>4.9</b>	1.4
Realistic			<b>6.1</b>	1.26
Likable			<b>4.57</b>	1.41
Biological movement			<b>5.88</b>	1.3
reassuring			<b>4.51</b>	1.46
Friendly	Without speaking	Stelarc	<b>4.35</b>	1.49
Familiar			<b>4.92</b>	1.69
Human-like			<b>6.37</b>	1.05
Convincing			5.20	1.73
Trustworthy			4.37	1.58
Realistic			5.78	1.57
Likable	With speaking	PH V2	4.16	1.43
Biological movement			5.41	1.55
Reassuring			3.9	1.58
Friendly			3.69	1.71
Familiar			4.69	1.72
Human-like			6.02	1.3
Convincing	Without speaking	PH V2	2.86	1.5
Trustworthy			2.94	1.64
Realistic			2.88	1.65
Likable			2.98	1.65
Biological movement			2.82	1.48
Reassuring			2.67	1.53
Friendly	With speaking	PH V1	2.94	1.56
Familiar			2.94	1.75
Human-like			2.92	1.69
Convincing			3.61	1.53
Trustworthy			3.49	1.66
Realistic			3.69	1.7
Likable	Without speaking	PH V1	3.47	1.62
Biological movement			3.53	1.89
Reassuring			3.16	1.45
Friendly			3.43	1.51
Familiar			3.59	1.78
Human-like			3.86	1.84
Convincing	With speaking	PH V1	2.45	1.42
Trustworthy			2.55	1.5
Realistic			2.31	1.47
Likable			2.63	1.52
Biological movement			2.51	1.36
Reassuring			2.24	1.38
Friendly	Without speaking	PH V1	2.47	1.63
Familiar			2.49	1.54
Human-like			2.41	1.58



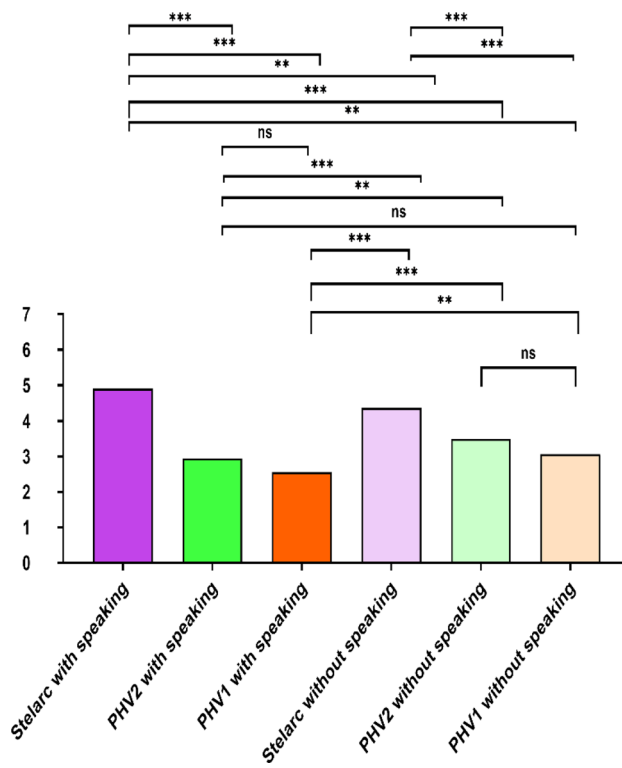
**Table 1** (continued)

	Speaking	Appearance	Mean	Standard deviation
Convincing	Without speaking	PH V1	2.94	1.51
Trustworthy			3.06	1.68
Realistic			2.86	1.63
Likable			3.22	1.66
Biological movement			3.04	1.55
Reassuring			2.69	1.47
Friendly			3.31	1.61
Familiar			3.16	1.61
Human-like			3.12	1.7

Bold signifies the highest mean rating



**Fig. 3** Result for paired t-test on convincing between speaking (absence and presence (Stelarc real voice, PH V2 machine-like voice, PH V1 more machine-like voice)) and (Stelarc, PH V2, PH V1) with 49 participants ( $p < .001^{***}$ ,  $p < .01^{**}$ , no significance(ns) (95 percentage CI))



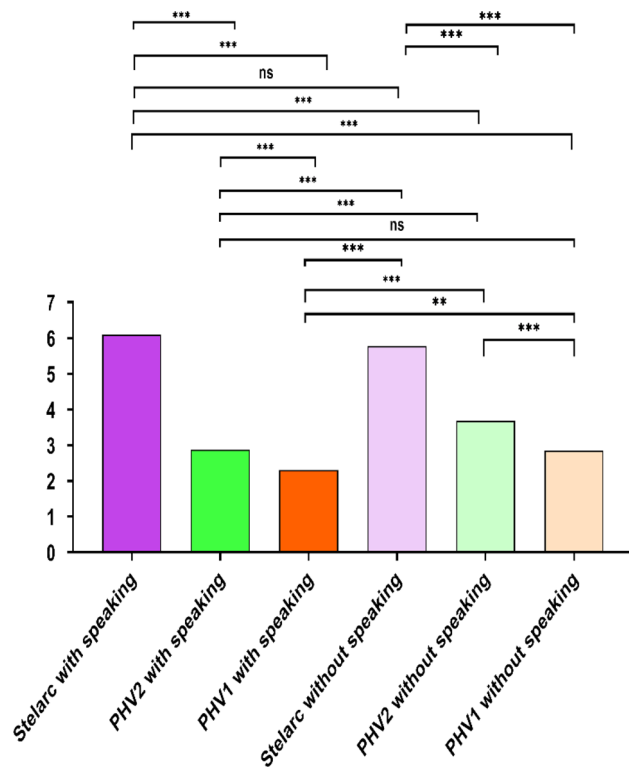
**Fig. 4** Result for paired t-test on trustworthy between speaking (absence and presence (Stelarc real voice, PH V2 machine-like voice, PH V1 more machine-like voice)) and (Stelarc, PH V2, PH V1) with 49 participants ( $p < .001^{***}$ ,  $p < .01^{**}$ , no significance(ns) (95 percentage CI))

difference between the other two conditions on ‘convincing.’ These results are illustrated in Fig. 3.

