Eyes on the Narrative: Exploring the Impact of Visual Realism and Audio Presentation on Gaze Behavior in AR Storytelling

Florian Weidner
Lancaster University
Lancaster, United Kingdom
f.weidner@lancaster.ac.uk

Jakob Hartbrich
Technische Universität Ilmenau
Ilmenau, Germany
jakob.hartbrich@tu-ilmenau.de

Stephanie Arévalo Arboleda
Technische Universität Ilmenau
Ilmenau, Germany
stephanie.arevalo@tu-ilmenau.de

Christian Kunert
Technische Universität Ilmenau
Ilmenau, Germany
christian.kunert@tu-ilmenau.de

Christian Schneiderwind
Technische Universität Ilmenau
Ilmenau, Germany
christian.schneiderwind@tu-ilmenau.de

Chenyao Diao
Technische Universität Ilmenau
Ilmenau, Germany
chenyao.diao@tu-ilmenau.de

Christoph Gerhardt
Technische Universität Ilmenau
Ilmenau, Germany
christoph.gerhardt@tu-ilmenau.de

Tatiana Surdu
Technische Universität Ilmenau
Ilmenau, Germany
tatiana.surdu@tu-ilmenau.de

Wolfgang Broll
Technische Universität Ilmenau
Ilmenau, Germany
wolfgang.broll@tu-ilmenau.de

Stephan Werner
Technische Universität Ilmenau
Ilmenau, Germany
stephan.werner@tu-ilmenau.de

Alexander Raake
Technische Universität Ilmenau
Ilmenau, Germany
alexander.raake@tu-ilmenau.de

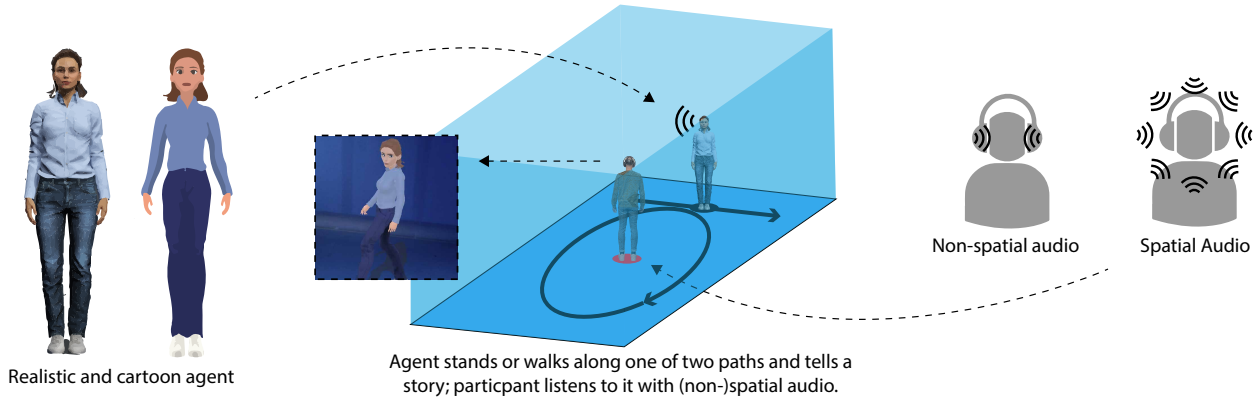


Figure 1: In this study, a participant stands in the centre of a room, and an AR agent, visualized by either a low-realism 3D model or a high-realism 3D model, is telling a story. The participant hears the story via spatial or non-spatial audio. We analyzed participants' gaze behaviour during the experience. We introduced two walking conditions (side-by-side and circle) to avert the participant from staring at a still-standing avatar.

ABSTRACT

Augmented Reality (AR) and Virtual Reality (VR) are essential tools for researchers and practitioners, serving purposes from training to entertainment: many of these applications rely on agents. This study explores the impact of agent characteristics on user reactions, focusing on gaze as a primary visual attention indicator in AR and VR. While existing research has investigated the agent's gaze and its influence on the user, it is unclear how the agent's auralization and visualization influence gaze behaviour. We investigate this by studying the impact of rendering style and type of audio on gaze

ETRA '24, June 04–07, 2024, Glasgow, United Kingdom

ACM ISBN 979-8-4007-0607-3/24/06
https://doi.org/10.1145/3649902.3653344

behaviour during a narrative AR experience. Participants listened to a story with the agent visualized as a cartoon-style or realistic virtual human and auralized with spatial or non-spatial audio. The results revealed that the agent's rendering style significantly influenced gaze behaviour, with cartoon-style agents capturing more visual attention. Audio variations did not yield significant differences. Together, our findings inform the design of AR user interfaces with agents, suggesting that low-realism visualizations are more captivating and, thus, more suitable for experiences where the user is supposed to look at the storyteller.

CCS CONCEPTS

• **Human-centered computing** → **Mixed / augmented reality**; **Empirical studies in HCI**.

