Effects of Hybrid and Synthetic Social Gaze in Avatar-Mediated Interactions

Daniel Roth[*]
HCI Group, University of Würzburg

Peter Kullmann[†]
HCI Group, University of Würzburg

Gary Bente[‡]
Michigan State University

Dominik Gall[§]
HCI Group, University of Würzburg

Marc Erich Latoschik[¶]
HCI Group, University of Würzburg

ABSTRACT

Human gaze is a crucial element in social interactions and therefore an important topic for social Augmented, Mixed, and Virtual Reality (AR,MR,VR) applications. In this paper we systematically compare four modes of gaze transmission: (1) natural gaze, (2) hybrid gaze, which combines natural gaze transmission with a social gaze model, (3) synthesized gaze, which combines a random gaze transmission with a social gaze model, and (4) purely random gaze. Investigating dyadic interactions, results show a linear trend for the perception of virtual rapport, trust, and interpersonal attraction, suggesting that these measures increase with higher naturalness and social adequateness of the transmission mode. We further investigated the perception of realism as well as the resulting gaze behavior of the avatars and the human participants. We discuss these results and their implications.

Index Terms: Human-centered computing—Visualization—Visualization techniques—;

1 INTRODUCTION

In face-to-face encounters, nonverbal behavior such as body movement, facial expressions and eye gaze, is of paramount importance for person perception, impression formation, emotion inferences and rapport [8, 35]. Our appearance and our behavior is not mutable, like spoken language. In fact, it is analogue and "always on". During virtual social interactions however, we interact with avatars driven by human behaviors, or with agents driven by algorithms [4]. Both can vary in the degree to which behavior is replicated and expressed. Consequently, a lack of variation and contingency in displayed behavior, or distortions in its dynamics, will be interpreted as a meaningful signal, no matter whether intentional or not.

To this regard, gaze behavior is especially important, due to its use for initiating conversations, understanding others and expressing ourselves, disclosing human comprehension, and means of turn taking [1, 13, 20, 22]. Kleinke [22] describes face-gaze, eye-gaze, mutual gaze, eye contact, looking/gazing, gaze avoidance, gaze omission, staring and glancing as some phenomena of gaze behavior which can be present in social interactions. Humans without visual or social disorders can typically experience, disambiguate, and interpret gaze behaviors, and adequately react to it.

This has major implications regarding the inclusion of gaze in future applications for virtual interactions [33]. First, imperfections in capture, transmission or rendering can lead to subtle variations in the observable behavior and potentially cause mis-attributions of intentions and emotions, or induce undesirable social impressions. Second, the lack or disturbance of gaze transmission could be either compensated by artificial social intelligence or completely transformed [2], for example via visual augmentations [31] triggered by social signals and phenomena. Third, the technology can serve as a monitoring and repair system which detects inadequate nonverbal behavior of humans (e.g. lack of eye contact) and compensates for those by synthesizing more appropriate behavior [32]. Against this background, the potential role of virtual gaze behavior, including opportunities [26] and risks, has to be re-asked.

The present study therefore aims at a systematic investigation of social gaze in dyadic virtual interactions, assessing effects on social perception. Based on a social gaze model respecting for speaker-listener coherences that was derived from previous literature, we compare 1) a pure *natural gaze* transmission that is based on eye tracking data, 2) a *hybrid gaze* model based on a natural gaze default transmission augmented with a social gaze model, 3) a *synthesized gaze* model based on a random gaze default transmission augmented with the social gaze model, and 4) a pure *random gaze* transmission that is merely based on statistically distribution without taking into account any social affordances.

Conceiving natural gaze behavior as the gold standard, we hypothesize that non-contingent random gaze, which just follows general rules of frequency and duration, falls short from being perceived as natural and meaningful. We further hypothesize that a dynamic augmentation of this behavior, i.e. making it contingent to the partners' speech activity using a synthesized gaze model will be evaluated more similar to natural gaze. Applying the same algorithms to augment the natural behavior could result in negative effects as it could destroy the subtle dynamics of natural interactions. To our best knowledge, this is the first study to provide a systematic comparison in this manner.

2 RELATED WORK

2.1 Gaze Behavior Models

Lee et al. [24] present a gaze model based on empirical models of saccades and statistical models of gaze data retrieved from video analysis. They included an attention monitor module that accounts for statuses such as talking vs. listening, head rotation, mutual gaze or averted gaze. They found indicators for higher engagement and friendliness compared to random gaze, and higher perceived liveliness and friendliness compared to static gaze.