### 4.2 Trustworthy

The result showed a significant difference in ‘trustworthy’ across the six conditions,  $F(5,240) = 32.8$ ,  $p < .001$ .

We followed up the significant difference with paired t-tests and found the Stelarc speaking conditions were rated significantly higher than all other conditions,  $p < .01$  on ‘trustworthy.’ The next highest was Stelarc in the non-speaking condition, which was significantly higher than the remaining five conditions,  $p < .01$  on ‘trustworthy.’ There was no significant difference among other conditions on ‘trustworthy.’ These are illustrated in Fig. 4.



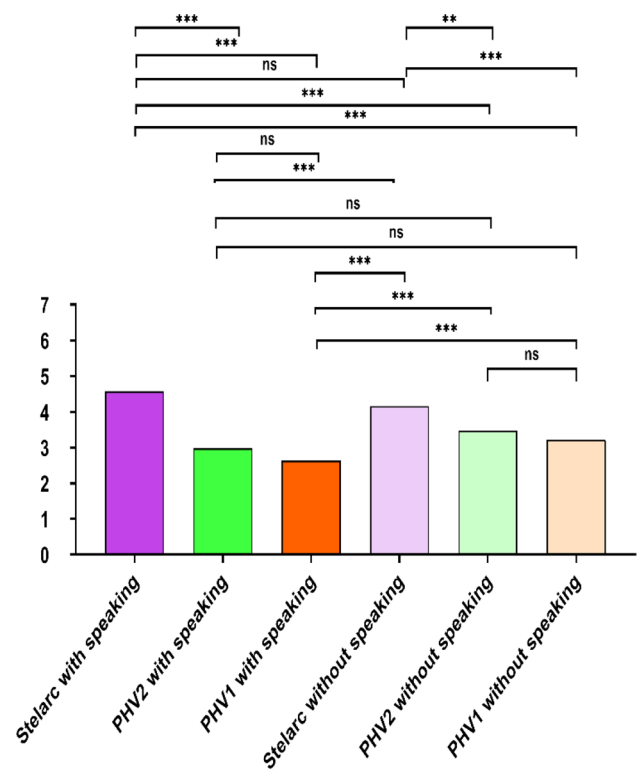
**Fig. 5** Result for paired t-test on realistic between speaking (absence and presence (Stelarc real voice, PH V2 machine-like voice, PH V1 more machine-like voice)) and (Stelarc, PH V2, PH V1) with 49 participants ( $p < .001^{***}$ ,  $p < .01^{**}$ , no significance(ns) (95 percentage CI))

### 4.3 Realistic

The result showed a significant difference in ‘realistic’ across the six conditions,  $F(5,240) = 93.0$ ,  $p < .001$ . We followed up the significant difference with paired t-tests and found the Stelarc speaking and non-speaking conditions were rated significantly higher than all other conditions,  $p < .01$  on ‘realistic.’ The next highest was PH V2 without speaking, which was rated significantly higher than the remaining three conditions,  $p < .01$  on ‘realistic.’ There was no significant difference among other conditions on ‘realistic.’ These results are illustrated in Fig. 5.

### 4.4 Likable

Stelarc speaking condition was rated significantly higher than all other conditions,  $p < .01$  on ‘likable.’ The next highest was the Stelarc non-speaking condition, which was significantly higher than the remaining four conditions,  $p < .01$  on ‘likable.’ There was no significant difference among other conditions on ‘likable.’ These results are illustrated in Fig. 6.



**Fig. 6** Result for paired t-test on likable between speaking (absence and presence (Stelarc real voice, PH V2 machine-like voice, PH V1 more machine-like voice)) and (Stelarc, PH V2, PH V1) with 49 participants ( $p < .001^{***}$ ,  $p < .01^{**}$ , no significance(ns) (95 percentage CI))

### 4.5 Biological Movement

The result showed a significant difference in ‘likable’ across the six conditions,  $F(5,240) = 21.9$ ,  $p < .001$ . We followed up on the significant difference with paired t-tests. We found that

The result showed a significant difference in ‘biological movement’ across the six conditions,  $F(5,240) = 72.1$ ,  $p < .001$ . We followed up the significant difference with paired t-tests and found the Stelarc speaking and not speaking conditions were rated significantly higher than all other conditions,  $p < .01$  on ‘biological movement.’ There was no significant difference among other conditions on ‘biological movement.’ These results are illustrated in Fig. 7.

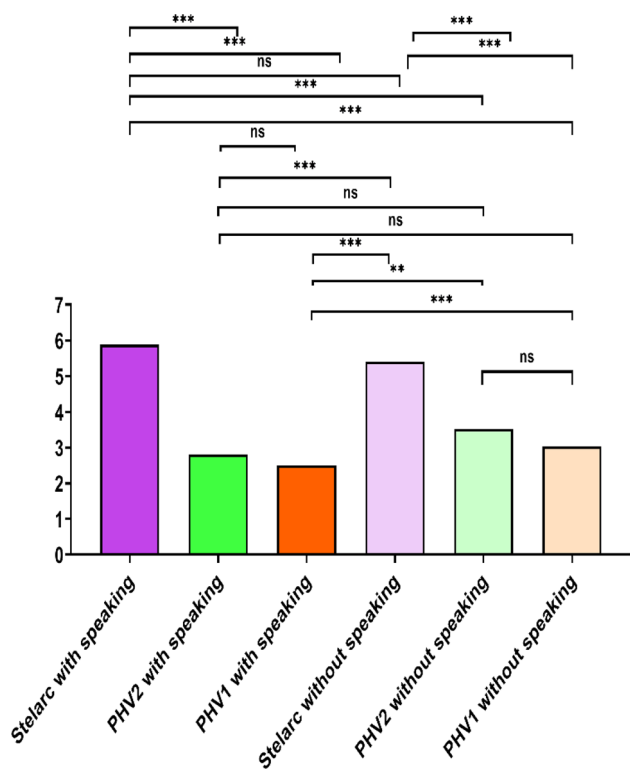
### 4.6 Reassuring

No significant difference among other conditions on ‘reassuring.’ These results are illustrated in Fig. 8.