KEYWORDS

Augmented reality, gaze behaviour, storytelling, agent, audio

ACM Reference Format:
Florian Weidner, Jakob Hartbrich, Stephanie Arévalo Arboleda, Christian Kunert, Christian Schneiderwind, Chenyao Diao, Christoph Gerhardt, Tatiana Surdu, Wolfgang Broll, Stephan Werner, and Alexander Raake. 2024. Eyes on the Narrative: Exploring the Impact of Visual Realism and Audio Presentation on Gaze Behavior in AR Storytelling. In *2024 Symposium on Eye Tracking Research and Applications (ETRA '24), June 04–07, 2024, Glasgow, United Kingdom.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3649902.3653344

1 INTRODUCTION

Gaze is crucial in human communication, conveying attention-related information and influencing interactions between people and their environment. For example, the time a person looks at another person correlates with emotional evaluations [Goldberg et al. 1969]. On the contrary, gaze aversion — avoiding looking at someone while speaking — shows adverse effects on communication, attraction, and credibility [Burgoon et al. 1986]. Given that virtual humans in the form of avatars (controlled by humans) and agents (controlled by a computer) are becoming integral to many augmented reality (AR) and virtual reality (VR) applications and that gaze is an integral part of human communication, understanding the factors behind how and why people gaze at virtual humans is essential for a compelling experience.

A substantial body of research has explored the impact of virtual humans' characteristics on user reactions. Visualization received a lot of attention with research investigating the effect of various rendering styles (e.g., Roth et al. [2016]), as well as their effect on behaviour [Nelson et al. 2022] and perception [McDonnell et al. 2012]. Although visual effects are being researched more often, auralization received some attention as well, with research showing that auralization plays a vital role during customization [Kao et al. 2022; Zempo et al. 2022] and impacts emotional responses [Warp et al. 2022]. Considering the interaction with agents, the user's gaze behaviour and its dependency on visualization and auralization of an agent have been given scant attention. For conversational tasks with an avatar, Wang et al. [2019] and Aseeri and Interrante [2021] analyzed rendering style only and concluded that realistic styles lead to more and more prolonged glances at the avatars. No analysis exists for agent-based storytelling experiences (and audio as an independent variable). This study addresses this gap by investigating the impact of rendering style and audio on gaze behaviour during a narrative experience delivered by an agent in AR, guided by the following research question:

Does the type of auralization or the rendering style influence participants' gaze behaviour in an AR storytelling experience?

In this work, gaze behaviour, operationalized as the total time users spent looking at the agent and the number of times they looked at it, was studied as participants repeatedly listened to a story narrated by an AR agent, visualized as a cartoon-style (low realism) or realistic (high realism) virtual human, auralized with spatial or non-spatial audio. To prevent participants from only staring at a static agent, additional conditions were introduced, where the agent moved from side to side or walked around the participant in a circle.

Results revealed that the type of agent visualization significantly affected the overall time and total number of times users looked at the agent, with the cartoon-style agent attracting longer and fewer glances than the realistic one. However, audio did not lead to significant differences. In our experience, this implies that participants paid more attention to the cartoon agent than the realistic one, suggesting them for narrative experiences. This inspires and guides applications like teaching, training, tours, and narrative experiences in entertainment, where users should look at the storyteller (the agent) rather than the environment.

2 RELATED WORK

2.1 Gaze and Agents

Considering gaze, the relationship between an agent and a user is similar to that between humans. As in real life, the agent's gaze is an effective and often essential non-verbal cue in various situations [Kevin et al. 2018]. Similarly, gaze has been demonstrated to predict attention in conversational multi-agent systems [Vertegaal et al. 2001]. This implies a high probability that a user is looking at the person speaking or someone speaking to. Gaze is also an effective communication cue in other, non-speech-based social situations, e.g., in pre-touch situations [Cuello Mejía et al. 2023]. Further, adding gaze cues to tasks where humans and agents collaborate has increased the human-agent team's effectiveness and human performance [Andrist et al. 2017; Wang and Ruiz 2021]. To achieve that, many systems simulate realistic gaze in agents [Le et al. 2012; Randhavane et al. 2019], striving for realism to invoke — at best — realistic responses from the human conversation partner [Bailly et al. 2010]. Overall, research on gaze and virtual humans has a long history, focusing on the virtual human having realistic gaze behaviour and the impact of this gaze behaviour on variables such as task performance or user behaviour. Little research has been done to see how other variables, such as the rendering style of the virtual human, affect gaze behaviour. In avatar-mediated VR applications (not about agents), findings indicate that utilizing a generic yet realistic avatar [Gonzalez-Franco et al. 2020; Wang et al. 2019] and a video avatar [Aseeri and Interrante 2021] results in a participant's gaze behaviour that closely resembles human-like interactions and increased trust. This effect holds even compared to the gaze behaviour induced by a scanned realistic avatar [Aseeri and

Interrante 2021]. The importance of gaze and the research outcomes on visualization with avatars in interactive scenarios motivate further research into human-agent interaction within non-interactive scenarios.

