electronics

Article

The Effect of Eye Contact in Multi-Party Conversations with Virtual Humans and Mitigating the Mona Lisa Effect

Junyeong Kum [1], Sunghun Jung [1] and Myungho Lee [2,*]

[1] Department of Information Convergence Engineering, Pusan National University, Busan 46241, Republic of Korea; kum.junyeong@pusan.ac.kr (J.K.); sunghun@pusan.ac.kr (S.J.)
[2] School of Computer Science and Engineering, Pusan National University, Busan 46241, Republic of Korea
* Correspondence: myungho.lee@pnu.edu; Tel.: +82-51-510-2347

Abstract: The demand for kiosk systems with embodied conversational agents has increased with the development of artificial intelligence. There have been attempts to utilize non-verbal cues, particularly virtual human (VH) eye contact, to enable human-like interaction. Eye contact with VHs can affect satisfaction with the system and the perception of VHs. However, when rendered in 2D kiosks, the gaze direction of a VH can be incorrectly perceived, due to a lack of stereo cues. A user study was conducted to examine the effects of the gaze behavior of VHs in multi-party conversations in a 2D display setting. The results showed that looking at actual speakers affects the perceived interpersonal skills, social presence, attention, co-presence, and competence in conversations with VHs. In a second study, the gaze perception was further examined with consideration of the Mona Lisa effect, which can lead users to believe that a VH rendered on a 2D display is gazing at them, regardless of the actual direction, within a narrow range. We also proposed the camera rotation angle fine tuning (CRAFT) method to enhance the users' perceptual accuracy regarding the direction of the VH's gaze.The results showed that the perceptual accuracy for the VH gaze decreased in a narrow range and that CRAFT could increase the perceptual accuracy.

Keywords: virtual reality; virtual human; Mona Lisa effect; non-verbal cues; interaction; eye contact

Citation: Kum, J.; Jung, S.; Lee, M. The Effect of Eye Contact in Multi-Party Conversations with Virtual Humans and Mitigating the Mona Lisa Effect. *Electronics* **2024**, *13*, 430. https://doi.org/10.3390/electronics13020430

Academic Editor: Beiwen Li

Received: 7 December 2023
Revised: 14 January 2024
Accepted: 17 January 2024
Published: 19 January 2024

1. Introduction

Virtual humans (VHs) are increasingly being utilized in diverse fields, such as healthcare and customer service. The SimSensei, for instance, exploited a VH as a counselor, providing comfortable conversations with patients, while monitoring their emotional reactions [1]. Additionally, VHs like Ada and Grace served as informative guides for tourists in museums [2]. Another notable example is the VH Mack at the MIT Media Lab, which responds to queries about research groups, projects, and staff, and offers navigational assistance within the facility [3]. This burgeoning trend has been fueled by advancements in the creation of realistic, human-like representations and by the capabilities of large language models like GPT [4]. Further bolstered by progress in computer vision, which enables VHs to detect and respond to users' emotions and non-verbal cues [5], these technologies have collectively expanded the deployment of VHs in kiosk systems.

There have been several studies focused on interaction with VHs. In general, users anticipate the same level of social interaction with VHs as they would with real humans [6]. To effectively mimic this interaction, VHs should provide verbal and nonverbal cues simultaneously. While 2D displays equipped with a sound system are proficient for conveying verbal cues, a significant challenge arises when attempting to deliver nonverbal cues such as gestures, facial expressions, and eye contact [7,8]. Among these cues, eye contact is a powerful way to express interest and attention during a conversation and serves as a strong indicator of turn-taking in multi-party environments [9].

However, displaying a 3D VH on a 2D display creates an optical illusion known as the Mona Lisa effect, where users within a specific range may misinterpret the gaze direction

of the VH [10]. In real-world conversations, individuals can clearly distinguish the gazes of others. In contrast, due to the Mona Lisa effect, the users may think that the VH is looking at them, even when the VH looks elsewhere. This issue is particularly pronounced in multi-party conversations when using a 2D display kiosk, as maintaining eye contact in a relatively narrow area becomes challenging. This paper explores the need for eye contact in multi-party conversations using VHs through user experience. Furthermore, a simple way to reduce the Mona Lisa effect is proposed.

This paper is structured as follows: Section 2 summarizes related works on embodied conversational agents (ECAs) and non-verbal cues in interactions with ECAs, especially eye contact; Section 3 details the first experiment exploring the effects of the gaze behavior of 2D display ECAs in multi-party conversations; Section 4 proposes and validates our proposed method to mitigate the Mona Lisa effect within a narrow range; finally, Section 5 concludes the paper, addressing future research directions.

2. Related Works

2.1. Embodied Conversational Agents (ECAs)

Text-based chatbots are used in various fields, such as customer assistance, education, and healthcare [11,12]. A large amount of research has focused on the user experience when interacting with chatbots [13]. The advantages of a text-based chatbot are that it is easy to create, can respond fast, and has the ability to answer simple questions. According to research, users evaluated text-based chatbots as efficient, time-saving, and appropriate for simple and intuitive tasks [14].

However, many users were disappointed with text-based chatbots, due to their lack of conversational capabilities such as natural language processing and limited interaction [15]. For responsive and sensitive tasks, the users tended to demand the satisfaction of their emotional needs such as trust and authenticity [16]. With only text, it was difficult to satisfy these needs. Therefore, by providing an ECA with a human-like appearance, researchers have tried to mimic interactions with real people.

ECAs can use additional communication channels, such as voice, facial expressions, and gestures during communication [17]. The research of Broadbent et al. showed how the appearance of a robot's face on a display could affect user perception [18]. Their study compared a faceless robot, a silver face, and a human face and found that users felt more comfortable when the robot's face appeared more human-like.Luo et al. found that human-like characters were perceived as more likable, appropriate, and trusted than non-human characters [19]. They further demonstrated that the online payment process could benefit from employing ECAs by increasing user trust in situations concerning privacy. Due to these benefits, researchers have applied ECAs in various contexts.

2.2. Nonverbal Cues

The human-like appearance of ECAs has enabled the use of additional communication channels, known as non-verbal cues. These cues include gestures, facial expressions, posture, eye contact, and prosody. According to Freigang et al., users tend to perceive the utterances of VHs as more competent and significant when they utilize gestures, facial expressions, head nods, and prosody [20]. Similarly, several advantages of non-verbal cues have been reported. For instance, Yuan et al. reported that the gestures of ECAs have positive effects on perceived human likeness, animacy, and intelligence, which in turn led to participants paying more attention to the ECA [21]. Greta enhanced her expression and credibility through a variety of gestures and facial expressions, making her more believable and natural [22]. In terms of facial expressions, Max had the ability to provide context-appropriate expressions, making him more appropriate and persuasive [23]. Regarding posture, Ellie could increase intimacy through proper body positioning, with these adjustments enhancing intimacy with the user and making the agent appear more natural [1]. The importance of appropriate prosody for perceived naturalness was emphasized in the work of Ehret et al. [24].

