Exploring the Appearance and Voice Mismatch of Virtual Characters

Minsoo Choi* Purdue University
Alexandros Koilias† University of the Aegean
Matias Volonte‡ Clemson University
Dominic Kao§ Purdue University
Christos Mousas¶ Purdue University

Abstract

We know that a voice that matches the appearance of a virtual character can positively impact how people interact with that character. However, although virtual characters and the human interactions with them have become more common, how mismatches in the appearance and voice of virtual characters could impact aspects of human perception toward virtual characters is underexplored. Thus, we conducted a 2 (appearance: human vs. robot virtual character) $\times$ 2 (voice: human vs. robot voice) within-group virtual reality study ($N = 21$) to explore how appearance and voice factors could impact how study participants rate the anthropomorphism, animacy, likability, perceived intelligence, and perceived safety of virtual characters. In our study, we instructed our participants to co-solve a jigsaw puzzle with the help of a virtual character, which we scripted to solve the jigsaw puzzle as efficiently as possible. Our results showed that in the presence of a human voice, the mismatching in the appearance and voice of the virtual character could still provide positive results in anthropomorphism, likability, and perceived safety, indicating that the human voice plays a critical enhancing role in how humans perceive a virtual character that is non-human in appearance.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality

1 Introduction

We are in an era in which humans, more often than before, interact with virtual characters (i.e., intelligent agents, virtual humans, avatars) in numerous applications, including intelligence and remote assistance, virtual instruction, and games. Several applications in these domains use virtual characters to interact and communicate with users in various ways. For example, virtual characters could be part of a virtual reality narrative in which they can either tell or be part of a story; intelligent entities (controlled through artificial intelligence algorithms) that can, for example, help users solve problems; and avatars or self-avatars that are controlled by the user using controllers or motion capture devices.

The intelligence assigned to a virtual character plays a key role in achieving effective interpersonal communication between a user and a virtual character [27]. However, we considered that additional factors could impact how humans perceive the virtual characters they interact with. The factor that researchers have examined most extensively is the appearance of the virtual characters [33,34,53,55]. This is because the appearance of a virtual character can range from unpleasant to pleasant [24], from machine-like to human-like [34], from aversive to rapport [53], and from unrealistic to realistic [43], and as a result could trigger different emotional responses from people. We also know from previous studies that the voice of the virtual character can impact humans' perception of the realism of the virtual character [45,52]. Moreover, some evidence suggests that voices that match the appearance of a virtual character positively impact how humans perceive and interact with virtual characters [16,25].

Considering that effective communication with a virtual character should involve the user being able to both observe and listen to that character, a reasonable direction for examination is the mismatch between the appearance and voice of virtual characters and how such a mismatch could impact how humans perceive them. Thus, in this paper, we aimed to extend such knowledge by exploring how virtual characters' appearance and voice combinations could impact anthropomorphism, animacy, likability, perceived intelligence, and perceived safety, which are concepts that, to the best of our knowledge, have not yet been explored in such a setting by the virtual reality community.

We approached the aforementioned direction by combining virtual characters with voices in a within-group study design. Specifically, in our 2 (appearance: human vs. robot virtual character) $\times$ 2 (voice: human vs. robot voice) study, we aimed to understand how the four examined combinations impact study participants' perceptions when we instructed them to work collaboratively on the same task with an intelligent virtual character. The two virtual characters we used in our study are shown in Figure 1. The findings from this research could help researchers, practitioners, and developers of virtual reality applications to implement more engaging and effective interpersonal communications between humans and virtual characters. As a result, both the engagement and the immersion of users could increase while simultaneously achieving the necessary perceptual goals.

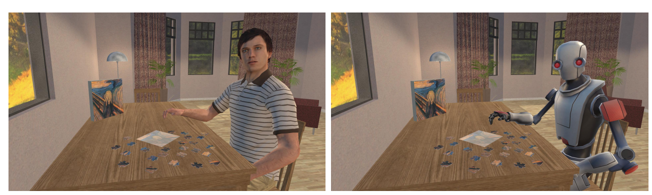

Figure 1: We explored the impact of the appearance and voice of virtual characters on anthropomorphism, animacy, likability, perceived intelligence, and perceived safety. We illustrate the human and robot virtual characters in our experimental virtual reality application.

Based on all of the above, in this paper we aimed to answer the following research questions:

- **RQ1:** How does appearance-voice mismatching impact study participants' anthropomorphism ratings?
- **RQ2:** How does appearance-voice mismatching impact study participants' animacy ratings?
- **RQ3:** How does appearance-voice mismatching impact study participants' likability ratings?
- **RQ4:** How does appearance-voice mismatching impact study participants' perceived intelligence ratings?
- **RQ5:** How does appearance-voice mismatching impact study participants' perceived safety ratings?