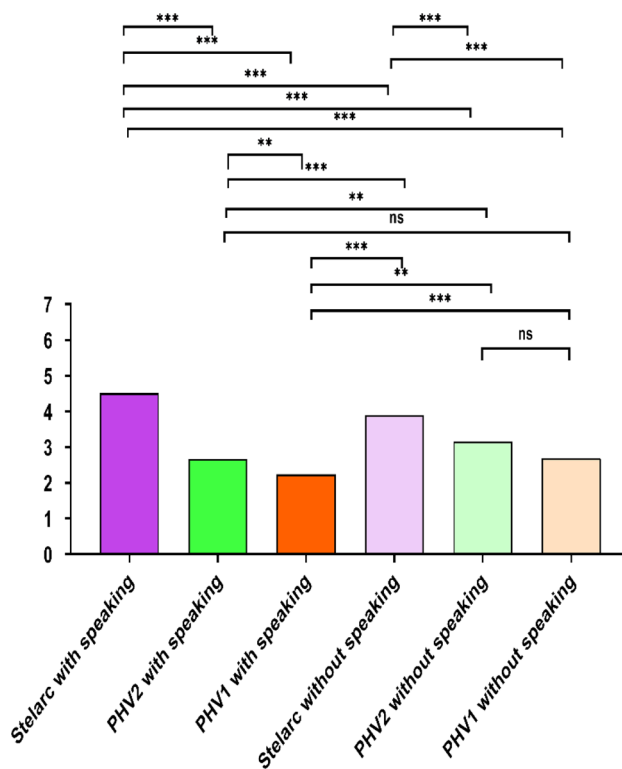
### 4.7 Friendly

The result showed a significant difference in ‘reassuring’ across the six conditions,  $F(5,240)$





**Fig. 7** Result for paired t-test on biological movement between speaking (absence and presence (Stelarc real voice, PH V2 machine-like voice, PH V1 more machine-like voice)) and (Stelarc, PH V2, PH V1) with 49 participants ( $p < .001^{***}$ ,  $p < .01^{**}$ , no significance(ns) (95 percentage CI))



**Fig. 8** Result for paired t-test on reassuring between speaking (absence and presence (Stelarc real voice, PH V2 machine-like voice, PH V1 more machine-like voice)) and (Stelarc, PH V2, PH V1) with 49 participants ( $p < .001^{***}$ ,  $p < .01^{**}$ , no significance(ns) (95 percentage CI))

= 34.2,  $p < .001$ . We followed up the interaction with paired t-tests and found the Stelarc speaking condition was rated significantly higher than all other conditions,  $p < .01$  on ‘reassuring.’ The next highest was Stelarc in the non-speaking condition, which was significantly higher than the remaining four conditions,  $p < .01$  on ‘reassuring.’ There was

The result showed a significant difference in ‘friendly’ across the six conditions,  $F(5,240) = 16.7$ ,  $p < .001$ . We followed up on the significant difference with paired t-tests. We found the Stelarc speaking condition was rated significantly higher than all other conditions,  $p < .01$  on ‘friendly.’ There was no significant difference among other conditions on ‘friendly.’ These results are illustrated in Fig. 9.

### 4.8 Familiar

The result showed a significant difference in ‘familiar’ across the six conditions,  $F(5,240) = 32.9$ ,  $p < .001$ .

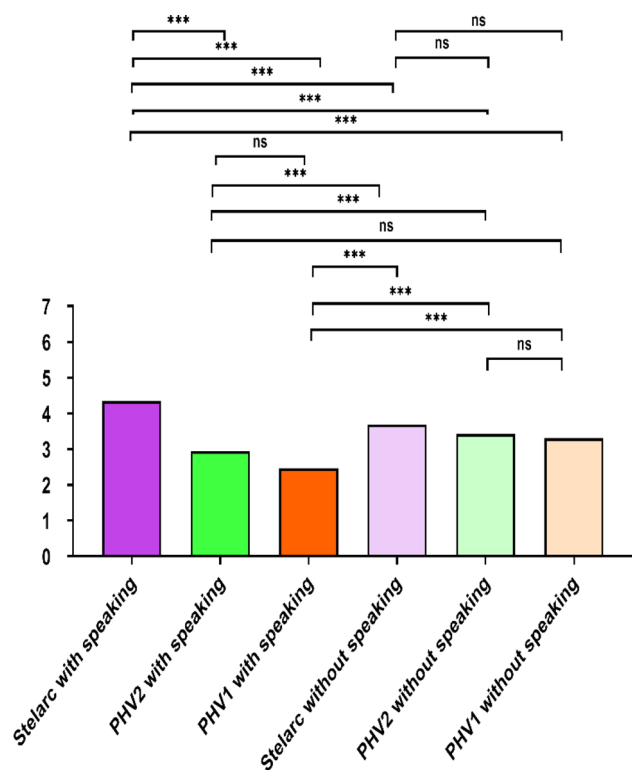
We followed up on the significant difference with paired t-tests. We found the Stelarc speaking and non-speaking conditions were rated significantly higher than all other

conditions,  $p < .01$  on ‘familiar.’ There was no significant difference among other conditions on ‘familiar.’ These results are illustrated in Fig. 10.

### 4.9 Human-like

The result showed a significant difference in ‘human-like’ across the six conditions,  $F(5,240) = 112$ ,  $p < .001$ . We followed up on the significant difference with paired t-tests. We found the Stelarc speaking condition was rated significantly higher than all other conditions,  $p < .01$  on ‘human-like.’ The next highest was Stelarc in the non-speaking condition, which was significantly higher than the remaining four conditions,  $p < .01$  on ‘human-like.’ It was followed by PH V2 without speaking, which was rated significantly higher than the remaining three conditions,  $p < .01$  on ‘human-like.’ There was no significant difference among other conditions on ‘human-like.’ These results are illustrated in Fig. 11.