2.3. Eye Contact in 2D Display

Eye contact with ECAs has been researched in various ways due to its importance [9]. Eye contact is considered an important factor as it has many functions, such as gathering information, showing attention, and expressing emotion [25,26]. The research of Ding et al. showed that users could identify the exact speaker in multi-party conversation using gaze information alone, without auditory cues [27]. This fact highlights the importance of gaze as a strong cue for identifying speakers in conversations. According to Kendon, gaze direction can be used to manage turn-taking in conversations, as well as to prevent and repair disruptions in dialogue [28]. They emphasized the role of gaze in maintaining conversational flow. The research of Abele indicated the importance of gaze in interpreting non-verbal cues from others and adjusting behavior accordingly [29]. Additionally, direct eye contact can express interest and respect for the interlocutor, while indirect eye contact can express discomfort or a lack of interest. This kind of research has shown the importance of eye contact, but there is a problem in the 2D display environment: the Mona Lisa effect. The Mona Lisa effect is a visual illusion that can occur within a narrow range, leading users to believe that a VH rendered on a 2D display is gazing at them, regardless of the actual direction of the gaze. Moubayed et al. compared the accuracy of users' recognition of gaze direction between a 3D talking head and 2D display [10,30]. The results showed that the users failed to accurately detect the gaze direction of the VH in 2D displays. The researchers explained that these results were due to the Mona Lisa effect.

In order to mitigate the Mona Lisa effect on a 2D display, research has focused on attaching special devices, to allow more accurate eye contact. To implement realistic group conversational situations, Otsuka et al. changed the position and rotation of a 2D display by mirroring head movement [31]. This was proven to enhance gaze awareness, eye contact, and other non-verbal behaviors, as well as telepresence. In the research of Vázquez et al., the eyes of VHs were displayed on a small 2D display, and the small display itself was rotated to maintain eye contact with the user [32]. Their experiment showed that the rotating display could mitigate the Mona Lisa effect more effectively than a fixed display. In addition, Hsueh-Han Wu et al. suggested a method to create this effect without any special devices [33]. By rotating the VH's iris and head, users could more accurately recognize the eye gaze of VHs.

There are two main experiments discussed in this paper. First, the impact of a VH's gaze behavior on user perception in multi-party conversations was compared in a 2D display setting. Second, a new method, camera rotation angle fine tuning (CRAFT), was proposed to mitigate the Mona Lisa effect, and its performance was evaluated .

3. Experiment 1

In this section, a user study was conducted to examine the influence of VH eye gaze in multi-party conversations.

3.1. Method

The participants experienced one of the following in a single conversation:

- Speaker eye contact: The VH answered the question by looking at the participant who asked it.
- Non-speaker eye contact: The VH answered the question by looking at the opposite participant. For example, if the right participant asked the question, the VH stared at the left participant.
- No eye contact: The VH answered the question by staring straight ahead, irrespective of who asked it.
- Right-participant eye contact: The VH answered the question by looking at the right participant, regardless of who asked it.
- Left-participant eye contact: The VH answered the question by looking at the left participant, regardless of who asked it.

Two participants experienced one of these conditions together in one conversation. However, in the left-participant eye contact case, each participant perceived a different eye gaze from the VH. The right-side participant might feel ignored, while the left-side participant might think the VH was focused only on him/her. To eliminate this discrepancy, the results of the right-side participants in the right-participant eye contact case and the left-side participants in the left-participant eye contact case were categorized as the looking me (LM) condition. Additionally, the results of the right-side participants in the left-participant eye contact case and the left-side participants in the right-participant Eye contact case were categorized as the looking partner (LP) condition. The cases of speaker eye contact, non-speaker eye contact, and no eye contact were named active speaker (AS), reverse speaker (RS), and looking center (LC), respectively. All participants experienced all conditions, and the sequence was randomized.

In all conditions, participants engaged in conversations with different VHs, and the order of the VHs was randomized and counter-balanced. The experimenter guided participants to ask questions based on a conversation card. The conversation card contained two questions per page. Participants on the left were instructed to ask the upper question first in all conditions. Two participants consecutively asked the VH questions, and the VH responded by adjusting its gaze according to the conditions. Each participant posed four questions per condition, and none of the questions overlapped. Each condition comprised eight turns of questions and answers: four for informative conversations and four for casual conversations. The informative questions focused on topics related to the university, as all participants were recruited from a local university. The casual conversations revolved around personal tastes or experiences.

The Wizard of Oz paradigm [34] was employed to eliminate the risk of improper answers due to incorrect speech recognition or unintended time delays caused by natural language processing or network issues. The experimenter controlled the VH's eye gaze and speech using a graphical user interface (GUI) (See Figure 1). A Intel RealSense Depth Camera D455 was used to capture the GUI image. The user's face was tracked with the S3FD object detection model [35] and highlighted with a red box. Additionally, the experimenter clicked on one of the participants' faces to control the VH's eye gaze using the GUI. By clicking on any point within the red box, the target object in the virtual environment in Unity moved linearly to the position corresponding to the center of the red box. The position was calculated using Pyrealsense2 (version 2.53.1) in Python (version 3.9) library, which matched the pixel point of the image to the 3D position using the camera as the origin. The GUI and Unity program for the experiments were run on a computer with an Intel Core i7-10700 Processor CPU and an NVIDIA GeForce RTX 3070 graphics card. The deep learning model was computed using CUDA (version 11.3), a graphics card environment. When one participant began asking a question, the experimenter clicked the target position (left-participant's face, right-participant's face, or center) according to the condition. Moreover, immediately after the question had been asked, the experimenter played the relevant prerecorded audio clip by clicking a button on the GUI. For a natural interaction, the VH answered the question after a delay of 0.9 s [36]. The VHs maintained their gaze while answering the question, and once the answer had been concluded, the VHs returned their gaze to the center.

3.1.1. Materials

The experiment was conducted in a 3.7 m by 5.5 m room with a 65-inch display, two tables, and two chairs. The VH was rendered on the display, and the virtual space of the VH was constructed similarly to the experimental space (see Figure 2). To simulate a scenario of sitting face-to-face with someone in an office setting, the VH was sitting in front of a black table with a computer, office supplies, and files arranged on the table. A black table was placed in front of the display. Two participants sat face-to-face with the VH at a distance of 2.4 m. They sat side by side in fixed seats, 1.28 m apart (See Figure 3).

Figure 1. Graphical user interface used in Experiment 1 to control the gaze behavior of the virtual human. The red boxes indicate the participants' faces.

Regarding the VHs, 3D human models generated using Character Creator 3 (https://www.reallusion.com/character-creator/, accessed on 16 January 2024) were utilized. The input reference images for these VHs were obtained from an open dataset featuring Asian women's faces (See Figure 4). To ensure distinctiveness, the VHs' clothes, hairstyles, facial features, and skin colors were varied, allowing participants to perceive different VHs based on the condition. The sizes of the VHs were similar to each other. To eliminate gender biases, all VHs were female.

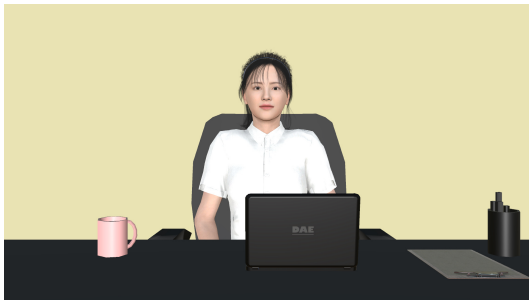
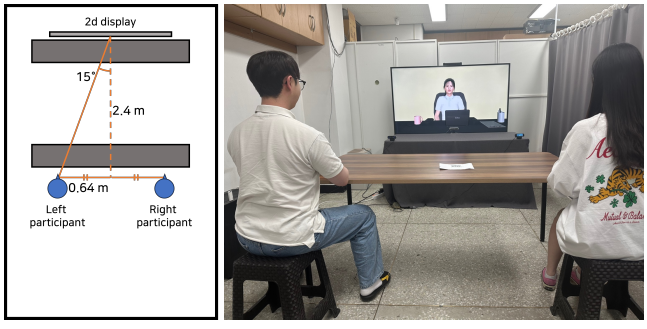


Figure 2. Virtual environment rendered in the 2D display.