With regard to conversational interaction, Garau et al. [15] compared random gaze to inferred gaze based on audio input. In accordance with literature on gaze behavior [1], they based their inference model on the assumption that people gaze more at their interaction partner while listening [14, 15] and included an audio trigger "while speaking"/"while listening" to infer gaze. They found that inferred gaze affects the evaluation of the interaction, including co-presence, the perception of similarity to a face-to-face interaction and results in a more positive evaluation of the partner.

Bente et al. [6] investigated different patterns of randomized gaze behaviors in three conditions: short periods of directed gaze (two

[*]e-mail: daniel.roth@uni-wuerzburg.de
[†]e-mail: peter.kullmann@stud-mail.uni-wuerzburg.de
[‡]e-mail: gabente@msu.edu
[§]e-mail: dominik.gall@uni-wuerzburg.de
[¶]e-mail: marc.latoschik@uni-wuerzburg.de

seconds averted, two seconds directed) vs. long periods of directed gaze (two seconds averted, four seconds directed) vs. real gaze. They found that longer periods of gaze direction led to higher co-presence and a better evaluation compared to shorter gaze periods.

Vinayagamoorthy et al. [36] expanded the model from [15] to inferred gaze, assuming that mean saccade magnitudes are shorter, that people tend to focus on their interaction partner more, and that inter-saccadic intervals between the focus positions are longer when listening [36]. They found that inferring gaze with more realistic characters can lead to increased effectiveness whereas applying the gaze model on a cartoonish character did not show any difference.

Ma and Deng [25] synthesize gaze by building a dynamic coupled component analysis-based statistical model, trained with eye-head motion data which seemed to be superior to the model from Lee and colleagues [24], but the evaluation was limited to the assessment of animation clips. Furthermore, their model did not include speech, which was identified as a modifier of gaze behavior [1, 14, 15] as a trigger of gaze patterns. Le et al. [23] further extended this work with a fully automated framework to generate head motions, eye gaze, and eyelid motion based on live or recorded speech input. Although extending the dimension of behavioral realism, live avatar-based interactions were not investigated in their work, nor in other related work that used a multi-model approach [29], or reactive gaze behavior based on head position tracking [21].

Overall, previous works show that artificial gaze models can result in positive impacts on the perception of an animation or a communication scenario. However, previous works did not systematically compare gaze models to natural gaze or hybrid forms, and thus indications about the level of adequateness and perceived realism of these models are still an open question, which will be tackled in the present work.

2.2 Gaze Transformation and Hybrid Approaches

Previous research investigating social gaze mostly focused on the replication of gaze behavior to avatars, e.g. [15], or artificial gaze models for agents solely driven by algorithms [34]. An alternative line of research investigated whether transforming the social interaction in virtual environments by decoupling the visual representation from the physical behavior can affect the interaction. Bailenson, Beall, et al., [3, 5] investigated the idea of non-zero-sum gaze, i.e. the user-dependent simulation of another person's gaze. By augmenting gaze behavior, virtual environments can surrogate individual points of attention to each participant of multi-user simulations, and therefore foster a more positive perception towards the interaction. For instance, one user can direct his/her gaze on multiple interaction partners simultaneously as each simulation can be rendered individually, which in turn can express increased attentiveness to each interactant. In a study with two participants and one presenting experimenter, Beall, et al. [5] found that participants directed their gaze more often towards the experimenter when non-zero-sum gaze was active (only directed gaze for each participant) compared to reduced gaze (only averted gaze) or natural gaze (approximated via head movements).

Although studies of transformed social interaction give indications about the potential impacts of modified behavior such as increased agreement [3, 5], previous models rely on either linear manipulations or algorithms that do not respect for interpersonal adaptations within the virtual interaction. Doing so however allows to merge both forms of embodied representations, avatars and agents, to hybrid forms [32]. Hybrid models have been investigated with regard to their potential to evoke continuous presence [17], and in the context of establishing alternative communication channels on the basis of behavioral phenomena, such as eye contact, joint attention, and grouping [31].

The benefits of hybrid technologies could be manifold. First, the introduction of augmented behaviors could compensate for the lack of sensor inputs and transmission errors by compensating for missing or interrupted behaviors based on a underlying social AI. Second, socio-communicative deficits could be studied, further extending "offline" paradigms [16] and allow to study these behaviors in interactions. Third, gaining insights into the effects of hybrid systems and their further development could foster the inclusion and training of individuals with social disorders.