As a result, we found that speaking Stelarc was rated highest, followed by non-speaking Stelarc and then PH V2 without speaking, on ‘convincing.’ We found that speaking and non-speaking Stelarc was rated highest on ‘trustworthy.’ We found that speaking and non-speaking Stelarc was rated



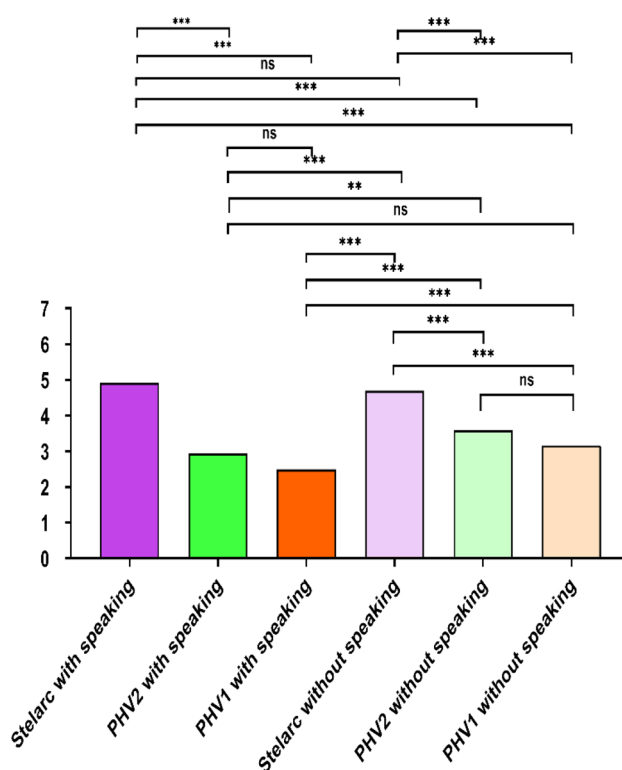
**Fig. 9** Result for paired t-test on friendly between speaking (absence and presence (Stelarc real voice, PH V2 machine-like voice, PH V1 more machine-like voice)) and (Stelarc, PH V2, PH V1) with 49 participants ( $p < .001^{***}$ ,  $p < .01^{**}$ , no significance(ns) (95 percentage CI))

highest, followed by PH V2 without speaking, on ‘realistic.’ We found that speaking was rated highest, followed by non-speaking Stelarc on ‘likable.’ We found that speaking and non-speaking Stelarc was rated highest on ‘biological movement.’ We found that speaking was rated highest, followed by non-speaking Stelarc on ‘reassuring.’ We found that speaking and non-speaking Stelarc was rated highest on ‘friendly.’ We found that speaking and non-speaking Stelarc was rated highest on ‘familiar.’ Lastly, we found that speaking Stelarc was rated highest, followed by non-speaking Stelarc and then PH V2 without speaking, on ‘human-like.’ These are illustrated in Tables 2, 3 and 4.

### 5 Results II—Effects of Speaking and Appearances

We hypothesized that there would be a difference between speaking Stelarc and Stelarc without speaking, on whether the interaction was convincing, trustworthy, realistic, likable, showed biological movement, reassuring, friendly, familiar, and human-like.

We hypothesized that there would also be differences in those measures between Prosthetic Head Version 2 (PH V2)



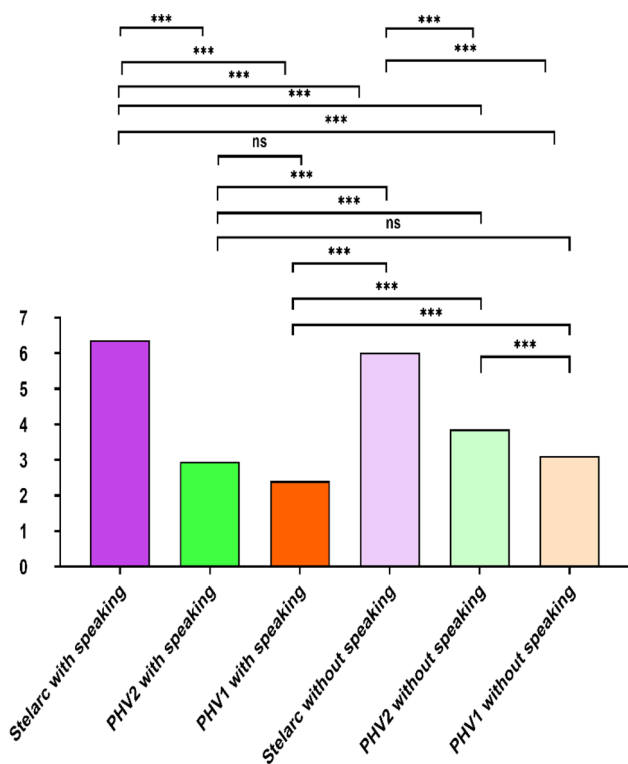
**Fig. 10** Result for paired t-test on familiar between speaking (absence and presence (Stelarc real voice, PH V2 machine-like voice, PH V1 more machine-like voice)) and (Stelarc, PH V2, PH V1) with 49 participants ( $p < .001^{***}$ ,  $p < .01^{**}$ , no significance(ns) (95 percentage CI))

while speaking and PH V2 without speaking, as well as Prosthetic Head Version 1 (PH V1) with speaking and PH V1 without speaking.

To test the hypotheses, we used a 2\*3 repeated measures factorial ANOVAs ( $\alpha=.05$ ) to investigate the effects of appearance and speaking on all nine dependent variables and followed up any significant interactions with separate paired samples t-tests ( $\alpha=.01$ ). Shapiro-Wilk tests were used to evaluate the assumptions of normality, and there were some violations so findings should be interpreted with caution. Below we present the individual results for each dependent variable.