2.2 Auralization, Visualization, Spatial Coherence, and Behavior

Similar to gaze and agents, research on agents and how to auralize and visualize them has a long history. Both auralization and visualization profoundly impact the experience and the user's reaction. For visualization, this entails, among others, social presence, presence, task performance, communication behaviour, and trust (for a comprehensive review, see Weidner et al. [2023]). Regarding audio, recent research showed that spatial audio increases immersion [Geronazzo et al. 2019; Tsepapadakis and Gavalas 2023] and leads to stronger emotional reactions [Geronazzo et al. 2019]. However, other variables, such as social presence, seem unaffected by the audio type (binaural, audio, or face-to-face) [Immohr et al. 2023]. Gaze behaviour and its dependency on audio type, although there is a relationship between both [Mendonça and Korshunova 2020], has not been researched yet within the context of agents and AR.

Next to research on the individual factors, audiovisual spatial coherence is relevant for virtual humans [Hayes 2015; Latoschik and Wienrich 2022], especially regarding speech. Audiovisual spatial coherence means that audio cues come from the location of the visual cues, which supports multisensory integration and, by that, understanding of the spoken word [Hendrickx et al. 2015; Stenzel et al. 2019]. Related to virtual humans and AR/VR, it has been shown that a perceived mismatch in voice and appearance is unappealing [Higgins et al. 2022; Zibrek et al. 2021], that a fine-tuned auralization paired with customization can enhance identification and immersion [Kao et al. 2021], and that audiovisually coherent experiences are more believable [Lam et al. 2023]. Interestingly, a lack of coherence seems not to affect social presence [Higgins et al. 2022; Zibrek et al. 2021]. Overall, the impact of audiovisual spatial coherence on gaze in AR has not received attention yet, although a sensory mismatch, e.g. the voice coming not from where the agent is standing, could lead to different gaze behaviour, similar to the Ventriloquist effect [Lavan et al. 2022]).

Overall, research shows that, while existing work in interactive scenarios has already investigated the impact of visualization and auralization on users' experiences, there remains a gap in research focused on narrative experiences, such as teachings, training, and tours where the agent is uni-directionally talking and the user listening. Moreover, there is a noticeable absence of studies examining the impact of audiovisual spatial coherence and the influence of audio alone on gaze behaviour, further underscoring the need for an initial exploration of these domains to understand better how to build narrative experiences.

3 STUDY OVERVIEW

We executed a within-subjects experiment (2x2x3 design) with 12 conditions to assess the impact of audio (spatial audio, non-spatial audio), the rendering style of the agent (cartoon, realistic), and the type of animation (standing, walking side to side, walking in a circle) on users gaze behaviour (all counterbalanced using a balanced Latin Square).

3.1 Apparatus

Participants wore a HoloLens2 (52° diagonal field of view) and stood in an empty laboratory room. They also wore wired headphones (Sennheiser HD600) connected to a PC. The HoloLens2 and the PC communicated via WebSockets within a dedicated WiFi network without noticeable latency. This was necessary to use spatial audio beyond the official Microsoft spatializer (cf. Section 3.1.1). The application was built with Unity 2021.2.7f1.

3.1.1 Auralization. Based on measured binaural room impulse responses (BRIR), we mimic the real room acoustics for spatial audio. Note *the Microsoft spatializer*[1] utilizes only Head-Related Transfer Functions (HRTFs). HRTFs describe the directional filtering of sound reaching a listener's ears but do not include room information or sound source-specific directivity. While reverberation can be adjusted, this does not guarantee an accurate match to the natural environment.

We used two sound samples for the auralization: female speech and footsteps. Both are dry and devoid of any noticeable room acoustic characteristics. The female speech was recorded within a dry speaker booth, utilizing a Neumann-U87 microphone with a cardioid pattern and a popkiller. Footstep samples were sourced from a database[2]. The speech sample and the footsteps were calibrated in a testroom to match their volume for a plausible experience.

3.1.2 Agents. Figure 1 (left) shows the realistic and cartoon visualization of the agent. The agents (about 80,000 vertices) were created and rigged with the help of the Character Creator 4 Software Suite[3]. The height of the agents (approximately 1.90 m) was not modified to the viewer's height. We ensured that viewers could always see the agent's face without looking up or down uncomfortably.

For lip movement and facial animation, we relied on a custom C++ implementation of *Oculus LipSync*, as the original does not work on the HoloLens 2 because it does not support x86 ARM platforms. We used the 15 visemes — visemes depict the animation of the mouth shape for specific sets of phonemes — specified in the *MPEG-4 Face and Body Animation (FBA)* standard [Pandzic and Forchheimer 2003].

For the footstep sound, each agent has triggers attached to the feet and plays the footsteps on collision with the floor.

3.1.3 Movement. We recorded three movement patterns to avoid participants only staring at the agent while in front of them. The movement patterns are circle (like an agent walking among the audience), side-to-side (like an agent walking on a stage), and standing. We captured a natural person via motion tracking using an OptiTrack system with ten cameras and 41 markers to capture the movement.