2.3 Discussion

Previous work focused on artificial gaze models for the animation of and interaction with virtual characters [21, 23–25, 29, 34], as well as for the compensation of missing sensory input in shared virtual environments [6, 14, 15, 36]. Two major approaches to construct artificial gaze can be derived: i) a randomized gaze model based on statistical distributions of fixations and saccades, and ii) a randomized gaze behavior default transmission which can be augmented using a model respecting for social affordances to a synthesized gaze behavior, which could also compensate for transmission interrupts. These approaches assume a complete lack of gaze information, and could be considered "agent" approaches. For the compensation of non-adequate social gaze (e.g., in inter-cultural contexts, or with regard to social disorders), and to foster affinity and liking, iii) a third "hybrid" solution that uses natural gaze as default transmission but also reacts upon social affordances based on an underlying AI can be constructed with hybrid avatar-agent technologies [32]. The present work systematically assesses to what degree adequate social gaze behavior in avatar-mediated communication can be established through such models.

3 Materials and Methods

We adopted a dyadic design with four conditions: "natural gaze, non-augmented" (natural/gold standard), "natural gaze, augmented" (hybrid), "random gaze, augmented" (synthesized), and "random gaze, non-augmented" (random/baseline control). In each dyad, participants were assigned to the same condition, and indicated the same biological sex.

3.1 Apparatus

A client-server architecture was developed in Unity3D (v.5.6.0f3). We tracked participants gaze using Tobii 4C eyetrackers (90 Hz), attached to a 28" screen (1920x1200px). We used Sennheiser PC310 headsets for audio transmission. A motion-to-photon latency measurement (240 Hz camera) of the eye movement resulted in $M = 308$ *ms* ($SD = 33$ *ms*, $N = 50$) latency. Audio latency was measured with source and client end-to-end audio recordings capturing impulse responses ($M = 281$ *ms*, $SD = 22$ *ms*). Participants were placed in remote rooms, see Fig. 1. We used cartoon-like 3D avatars (faceshift AG, adapted) displayed in front of a black background, that matched the biological gender of participants (Fig. 1). We chose this avatar type as it provides a clear gaze indication.

The application is based on five modules. (A) Speech to animation (SALSA, Crazy Minnow Studio, LLC) is used to process audio input and approximate mouth movements. (B) voice communication is established via voice over IP (Dissonance Voice Chat, Placeholder Software). (C) Natural gaze/head movement tracking is performed using Tobii AB's Unity SDK. (D) A model-based reflex agent (social AI) to be able to transition between the default gaze state and the social gaze model (Fig. 2). (E) An experimenter control interface.

3.2 Computed Gaze Models

The simulation transmitted speech directly to the interlocutors and rendered gaze according to one of the conditions onto their digital representation. For the *random gaze model*, fixations lasted between 1 second and 3.3 seconds. To indicate a fixation, a random (normal distribution) screen coordinate was generated by projecting ray from a fixed (user spawn) position to a random coordinate on the near

Figure 1: ***Left:*** Female and male avatars used in the study. ***Center:*** Study setup with users at maximum distance to the screen. The chairs were fixed to assure tracking quality. ***Right:*** The target head area of interest (head AOI) and eyes AOI used to assess the avatar/agent/hybrid gaze behavior as well as the subjects' real gaze behaviors during the interaction.

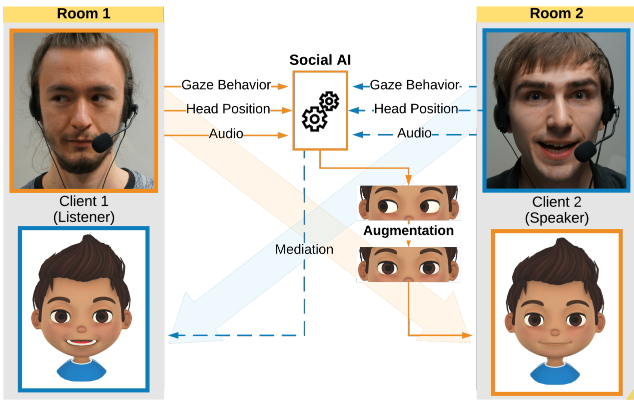

Figure 2: Hybrid gaze model. The model transmits natural gaze and intervenes according to social adequateness. Depicted by the orange pathway, the social AI detects the status of a speaker via audio input. As an adequate reaction, the gaze behavior of the listener is augmented (directed gaze), altering the natural gaze behavior.

camera frustum. The gaze direction of the avatar was then modified to this direction. To approximate saccades, we used a linear interpolation of 20 frames (i.e. 0.33 seconds at 60 Hz render refresh rate).