*e-mail: choi714@purdue.edu
†e-mail: alex.kilias@go.uop.gr
‡e-mail: mvolont@clemson.edu
§e-mail: kaod@purdue.edu
¶e-mail: cmousas@purdue.edu

Our paper is structured as follows. We present related works in Section 2. We describe our methodology in Section 3. We report our results in Section 4 and discuss them in Section 5. We note our study's limitations in Section 6. Finally, we address conclusions and potential future directions in Section 7.

2 Related Work

In the below subsections, we discuss work related to our project.

2.1 Virtual Character Appearance

The appearance of virtual characters has been of concern to the scientific community; thus, researchers have conducted studies to explore the impact of virtual character appearance on human behavior. Among other effects, researchers have studied the impact of a virtual character's appearance to understand how aspects such as render style [35,55], gender [26,42], age [6,40,48], anthropomorphism [12,28,34], skin color [2,37], and more, could impact human's perceptions [54] and emotional reactions [33] toward that virtual character and provoke a change in human behavior and attitude [36,41]. As a result, researchers have made available significant knowledge on the effects of virtual character appearance on the psychophysical aspects of human behavior.

Bailenson et al. [5] found a correlation between a character's appearance with realism and proximity; they also reported several complex implications for the behavior and appearance of virtual characters. McDonnell et al. [29] and Zibrek et al. [53] explored the render styles of a virtual character. They found that the rendering style could impact human perception, including appeal ratings and perceived realism [29]. They also found that rendering style and personality traits had the most significant impact on the appeal of virtual characters [53]. Nelson et al. [35] and Mousas et al. [34] explored avoidance behavior toward virtual characters and found that more aversive stimuli (i.e., zombie-like virtual characters) significantly impacted humans' avoidance behavior.

However, regarding robots, several findings have been attributed to the uncanny valley effect initially hypothesized by Mori [32]. We know from prior research that when the appearance of a robot becomes more human-like, it elicits more positive and empathetic responses until it reaches a peak or realism, at which point subtle shortcomings in human-likeness cause feelings of eeriness or even disgust and fear [16,18,49]. Moreover, Bartneck et al. [7] reported no difference in the likability of a human and a highly humanoid physical robot and found that realistic movement did not significantly increase the robot's human-likeness or likability. Piwek et al. [38] found that natural motion improves the acceptability of virtual characters. Still, Urgen et al. [47] reported that a mismatch in the realism of appearance and movement elicited an uncanniness effect for a visually highly realistic physical robot. Lastly, Ferstl et al. [16] found that maximizing the realism of voice and motion is preferable, even when this leads to realism mismatches, although lower realism may be preferable for visual appearance.

2.2 Virtual Character Voice

One way to enhance the realism of a virtual character is to assign it a voice. In general, we can assign to virtual characters the prerecorded voice of a human voice actor [21], synthesized voices using text-to-speech solutions [39], or a mixture of both [22]. Researchers have reported that humans perceive synthesized voices as less sympathetic [45] and less preferable [9] when compared to real human voices. However, researchers also mentioned that the naturalness of a voice is related to the speaker's distinct characteristics rather than the realism of the voice [1,20].

Regarding voices assigned to virtual characters, researchers reported that unnatural voices do not necessarily impact the character's social presence or empathetic responses from the audience [4,19]. However, a mismatch in realism between the voice and the character's appearance could create a sense of uncanniness in humans, resulting in discomfort [19,31]. Among others, Ferstl et al. [16] highlighted the importance of believable voices in virtual characters by illustrating that realism in voice is preferable over realism of appearance when they produce perceptual mismatches. Moreover, Lam et al. [25] found that it is possible to enhance the believability of characters by generating appropriate voices through pitch manipulation of voices.

2.3 Contributions

So far, researchers have explored both the appearance and voice of virtual characters individually and in various combinations and demonstrated that matching appropriate voices to virtual characters is crucial. This is because vocal characteristics and their appropriateness can influence human perception of a character. Less explored, however, is how voice and appearance mismatching could impact other aspects of human perception, such as anthropomorphism, animacy, likability, perceived intelligence, and perceived safety toward virtual characters. Thus, we consider this to be the main contribution of this paper, as it expands current knowledge on appearance and voice mismatching in virtual characters.

3 Materials and Methods

In this section, we present this study's methodology and implementation details.

3.1 Participants

We conducted an a priori power analysis using the G*Power software to determine the appropriate sample size for our study [15]. We used the following settings for our power analysis: a medium effect size of $f = .30$, an $\alpha = .05$ error probability, one group with four repeated measurements, an $r = .50$ correlation among repeated measures, and a $\varepsilon = .70$ for non-sphericity correction. Based on these settings, to achieve a .80 power ($1-\beta$ error probability), the power analysis recommended an $N = 21$ sample size.