#### 5.1 Convincing

We found a statistically significant main effect on the ‘convincing’ variable for appearance,  $F(2,96) = 99.26, p < .001$  but no significant effect for speaking. We found a statistically significant interaction between appearance and speaking on ‘convincing,’  $F(2,96) = 20.53, p < .001$ . We followed up the interaction with paired t-tests and found a significant difference between the Stelarc speaking and Stelarc not speaking on ‘convincing’  $p < .01$ . We found a significant difference



**Fig. 11** Result for paired t-test on human-like between speaking (absence and presence (Stelarc real voice, PH V2 machine-like voice, PH V1 more machine-like voice)) and (Stelarc, PH V2, PH V1) with 49 participants ( $p < .001^{***}$ ,  $p < .01^{**}$ , no significance (ns) (95 percentage CI))

**Table 2** Significant ratings and the highest order against the six conditions

Measures	First
Convincing	Stelarc speaking
Trustworthy	Stelarc Speaking and non-speaking
Realistic	Stelarc speaking and non-speaking
Likable	Stelarc speaking
Biological	Movement Stelarc speaking and non-speaking
Reassuring	Stelarc speaking
Friendly	Stelarc speaking and non-speaking
Familiar	Stelarc speaking and non-speaking
Human-like	Stelarc speaking

**Table 3** Significant ratings and the second highest order against the six conditions

Measures	Second
Convincing	Stelarc non-speaking
Realistic	PH V2 non-speaking
Likable	Stelarc non-speaking
Reassuring	Stelarc non-speaking
Human-like	Stelarc non-speaking

**Table 4** Significant ratings and the third highest order against the six conditions

Measures	Third
Convincing	PH V2 non-speaking
Human-like	PH V2 non-speaking

between PH V2 and PH V2 without speaking on ‘convincing’  $p < .001$ . We found a significant difference between PH V1 with speaking and PH V1 without speaking on ‘convincing’  $p < .01$ . The result suggests that Stelarc, while speaking, positively affects the impression formed of Stelarc on the measures of how convincing he seems. In contrast, speaking negatively affects the impression formed of PH V2 and PH V1 on how convincing they seem. These results are illustrated in Fig. 3.

### 5.2 Trustworthy

We found a statistically significant main effect of the ‘trustworthy’ variable for appearance,  $F(2,96) = 48.44$ ,  $p < .001$  but no significant effect for speaking. We found a statistically significant interaction between appearance and speaking on ‘trustworthy,’  $F(2,96) = 13.90$ ,  $p < .001$ . We followed up the interaction with paired t-tests and found a significant difference between the Stelarc speaking and not speaking on ‘trustworthy’  $p < .01$ . We found a significant difference between PH V2 with speaking and PH V2 without speaking on ‘trustworthy’  $p < .01$ . We found a significant difference between PH V1 with speaking and PH V1 without speaking on ‘trustworthy’  $p < .01$ . The result suggests that Stelarc, while speaking, positively affects the impression formed of Stelarc on the measures of how trustworthy he seems. In contrast, speaking negatively affects the impression formed of PH V2 and PH V1 on the measures of how trustworthy they seem. These results are illustrated in Fig. 4.

### 5.3 Realistic

We found a statistically significant main effect of the variable ‘realistic’ for appearance,  $F(2,96) = 148.64$  and  $p < .001$ , and a significant effect for speaking  $F(1,48) = 6.04$  and  $p = .018$ . We found a statistically significant interaction that indicated the effects of appearance and speaking on impression formation of ‘realistic,’  $F(2,96) = 13.62$ ,  $p < .001$ . We followed up the interaction with paired t-tests and did not find a significant difference between Stelarc speaking and not speaking on ‘realistic.’ We found a significant difference between PH V2 with speaking and PH V2 without speaking on ‘realistic’  $p < .01$ . We found a significant difference between PH V1 with speaking and PH V1 without speaking on ‘realistic’  $p$

$< .01$ . The result suggests that Stelarc while speaking, did not affect the impression formed of Stelarc on the measures of how realistic he seems. In contrast, speaking negatively affects the impression formation of PH V2 and PH V1 on how realistic they seem. These are illustrated in Fig. 5.

#### 5.4 Likable

We found a statistically significant main effect of ‘likable’ for appearance,  $F(2,96) = 29.65$  and  $p < .001$ . There was no significant effect on speaking. We found a statistically significant interaction between appearance and speaking on ‘likable,’  $F(2,96) = 15.43$ ,  $p < .001$ . We followed up the interaction with paired t-tests and did not find a significant difference between Stelarc speaking and Stelarc not speaking on ‘likable,’  $p = .017$ . We did not find a significant difference between PH V2 speaking and not speaking on ‘likable’  $p = .016$ . We found a significant difference between PH V1 with speaking and PH V1 without speaking on ‘likable’  $p < .001$ . The result suggests that speaking Stelarc and PH V2 did not affect how likable they seemed. In contrast, speaking negatively affects how likable PH V1 appears to participants. These results are illustrated in Fig. 6.

#### 5.5 Biological Movement

We found a statistically significant main effect of ‘biological movement’ for appearance,  $F(2,96) = 121.80$  and  $p < .001$  but no significant effect for speaking. We found a statistically significant interaction between appearance and speaking on impression formation of ‘biological movement,’  $F(2,96) = 12.65$ ,  $p < .001$ . We followed up the interaction with paired t-tests and did not find a significant difference between Stelarc speaking and Stelarc not speaking on ‘biological movement,’  $p = .018$ . We found a significant difference between PH V2 with speaking and PH V2 without speaking on ‘biological movement’  $p < .01$ . We found a significant difference between PH V1 with speaking and PH V1 without speaking on ‘biological movement’  $p < .01$ . The result suggests that Stelarc speaking did not affect impression formation on the measure of biological movement. In contrast, speaking negatively affects the impression formed of PH V2 and PH V1 on the measure of biological movement. These results are illustrated in Fig. 7.