(**a**) (**b**)

Figure 3. (**a**) The experimental setup and the schematics of the experimental space and (**b**) actual experimental space.

Figure 4. Three-dimensional human models used in the experiment.

All VHs had distinct voices created using the Naver CLOVA Speech Synthesis API (https://developers.naver.com/products/clova/tts/, accessed on 16 January 2024). Additionally, SALSA LipSync asset (https://crazyminnowstudio.com/unity-3d/lip-sync-salsa/, accessed on 16 January 2024) was implemented to generate natural lip motion and facial expressions for the VHs. Final IK asset (http://root-motion.com/, accessed on 16 January 2024) was used to control the VH's eye gaze, adjusting the 3D human model's gesture by moving a target object in a 3D virtual environment. By adjusting the weights of the spine, neck, head, and pupil, the gaze control of the VHs was specified.In this experiment, weights that were considered natural for VHs, to move in accordance with the target object's position, were used. The experimenter clicked the target position (left-participant's face, right-participant's face, or the center) using the GUI. The target position of the VH's eye gaze was indicated in the GUI, according to the trials. The target object was initially placed in the center of the two participants and moved linearly to the clicked position once the experimenter had clicked a participant's face or the reset button for the center.

3.1.2. Procedure

The two participants individually answered a pre-questionnaire using a laptop, after a brief explanation of the experiment. The experimenter then directed the participants to their designated seats and instructed them not to move arbitrarily. Whether each participant sat on the left or right side was randomized and remained fixed throughout the experiment. The experimenter guided the left-side participant to ask the upper question on the query card first, and the right-side participant was instructed to ask the lower question after the VH had answered the left-side participant's question.

Initially, the VH stared at the center of the two participants and said, "Hello, nice to meet you. Ask whatever you want". The participants then asked questions in sequence as guided. Each condition comprised eight turns of questions and answers. After the participants had asked all the questions on the query cards, the VH concluded the conversation with, "I hope it helps. Goodbye". Subsequently, the experimenter entered the experimental space and guided the participants to answer the post-questionnaire. While the participants completed the post-questionnaire with their backs to the display, the experimenter set up the conditions for the subsequent round. This procedure was repeated until all five conditions had been performed. After completing all conditions, the experimenter asked for the participants' overall evaluation of the VHs. All processes were recorded, and informed consent was obtained from all participants.

3.2. Measurements

We collected the participants' subjective and objective measurements for each condition.

3.2.1. Subjective Measurements

In the pre-questionnaire, the participants answered questions pertaining to demographics, prior conversation experiences with VHs and robots, and the negative attitudes towards robot scale (NARS [37]). Prior conversation experiences with VHs and robots were measured on a 7-point Likert scale (1: not at all to 7: every day). NARS was measured with six items on a 5-point Likert scale (1: strongly disagree to 5: strongly agree). At the end of each condition, the participants filled out the post-questionnaire on a 5-point Likert scale (1: strongly disagree to 5: strongly agree). For constructs with more than one item, the ratings per construct were averaged.

- Eye-gaze of VHs: The participants were asked whether the VH looked between them, in the center, or at the other participants when the participants asked a question. This item aimed to determine whether the participants were adequately aware of the VH's gaze.
- Interpersonal skill: The VH's interpersonal skills as perceived by the participants were measured.This construct contained six items from Oh et al.'s research [38].

- Social presence: How the participants perceived the social presence of the VHs was measured. This construct contained five items from Bailenson et al.'s research [39].
- Attention: How the participants perceived that the participants and the VHs paid attention to each other during the conversation was measured. This construct contained six items from Harms and Biocca's research [40].
- Co-presence: How the participants perceived the co-presence of the VHs was measured. This construct contained six items from Harms and Biocca's research [40].
- Competence: RoSAS [41] was used to measure these constructs. The participants rated the competence of the VHs using six items.
- Positive attitude and negative attitude: The participants' positive and negative attitudes were measured during the conversation with the VHs using PANAS [42].

3.2.2. Objective Measurements

The participants' gaze behavior was measured according to the VHs' eye gaze by calculating the average time looking at the VHs per each condition. Moreover, participant comments were collected at the end of each condition. Using these comments, it was possible to analyze the participants' positive and negative sentiments according to the VHs' eye gaze using Naver CLOVA Sentiment API (https://www.ncloud.com/product/aiService/clovaSentiment, accessed on 16 January 2024). The API presents the input sentence's positive, neutral, and negative sentiment probabilities. The highest sentiment probability of a comment was regarded as the participant's sentiment, according to the condition. Blank comments were excluded from the analysis.

3.3. Participants

We recruited 30 participants (18 females and 12 males) from a local university. All were native Korean speakers with a mean age of 22.56 (SD = 2.17). Three participants were excluded from the analysis due to problems during the video recording process. Therefore, data from 27 participants were used. In the pre-questionnaire, the participants reported little experience of conversation with VHs and robots (M = 1.59, SD = 0.93). The participants' majors included engineering, education, and medicine. In this experiment, two participants had to be recruited at the same time. If one of the participants suddenly canceled or did not come, one of the experimenters participated instead. In this case, the questionnaire completed by the experimenter was not used for analysis, and the other participant was unaware of the situation.

3.4. Results

This section outlines the methods and results of our analysis. We utilized IBM SPSS version 27 for our statistical analysis, setting the significance level at 5%. Friedman tests [43] were performed on each subjective measurement and average gaze time of the participants, and Wilcoxon signed-rank tests were used for pairwise comparisons with the Bonferroni adjustment applied to the p-values. The analysis is detailed in the following subsections, which are separately focused on subjective measures and objective measures.

3.4.1. Subjective Measurements

Ratings for the subjective measurements were averaged and the value for Cronbach's α [44] of each construct is provided in Table 1. The results are summarized in Figure 5 and Table 1.

- Interpersonal skill: There were significant differences between AS and LC conditions and AS and RS conditions (p = 0.005 and p = 0.003, respectively). However, there were no statistical differences between other conditions (See Figure 5a).
- Social presence: There were significant differences between AS and RS conditions and AS and LP conditions (p = 0.004 and p = 0.001 respectively). However, there were no statistical differences between other conditions (See Figure 5b).

- Attention: There were significant differences between AS and RS conditions and AS and LP conditions ($p = 0.004$ and $p = 0.001$ respectively). However, there were no statistical differences between other conditions (See Figure 5c).
- Co-presence: There were significant differences between AS and RS conditions, AS and LP conditions, and AS and LC conditions ($p < 0.001$, $p < 0.001$, and $p = 0.004$ respectively). However, there were no statistical differences between other conditions (See Figure 5d).
- Competence: There was a significant difference between AS and LC conditions ($p = 0.003$). However, there were no statistical differences between other conditions (See Figure 5e).
- Positive attitude and Negative attitude: In these constructs, no differences between conditions were found.