For our study, we recruited our participants based on emails sent out to all undergraduate and graduate students at our university, as well as through class announcements. Our body of participants (age range: $19-32$ years old) comprised seven female (age: $M = 22.85$, $SD = 2.11$) and 14 male (age: $M = 24.14$, $SD = 3.48$) students. All 21 participants were volunteers without the expectation of credit or compensation, and all finished the study successfully. None of the participants reported visual or auditory processing disorders.

3.2 Virtual Reality Application

We developed our virtual reality application in the Unity (version 2020.3.20) game engine using the Oculus Integration Toolkit. Both for our implementation and study, we used a Dell Alienware Aurora R7 desktop computer (Intel Core i7, NVIDIA GeForce RTX 2080, 32GB RAM). Also, we used Meta's Quest 1 as our virtual reality head-mounted display.

We designed a living room environment for our application in which to immerse our participants (see Figure 2). We sited the virtual character at the table in the virtual living room and placed our participants sitting next to the virtual character. We placed a semitransparent puzzle board so that participants could indicate where the jigsaw puzzle pieces should be placed. We also positioned the jigsaw puzzle box on the table so that the participants could observe a clear picture of the finished puzzle. We placed the jigsaw puzzle pieces randomly on the table at a reachable distance from the participant and the virtual character; thus, both could easily grab them.

During the runtime of our application, we instructed our participants to use the virtual reality controller to grab a jigsaw puzzle piece (by pressing the grab button) and then leave it on the puzzle

(a) (b) (c)

Figure 2: The virtual reality environment we used in our study (a) from a perspective view and (b) a top view of the table area in which our participants interacted with the puzzle. The red asterisk in (b) indicates where we placed our participants. (c) Still captured during the jigsaw puzzle co-solving process in which our participant's virtual hand grabs a jigsaw puzzle piece.

board (by releasing the grab button). In total, there were 25 jigsaw puzzle pieces (our puzzle is a 5×5 grid). The size of each jigsaw puzzle piece is 4×4 cm. On the semitransparent puzzle board, we placed transparent target spots indicating where each piece should be placed. Once a jigsaw puzzle piece was at an appropriate distance from the target spot, the jigsaw puzzle piece snapped to the target spot once the participants released the grab button. Figure 2(c) illustrates, from a first-person view, the participant picking up a jigsaw puzzle piece.

We downloaded the human male virtual character (Male_Adult_01) from Microsoft's Rocketbox Avatar[1] library and the robot virtual character (Space Robot Kyle — URP[2]) from Unity's asset store. In our application, we assigned an idle sitting animation to the virtual characters, and we used an inverse kinematics solution (the forward and backward inverse kinematics solver [3]) to edit the hand reaching and grabbing animation of the virtual characters. We developed a script that allows the character to solve the jigsaw puzzle game. In our script, the system chooses a jigsaw puzzle piece that has been placed randomly on the table (has not been placed on the puzzle board) and generates the grabbing animation so the virtual character grabs it. The script then lets the character place the jigsaw puzzle piece in the appropriate target spot. So that the participants could perceive the virtual character as an intelligent one, we made our virtual character solve the puzzle efficiently by always placing the selected jigsaw puzzle piece in the correct target spot on the puzzle board.

We implemented small talk between the participant and the virtual character that appeared during the puzzle co-solving process. The participants of our study were able to respond to the virtual characters by choosing options from a GUI that appeared during the small talk. For the human virtual character, we generated the speech assigned to the virtual character using Microsoft's Azure text-to-speech[3] service using the Tony actor (male adult from the United States voice model) with a cheerful speaking style. For the robot virtual character, we used the Audacity (version 3.3.1) software to create the generated robotic version of the synthesized human voice. In Audacity, we created three tracks with the initial synthesized human voice. For the first audio track, we applied the `echo` effect with a .04 `delay time` and a .60 `decay factor`. For the second audio track, we changed the `pitch` by applying a -10% change. For the third audio track, we changed the `tempo` by applying a -3% change. We want to note that in our study we decided to use synthetic voices since the voice synthesizer and the sound editing software permitted more control over voice pitch levels. We used the Salsa LipSync Suite[4] from Unity's asset store to lip-synch the human virtual character. Moreover, we applied eye blinks and head movement to increase realism in our human virtual character. The lip-sync and facial animations did not apply to the robot virtual character, as the robot's eye cannot blink and it does not have a mouth.