We conceptualized our *social gaze model* similar to the one proposed by [24]. Drawing from previous findings that listeners make more eye contact than speakers, we display behavior depicting an attentive listener, see Fig. 2. Our idle states are the constructed random gaze behavior or the tracked natural behavior, respectively. We detect a potential adequate listening behavior through audio input, similar to [15]. Once an audio threshold is reached, an event triggers the augmentation state. In the augmentation state, the gaze of the listener is directed to the screen center, in order to induce a "listening focus" (i.e. directed gaze) to the speaker. Starting with this directed gaze, the inter-saccadic interval is randomly selected from a normal distribution ($M = 3.97$ *s*, $SD = 0.78$ *s*) in order to prevent a "stare" effect and simulate a more naturalistic behavior. Upon reaching the directed gaze duration limit, a saccade is introduced, following a fixation on a random fixation point on the interlocutor's screen for a shorter interval (normal distribution, $M = 0.22$ *s*, $SD = 0.12$ *s*). All fixations are held stable for the time of the inter-saccadic interval. We chose this approach based on findings suggesting a preferred mutual gaze duration for dyadic settings at around three seconds [7].

This model results in virtual characters who face a speaking human to make eye contact and occasionally look away. In the two augmented conditions, it is thus combined with either a natural gaze behavior or randomized gaze behavior default transmission

3.3 Procedure

The study was conducted at the University of Würzburg and took about 45 minutes. Participants were welcomed and quickly introduced to each other. Each participant was guided into a separate room and handed the study information and the consent form. Once participants agreed to participate, they answered the pre-study questionnaire. Participants were then quickly instructed about the apparatus, equipped with headphones, and calibrated for eye tracking. The experimenter then started the simulation, which gave written instructions to the participants about their task for approximately 30 seconds ("In the following, you will see your communication partner visualized as virtual character. Please have a conversation for the next five minutes and get to know each other, as if it was a normal conversation. When getting to know each other, you are free to choose any topics you want to talk about."), followed by the interaction which lasted for five minutes. Afterwards, the audio stream was cut followed by a visual text box asking to wait for the experimenter. The participants were then presented with the questionnaire assessing dependent measures and finally debriefed in detail. The study was approved by the Ethical Committee of [Institution]. Student participants were compensated with study credit points.

3.4 Measures

We introduced a *manipulation check*, assessing the amount of time (dwell time) the avatars established directed gaze (head AOI), eye gaze (eyes AOI), and background focus (background AOI). To do so, we calculated gaze vector ray cast hits of dynamic areas of interest (AOI, see Fig. 1) in the simulation (60 Hz). Dwell times for the analysis were calculated exclusively (i.e. an eyes AOI hit would not change the head AOI statistics). As this check was introduced during the cause of data collection, the sample size is limited to $N = 58$.

We asked participants to judge their *(virtual) rapport* with their interlocutors [18], e.g., "I felt I had a connection with my interaction partner."). The scale (1 strongly disagree, 7 strongly agree) included 11 items (Cronbach's $\alpha = .85$).

We measured *social presence* with six items ($\alpha = .83$) slightly adapted from [28] to fit the same response format, e.g., "To what extent did you feel able to assess your partner's reactions to what you said?" = "I was able to assess my interaction partner's reactions to my statements" (1 strongly disagree, 7 strongly agree).

We assessed *interpersonal attraction* [9] using six agreement statements (1 strongly disagree, 7 strongly agree), e.g. "I like my interaction partner", "My interaction partner is friendly" ($\alpha = .86$).

As trust was shown to be affected by gaze in multiple studies (e.g. [27]), the *perceived trust* was measured with three items ("I think, my interaction partner has good intentions", "I would rely on my interaction partner", "I would trust my interaction partner"; 1 does not apply at all, 7 totally applies; $\alpha = .745$)

To investigate the impact of perceived gaze *movement realism and movement naturalness*, we presented participants two statements (1 doesn't apply at all, 7 totally applies): "The eyes of the virtual character moved realistically", and "The eyes of the virtual character moved naturally". To check for any impact on the avatar perception, we assessed the avatar's appearance regarding *humanness* ($\alpha = .816$), and *eeriness* ($\alpha = .736$) using a semantic differential [19].