For the standing position, the minimum distance between the participant and the agent is 1.95 m, and the maximum is 2.06 m, with an average of M = 1.99 m, SD = 0.021 m (aka, the agent is about

[1] https://github.com/microsoft/spatialaudio-unity/, 28-01-2024
[2] https://www.fesliyanstudios.com/royalty-free-sound-effects-download/footsteps-31, 28-01-2024
[3] https://www.reallusion.com/character-creator/, 28-01-2024

2 m away from the participant). The standing agent is not stiff and static but has idle movements (swaying, subtle arm and head movements). For the side-to-side movement, the distance spans from 1.95 m to 3.03 m, with an average distance of M = 2.41 m, SD = 0.30 m (aka, the agent walks from one side to the other side, appr. 2 m away from the participant; cf. Figure 1). For the circular motion, the distances range from 0.72 m to 2.63 m, averaging M = 2.02 m, SD = 0.55 m (aka, the agent walks in a slightly oval shape around the participant). During the recording, the actor looked at the position where the participant would be in the experiment to simulate eye contact.

3.2 Task & Procedure

This study employed a blend of free-viewing and story-listening tasks. Participants observed the agent and listened to a short fictional story. The selected stimulus was Frank Kafka's "Give it up" (128 words, 52 seconds [Kafka and Glatzer 1971, p366]). Consistency in story usage across all conditions mitigated potential bias from participant preferences.

The experiment, lasting approximately 50 minutes, comprised the following steps:

- Participants received a brief introduction to the experiment's purpose and completed a consent form (5 minutes).
- They trained on HoloLens 2 usage and performed eyetracking calibration (5 minutes).
- Participants, looking forward and standing within a confined space marked on the floor (0.4 m x 0.4 m), observed and listened during the 12 experimental conditions (52 seconds each).

3.3 Measures

During the listening task, we recorded eye-tracking data and calculated the time t participants looked at the agent. We did this by calculating the overall intersection time between the combined gaze ray and the agent's 3D model. We also calculate the number of times participants look at the agent by calculating how often the eye gaze ray enters the 3D model of the agent.

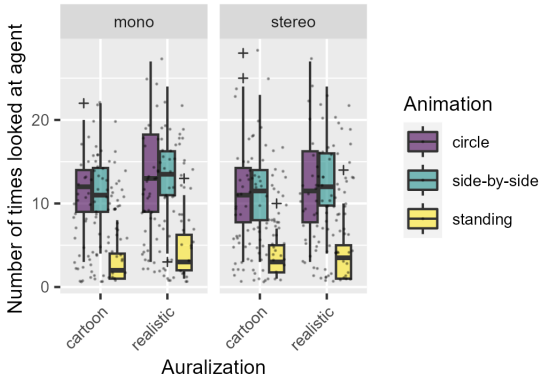
3.4 Sample

We enlisted 36 participants (21 male, 14 female, and one non-binary), aged 22 to 32 years (M = 26.53, SD = 2.65), without reported hearing issues, all affiliated with the university as students or staff. Of the participants, three wore contact lenses, ten used eyeglasses, and the rest had no visual aids. Among them, 24 had prior AR/VR experience, including two with Unity programming skills, while 12 were new to AR/VR.

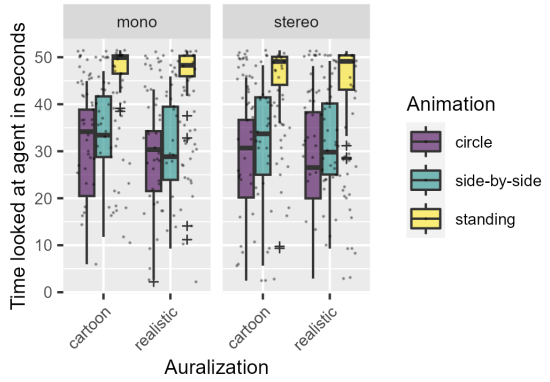
The study received prior approval from the university's Ethical Committee and adhered to national research organization guidelines and the Helsinki Declaration. Participants provided written consent and were compensated with 12 euros for their involvement.

4 RESULTS

We applied the Aligned Rank Transform (ART) [Kay et al. 2021] to analyze eye tracking data, as it did not follow a normal distribution. The assessment of normality involved visual inspection using QQ-plots and Shapiro-Wilk tests. In cases where significant differences were identified, we performed post-hoc pairwise comparisons using ART-C with Bonferroni correction.



(a) Number of times participants looked at agent. Cartoon agent received fewer glances than the realistic one ($p < .05$).



(b) Time participants looked at the agent in seconds. Cartoon agent was looked at longer than the realistic agent ($p < .05$).

Figure 2: Box plot of gaze behaviour grouped by auralization, visualization, and animation.

4.1 Number of times participants looked at the agent

Figure 2a illustrates the number of times people looked at the 3D model of the agent, grouped by auralization, visualization, and animation. *Agent* had a medium ($\eta_p^2 = .02$) and significant effect ($F(1, 385) = 9.205, p = .0026$). Pairwise comparisons indicate that the cartoon agent (M = 8.73, SD = 5.66) attracted significantly fewer glances than the realistic agent (M = 9.94, SD = 6.37, $p = .003$). This indicates that participants looked away and back at the realistic agent more often than the cartoon agent. *Animation* also had a large ($\eta_p^2 = .57$) and significant effect ($F(1, 385) = 252.4, p < .0001$). Bonferonni-corrected post hoc pairwise comparisons state that the standing animation (M = 3.58, SD = 2.76) was looked at less often than the circle animation (M = 12.3, SD = 5.58, $p < .0001$) and the side-by-side animation (M = 12.1, SD = 4.63, $p < .001$). This tells us that participants look more often elsewhere and then

back at the agent when walking in circles or from side to side compared to the standing condition. Note we analyze this to confirm that participants actively looked at the agent and did not ignore it completely (circle, side-by-side) or just stared at it (standing).