[1]https://www.microsoft.com/en-us/research/blog/microsoft-rocketbox-avatar-library-now-available-for-research-and-academic-use/

[2]https://assetstore.unity.com/packages/3d/characters/robots/space-robot-kyle-urp-4696

[3]https://azure.microsoft.com/en-us/products/cognitive-services/text-to-speech

3.3 Experimental Conditions

We developed four experimental conditions: 1) human virtual character with human voice, 2) human virtual character with robot voice, 3) robot virtual character with human voice, and 4) robot virtual character with robot voice, following our 2 (appearance: human vs. robot virtual character) $\times$ 2 (voice: human vs. robot voice) within-group design. We provide examples of the four experimental conditions in our supplementary materials video.

3.4 Contributions

For our study, we used the instrument developed by Bartneck et al. [8] to assess the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of virtual characters. Initially, this instrument was developed to evaluate robots; however, it has been adopted widely in several related fields in the past few years. Thus, considering that in our study we assess robot appearance and voice against human appearance and voice, we realized that such an instrument could be quite appropriate. We want to mention that we slightly modified the developed instrument to reflect our study. Participants rated their experiences on a 5-point scale as initially proposed in the Bartneck et al. [8] instrument.

3.5 Procedure

Our participants selected the day and time that best fit their schedule. Once they arrived in our lab, the research team provided the consent form approved by the Institutional Review Board of our university and then instructed them to read it carefully and, if they agreed, to sign it and participate in the study. The research team was also willing to answer any participants' questions about the study and the procedure.

After signing the consent form, the research team provided our participants with a virtual reality head-mounted display (the Oculus Quest 1) and helped them to set it up. In our lab space, we asked the participants to sit in a chair. Note that no obstacles were near the participants; thus, they could perform the reaching tasks without thinking or hurting themselves. For our experiment, the sequence in which each participant experienced the four conditions was decided by the Latin square method [51], which balanced the conditions for first-order carryover effects.

[4]https://assetstore.unity.com/packages/tools/animation/salsa-lipsync-suite-148442

Once each condition was completed, we asked our participants to fill out our survey in the Qualtrics survey tool. We also informed them that they could have additional break time between the conditions. After completing all four conditions, we let the participants ask any questions and encouraged them to provide feedback about the experiment and their experiences with the four different conditions. Our participants spent less than 40 minutes to complete the whole study.

4 Results

For our statistical analyses, we used the four experimental conditions as independent variables and the self-reported ratings as dependent variables. We examined the normality of our data graphically using Q-Q plots of the residuals and with the Shapiro-Wilk test at the 5% level. The collected data fulfilled the normality criteria. Thus, for our statistical analyses we used a two-way repeated measures analysis of variance (RM-ANOVA) for each measurement.

Anthropomorphism. Our simple main effect analysis on the appearance factor showed that participants rated the anthropomorphism of the human ($M = 3.33$, $SE = .09$) higher than the robot ($M = 2.43$, $SE = .06$) virtual character (Wilk's $\Lambda = .199$, $F[1,20] = 80.616$, $p < .001$, $\eta_p^2 = .801$). Similarly, simple main effects on the voice factor showed participants rated the anthropomorphism of the human ($M = 3.18$, $SE = .07$) higher than the robot ($M = 2.58$, $SE = .07$) voice (Wilk's $\Lambda = .321$, $F[1,20] = 42.402$, $p < .001$, $\eta_p^2 = .679$). We also found a statistically significant appearance $\times$ voice interaction effect (Wilk's $\Lambda = .673$, $F[1,20] = 9.700$, $p = .005$, $\eta_p^2 = .327$), indicating that, in the presence of a human voice, participants rated the anthropomorphism of the two virtual characters higher.

Animacy. We did not find a statistically significant result for either the appearance factor (Wilk's $\Lambda = .965$, $F[1,20] = .729$, $p = .403$, $\eta_p^2 = .035$) or voice factor (Wilk's $\Lambda = .691$, $F[1,20] = .802$, $p = .381$, $\eta_p^2 = .039$), or appearance $\times$ voice interaction (Wilk's $\Lambda = .928$, $F[1,20] = 1.562$, $p = .226$, $\eta_p^2 = .072$).

Likability. We did not find a statistically significant result for the appearance factor (Wilk's $\Lambda = .999$, $F[1,20] = .127$, $p = .872$, $\eta_p^2 = .201$). However, our simple main effect analysis on the voice factor showed participants rated the likability of the human ($M = 3.75$, $SE = .06$) higher than the robot ($M = 2.99$, $SE = .08$) voice (Wilk's $\Lambda = .361$, $F[1,20] = 56.575$, $p < .001$, $\eta_p^2 = .739$). We also found a statistically significant appearance $\times$ voice interaction effect (Wilk's $\Lambda = .714$, $F[1,20] = 8.026$, $p = .10$, $\eta_p^2 = .286$), indicating that, in the presence of a human voice, participants rated the likability of the two virtual characters higher.