Similarly to the manipulation check, we assessed the actual *users' true gaze behavior* by dwell times for the head AOI, eyes AOI, and background AOI (see Fig. 1). In addition to these measures, we assessed affect and the Big Five inventory. The results of these data are not subject to the current paper.

3.5 Participants

148 participants were recruited at the University of Würzburg. We excluded dyads when tracking failed for longer periods ($N = 28$), and when participants knew each other before ($N = 28$). The final sample consisted of $N = 90$ participants, $M_{age} = 22.01$ ($SD_{age} = 3.20$). Of these, 48 were female, 89 were german, and 88 were students. Dyads were randomly assigned to one condition. A chi square test showed that the distribution of gender was not significantly different.

3.6 Considerations for the Analysis

We assessed the need for a multi-level model [12] and tested for a change in the likelihood ratio. None of the tests revealed significant decreases, thus denying the need for multi-level modeling. To analyze the manipulation check and the participants' gaze behavior, we calculated one-way ANOVAs with pos-hoc comparisons (Tukey HSD). To analyze the subjective outcomes, we calculated Kruskal-Wallis tests. The assumption of equal distributions was assessed by visual inspection of a boxplot. Subsequently, we conducted pairwise comparisons using Dunn's [11] procedure with a Bonferroni correction for multiple comparisons. Furthermore, we conducted Jonckheere-Terpstra tests for ordered alternatives to identify linear trends, ordered from the most natural to the most synthetic condition: "natural-gaze, non augmented" (*natural gaze*, $N = 24$, 12 female) "natural gaze, augmented" (*hybrid gaze*, $N = 22$, 14 female), "random gaze, augmented" (*synthetic gaze*, $N = 24$, 10 female) and "random gaze, non augmented" (*random gaze*, $N = 20$, 12 female).

4 Results

4.1 Manipulation Check

One-way ANOVAs showed significant main effects for the the background AOI $F(3,57) = 96.63$, $p < .001$, $\eta_p^2 = .843$, the head AOI $F(3,57) = 9.60$, $p < .001$, $\eta_p^2 = .348$, and the eyes AOI $F(3,57) = 6.84$, $p = .001$, $\eta_p^2 = .275$. Pairwise comparisons showed that the random gaze evoked significantly more background attention than all other conditions, and that hybrid gaze showed significantly less attention to the background compared to all other conditions ($ps <= .010$). Natural gaze was significantly lower in head AOI dwell time than synthetic gaze ($p = .019$). Random gaze, in turn, was significantly lower in head AOI dwell time than hybrid gaze and synthetic gaze ($ps <= .001$). Natural gaze was significantly higher in eye AOI dwell time than random gaze ($p = .001$) and random gaze was significantly lower in eye AOI dwell time than natural and hybrid gaze ($ps <= 0.03$). No further significant effects were found. The social gaze model seems to have successfully altered the gaze behavior, i.e. directed gaze if appropriate (Fig. 3). The manipulation check was therefore deemed successful.

4.2 Virtual Rapport

A Kruskal-Wallis test (equal distributions) showed that rapport scores were statistically significantly different between the conditions, ($\chi^2(3) = 9.13$), $p = .028$. Pairwise comparisons revealed a statistically significant difference in rapport scores between the random gaze condition ($Mdn = 5.14$) and the natural gaze condition ($Mdn = 5.73$, $p = .024$, see Table 1). A Jonckheere-Terpstra test showed that there was a statistically significant trend for rapport scores, $TJT = 1109.00$, $z = -2.941$, $p = .003$.

4.3 Social Presence

A Kruskal-Wallis test (equal destributions) did not yield to significant results ($\chi^2(3) = 3.77$), $p = .287$. A Jonckheere-Terpstra test did not show a significant trend $TJT = 1295.00$, $z = -1.593$, $p = .111$.

4.4 Interpersonal Attraction

A Kruskal-Wallis test (equal distributions) revealed that median scores were significantly different between the conditions, ($\chi^2(3) = 9.13$), $p = .028$. Pairwise comparisons revealed statistically significant differences in interpersonal attraction scores between the random gaze condition ($Mdn = 5.14$) and natural gaze condition ($Mdn = 5.73$, $p = .024$), but not between any other conditions. A Jonckheere-Terpstra test showed that there was a statistically significant trend, $TJT = 1029.50$, $z = -3.524$, $p < .001$.

4.5 Trust

A Kruskal-Wallis test (equal destributions) did not yield to significant results ($\chi^2(3) = 5.25$), $p = .137$. A Jonckheere-Terpstra test did however show a significant trend $TJT = 1211.00$, $z = -2.230$, $p = .026$.