Regarding the NARS, the ratings were averaged, and Pearson correlation tests [45] were performed between the NARS score and each construct's mean values. A negative impact of the NARS on the Co-presence and Competence of the VHs was found. The participants with high NARS scores tended to evaluate the VHs' co-presence (Pearson's $r = -0.424$, $p = 0.027$) and competence (Pearson's $r = -0.467$, $p = 0.014$) slightly higher than participants with low NARS score, regardless of the eye gaze.

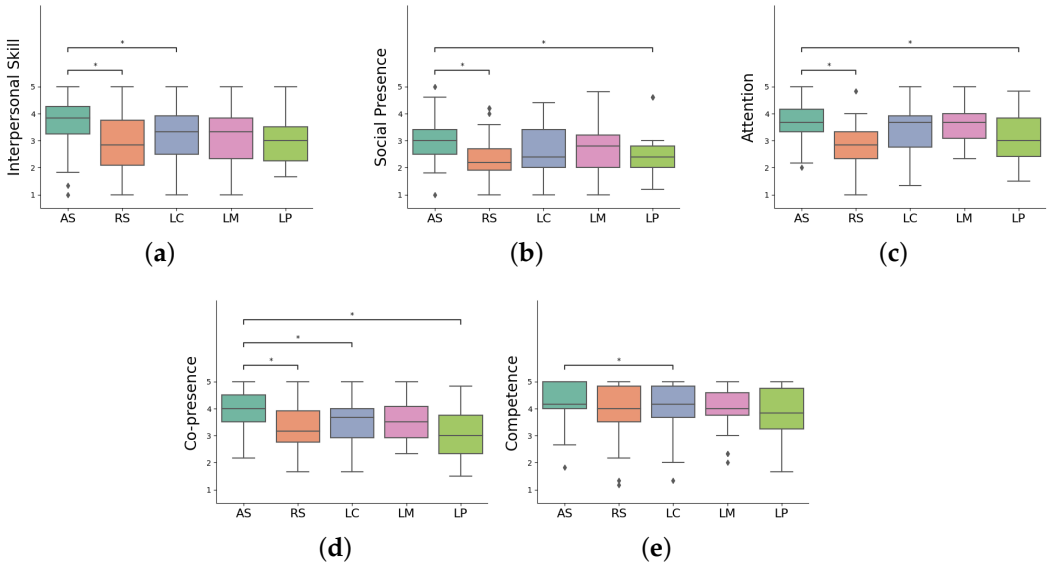


Figure 5. The results of (**a**) Interpersonal skill, (**b**) Social presence, (**c**) Attention, (**d**) Co-presence, and (**e**) Competence (*: $p < 0.05$, ⬥: outlier).

Table 1. Cronbach's α of subjective measurements using mean value and summary of the Friedman test results.

	Cronbach's α	χ^2	***p*-Value**
NARS	0.687	-	-
Interpersonal skill	0.945	20.511	<0.001
Social presence	0.859	17.142	0.002
Co-presence	0.883	22.604	<0.001
Attention	0.812	27.182	<0.001
Competence	0.925	9.109	0.058
Positive attitude	0.936	1.247	0.870
Negative attitude	0.854	9.896	0.042

3.4.2. Objective Measurements

Two labelers reviewed the video recordings of the participants frame by frame and computed the average duration spent observing the VHs for each condition. The correlation between the results of the two labelers was confirmed (Pearson's $r = 0.909$, $p < 0.001$). The proportion of frames participants visually focused on the VH per total frame was calculated per condition (See Figure 6). The average of the data from the two labelers was used as the results. Results showed a significant difference between AS and LC conditions ($p = 0.036$).

Regarding the sentiments of the participants' comments, the sentiments of 17 comments were analyzed per condition. Figure 7 represents the sentiments (positive, neutral, and negative) per condition. The numbers of negative sentiments for LM and RS conditions were greater (10 votes and 9 votes, respectively) than for other conditions. Also, it was found that the participants wrote their comments about the experiment and the VH more positively in the AS condition.

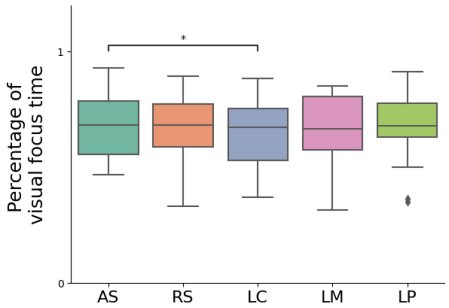


Figure 6. The percentage of time the participant spent visually focusing on the VH during the entire conversation (*: $p < 0.05$, ◆: outlier).

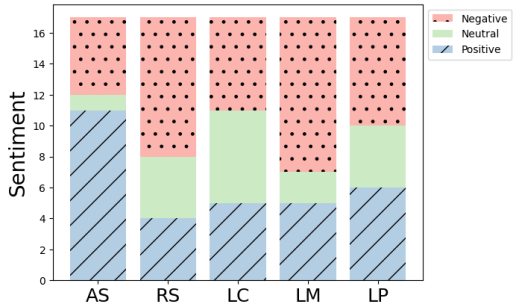


Figure 7. Results of analyzing the comments' sentiment.

3.5. Discussion

The results show that the VHs that looked at the exact speaker in the conversation were more highly evaluated than others in the subjective measurements. Between AS and RS conditions, there were statistically significant differences in Interpersonal skill, Social presence, Attention, and Co-presence. In the RS condition, participants P8 and P20 commented, "When I started the conversation, she turned her eyes, so I felt unpleasant and bizarre" and "I thought that two-way communication was not good because the VH looked at the other person during my turns", respectively. The participants considered it awkward and offensive for VHs to look at the opposite participant.

According to the LC condition, there were statistically significant differences in Interpersonal skills and Co-presence with tne AS condition. In the LC condition, participant P19 said, "Even when the participant spoke, the VH's gaze was fixed in the center, so I did not think the VH conversed with me or the VH was in the same place with me". It seems that participants perceived the VH as not being in the same space and underestimated the VH's interpersonal abilities, as the VH did not exhibit any physical feedback during the conversation. Although it is a subtle difference, the percentage of time participants visually focused on the VHs in the LC condition was the smallest.

Regarding the LP condition, there were statistically significant differences in Interpersonal skill, Social presence, Attention, and Co-presence. Participant P16 said, "When compared to the last experiment, the credibility of the answer was diminished in that the VH did not look at me at all during the conversation". It is expected that user satisfaction is reduced if participants perceive that the VH only looks at others in multilateral conversations.

However, the subjective measurements showed no difference between AS and LM conditions. In other words, the participants evaluated the VHs similarly regardless of whether they gazed at the actual speaker or only at themselves. Therefore, it can be assumed that it is okay for users to mistake VHs for continuing to see themselves in a multi-party conversation due to the Mona Lisa effect [10]. In the objective measurements, however, it was found that participants exhibited a higher level of negative sentiment during the LM condition of the conversation compared to the AS condition. For instance, regarding the LM condition, participant P14 commented, "I could not feel like talking to a real person because she looked at me for all the questions". Also, participant P19 commented, "I felt a bit burdened and being watched as she kept looking at me during the conversation". Regarding the AS condition, participants P16 and P7 commented, "Just looking at the exact speaker could enhance the credibility of her words" and "It was impressive that she looked at me when she answered my question and another participant when she answered another participant's question," respectively.