Perceived Intelligence. Our simple main effect analysis on the appearance factor showed that participants provided higher ratings in perceived intelligence for the robot ($M = 3.14$, $SE = .06$) than for the human ($M = 2.95$, $SE = .05$) virtual character (Wilk's $\Lambda = .778$, $F[1,20] = 5.707$, $p = .027$, $\eta_p^2 = .222$). However, neither the simple main effect analysis on the voice factor (Wilk's $\Lambda = .915$, $F[1,20] = 1.862$, $p = .188$, $\eta_p^2 = .085$) nor the appearance $\times$ voice interaction effect (Wilk's $\Lambda = .890$, $F[1,20] = 2.466$, $p = .132$, $\eta_p^2 = .110$) revealed statistically significant results.

Perceived Safety. We did not find a statistically significant result for the appearance factor (Wilk's $\Lambda = .996$, $F[1,20] = .088$, $p = .769$, $\eta_p^2 = .104$). However, our simple main effect analysis on the voice factor showed that participants rated the perceived safety of the human ($M = 3.57$, $SE = .09$) higher than the robot ($M = 2.73$, $SE = .11$) voice (Wilk's $\Lambda = .343$, $F[1,20] = 38.321$, $p < .001$, $\eta_p^2 = .457$). We also found a statistically significant appearance $\times$ voice interaction effect (Wilk's $\Lambda = .466$, $F[1,20] = 22.873$, $p < .001$, $\eta_p^2 = .534$), indicating that, in the presence of a human voice, participants rated the perceived safety of the two virtual characters higher.

5 Discussion

Our study revealed several interesting findings. Regarding the anthropomorphism of the virtual characters (**RQ1**), we can argue that our participants distinguished the different appearances of the two characters and rated them accordingly, rating the anthropomorphism of the human character higher than the robot virtual character. We consider this an expected finding that is in line with a previous study [16]. Similarly, our participants rated the human voice we assigned to the virtual characters as more anthropomorphic than the robot voice, confirming prior work [14, 44]. Interestingly, we found an interaction effect on anthropomorphism, indicating that a human voice prompted our participants to rate our virtual character as more anthropomorphic compared to the same character assigned a robot voice. This is a result that extends Ferstl et al. [16] in terms of the specific metric used, who indicated that realism in voice is preferable to the realism of appearance, and Piwek et al. [38] and Thompson et al. [46], who found that increasing motion realism consistently increased ratings of anthropomorphism. Our finding shows that, despite a mismatch between appearance and voice, the human voice could increase how humans perceive the anthropomorphism of a virtual character. We consider this a relatively novel finding, indicating that a human-like voice can impact how we perceive a non-human and mechanical-like virtual character.

Our findings regarding animacy (**RQ2**) were not statistically significant. We partially expected these results. We know from previously published work that motions can impact how humans perceive virtual characters [33]; body motion is a crucial factor in the perceived realism and likability of virtual humans [16], and humans are more sensitive to temporal misalignment of body motions [13]. However, in our study, the two virtual characters were animated identically in all conditions. Our participants realized that there was no difference in the examined conditions. Thus, we can conclude that neither the appearance nor the voice of a virtual character could impact how our participants perceived a virtual character's animacy, extending previous findings [16, 47] by indicating that a mismatch in the realism of appearance and voice did not impact participants' perception of the animacy of the virtual characters.

In terms of how our participants rated the likability of the virtual characters (**RQ3**), we found that, interestingly, they did not differentiate their ratings in terms of the appearance of the virtual character. Thus, although we know from previous research that different virtual characters' appearance [16, 19, 34] can impact how humans perceive them, we think that our study participants decided to assign similar ratings to the two virtual characters because of the less aversive appearance of the robot virtual character [16]. Contrarily, our participants preferred the human voice that we used for our virtual character, in accordance with similar research in the human-robot interaction domain [50]. However, we obtained another interesting result. Our participants increased their likability rating when we applied a human voice to a virtual character, which is a finding that expands Ferstl et al.'s [16] study, indicating that a human voice alone may be enough to increase the likability of a virtual character, even if there is a mismatch between appearance and voice.

Our participants rated the perceived intelligence higher (**RQ4**) when interacting with the robot virtual character than with the human virtual character conditions. Although we know from prior research that perceived intelligence is related to an agent's ability to complete tasks, it is also related to a range of human-agent interaction design elements, such as appearance [23] and interaction modality [10]. Moreover, evidence shows that humans prefer teammates they believe to be human-controlled, even if they are actually intelligent virtual characters [30]. Thus, considering that the human virtual character was computer-controlled, and taking into account the simplicity of the interaction modality we implemented, our study participants perceived the robot virtual character as more intelligent. Lastly, regarding the non-significant results of appearance and voice interaction effect, we argue that the mismatch in appearance and voice was not enough to change our participants' perception of the intelligence of the virtual character.