There were no other significant main or interaction effects ($p > .165$). Repeated measures ANOVA on the rank-aligned data did not detect an effect of the trial number (1 to 12; $F(11, 385) = .750, p < .690, \eta_p^2 = .02$).

4.2 Time participants looked at the agent

Figure 2b shows the time participants looked at the agent grouped by visualization, auralization, and animation (all values in seconds). *Agent* yielded a statistically significant effect ($F(1, 385) = 8.70, p = .0034$) on the time the participant looked at the agent. The effect size ($\eta_p^2 = 0.022$) indicates a small impact. Bonferonni-corrected pairwise post hoc comparisons revealed a significant difference between the cartoon (M = 36.6 s, SD = 12.5 s) and realistic agent visualization (M = 34.4 s, SD = 12.7 s; $p = .003$). This means the cartoon visualization attracted significantly longer glances than the realistic visualization. Similarly, *Animation* exhibited a significant effect ($F(2, 385) = 291.91, p < .0001$), $\eta_p^2 = 0.603$) on the time participants looked at the agent. Bonferonni-corrected pairwise post hoc comparisons and revealed significant differences. Regardless of visualization or auralization, the agent walking in circles (M = 28.0 s, SD = 11.5 s) attracted shorter glances than the one walking from side to side (M = 31.9 s, SD = 10.2 s) and the one standing (M = 46.1 s, SD = 7.77 s; $p < .002$). In addition to that, the agent walking from one side to the other was looked at longer than the one standing ($p < .0001$). That means the standing agent was looked at longest, and the side-by-side-walking agent was looked at longer than the one walking in circles. Next to that, there was no significant effect of *Audio* ($p = .281$) and no significant interactions ($p > .201$). Repeated measures ANOVA on the rank-aligned data did not detect an effect of the trial number (1 to 12; $F(11, 420) = .457, p < .93, \eta_p^2 = .01$).

4.3 Correlation of Time and Number

The Spearman's rank correlation between the total time and the number of times participants looked at the agent was statistically significant (ρ = -.6253., S = 21838946, $p < .001$). The correlation coefficient suggests a strong relationship between the two variables: the more overall glances the agent accumulates, the less participants tend to look at the agent. This is expected as many glances suggest they look not at the agent but elsewhere and back again.

5 DISCUSSION

We were interested in whether the type of audio or the agent's rendering style influences gaze behaviour. We tested this with an agent that walked in circles, walked from side to side, and stood in one place to induce dynamics in the gaze behaviour.

Our study revealed the impact of visualization on participants' viewing behaviour, encompassing the overall duration of gaze directed at the agent and the total number of gaze shifts. Specifically, participants diverted their gaze less often and looked longer at the cartoon agent, unlike the more averted and overall shorter gaze duration at the realistic agent. It is possible that an uncanny valley effect influenced gaze behaviour [Grebot et al. 2022]. However, the 3d models have been used in prior studies without adverse comments from participants [Arboleda et al. 2024; Mikhailova et al. 2024]. Thus, we believe both to have minimal influence. The reason for this could be that the realistic agent seemed less attractive simply because it was well-embedded in reality, whereas the cartoon agent stood out due to its (purposefully) unrealistic design. This design might have invited participants to look and explore the agent longer (and, by that, divert their gaze less often). This aligns with the fact that "oddball stimuli", which are visually very different from the background, attract more visual attention [Amit et al. 2018; Corbetta and Shulman 2002]. Our results suggest that narrative experiences that require the user to look at the storyteller might benefit from an agent with a less realistic rendering style as it appears more salient.

In previous research on communication, non-realistic avatars got fewer and shorter gazes [Wang et al. 2019]. Aseeri and Interrante [2021] also concluded that a realistic avatar's face (3D video embedded in VR) attracts more gazes than a scanned avatar (3D model). Together with our results (which are contradictory), this suggests that realism might play more of a role in an interactive task, especially in communication tasks where facial expressions are relevant. In storytelling applications, the playful character of a non-realistic (cartoon) agent seems to attract more visual attention than a plain, realistic, well-embedded virtual human. The difference in applications is also highlighted by Korre [2023], who conclude that realistic avatars are perceived as more serious, whereas cartoon agents add to the feeling of the experience being a game.

We detected no significant differences regarding audiovisual spatial coherence (no interaction effects) and not for audio alone. This suggests that in simple storytelling applications, both coherence and type of audio play less of a role when considering gaze behaviour. The type of animation led to significant differences, although this was expected: The HoloLens2 has a relatively small field of view (FoV), making it more likely that the walking agent will walk out of the FoV before the user adjusts their gaze. Also, for the circle animation, the agent was sometimes behind the participants and, while in the beginning, most turned around to follow, not all of them turned fully around later in the experiment but waited for the agent to walk back in front of them.