Through these results, VHs should correctly identify the speaker in a multi-party conversation. However, in a multi-party conversation with VH displayed in 2D, where several people stand and use it within a relatively narrow range, the Mona Lisa effect comes into play. This effect causes users within the limited range to perceive that the VH is looking at them, regardless of the actual person the VH is focused on. According to the findings, it will have a negative impact on the user experience. Therefore, there is a need for a methodology that enables users to accurately discern VH's gaze, even in the narrow range where the Mona Lisa effect occurs.

In the following section, CRAFT is proposed to enhance the users' perceptual accuracy of VH's eye gaze within a narrow range and its effectiveness is validated.

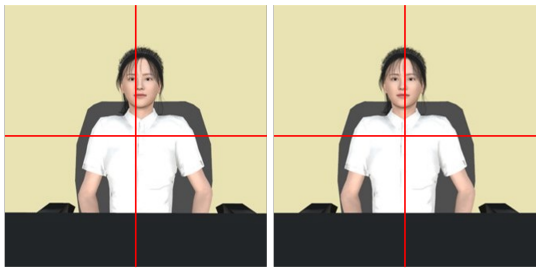
4. Experiment 2

In this section, the user study performed to confirm the effectiveness of CRAFT for increasing the participants' perceptual accuracy of VH's eye gaze behavior within a narrow range is detailed.

4.1. CRAFT

In Experiment 2, two main ideas were used to design CRAFT. First, the Mona Lisa effect of the user's perceived range is reduced when the 2D display is rotated, as discussed in Vázquez et al.'s research [32]. This reduction allows for more accurate eye contact with the user. Second, the Fish Tank VR [46] method is used. Fish Tank VR is an optical illusion technique that shows the user's perspective on a 2D display, taking into account the user's position and angle. CRAFT provides a fixed display with a rotating effect. The main camera will rotate based on the user's position where the VH is looking. Then, the user could feel the display rotating (See Figure 8).

The main camera is positioned in the center of the virtual environment (See Figure 9). When the VH looks at the target user, the main camera rotates in the same direction as the target user. Also, the rotation of the main camera is limited to one degree because the users noticed the rotation when it was greater than one degree during pilot test. By providing a screen similar to the viewpoint of a specific user, it is expected that the accuracy of gaze recognition of both users may be improved.



(**a**) (**b**)

Figure 8. The VH looking at the left participant (**a**) with and (**b**) without CRAFT in the 5-degree condition. Red crosses at the center have been added to better illustrate the subtle changes.

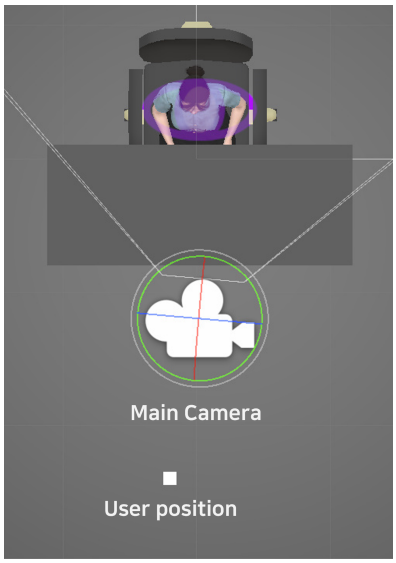
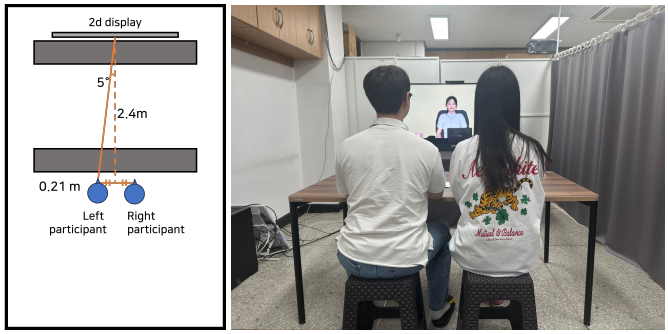


Figure 9. The top view in the Unity 3D virtual environment in the 5-degree condition with CRAFT, when VH looking at the left participant. In order to show the difference clearly, the main camera was rotated 5 degrees in this figure.

4.2. Method

We employed a within-subjects design for this experiment, comparing two main factors: angles representing within and outside the Mona Lisa effect range, specifically 5-degree and 15-degree, and the deployment of CRAFT. As CRAFT was designed to mitigate the Mona Lisa effect, it was only applied in the 5-degree condition. Detailed configurations of the experimental settings are shown in Figure 10 (Cf. Figure 3).



(**a**) (**b**)

Figure 10. (**a**) The experimental setup and the schematics of the experimental space and (**b**) actual experimental space in a 5-degree condition.

The experiment was divided into three sub-phases. In the first phase, participants experienced the 15-degree condition, and in the remaining two phases, they experienced the 5-degree condition. Each phase consisted of 30 trials. In each trial, participants were asked to identify the perceived gaze direction of the VH appearing on the 2D display. Similar to Experiment 1, we recruited two participants simultaneously, and the gaze directions considered were left participant, right participant, and center. The VH exhibited each gaze direction ten times, totaling 30 for each phase. In the second and third phases, where the 5-degree condition was examined, half of the trials were conducted with CRAFT applied and half without. Therefore, we collected a total of 90 responses from each participant.

Within each phase, the order of gaze directions was randomized. Similarly, five VHs from Experiment 1 were used, with their appearance order also randomized.

4.2.1. Materials

The experiment was controlled with the Wizard of Oz method. The experimenter controlled the VHs' eye gaze by using GUI. Through the GUI, the experimenter could see two participants and red boxes on their faces. If the experimenter clicked on any point in the red box, the target object in the virtual environment in Unity moved linearly to the position corresponding to the center of the red box. The target position of the VH's eye gaze per trial was marked in the GUI. Through the program, the application of CRAFT was automatically controlled.

Although the experimenter fixed the position of the two participants, the distance between the center and the participant could be different depending on the participants' height and posture. To remove this variation, if the distance between the center and the participant was 0.2 m or less, the camera's rotation was linearly changed up to 1 degree, and if over 0.2 m, the camera's rotation was fixed to 1 degree while the VH's head followed the target. All other parameters are as described in Experiment 1.

4.2.2. Procedure

After a brief explanation, the experimenter guided the two participants to designated positions for the 15-degree condition. The participants clicked one of three items (me, the other participant, or the center) that the VH was looking at per trial using their smartphones. Before starting the experiment, the TV was rendered only a desk and chair. At the beginning, after rendering a black display for three seconds, the VH looking at a specific target according to the trial was rendered. After both participants clicked the item they thought the VH was looking at, the experimenter moved on to the subsequent trial by clicking the button on GUI. The black display was rendered for three seconds between each trial to eliminate the influence of previous trials. Also, when the black display was rendered, the experimenter clicked the next target, which the VH would look at according to the trial. After the 15-degree condition, the experimenter guided the participants to the positions for the 5-degree conditions. The 5-degree condition was conducted in the same way, but twice.

4.3. Measurements

To check the perceived gaze direction of the VH, the participants selected whether they thought the VH was looking at them, the center, or the other participant in each trial. The correct responses were counted.