We found that when we assigned a human voice to our virtual characters, the participants rated the perceived safety higher (**RQ5**). This is quite an interesting result, showing that even when there is a mismatch between appearance and voice, humans tend to feel safer when the voice is human-like. We know from previous studies of autonomous vehicles that humans prefer female voices [11]. Moreover, we know from virtual assistant research that vocal characteristics engender higher trust in humans when interacting with such systems [17]. Thus, we could say that our results extend these findings, showing that humans generally prefer human voices when interacting with virtual characters. However, further experimentation is needed to conclude whether female voices and other vocal characteristics could potentially increase feelings of safety and trust toward virtual characters.

6 Limitations

We report our limitations so that researchers willing to conduct studies on virtual characters and voices can consider them in the future. Note that the listed limitations do not invalidate our study and its aim of understanding the mismatch between the appearance and voice of virtual characters.

First, our study considered only a single intelligence level assigned to our virtual character. Thus, we think that exploring how study participants respond to the scales we used when implementing a less intelligent virtual character is critical. Such a study might reveal insightful results as to how the intelligence of the virtual character could impact how humans perceive that character.

Second, we implemented a relatively small amount of talk with the virtual character. Thus, although the participants could observe the virtual character continuously, they only listened to them at the beginning, middle, and end of their interaction. Although we do not believe that this greatly impacted our participants' responses, examining how a more extensive and active dialog could impact study participants' responses is essential.

Third, our robot virtual character had a relatively simple face with only eyes attached to it. It was not possible to apply lip-sync and eye blinks. We think that facial animations could enhance the robot virtual character's realism and positively impact our participants' anthropomorphism ratings. Thus, we will need to reexamine our findings when both an even more abstract and a more detailed robot appearance are used.

7 Conclusions and Future Work

In this paper, we examined the effects of the appearance and voice of virtual characters on several human perceptions of virtual characters. Overall, we found that a virtual character's appearance and voice can induce positive reactions when the voice is human-like, even if there is a mismatch. Although this study aligns partially with previously published work which showed that the appearance and voice of a virtual character could impact human behavior [9, 21, 31, 52], we expanded that knowledge, showing that appearance and voice can also impact other perceptual aspects regarding virtual characters. Moreover, our results suggest that even a voice that can be considered less common and natural (in our case, the robot voice) is sufficient to negatively impact humans' perception toward virtual characters.

In future studies, we plan to focus on aspects of intelligence and behavior assigned to our virtual characters to understand how such factors could impact how humans perceive them. The aspects that we would like to focus on include different intelligent levels, more advanced dialog systems, and more advanced puzzle-solving capabilities. We think these directions could help us build a more intelligent and socially engaging virtual character that could help us further explore human perceptions toward virtual characters.

References

[1] T. B. Amin, P. Marziliano, and J. S. German. Glottal and vocal tract characteristics of voice impersonators. *IEEE Transactions on Multimedia*, 16(3):668–678, 2014.

[2] A. D. Andrade, J. G. Ruiz, M. J. Mintzer, P. Cifuentes, R. Anam, J. Diem, O. Gómez-Marín, H. Sun, and B. A. Roos. Medical students' attitudes toward obese patient avatars of different skin color. In *Medicine Meets Virtual Reality 19*, pp. 23–29. IOS Press, 2012.

[3] A. Aristidou and J. Lasenby. Fabrik: A fast, iterative solver for the inverse kinematics problem. *Graphical Models*, 73(5):243–260, 2011.

[4] M. P. Aylett, A. Vinciarelli, and M. Wester. Speech synthesis for the generation of artificial personality. *IEEE transactions on affective computing*, 11(2):361–372, 2017.

[5] J. N. Bailenson, K. Swinth, C. Hoyt, S. Persky, A. Dimov, and J. Blascovich. The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence*, 14(4):379–393, 2005.

[6] D. Banakou, R. Groten, and M. Slater. Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences*, 110(31):12846–12851, 2013.

[7] C. Bartneck, T. Kanda, H. Ishiguro, and N. Hagita. My robotic doppelgänger-a critical look at the uncanny valley. In *RO-MAN 2009-The 18th IEEE international symposium on robot and human interactive communication*, pp. 269–276. IEEE, 2009.

[8] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1:71–81, 2009.

[9] J. Cabral. The influence of synthetic voice on the evaluation of a virtual character. ISCA, 2017.

[10] A. Deshmukh, B. Craenen, M. E. Foster, and A. Vinciarelli. The more i understand it, the less i like it: The relationship between understandability and godspeed scores for robotic gestures. In *2018 27th IEEE International symposium on robot and human interactive communication (RO-MAN)*, pp. 216–221. IEEE, 2018.