Together, our findings imply that researchers and developers should especially pay attention to the visualization of the virtual human and that a simple non-spatial audio solution might be sufficient when primarily analyzing gaze in uni-directional storytelling scenarios.

The main strengths of the present study are the experimental design and the manipulation of various conditions, including auralization, visualization and animation (acting as a control). Nevertheless, it is important to acknowledge its main limitations. For this (first) analysis, we only investigated eye movements on the full body and not other types of movement, such as head gaze or body rotation and no subdivision (e.g., fixations on the face, which is a gaze attractor [Coutrot and Guyader 2014]). We did not analyze where exactly participants were looking at while looking at the agent but also not while looking away. We also did not have a condition featuring an actual human (actor) to see if the gaze behaviour with AR agents is the same as with real humans. Future research should investigate these variables and the influence of different types of

experiences, such as different stories and educational content. In addition, having qualitative feedback of participants' experiences would further explain differences in gaze behaviour.

6 CONCLUSION

This study investigated the impact of rendering style and auralization on gaze behaviour during an AR storytelling experience delivered by an agent. The results indicate that the type of agent visualization has a significant effect on participants' gaze behaviour, with the cartoon-style agent attracting longer gaze duration in total and fewer glances compared to the realistic virtual human. However, the audio conditions, spatial or non-spatial, and audiovisual coherence (or the lack thereof) did not lead to significant differences in gaze behaviour. Together, our results emphasize the significance of visualization in shaping users' gaze behaviour. Our findings highlight that if visual attention is important, a cartoon agent might be more beneficial than a realistic visualization, whereas auralization does not affect gaze behaviour significantly in storytelling experiences. In the future, further exploration of the influence of task (storytelling, conversation, collaboration, learning) and the validity of the gaze behaviour (comparison with actual humans) can shed further light on the gaze behaviour during the interaction with AR agents.

ACKNOWLEDGMENTS

This research is funded by the Carl-Zeiss-Foundation ("Breakthroughs 2020" program, https://www.carl-zeiss-stiftung.de/programm/czs-durchbrueche), in the CO-HUMANICS project, by the German Federal Ministry of Education and Research (BMBF) through the MULTIPARTIES project (Grant No. 16SV8922), and by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant No. 101021229, GEMINI: Gaze and Eye Movement in Interaction). Supplemental material available at https://zenodo.org/doi/10.5281/zenodo.10911677.

REFERENCES

Roy Amit, Shlomit Yuval-Greenberg, et al. 2018. Oculomotor behavior during non-visual tasks: the role of visual saliency. *Journal of Vision* 18, 10 (2018), 1199–1199.

Sean Andrist, Michael Gleicher, and Bilge Mutlu. 2017. Looking Coordinated: Bidirectional Gaze Mechanisms for Collaborative Interaction with Virtual Characters. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 2571–2582. https://doi.org/10.1145/3025453.3026033

Stephanie Arévalo Arboleda, Christian Kunert, Jakob Hartbrich, Christian Schneiderwind, Chenyao Diao, Christoph Gerhardt, Tatiana Surdu, Florian Weidner, Wolfgang Broll, Stephan Werner, and Alexander Raake. 2024. Beyond Looks: A Study on Agent Movement and Audiovisual Spatial Coherence in Augmented Reality. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces*. Orlando, FL. https://doi.org/10.1109/VR58804.2024.00071

Sahar Aseeri and Victoria Interrante. 2021. The Influence of Avatar Representation on Interpersonal Communication in Virtual Social Environments. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (2021), 2608–2617. https://doi.org/10.1109/TVCG.2021.3067783

Gérard Bailly, Stephan Raidt, and Frédéric Elisei. 2010. Gaze, conversational agents and face-to-face communication. *Speech Communication* 52, 6 (2010), 598–612. https://doi.org/10.1016/j.specom.2010.02.015 Speech and Face-to-Face Communication.

Judee K Burgoon, Deborah A Coker, and Ray A Coker. 1986. Communicative effects of gaze behavior: A test of two contrasting explanations. *Human Communication Research* 12, 4 (1986), 495–524.

Maurizio Corbetta and Gordon L Shulman. 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience* 3, 3 (2002), 201–215.

Antoine Coutrot and Nathalie Guyader. 2014. How saliency, faces, and sound influence gaze in dynamic social scenes. *Journal of vision* 14, 8 (2014), 5–5.

Dario Alfonso Cuello Mejía, Hidenobu Sumioka, Hiroshi Ishiguro, and Masahiro Shiomi. 2023. Evaluating Gaze Behaviors as Pre-Touch Reactions for Virtual Agents. *Frontiers in Psychology* 14 (2023). https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1129677

Michele Geronazzo, Amalie Rosenkvist, David Sebastian Eriksen, Camilla Kirstine Markmann-Hansen, Jeppe Køhlert, Miicha Valimaa, Mikkel Brogaard Vittrup, and Stefania Serafin. 2019. Creating an audio story with interactive binaural rendering in virtual reality. *Wireless Communications and Mobile Computing* 2019 (2019), 1–14.