To compare the score based on angle, the scores from 5-degree and 15-degree condition were used. In the 5-degree condition, only trials without CRAFT were used. Also, to verify the effectiveness of CRAFT, the scores of trials with and without CRAFT were compared. Because there were no differences between trials with and without CRAFT in cases where the answer was the center, only the score that the answer was left-participant or right-participant was used. Table 2 shows the number of trials of each condition. The 5-degree condition was performed twice, and the corresponding number of times was written.

Table 2. Number of trials showing the VH looking at each of three positions for different angle settings.

	5-Degree		15-Degree
	With CRAFT	**Without CRAFT**	
Left participant	10	10	10
Right participant	10	10	10
center	10	10	10

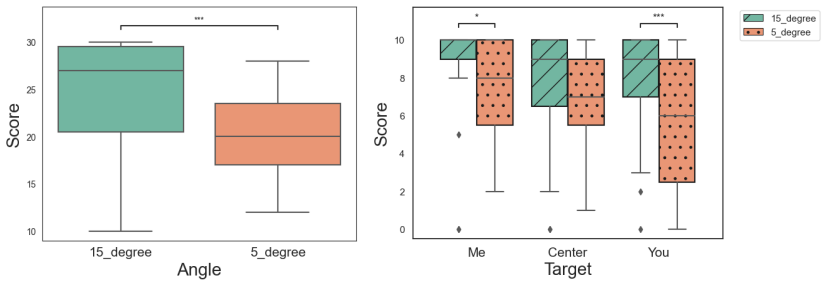
4.4. Participants

We recruited 30 participants from a local university with a mean age of 22.57 (SD = 2.28). Those who participated in Experiment 1 were also able to participate. In the same way as Experiment 1, two participants were recruited at the same time.

4.5. Results

A two-way ANOVA was performed to compare the conditions with two independent variables: the degree, and the target position. The significance level was set as 0.05. First, we compared the 15-degree and 5-degree conditions (Figure 11a). The results showed that the participants could distinguish more precisely where the VH looked in the 15-degree condition, outside the Mona Lisa effect ($p < 0.001$). The scores according to the target location where the VH gazed were compared.(Figure 11b). In the results, the participants distinguished the eye gaze of the VH more accurately in the 15-degree condition than in the 5-degree condition when the VH gazed at them or the other participant ($p = 0.008$ and 0.001, respectively). No statistically significant difference was found between conditions when the VH gazed at the center ($p = 0.061$).

Furthermore, to confirm the effectiveness of CRAFT, the scores of the trials with and without CRAFT applied in the 5-degree condition were compared (Figure 12a). We also used two-way ANOVA to compare the conditions with two independent variables: CRAFT and the target position. In the results, the participants were more accurate at distinguishing where the VH looked in the narrow range when CRAFT was applied ($p = 0.045$). In addition, we compared the scores according to the target location that the VH gazed at (Figure 12b). The results showed no statistically significant difference between with or without CRAFT when the VH gazed at them or the other participant ($p = 0.253$ and 0.060, respectively).

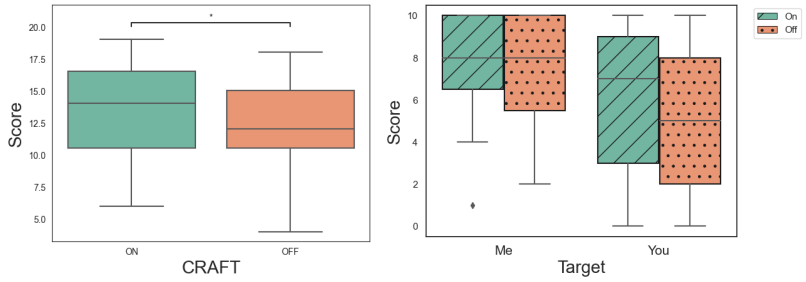


(**a**) (**b**)

Figure 11. The score (**a**) for the 15-degree and 5-degree conditions and (**b**) based on the target for each condition (*: $p < 0.05$, ***: $p < 0.001$, ⬥: outlier).

4.6. Discussion

The results indicate that the accuracy of the participants' perception of the VH's gaze behavior in the 15-degree condition, where the Mona Lisa effect did not occur, was statistically significantly higher than in the 5-degree condition ($p < 0.001$, Figure 11a). In addition, the participants tended to distinguish more accurately between Me and You ($p = 0.018$, Figure 11b). This appears to be attributable to the Mona Lisa effect.



(**a**) (**b**)

Figure 12. The score (**a**) depending on whether CRAFT was applied or not and (**b**) based on the target for each condition (*: $p < 0.05$, ⬥: outlier).

The results demonstrate that including supplementary visual cues effectively enhanced perceptual accuracy in a narrow area (see Figure 12a). In a previous study, attempts were made to mitigate the Mona Lisa effect by rotating the 2D display [10,30]. It was found that simply rotating the camera, which rendered the virtual environment without using any other devices, could give this effect. The difference in the effectiveness of the model was more pronounced when the VH was looking at another participant ($p = 0.060$) compared to looking at the user ($p = 0.253$, Figure 12b).

These findings suggest that CRAFT could improve the accuracy of perceiving that the VH was looking at the other user rather than oneself. Building on the results of Experiment 1, where the illusion that the VH gazed exclusively at the user in a multi-party conversation led to negative sentiments, the proposed method provided a straightforward means to enhance the accuracy of the VH's gaze behavior, without the need for additional devices and in a real kiosk conversation scenario with multiple people in a confined space.

5. Conclusions

In response to the growing deployment of virtual humans (VHs) in 2D kiosks, this paper investigated the impact of VHs' gaze behavior on user perception within a multi-lateral conversation context. The study comprised two experiments: the first examining users' perceptions of VHs, and the second specifically focusing on the discernibility of gaze direction and the mitigation of the perceptual error known as the Mona Lisa effect in a 2D setting.

In Experiment 1, the impact of the VH's gaze behavior on the users' perceptions in a multi-party conversation was identified. Five cases were considered: looking at the actual speaker (AS), looking at the reverse speaker (RS), looking only at me (LM), looking only at the other participant (LP), and looking only at the center (LC). In the results, the VHs looking at the actual speaker received higher scores for competence, co-presence, interpersonal skill, and social presence. Additionally, the participants tended to write negative comments when the VHs looked only at them, regardless of the speaker. These findings align with prior research indicating that a fixed stare from VHs towards participants can negatively influence their experience, heightening feelings of tension and giving the unsettling impression of being intently watched, compared to when a VHs demonstrates responsive gaze behavior [47]. In addition, previous robotics studies confirmed that in multi-party conversations, the robot's gaze at the correct speaker had a positive effect on the perception of the robot [48]. In this study, it was found that eye contact with VHs could positively impact multi-party conversations, not only in 3D but also in 2D environments. However, the visual illusion the Mona Lisa effect can frequently occur in kiosk environments used by multiple people within a narrow range. Therefore CRAFT, which has been shown to mitigate the Mona Lisa effect and enhance users' perceptual accuracy regarding VH gaze behavior in a narrow range, was proposed.

Experiment 2 demonstrated the effectiveness of CRAFT. Participants answered whether the VH appeared to be looking at the left participant, the center, or the right participant

within a narrow range, both with and without CRAFT. The results confirmed that CRAFT significantly increased the accuracy of the perceived gaze direction. It was demonstrated that a subtle rotation of the virtual camera, which rendered the VH and the virtual environment on a 2D screen, successfully provided visual cues to the participants. Building on this, it is anticipated that giving multiple visual cues could further enhance the likelihood of accurately perceiving eye contact.