[11] J. Dong, E. Lawson, J. Olsen, and M. Jeon. Female voice agents in fully autonomous vehicles are not only more likeable and comfortable, but also more competent. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 64, pp. 1033–1037. SAGE Publications Sage CA: Los Angeles, CA, 2020.

[12] C. Dubosc, G. Gorisse, O. Christmann, S. Fleury, K. Poinsot, and S. Richir. Impact of avatar facial anthropomorphism on body ownership, attractiveness and social presence in collaborative tasks in immersive virtual environments. *Computers & Graphics*, 101:82–92, 2021.

[13] C. Ennis, R. McDonnell, and C. O'Sullivan. Seeing is believing: body motion dominates in multisensory conversations. *ACM Transactions on Graphics (TOG)*, 29(4):1–9, 2010.

[14] F. Eyssel, D. Kuchenbrandt, S. Bobinger, L. De Ruiter, and F. Hegel. 'if you sound like me, you must be more human' on the interplay of robot and user features on human-robot acceptance and anthropomorphism. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pp. 125–126, 2012.

[15] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner. G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2):175–191, 2007.

[16] Y. Ferstl, S. Thomas, C. Guiard, C. Ennis, and R. McDonnell. Human or robot? investigating voice, appearance and gesture motion realism of conversational social agents. In *Proceedings of the 21st ACM international conference on intelligent virtual agents*, pp. 76–83, 2021.

[17] K. L. Goodman and C. B. Mayhorn. Pitch perfect: Influence of perceived voice agent gender and vocal pitch on trust and reliance. In *Proceedings of the Human Factors and Ergonomics Society Annual*