Gordon N Goldberg, Charles A Kiesler, and Barry E Collins. 1969. Visual behavior and face-to-face distance during interaction. *Sociometry* (1969), 43–53.

Mar Gonzalez-Franco, Anthony Steed, Steve Hoogendyk, and Eyal Ofek. 2020. Using Facial Animation to Increase the Enfacement Illusion and Avatar Self-Identification. *IEEE Transactions on Visualization and Computer Graphics* 26, 5 (2020), 2023–2029. https://doi.org/10.1109/TVCG.2020.2973075

Ivan Bouchardet da Fonseca Grebot, Pedro Henrique Pinheiro Cintra, Emilly Fátima Ferreira de Lima, Michella Vaz de Castro, and Rui de Moraes Jr. 2022. Uncanny Valley Hypothesis and Hierarchy of Facial Features in the Human Likeness Continua: An Eye-Tracking Approach. *Psychology & Neuroscience* 15, 1 (2022), 28–42. https://doi.org/10.1037/pne0000281

Aleshia Hayes. 2015. The experience of physical and social presence in a virtual learning environment as impacted by the affordance of movement enabled by motion tracking. *Electronic Theses and Dissertations, University of Central Florida* (2015).

Etienne Hendrickx, Mathieu Paquier, and Vincent Koehl. 2015. Audiovisual Spatial Coherence for 2D and Stereoscopic-3D Movies. *J. Audio Eng. Soc* 63, 11 (2015), 889–899. http://www.aes.org/e-lib/browse.cfm?elib=18049

Darragh Higgins, Katja Zibrek, Joao Cabral, Donal Egan, and Rachel McDonnell. 2022. Sympathy for the digital: Influence of synthetic voice on affinity, social presence and empathy for photorealistic virtual humans. *Computers & Graphics* 104 (2022), 116–128.

Felix Immohr, Gareth Rendle, Annika Neidhardt, Steve Göring, Rakesh Rao Ramachandra Rao, Stephanie Arevalo Arboleda, Bernd Froehlich, and Alexander Raake. 2023. Proof-of-Concept Study to Evaluate the Impact of Spatial Audio on Social Presence and User Behavior in Multi-Modal VR Communication. In *Proceedings of the 2023 ACM International Conference on Interactive Media Experiences*. 209–215.

Franz Kafka and Nahum N. Glatzer. 1971. *Franz Kafka, the Complete Stories*. Schocken Books, New York.

Dominic Kao, Rabindra Ratan, Christos Mousas, Amogh Joshi, and Edward F. Melcer. 2022. Audio Matters Too: How Audial Avatar Customization Enhances Visual Avatar Customization. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA,) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 165, 27 pages. https://doi.org/10.1145/3491102.3501848

Dominic Kao, Rabindra Ratan, Christos Mousas, and Alejandra J Magana. 2021. The effects of a self-similar avatar voice in educational games. *Proceedings of the ACM on Human-Computer Interaction* 5, CHI PLAY (2021), 1–28.

Matthew Kay, Lisa A. Elkin, James J. Higgins, and Jacob O. Wobbrock. 2021. *mjskay/ARTool: ARTool 0.11.0*. https://doi.org/10.5281/zenodo.4721941

Stevanus Kevin, Yun Suen Pai, and Kai Kunze. 2018. Virtual gaze: exploring use of gaze as rich interaction method with virtual agent in interactive virtual reality content. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology* (Tokyo, Japan) *(VRST '18)*. Association for Computing Machinery, New York, NY, USA, Article 130, 2 pages. https://doi.org/10.1145/3281505.3281587

Danai Korre. 2023. Comparing Photorealistic and Animated Embodied Conversational Agents in Serious Games: An Empirical Study on User Experience. In *International Conference on Human-Computer Interaction*. Springer, 317–335.

Luchcha Lam, Minsoo Choi, Magzhan Mukanova, Klay Hauser, Fangzheng Zhao, Richard Mayer, Christos Mousas, and Nicoletta Adamo-Villani. 2023. Effects of Body Type and Voice Pitch on Perceived Audio-Visual Correspondence and Believability of Virtual Characters. In *ACM Symposium on Applied Perception 2023*. 1–11.

Marc Erich Latoschik and Carolin Wienrich. 2022. Congruence and plausibility, not presence: Pivotal conditions for XR experiences and effects, a novel approach. *Frontiers in Virtual Reality* 3 (2022), 694433.

Nadine Lavan, Wing Yue Chan, Yongping Zhuang, Isabelle Mareschal, and Sukhwinder S Shergill. 2022. Direct eye gaze enhances the ventriloquism effect. *Attention, Perception, & Psychophysics* 84, 7 (2022), 2293–2302.