These findings have practical implications for diverse scenarios involving multiple participants in a conversation with a VH on a 2D display. For instance, in a group counseling setting, a VH serving as a psychological counselor should direct its gaze towards the current speaker. Similarly, a VH functioning as a guide in settings like museums or hospitals must have the capability to focus on the individual posing a question, even in the presence of several users. We hope that using CRAFT in such situations will improve the perception of VH gaze behavior.

Our experiments had several limitations worth noting. In Experiment 1, the homogeneity of the participants—all being of a similar age, from the same local university, and sharing the same cultural background—presented a limitation, as eye contact can carry varied meanings across different cultures [49]. The experimental design also created a constrained multi-party conversation, where two participants alternately asked predetermined questions in fixed positions, and the VHs responded accordingly. Additionally, the manual control of the direction of the VHs' gaze by the experimenter using a GUI was a limitation. To overcome this, we developed a model capable of real-time speaker identification and automatic control of VH gaze direction. However, further research is required to validate this model in free-flowing, multilateral conversations. Another limitation was that the VHs focused their gaze on only one spot during each conversational turn. In natural multilateral conversations, individuals often shift their gaze between multiple people. Therefore, implementing an eye gaze paradigm where VHs primarily focus on the speaker, while occasionally shifting their gaze to others, would more accurately simulate real-world conversations. In Experiment 2, we rendered only the upper body of the VH in the 2D display. Rendering the full body could offer more varied visual cues, like the direction of the waist or feet, enhancing the accuracy of participants' perception of the VH's gaze. Moreover, we only established the effectiveness of CRAFT in a fixed setting. Future research should explore the effectiveness of CRAFT in dynamic environments with multiple users and in an actual kiosk.

Author Contributions: Conceptualization, J.K.; Methodology, S.J.; Software, S.J.; Formal analysis, J.K.; Writing—original draft, S.J.; Writing—review & editing, J.K. and M.L.; Supervision, M.L.; Funding acquisition, M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by Year 2021 Culture Technology R&D Program by Ministry of Culture, Sports and Tourism and Korea Creative Content Agency (Project Number: R2021040269).

Institutional Review Board Statement: Our studies were conducted under IRB approval, and the following is the information. Approval Code: PNU IRB/2023_35_HR, Approval Date: 2 March 2023.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. DeVault, D.; Artstein, R.; Benn, G.; Dey, T.; Fast, E.; Gainer, A.; Georgila, K.; Gratch, J.; Hartholt, A.; Lhommet, M.; et al. SimSensei Kiosk: A virtual human interviewer for healthcare decision support. In Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems, Paris, France, 5–9 May 2014; pp. 1061–1068.
2. Swartout, W.; Traum, D.; Artstein, R.; Noren, D.; Debevec, P.; Bronnenkant, K.; Williams, J.; Leuski, A.; Narayanan, S.; Piepol, D.; et al. Ada and Grace: Toward realistic and engaging virtual museum guides. In Proceedings of the Intelligent

Virtual Agents: 10th International Conference, IVA 2010, Philadelphia, PA, USA, 20–22 September 2010; Proceedings 10; Springer: Berlin/Heidelberg, Germany, 2010; pp. 286–300.
3. Cassell, J.; Stocky, T.; Bickmore, T.; Gao, Y.; Nakano, Y.; Ryokai, K.; Tversky, D.; Vaucelle, C.; Vilhjálmsson, H. Mack: Media lab autonomous conversational kiosk. In Proceedings of the IMAGINA 2002, Monte Carlo, Monaco, 12–15 February 2002; Volume 2, pp. 12–15.
4. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1877–1901.
5. Turabzadeh, S.; Meng, H.; Swash, R.M.; Pleva, M.; Juhar, J. Facial expression emotion detection for real-time embedded systems. *Technologies* **2018**, *6*, 17. [CrossRef]
6. Jun, H.; Bailenson, J. Effects of behavioral and anthropomorphic realism on social influence with virtual humans in AR. In Proceedings of the 2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Recife, Brazil, 9–13 November 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 41–44.
7. Cassell, J.; Thorisson, K.R. The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Appl. Artif. Intell.* **1999**, *13*, 519–538. [CrossRef]
8. Aneja, D.; Hoegen, R.; McDuff, D.; Czerwinski, M. Understanding conversational and expressive style in a multimodal embodied conversational agent. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 8–13 May 2021; pp. 1–10.
9. Ruhland, K.; Peters, C.E.; Andrist, S.; Badler, J.B.; Badler, N.I.; Gleicher, M.; Mutlu, B.; McDonnell, R. A review of eye gaze in virtual agents, social robotics and hci: Behaviour generation, user interaction and perception. In *Computer Graphics Forum*; Wiley Online Library: Hoboken, NJ, USA, 2015; Volume 34, pp. 299–326.
10. Moubayed, S.A.; Edlund, J.; Beskow, J. Taming Mona Lisa: Communicating gaze faithfully in 2D and 3D facial projections. *Acm Trans. Interact. Intell. Syst. (TiiS)* **2012**, *1*, 1–25. [CrossRef]
11. Fitzpatrick, K.K.; Darcy, A.; Vierhile, M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Ment. Health* **2017**, *4*, e7785. [CrossRef] [PubMed]
12. Denecke, K.; Vaaheesan, S.; Arulnathan, A. A mental health chatbot for regulating emotions (SERMO)-concept and usability test. *IEEE Trans. Emerg. Top. Comput.* **2020**, *9*, 1170–1182. [CrossRef]
13. Rapp, A.; Curti, L.; Boldi, A. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *Int. J. Hum.-Comput. Stud.* **2021**, *151*, 102630. [CrossRef]
14. Følstad, A.; Skjuve, M.; Brandtzaeg, P.B. Different chatbots for different purposes: Towards a typology of chatbots to understand interaction design. In Proceedings of the Internet Science: INSCI 2018 International Workshops, St. Petersburg, Russia, 24–26 October 2018; Revised Selected Papers 5; Springer: Cham, Switzerland, 2019; pp. 145–156.
15. Jain, M.; Kumar, P.; Kota, R.; Patel, S.N. Evaluating and informing the design of chatbots. In Proceedings of the 2018 Designing Interactive Systems Conference, Hong Kong, China, 9–13 June 2018; pp. 895–906.
16. Zamora, J. I'm sorry, dave, i'm afraid i can't do that: Chatbot perception and expectations. In Proceedings of the 5th International Conference on Human Agent Interaction, Bielefeld, Germany, 17–20 October 2017; pp. 253–260.
17. Laranjo, L.; Dunn, A.G.; Tong, H.L.; Kocaballi, A.B.; Chen, J.; Bashir, R.; Surian, D.; Gallego, B.; Magrabi, F.; Lau, A.Y.; et al. Conversational agents in healthcare: A systematic review. *J. Am. Med. Inform. Assoc.* **2018**, *25*, 1248–1258. [CrossRef] [PubMed]
18. Broadbent, E.; Kumar, V.; Li, X.; Sollers, J., 3rd; Stafford, R.Q.; MacDonald, B.A.; Wegner, D.M. Robots with display screens: A robot with a more humanlike face display is perceived to have more mind and a better personality. *PLoS ONE* **2013**, *8*, e72589. [CrossRef]
19. Luo, J.; McGoldrick, P.; Beatty, S.; Keeling, K.A. On-screen characters: Their design and influence on consumer trust. *J. Serv. Mark.* **2006**, *20*, 112–124. [CrossRef]
20. Freigang, F.; Klett, S.; Kopp, S. Pragmatic multimodality: Effects of nonverbal cues of focus and certainty in a virtual human. In Proceedings of the Intelligent Virtual Agents: 17th International Conference, IVA 2017, Stockholm, Sweden, 27–30 August 2017; Proceedings 17; Springer: Cham, Switzerland, 2017; pp. 142–155.
21. He, Y.; Pereira, A.; Kucherenko, T. Evaluating data-driven co-speech gestures of embodied conversational agents through real-time interaction. In Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents, Würzburg, Germany, 19–22 September 2022; pp. 1–8.
22. Poggi, I.; Pelachaud, C.; de Rosis, F.; Carofiglio, V.; De Carolis, B. Greta. a believable embodied conversational agent. In *Multimodal Intelligent Information Presentation*; Springer: Dordrecht, The Netherlands, 2005; pp. 3–25.
23. Becker, C.; Kopp, S.; Wachsmuth, I. Simulating the emotion dynamics of a multimodal conversational agent. In Proceedings of the Tutorial and Research Workshop on Affective Dialogue Systems, Kloster Irsee, Germany, 14–16 June 2004; Springer: Berlin/Heidelberg, Germany, 2004; pp. 154–165.
24. Ehret, J.; Bönsch, A.; Aspöck, L.; Röhr, C.T.; Baumann, S.; Grice, M.; Fels, J.; Kuhlen, T.W. Do prosody and embodiment influence the perceived naturalness of conversational agents' speech? *ACM Trans. Appl. Percept. (TAP)* **2021**, *18*, 1–15. [CrossRef]
25. Argyle, M.; Cook, M. *Gaze and Mutual Gaze*; Cambridge University Press: Cambridge, UK, 1976.
26. Kendon, A. *Conducting Interaction: Patterns of Behavior in Focused Encounters*; CUP Archive: Cambridge, UK, 1990; Volume 7.