4.6 Behavioral Realism and Behavioral Naturalness

A Kruskal-Wallis test showed that judgments of behavioral realism were significantly different between the conditions, ($\chi^2(3) = 9.66$), $p = .022$. Pairwise comparisons revealed statistically significant differences in scores between the random gaze condition ($Mdn = 4.50$) and the natural gaze condition ($Mdn = 6.0$, $p = .026$), but not between any other comparison. Similarly, behavior naturalism was judged significantly different between conditions ($\chi^2(3) = 11.950$), $p = .008$ and pairwise comparisons confirmed a significant difference between the random gaze condition ($Mdn = 5$) and natural gaze condition ($Mdn = 6.0$, $p = .005$). Jonckheere-Terpstra tests confirmed a significant linear trend for both, behavioral realism ($TJT = 1212.50$, $z = -2.258$, $p < .024$) and behavioral naturalness ($TJT = 1138.00$, $z = -2.849$, $p < .004$). Interestingly, the hybrid condition was slightly inferior to the synthesized condition (non-significant).

4.7 Humanness and Eeriness

No significant effects were found for humanness or eeriness.

4.8 Resulting Participant Behavior

A one way ANOVA revealed a significant main effect for condition $F(3,86) = 3.44$, $p = .020$. Pairwise comparisons showed a significant difference of random gaze, with the longest overall dwell times, from natural gaze ($p = .047$), as well as of random gaze from random augmented gaze ($p = .044$) for the head AOI dwell times, see Fig. 3. No further significant effects were found.

5 Discussion

Our goal was to systematically investigating gaze models in avatar-mediated communication. To do so, we implemented a model based on earlier approaches to augment social gaze [15, 24] and compared four gaze conditions: a) natural gaze b) hybrid gaze, c) synthesized, and d) random gaze and measured the impact of these models on five minute long social interactions on perceived virtual rapport, social presence, interpersonal attraction, trust, behavioral realism and naturalness of same sex dyads. Furthermore, we evaluated the resulting visualized gaze behavior of each condition as well as the resulting gaze behavior of the participants.

Our results are in line with previous findings. Supporting our hypotheses, natural gaze was superior and random gaze was inferior to all other models, with regard to the subjective measures. The linear trends found for virtual rapport, interpersonal attraction, and trust indicate that natural gaze suffers from artificial manipulation whereas random gaze behavior constructed based on statistical distributions

Table 1: Subjective results.

	Natural gaze		Hybrid gaze		Synthesized gaze		Random gaze	
Measure	*Mdn*	*M* (*SD*)	*Mdn*	*M* (*SD*)	*Mdn*	*M* (*SD*)	*Mdn*	*M* (*SD*)
Virtual Rapport	5.73	5.72 (0.58)	5.64	5.60 (0.58)	5.64	5.38 (0.64)	5.14	5.05 (0.81)
Social Presence	3.83	3.75 (1.10)	3.75	3.65 (1.11)	3.67	3.65 (1.29)	3.17	3.13 (0.99)
Interpersonal Attraction	5.75	5.67 (0.50)	5.50	5.56 (0.65)	5.25	5.24 (0.56)	5.08	4.95 (0.80)
Trust	5.33	5.35 (0.63)	5.33	5.24 (0.64)	5.17	5.13 (0.73)	5.00	4.83 (0.74)
Gaze Behavior Realism	6.00	5.25 (1.29)	5.00	4.55 (1.60)	5.00	5.04 (1.46)	4.50	4.00 (1.49)
Gaze Behavior Naturalism	6.00	5.33 (1.20)	5.00	4.55 (1.68)	5.00	4.88 (1.48)	5.00	4.15 (1.18)
Humanness	3.08	2.97 (0.69)	2.67	2.64 (0.77)	2.83	2.81 (0.79)	2.75	2.61 (0.70)
Eeriness	2.75	2.71 (0.33)	2.88	2.72 (0.62)	2.81	2.70 (0.45)	2.75	2.73 (0.33)