Binh H. Le, Xiaohan Ma, and Zhigang Deng. 2012. Live Speech Driven Head-and-Eye Motion Generators. *IEEE Transactions on Visualization and Computer Graphics* 18, 11 (2012), 1902–1914. https://doi.org/10.1109/TVCG.2012.74

Rachel McDonnell, Martin Breidt, and Heinrich H. Bülthoff. 2012. Render Me Real? Investigating the Effect of Render Style on the Perception of Animated Virtual Humans. *ACM Trans. Graph.* 31, 4, Article 91 (jul 2012), 11 pages. https://doi.org/10.1145/2185520.2185587

Catarina Mendonça and Victoria Korshunova. 2020. Surround sound spreads visual attention and increases cognitive effort in immersive media reproductions. In *Proceedings of the 15th International Audio Mostly Conference* (Graz, Austria) *(AM '20)*. Association for Computing Machinery, New York, NY, USA, 16–21. https:

//doi.org/10.1145/3411109.3411118

Veronika Mikhailova, Christoph Gerhardt, Christian Kunert, Tobias Schwandt, Florian Weidner, Wolfgang Broll, and Nicola Döring. 2024. Age and Realism of Avatars in Simulated Augmented Reality: Experimental Evaluation of Anticipated User Experience. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces*. https://doi.org/10.1109/VR58804.2024.00032

Michael G. Nelson, Alexandros Koilias, Christos-Nikolaos Anagnostopoulos, and Christos Mousas. 2022. 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)*. 594–599. https://doi.org/10.1109/ISMAR-Adjunct57072.2022.00123

I.S. Pandzic and R. Forchheimer. 2003. *MPEG-4 Facial Animation: The Standard, Implementation and Applications*. Wiley. https://books.google.de/books?id=cMAS4gaUWVAC

Tanmay Randhavane, Aniket Bera, Kyra Kapsaskis, Rahul Sheth, Kurt Gray, and Dinesh Manocha. 2019. EVA: Generating Emotional Behavior of Virtual Agents using Expressive Features of Gait and Gaze. In *ACM Symposium on Applied Perception 2019* (Barcelona, Spain) *(SAP '19)*. Association for Computing Machinery, New York, NY, USA, Article 6, 10 pages. https://doi.org/10.1145/3343036.3343129

Daniel Roth, Jean-Luc Lugrin, Dmitri Galakhov, Arvid Hofmann, Gary Bente, Marc Erich Latoschik, and Arnulph Fuhrmann. 2016. Avatar realism and social interaction quality in virtual reality. In *2016 IEEE Virtual Reality (VR)*. 277–278. https://doi.org/10.1109/VR.2016.7504761

Hanne Stenzel, Jon Francombe, and Philip JB Jackson. 2019. Limits of perceived audio-visual spatial coherence as defined by reaction time measurements. *Frontiers in neuroscience* 13 (2019), 451.

Michalis Tsepapadakis and Damianos Gavalas. 2023. Are you talking to me? An Audio Augmented Reality conversational guide for cultural heritage. *Pervasive and Mobile Computing* 92 (2023), 101797.

Roel Vertegaal, Robert Slagter, Gerrit van der Veer, and Anton Nijholt. 2001. Eye Gaze Patterns in Conversations: There is More to Conversational Agents than Meets the Eyes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) *(CHI '01)*. Association for Computing Machinery, New York, NY, USA, 301–308. https://doi.org/10.1145/365024.365119

Isaac Wang and Jaime Ruiz. 2021. Examining the Use of Nonverbal Communication in Virtual Agents. *International Journal of Human–Computer Interaction* 37, 17 (2021), 1648–1673. https://doi.org/10.1080/10447318.2021.1898851 arXiv:https://doi.org/10.1080/10447318.2021.1898851

Isaac Wang, Jesse Smith, and Jaime Ruiz. 2019. Exploring Virtual Agents for Augmented Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300511

Richard Warp, Michael Zhu, Ivana Kiprijanovska, Jonathan Wiesler, Scot Stafford, and Ifigeneia Mavridou. 2022. Validating the effects of immersion and spatial audio using novel continuous biometric sensor measures for Virtual Reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 262–265. https://doi.org/10.1109/ISMAR-Adjunct57072.2022.00058

Florian Weidner, Gerd Boettcher, Stephanie Arevalo Arboleda, Chenyao Diao, Luljeta Sinani, Christian Kunert, Christoph Gerhardt, Wolfgang Broll, and Alexander Raake. 2023. A Systematic Review on the Visualization of Avatars and Agents in AR & VR displayed using Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2596–2606. https://doi.org/10.1109/TVCG.2023.3247072

Keiichi Zempo, Azusa Yamazaki, Naoto Wakatsuki, Koichi Mizutani, and Yukihiko Okada. 2022. Mouth-in-the-Door: The Effect of a Sound Image of an Avatar Intruding on Personal Space That Deviates in Position From the Visual Image. *IEEE Access* 10 (2022), 125772–125791. https://doi.org/10.1109/ACCESS.2022.3222804

Katja Zibrek, Joao Cabral, and Rachel McDonnell. 2021. Does Synthetic Voice Alter Social Response to a Photorealistic Character in Virtual Reality?. In *Proceedings of the 14th ACM SIGGRAPH Conference on Motion, Interaction and Games* (Virtual Event, Switzerland) *(MIG '21)*. Association for Computing Machinery, New York, NY, USA, Article 11, 6 pages. https://doi.org/10.1145/3487983.3488296