27. Ding, Y.; Zhang, Y.; Xiao, M.; Deng, Z. A multifaceted study on eye contact based speaker identification in three-party conversations. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; pp. 3011–3021.
28. Kendon, A. Some functions of gaze-direction in social interaction. *Acta Psychol.* **1967**, *26*, 22–63. [CrossRef] [PubMed]
29. Abele, A. Functions of gaze in social interaction: Communication and monitoring. *J. Nonverbal Behav.* **1986**, *10*, 83–101. [CrossRef]
30. Al Moubayed, S.; Skantze, G. Turn-taking control using gaze in multiparty human-computer dialogue: Effects of 2d and 3d displays. In Proceedings of the International Conference on Audio-Visual Speech Processing 2011, Volterra, Italy, 31 August–3 September 2011; KTH Royal Institute of Technology: Stockholm, Sweden, 2011; pp. 99–102.
31. Otsuka, K. MMSpace: Kinetically-augmented telepresence for small group-to-group conversations. In Proceedings of the 2016 IEEE Virtual Reality (VR), Greenville, SC, USA, 19–23 March 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 19–28.
32. Vázquez, M.; Milkessa, Y.; Li, M.M.; Govil, N. Gaze by Semi-Virtual Robotic Heads: Effects of Eye and Head Motion. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 24 October 2020–24 January 2021; IEEE: Piscataway, NJ, USA, 2020; pp. 11065–11071.
33. Wu, H.H.; Mitake, H.; Hasegawa, S. Eye-Gaze Control of Virtual Agents Compensating Mona Lisa Effect. In Proceedings of the HAI: Human-Agent Interaction Symposium, Southampton, UK, 15–18 December 2018.
34. Green, P.; Wei-Haas, L. The rapid development of user interfaces: Experience with the Wizard of Oz method. In Proceedings of the Human Factors Society Annual Meeting, Baltimore, MD, USA, 29 September–3 October 1985; SAGE Publications Sage CA : Los Angeles, CA, USA, 1985; Volume 29, pp. 470–474.
35. Zhang, S.; Zhu, X.; Lei, Z.; Shi, H.; Wang, X.; Li, S.Z. S3fd: Single shot scale-invariant face detector. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 192–201.
36. Kanda, T.; Kamasima, M.; Imai, M.; Ono, T.; Sakamoto, D.; Ishiguro, H.; Anzai, Y. A humanoid robot that pretends to listen to route guidance from a human. *Auton. Robot.* **2007**, *22*, 87–100. [CrossRef]
37. Syrdal, D.S.; Dautenhahn, K.; Koay, K.L.; Walters, M.L. The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study. In *Adaptive and Emergent Behaviour and Complex Systems, Proceedings of the 23rd Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour, Edinburgh, UK, 6–9 April 2009*; Society for the Study of Artificial Intelligence and the Simulation of Behaviour: Voluntari, Romania, 2009.
38. Oh, S.Y.; Bailenson, J.; Krämer, N.; Li, B. Let the avatar brighten your smile: Effects of enhancing facial expressions in virtual environments. *PLoS ONE* **2016**, *11*, e0161794. [CrossRef] [PubMed]
39. Bailenson, J.N.; Blascovich, J.; Beall, A.C.; Loomis, J.M. Interpersonal distance in immersive virtual environments. *Personal. Soc. Psychol. Bull.* **2003**, *29*, 819–833. [CrossRef] [PubMed]
40. Harms, C.; Biocca, F. Internal consistency and reliability of the networked minds measure of social presence. In Proceedings of the Seventh Annual International Workshop: Presence, Valencia, Spain, 13–15 October 2004; Volume 2004.
41. Carpinella, C.M.; Wyman, A.B.; Perez, M.A.; Stroessner, S.J. The robotic social attributes scale (RoSAS) development and validation. In Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6–9 March 2017; pp. 254–262.
42. Watson, D.; Clark, L.A.; Tellegen, A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *J. Personal. Soc. Psychol.* **1988**, *54*, 1063. [CrossRef] [PubMed]
43. Friedman, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Am. Stat. Assoc.* **1937**, *32*, 675–701. [CrossRef]
44. Cronbach, L.J. Coefficient alpha and the internal structure of tests. *Psychometrika* **1951**, *16*, 297–334. [CrossRef]
45. Cohen, I.; Huang, Y.; Chen, J.; Benesty, J.; Benesty, J.; Chen, J.; Huang, Y.; Cohen, I. Pearson correlation coefficient. In *Noise Reduction in Speech Processing*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 1–4.
46. Ware, C.; Arthur, K.; Booth, K.S. Fish Tank Virtual Reality. In Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems, Amsterdam The Netherlands, 24–29 April 1993; pp. 37–42. [CrossRef]
47. Wang, N.; Gratch, J. Don't just stare at me! In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, GA, USA, 10–15 April 2010; pp. 1241–1250.
48. Xu, Q.; Li, L.; Wang, G. Designing engagement-aware agents for multiparty conversations. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France, 27 April–2 May 2013; pp. 2233–2242.
49. Uono, S.; Hietanen, J.K. Eye contact perception in the west and east: A cross-cultural study. *PLoS ONE* **2015**, *10*, e0118094. [CrossRef] [PubMed]