#### 5.6 Reassuring

We found a statistically significant main effect of ‘reassuring’ for appearance,  $F(2,96) = 49.49$  and  $p < .001$  but no significant effect for speaking. We found a statistically significant interaction between appearance and speaking on impression formation of ‘reassuring,’  $F(2,96) = 16.63$ ,  $p < .001$ . We followed up the interaction with paired t-tests and found a

significant difference between Stelarc speaking and Stelarc not speaking on ‘reassuring’  $p < .001$ . We found a significant difference between PH V2 with speaking and PH V2 without speaking on ‘reassuring’  $p < .01$ . We found a significant difference between PH V1 with speaking and PH V1 without speaking on ‘reassuring’  $p < .01$ . The result suggests that Stelarc, while speaking, significantly positively affects impression formation on how reassuring he seems. In contrast, speaking negatively affects the impression formed of PH V2 and PH V1 on how reassuring they seem. These results are illustrated in Fig. 8.

#### 5.7 Friendly

We found a statistically significant main effect of ‘friendly’ for appearance,  $F(2,96) = 18.76$ ,  $p < .001$  but no significant effect for speaking. We found a statistically significant interaction between the effects of appearance and speaking on the impression formation of ‘friendly,’  $F(2,96) = 19.97$ ,  $p < .001$ . We found a significant difference between Stelarc speaking and Stelarc not speaking on ‘friendly’  $p < .001$ . We followed up the interaction with paired t-tests and found a significant difference between PH V2 with speaking and PH V2 without speaking on ‘friendly’  $p < .01$ . We found a significant difference between PH V1 with speaking and PH V1 without speaking on ‘friendly’  $p < .001$ . The result suggests that Stelarc, while speaking, positively affects the impression formation on the measures of how friendly he is. In contrast, speaking negatively affects the impression formed of PH V2 and PH V1 on the measures of how friendly they are. These results are illustrated in Fig. 9.

#### 5.8 Familiar

We found a statistically significant main effect of ‘familiar’ on appearance,  $F(2,96) = 46.98$ ,  $p < .001$ . There was a significant effect for speaking  $F(1,48) = 8.01$ ,  $p = .007$ . We found a statistically significant interaction between the effects of appearance and speaking on impression formation for ‘familiar,’  $F(2,96) = 9.25$ ,  $p < .001$ . We followed up the interaction with paired t-tests and did not find a significant difference between Stelarc speaking and Stelarc not speaking on ‘familiar.’ We found a significant difference between PH V2 speaking and PH V2 without speaking on ‘familiar’  $p < .01$ . We found a significant difference between PH V1 speaking and PH V1 without speaking on ‘familiar’  $p < .001$ . The result suggests that Stelarc while speaking, did not affect the impression formed of how familiar he seemed. In contrast, speaking negatively affects the impression formed of PH V2 and PH V1 on the measures of how familiar they seem. These results are illustrated in Fig. 10.

## 5.9 Human-Like

We found a statistically significant main effect of ‘human-like’ for appearance,  $F(2,96) = 169.7, p < .001$ , as well as for speaking  $F(1,47) = 13, p < .001$ . We found a statistically significant interaction between the effects of appearance and speaking on impression formation of ‘human-like,’  $F(2,96) = 18.8, p < .001$ . We followed up the interaction with paired t-tests and found a significant difference between the Stelarc speaking and Stelarc not speaking on ‘human-like’  $p < .001$ . We found a significant difference between PH V2 with speaking and PH V2 without speaking on ‘human-like’  $p < .001$ . We found a significant difference between PH V1 with speaking and PH V1 without speaking on ‘human-like’  $p < .001$ . The result suggests that Stelarc while speaking positively, affects the measures of how human-like he seems. In contrast, speaking negatively affects how ‘human-like’ PH V2 and PH V1 seem. These results are illustrated in Fig. 11.

As a result, we found a significant main effect for appearance (Stelarc, PH V2, PH V1) on whether the interaction was convincing, trustworthy, realistic, likable, showed biological movement, reassuring, friendly, familiar, and human-like. We also found a significant interaction between appearances and speaking for whether the interaction was convincing, trustworthy, realistic, likable, showed biological movement, reassuring, friendly, familiar, and human-like. Lastly, we found a significant effect for speaking on whether the interaction was realistic, familiar, and human-like.

Thus, we found that speaking positively affects the impression formed of Stelarc in terms of whether the interaction was convincing, trustworthy, reassuring, friendly, and human-like. However, speaking negatively affected the impression formed of PH V2 in terms of whether the interaction was convincing, trustworthy, realistic, biological movement, reassuring, friendly, familiar, and human-like. It also negatively affected the impression formed of PH V1 in terms of whether the interaction was convincing, trustworthy, realistic, likable, biological movement, reassuring, friendly, familiar, and human-like.

## 6 Discussion

### 6.1 Ranking of Heads

The results partly supported the hypothesis that an order of preference would be observed, such that speaking Stelarc videos (real person, real voice) would be rated highest in terms of whether the interaction was trustworthy, realistic, likable, showed biological movement, reassuring, friendly, familiar, and human-like, followed by PH V2 while speaking (PH V2, machine voice), PH V1 with speaking (PH V1, machine voice), the Stelarc without speaking, the PH V2

without speaking, and then the PH V1 without speaking. We found significant differences between Stelarc with speaking, PH V2 with speaking, PH V1 with speaking, Stelarc without speaking, PH V2 without speaking, and PH V1 without speaking. This result supports the earlier studies, which found that humans speaking and humans were rated both higher than human-like appearances and voices on human-like, confident, and trustworthy [11, 12, 19]. Therefore, speaking may affect the participants’ responses to different appearances, demonstrating that including a human voice positively affects the impression of faces.

We found that the heads’ fidelity dictated the ratings, indicating that the closer we get to reality, the participants responded positively in speaking and non-speaking conditions. As PH V2 is a high-fidelity photo rendering of the Stelarc’s face, we could argue that either there is plausibly no evidence to indicate an ‘uncanny effect’ concerning avatars. One could equally argue that we may have ‘missed’ the uncanny valley between PH V1 and V2 and PH V2 and the real head in our experiments. It mirrors a study conducted by McDonnell et al. [14], such that a realistic human-like avatar was rated higher than an abstract human-like appearance. We argue that our result might be possible because of the degree of fidelity of the heads, as it seems to push the face in the positive direction of the uncanny valley.