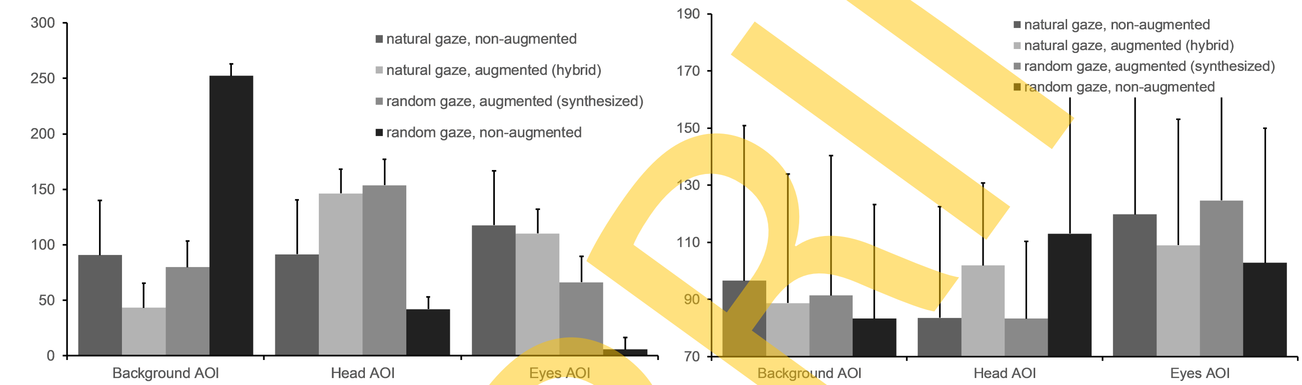

Figure 3: ***Left:*** Virtual character behavior. Mean dwell times evoked by the behavior ($N = 58$) displayed by the virtual character (respectively, each gaze model) in each condition. ***Right:*** Human behavior. Dwell times of the participants ($N = 90$) for each condition. Dwell times are displayed in seconds. Error bars denote standard deviations. Missing samples (human behavior) were discarded.

of saccades and fixations benefits from establishing adequate social gaze behavior. Exploratory ANOVAs conducted for these measures led to similar results. Therefore, we interpret that synthetic as well as hybrid gaze models that establish social gaze contingencies based on typical nonverbal patterns are superior to purely random models.

Some interpretations arise from these findings. First, it seems logical that the statistical distribution of resulting behavior using the synthesized (i.e. agent or random non-augmented) is imperfect, compared to the natural gaze. However, the results do indeed show that to some part, social gaze contingency can be established through a simple rule-based system, that is now quantified by the results of the study. Thus, supporting evidence is provided that displaying attention and interest (i.e. directing gaze to the interlocutor while listening) is an important social reaction.

Another interpretation of this finding is that the gaze model we used and the resulting behavioral blending and animation were not sophisticated enough to seamlessly blend into the "gold standard" natural gaze behavior. The results for the perceptual judgment of realism and naturalness interestingly give a hint that the hybrid gaze model, i.e. natural augmented gaze, led to a lower rating in realism and naturalness (non significant) compared to the pure natural gaze model as well as the sophisticated agent (i.e. synthesized) model. This could be a similar effect to the one mentioned in [36]: a consistency break in realism. However, the effect we found is limited to the behavioral level and might point at a problematic "break" in behavioral characteristics of gaze, i.e. the model that we applied to augment social gaze in the hybrid condition did not seamlessly blend into the natural gaze behavior. Further developments, not only for purely artificial gaze models but also for hybrid models, therefore should take consistency into account as an important factor. Furthermore, future developments for VR simulations of the model should also use the eye positions of the user as specific target point for induced eye gaze.

Results for the behavioral analyses indicate that participants in the random gaze condition tried to establish social gaze contacts via directed gaze, as shown by high dwell times in the head AOI. One interpretation of this is that participants tried to initiate social gaze contact (i.e. directed gaze) by reflex. As no adequate reaction resulted, this pattern was followed. Another way of interpreting this finding is that the participants tried to "interpret" sense into the random movements due to lack of variation, which could then have acted as a distractor and caused additional mental demand. However, we did not assess measures that could support this interpretation and the behavioral analyses do not provide a clear image due to a slightly reduced eyes AOI focus.

5.1 Limitations

Some limitations arise from our study. First, the sample was relatively small and populated with normal-developed students which does not allow to generalize findings. Second, we did not use virtual reality hardware, an immersive setup or a "fishtank VR" paradigm in our study. The findings should therefore be interpreted with care regarding more immersive systems as the mona lisa effect could have had an impact [30]. Third, our augmentation did only alter gaze direction, not the direction of the avatar's head, and did only aim to augment directed gaze and not eye contact. Fourth, we did not assess interactions with more than two users, such as e.g. Ding et al. [10], which provides room for further research. Fifth, we used cartoon-like characters that could have biased impression formation [36] which might change with more realistic characters.