Meeting, vol. 65, pp. 1529–1530. SAGE Publications Sage CA: Los Angeles, CA, 2021.
[18] D. Hepperle, C. F. Purps, J. Deuchler, and M. Wölfel. Aspects of visual avatar appearance: self-representation, display type, and uncanny valley. *The visual computer*, 38(4):1227–1244, 2022.
[19] D. Higgins, K. Zibrek, J. Cabral, D. Egan, and R. McDonnell. Sympathy for the digital: Influence of synthetic voice on affinity, social presence and empathy for photorealistic virtual humans. *Computers & Graphics*, 104:116–128, 2022.
[20] D.-Y. Huang, E. P. Ong, S. Rahardja, M. Dong, and H. Li. Transformation of vocal characteristics: A review of literature. *International Journal of Electrical and Computer Engineering*, 3(12):2288–2296, 2009.
[21] D. Kao, R. Ratan, C. Mousas, A. Joshi, and E. F. Melcer. Audio matters too: How audial avatar customization enhances visual avatar customization. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–27, 2022.
[22] D. Kao, R. Ratan, C. Mousas, and A. J. Magana. The effects of a self-similar avatar voice in educational games. *Proceedings of the ACM on Human-Computer Interaction*, 5(CHI PLAY):1–28, 2021.
[23] J. Kim and I. Im. Anthropomorphic response: Understanding interactions between humans and artificial intelligence agents. *Computers in Human Behavior*, 139:107512, 2023.
[24] O. Korn, L. Stamm, and G. Moeckl. Designing authentic emotions for non-human characters: A study evaluating virtual affective behavior. In *Proceedings of the 2017 conference on designing interactive systems*, pp. 477–487, 2017.
[25] L. Lam, M. Choi, M. Mukanova, K. Hauser, F. Zhao, R. Mayer, C. Mousas, and N. Adamo. Effects of body type and voice pitch on perceived audio-visual correspondence and believability of virtual characters. In *ACM Symposium on Applied Perception*, 2023.
[26] J.-E. R. Lee, C. I. Nass, and J. N. Bailenson. Does the mask govern the mind?: Effects of arbitrary gender representation on quantitative task performance in avatar-represented virtual groups. *Cyberpsychology, Behavior, and Social Networking*, 17(4):248–254, 2014.
[27] B. Lugrin, C. Pelachaud, and D. Traum. The handbook on socially interactive agents: 20 years of research on embodied conversational agents, intelligent virtual agents, and social robotics, volume 2: Interactivity, platforms, application. 2022.
[28] J.-L. Lugrin, I. Polyschev, D. Roth, and M. E. Latoschik. Avatar anthropomorphism and acrophobia. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, pp. 315–316, 2016.
[29] R. McDonnell, M. Breidt, and H. H. Bülthoff. Render me real? investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics (TOG)*, 31(4):1–11, 2012.
[30] T. Merritt, K. McGee, T. L. Chuah, and C. Ong. Choosing human team-mates: perceived identity as a moderator of player preference and enjoyment. In *Proceedings of the 6th International Conference on Foundations of Digital Games*, pp. 196–203, 2011.
[31] W. J. Mitchell, K. A. Szerszen Sr, A. S. Lu, P. W. Schermerhorn, M. Scheutz, and K. F. MacDorman. A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception*, 2(1):10–12, 2011.
[32] M. Mori. Bukimi no tani (the uncanny valley). *Energy*, 7(4):33–35, 1970.
[33] C. Mousas, D. Anastasiou, and O. Spantidi. The effects of appearance and motion of virtual characters on emotional reactivity. *Computers in Human Behavior*, 86:99–108, 2018.
[34] C. Mousas, A. Koilias, B. Rekabdar, D. Kao, and D. Anastaslou. Toward understanding the effects of virtual character appearance on avoidance movement behavior. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 40–49. IEEE, 2021.
[35] M. G. Nelson, A. Koilias, C.-N. Anagnostopoulos, and C. Mousas. Effects of rendering styles of a virtual character on avoidance movement behavior. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 594–599. IEEE, 2022.
[36] X. Pan, M. Gillies, and M. Slater. Virtual character personality influences participant attitudes and behavior–an interview with a virtual human character about her social anxiety. *Frontiers in Robotics and AI*, 2:1, 2015.
[37] T. C. Peck, S. Seinfeld, S. M. Aglioti, and M. Slater. Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and cognition*, 22(3):779–787, 2013.
[38] L. Piwek, L. S. McKay, and F. E. Pollick. Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition*, 130(3):271–277, 2014.
[39] L. Qiu and I. Benbasat. An investigation into the effects of text-to-speech voice and 3d avatars on the perception of presence and flow of live help in electronic commerce. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12(4):329–355, 2005.
[40] R. Reinhard, K. G. Shah, C. A. Faust-Christmann, and T. Lachmann. Acting your avatar's age: effects of virtual reality avatar embodiment on real life walking speed. *Media Psychology*, 23(2):293–315, 2020.
[41] M. Sakamoto, T. Nakajima, and T. Alexandrova. Enhancing values through virtuality for intelligent artifacts that influence human attitude and behavior. *Multimedia Tools and Applications*, 74:11537–11568, 2015.
[42] V. Schwind, P. Knierim, C. Tasci, P. Franczak, N. Haas, and N. Henze. " these are not my hands!" effect of gender on the perception of avatar hands in virtual reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 1577–1582, 2017.
[43] V. Schwind, K. Wolf, and N. Henze. Avoiding the uncanny valley in virtual character design. *interactions*, 25(5):45–49, 2018.
[44] V. K. Sims, M. G. Chin, H. C. Lum, L. Upham-Ellis, T. Ballion, and N. C. Lagattuta. Robots' auditory cues are subject to anthropomorphism. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 53, pp. 1418–1421. SAGE Publications Sage CA: Los Angeles, CA, 2009.
[45] S. Thomas, Y. Ferstl, R. McDonnell, and C. Ennis. Investigating how speech and animation realism influence the perceived personality of virtual characters and agents. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 11–20. IEEE, 2022.
[46] J. C. Thompson, J. G. Trafton, and P. McKnight. The perception of humanness from the movements of synthetic agents. *Perception*, 40(6):695–704, 2011.
[47] B. A. Urgen, M. Kutas, and A. P. Saygin. Uncanny valley as a window into predictive processing in the social brain. *Neuropsychologia*, 114:181–185, 2018.
[48] N. Vahle and M. J. Tomasik. The embodiment of an older avatar in a virtual reality setting impacts the social motivation of young adults. *Experimental Aging Research*, 48(2):164–176, 2022.
[49] M. L. Walters, D. S. Syrdal, K. Dautenhahn, R. Te Boekhorst, and K. L. Koay. Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Autonomous Robots*, 24:159–178, 2008.
[50] M. L. Walters, D. S. Syrdal, K. L. Koay, K. Dautenhahn, and R. Te Boekhorst. Human approach distances to a mechanical-looking robot with different robot voice styles. In *RO-MAN 2008-The 17th IEEE international symposium on robot and human interactive communication*, pp. 707–712. IEEE, 2008.
[51] E. J. Williams. Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Chemistry*, 2(2):149–168, 1949.
[52] K. Zibrek, J. Cabral, and R. McDonnell. 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*, pp. 1–6, 2021.
[53] K. Zibrek, E. Kokkinara, and R. McDonnell. The effect of realistic appearance of virtual characters in immersive environments-does the character's personality play a role? *IEEE transactions on visualization and computer graphics*, 24(4):1681–1690, 2018.
[54] K. Zibrek, S. Martin, and R. McDonnell. Is photorealism important for perception of expressive virtual humans in virtual reality? *ACM Transactions on Applied Perception (TAP)*, 16(3):1–19, 2019.
[55] K. Zibrek and R. McDonnell. Does render style affect perception of personality in virtual humans? In *Proceedings of the ACM symposium on applied perception*, pp. 111–115, 2014.