However, PH V2 with speaking and PH V1 with and without speaking were rated the same on how convincing, trustworthy, realistic, likable, showed biological movement, reassuring, friendly, familiar, and human-like. This result supports the earlier study, which found that asynchronization of lips and voice made the virtual appearance uncannier and that a machine-like voice caused eeriness [17, 20, 21]. In which case, our result might be possible because of a machine-like voice, demonstrating that speaking with a machine-like voice causes a negative impression of avatars. Thus, this finding suggests to the HRI community not to use a machine-like voice informing positive impression.

### 6.2 Effect of Speaking and not Speaking on (Stelarc, PH V2, PH V1)

The result supported the hypothesis that there would be a difference such that all speaking versions would be rated higher than the non-speaking versions on all measures (convincing, trustworthy, realistic, likable, showed biological movement, reassuring, friendly, familiar, and human-like) for each appearance type (real Stelarc, PH V2, and PH V1). There were significant differences observed between real Stelarc with speaking and without speaking conditions in terms of whether the interactions were convincing, trustworthy, reassuring, friendly, and human-like. Results reflected the findings of the previous studies, which found that a human and human voice was chosen more often than a



human-like appearance and human-like voice [11, 17]. Our result might be possible because authentic/familiar voice affected the impression formed of Stelarc, so participants rated the Stelarc while speaking higher than Stelarc non-speaking. Such a result may suggest to Human-Robot Interaction researchers that authentic/familiar positively affects the impression formed of a human appearance, suggesting the use of a human appearance and authentic/familiar voice when choosing an appropriate avatar.

Our finding did not show significant differences between speaking and non-speaking Stelarc as to whether the interaction was rated as showing biological movement or was realistic and likable. However, our result might be because there was a synchronization of lips and voice. An earlier study found that, as the synchronization of lips and voice increases, the uncanny valley effect decreases [20, 21]. Thus, participants rated speaking and non-speaking Stelarc the same on measures of biological movement, realistic, and likable. Our finding suggests that the synchronization of lips and voice might not cause a difference in the impression formation of speaking and non-speaking real faces.

The result supported the hypothesis that there would be a difference between the Prosthetic Head Version 2 (PH V2) and Version 1 (PH V1) with speaking and PH V2 and V1 without speaking on measures of convincing, trustworthy, realistic, likable, biological movement, reassuring, friendly, familiar, and human-like. There were significant differences between the PH V2 and V1 speaking and the non-speaking conditions, such that non-speaking avatars were rated higher than speaking avatars. These findings partly support the study conducted by McDonnell et al. [14] that found a realistic human-like appearance with lip motion did not affect how the interaction was rated on the measures of reassuring, realistic, friendly, and trustworthy. They moreover found that there was no interaction between lip motion and appearance. Our study found significant interactions between appearance and speaking in convincing, trustworthy, realistic, likable, biological movement, reassuring, friendly, familiar, and human-like. Appearance significantly affected all nine measures, and speaking significantly affected how realistic, familiar, and human-like the interaction was judged. Our finding suggests that including voice in the study causes the interaction between speaking and appearance. Further, our result illustrates that the factors such as the avatar's human likeness and familiarity contribute to forming the uncanny effect and uncanny valley effect.

Our results supported earlier studies in which they found that a human voice was preferred more than a machine-like voice [17], and which suggested that speech, audio, and speaking contribute to forming the uncanny valley effect [11, 12, 19]. They found that a human was preferred more than a human-like appearance on a human-like. Also, they found a human voice was preferred over a human-like voice on

human-like, confident, and trustworthy measures. The unfamiliarity of voice contributes to forming an uncanny effect [3, 9]. Our finding is that speaking affected the impression formed by the two Prosthetic Head Versions. It may suggest that unfamiliarity of voice decreased the positive perception of PH V2 and PH V1. Our finding might be because of audio perception—a machine-like voice causes an uncanny valley effect and uncanny effect, implying that speaking with a machine-like voice negatively affects the impression of a realistic human-like appearance.

Note that there were significant differences between the PH V2 and PH V1 with speaking and PH V2 and V1 without speaking conditions on perceptions of biological movement. Our study supports earlier studies, which found that as the asynchronization between lips and audio increases, the uncanny valley effect is observed more in virtual heads than in humans [20, 21]. Our result might be because asynchronization of lips and voice negatively affected the impression formed of the speaking avatars. In this case, factors such as synchronization of lips and voice might cause an uncanny valley effect in avatars.

## 7 Conclusions

The study observed whether incorporating a speaking voice means a robot face is perceived as more convincing, trustworthy, realistic, likable, showing biological movement, reassuring, friendly, familiar, and human-like. Additionally, it aimed to investigate the uncanny valley effect and the uncanny effect of speech. We found that including human(authentic/familiar) speech in a video of a real human positively affected impression formation as expected. Conversely, using a machine-like speaking voice with videos of avatars was found to negatively affect impression formation. We also found that a human (authentic/familiar) speaking in a human voice was rated higher than speaking avatars (machine-like) overall—this is the ideal scenario for which the technology is still not ready. This work demonstrates that familiarity and human-like effects of speaking voice and visual perceptions can play a strong role in forming the uncanny valley effect and uncanny effect of speech. It should be noted that our study primarily focused on speech's effect on the interaction's impression formation. Notwithstanding its limitations, that is, more variation in voices and faces, our findings suggest that the HRI community considers a familiar and human-like (voice and appearance) so that it makes a positive impression of the interaction and leads people to feel comfortable when interacting with a robot. Also, it suggests that HRI might benefit from minimal verbal cues.

The study also demonstrates the importance of future HRI research to consider speech in forming positive impression formation of interactions. For instance, further research is



warranted to investigate the effect of lips and voice synchronization (Stelarc, PH V2, PH V1) in creating the effect of the uncanny and uncanny valley. More research is needed to determine whether these effects persist over different appearances, ranging from cartoon-like to realistic characters, over a longer speaking duration with more variations in voices, and across gendered and non-gendered avatars. Also, it would be potential for future research to test the uncanny valley effect with “real” robots which include biological movements and face-to-face interaction.

**Funding** Open Access funding enabled and organized by CAUL and its Member Institutions.

**Data Availability** The data sets generated during and/or analyzed during the current study are available in the Data for IJSR repository, <https://drive.google.com/drive/folders/1WtuBMHFwqSAQoz1rMOYK9-fSARIJRsN1?usp=sharing>.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Brenton H, Gillies M, Ballin D, Chatting D (2005) The uncanny valley: does it exist. In: Proceedings of conference of human computer interaction, workshop on human animated character interaction
- Cavedon L, Kroos C, Herath D, Burnham D, Bishop L, Leung Y, Stevens CJ (2015) C’mon dude!: Users adapt their behaviour to a robotic agent with an attention model. *Int J Hum Comput Stud* 80:14–23. <https://doi.org/10.1016/j.ijhcs.2015.02.012>
- Freud S, Strachey J, Freud A (1955) The standard edition of the complete psychological works
- Gurung N, Grant JB, Herath D (2021) What’s in a face? The effect of faces in human robot interaction. In: 2021 30th IEEE International conference on robot & human interactive communication (RO-MAN). pp 1024–1030. <https://doi.org/10.1109/RO-MAN50785.2021.951537>
- Gurung N, Herath D, Grant JB (2021) Feeling safe: a study on trust with an interactive robotic art installation. In: Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction. pp 447–451. <https://doi.org/10.1145/3434074.3447211>
- Hanson D, Olney A, Prilliman S, Mathews E, Zielke M, Hammons D, Fernandez R, Stephanou H (2005) Upending the uncanny valley. *AAAI* 5:1728–1729
- Ho CC, MacDorman KF (2010) Revisiting the uncanny valley theory: Developing and validating an alternative to the godspeed indices. *Comput Hum Behav* 26(6):1508–1518. <https://doi.org/10.1016/j.chb.2010.05.015>
- Ho CC, MacDorman KF (2017) Measuring the uncanny valley effect. *Int J Soc Robot* 9(1):129–139. <https://doi.org/10.1007/s12369-016-0380-9>
- Jochum E, Goldberg K (2016) Cultivating the uncanny: the telegarden and other oddities. *Robots and Art*. Springer, Berlin, pp 149–175
- Kroos C, Herath D (2009) Stelarc the articulated head: An intelligent interactive agent as an artistic installation. In: Proceedings of international conference on intelligent robots and systems
- Kühne K, Fischer MH, 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. *Front Neurobot* 14:105. <https://doi.org/10.3389/fnbot.2020.593732>
- Laakasuo M, Palomäki J, Köbis N (2021) Moral uncanny valley: a robot’s appearance moderates how its decisions are judged. *Int J Soc Robot*. <https://doi.org/10.1007/s12369-020-00738-6>
- Łupkowski P, Rybka M, Dziedzic D, Włodarczyk W (2019) The background context condition for the uncanny valley hypothesis. *Int J Soc Robot* 11(1):25–33. <https://doi.org/10.3389/fpsyg.2015.00390>
- McDonnell R, Breidt M, Bühlhoff HH (2012) Render me real? Investigating the effect of render style on the perception of animated virtual humans. *ACM Trans Graph* 31(4):1–11. <https://doi.org/10.1145/2793107.2793116>
- Mori M, MacDorman KF, Kageki N (2012) The uncanny valley [from the field]. *IEEE Robotics Automation Magazine* 19(2):98–100. <https://doi.org/10.1109/MRA.2012.2192811>
- Paetzel M (2016) The influence of appearance and interaction strategy of a social robot on the feeling of uncanniness in humans. In: Proceedings of the 18th ACM International conference on multi-modal interaction. pp 522–526. <https://doi.org/10.1145/2993148.2997612>
- Romportl J (2014) Speech synthesis and uncanny valley. *International conference on text, speech, and dialogue*. Springer, Berlin, pp 595–602
- Schwind V, Wolf K, Henze N, Korn O (2015) Determining the characteristics of preferred virtual faces using an avatar generator. In: proceedings of the 2015 annual symposium on computer-human interaction in play. pp 221–230. <https://doi.org/10.1145/2793107.2793116>
- Thepsonthorn C, Ogawa K-I, Miyake Yo (2021) The exploration of the uncanny valley from the viewpoint of the robot’s nonverbal behaviour. *Int J Soc Robot* 13(6):1443–1455. <https://doi.org/10.1007/s12369-020-00726-w>
- Tinwell A, Grimshaw M, Nabi DA (2015) The effect of onset asynchrony in audio- visual speech and the uncanny valley in virtual characters. *Int J Mech Robot Syst* 2(2):97–110. <https://doi.org/10.1504/IJMRS.2015.068991>
- Tinwell A, Grimshaw M, Williams A (2010) Uncanny behaviour in survival horror games. *J Gaming Virt Worlds* 2(1):3–25

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Neelu Gurung** is a graduate of the University of Canberra. She is working towards growing further in her achievement.

**Janie Busby Grant** is a researcher in cognitive psychology at the University of Canberra. Her research foci are future-oriented thought, mental health, and social robotics. She works with a wide range of industry and government partners, focusing on collaborative, multi-disciplinary health projects. Recent robotics work includes research examining the implementation of robot systems in hospital and aged care settings, as well as human cognition and perceptions of human-robot interaction.

**Damith Herath** is an Associate Professor in Robotics and Art at the University of Canberra. Damith is a multi-award-winning entrepreneur and a roboticist with extensive experience leading multidisciplinary research teams on complex robotic integration and industrial and research projects for over two decades. He founded Australia's first collaborative robotics startup in 2011 and was named one of the most innovative young tech companies in Australia in 2014. Teams he led in 2015 and 2016 consecutively became finalists and, in 2016, a top-ten category winner in the coveted Amazon Robotics Challenge - an industry-focussed competition amongst the robotics research elite. In addition, Damith has chaired several international workshops on Robots and Art and is the lead editor of the book "Robots and Art: Exploring an Unlikely Symbiosis" - the first significant work to feature leading roboticists and artists together in the field of Robotic Art.