6 Conclusion

We conclude that social gaze models that are based on available modalities could be beneficial for the development of social virtual environments, such as to cope with the lack of sensory inputs to track and replicate gaze behavior in order to substitute behavioral channels or to compensate for dropouts in data transmission. Future research should improve present approaches to social gaze models. Although benefits could arise from hybrid models, research should also focus on debating what potential ethical and security issues. While these technologies could help to research gaze behavior and cope with problematic situations of individuals with social disorders (e.g., a neutral job interview), risks might arise if such technologies continuously manipulate and learn an individual's behavior.

Acknowledgments

We thank the anonymous reviewers for their feedback and Roman Eyck for his support during the project.

References

[1] M. Argyle. *Bodily communication*. Methuen, London, 1984.

[2] J. N. Bailenson, A. C. Beall, J. Loomis, J. Blascovich, and M. Turk. Transformed Social Interaction: Decoupling Representation from Behavior and Form in Collaborative Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 13(4):428–441, Aug. 2004. doi: 10.1162/1054746041944803

[3] J. N. Bailenson, A. C. Beall, J. Loomis, J. Blascovich, and M. Turk. Transformed social interaction, augmented gaze, and social influence in immersive virtual environments. *Human communication research*, 31(4):511–537, 2005.

[4] J. N. Bailenson and J. Blascovich. Avatars. In *Encyclopedia of Human-Computer Interaction*. Berkshire Publishing Group, 2004.

[5] A. C. Beall, J. N. Bailenson, J. Blascovich, and C. Rex. Non-zero-sum mutual gaze in collaborative virtual environments. In *Proceedings of HCI International*, 2003.

[6] G. Bente, F. Eschenburg, and L. Aelker. Effects of simulated gaze on social presence, person perception and personality attribution in avatar-mediated communication. In *Presence 2007: Proceedings of the 10th Annual International Workshop on Presence, October 25-27, 2007, Barcelona, Spain*, pp. 207–14, 2007.

[7] N. Binetti, C. Harrison, A. Coutrot, A. Johnston, and I. Mareschal. Pupil dilation as an index of preferred mutual gaze duration. *Open Science*, 3(7):160086, July 2016. doi: 10.1098/rsos.160086

[8] J. K. Burgoon, L. K. Guerrero, and V. Manusov. Nonverbal signals. *The SAGE handbook of interpersonal communication. London: SAGE*, 2011.

[9] D. Davis and W. T. Perkowitz. Consequences of responsiveness in dyadic interaction: Effects of probability of response and proportion of content-related responses on interpersonal attraction. *Journal of Personality and Social Psychology*, 37(4):534–550, 1979. doi: 10.1037/0022-3514.37.4.534

[10] Y. Ding, Y. Zhang, M. Xiao, and Z. Deng. 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*, pp. 3011–3021. ACM, 2017.

[11] O. J. Dunn. Multiple comparisons using rank sums. *Technometrics*, 6(3):241–252, 1964.

[12] A. Field, J. Miles, and Z. Field. *Discovering statistics using R*. Sage publications, 2012.

[13] A. Frischen, A. P. Bayliss, and S. P. Tipper. Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, 133(4):694–724, 2007. doi: 10.1037/0033-2909.133.4.694

[14] M. Garau, M. Slater, S. Bee, and M. A. Sasse. The impact of eye gaze on communication using humanoid avatars. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 309–316. ACM, 2001.

[15] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M. A. Sasse. The Impact of Avatar Realism and Eye Gaze Control on Perceived Quality of Communication in a Shared Immersive Virtual Environment. 2003.

[16] A. L. Georgescu, B. Kuzmanovic, D. Roth, G. Bente, and K. Vogeley. The Use of Virtual Characters to Assess and Train Non-Verbal Communication in High-Functioning Autism. *Frontiers in Human Neuroscience*, 8, Oct. 2014. doi: 10.3389/fnhum.2014.00807

[17] M. Gerhard, D. Moore, and D. Hobbs. Embodiment and copresence in collaborative interfaces. *International Journal of Human-Computer Studies*, 61(4):453–480, 2004.

[18] J. Gratch, D. DeVault, G. M. Lucas, and S. Marsella. Negotiation as a challenge problem for virtual humans. In *International Conference on Intelligent Virtual Agents*, pp. 201–215. Springer, 2015.

[19] C.-C. Ho and K. F. MacDorman. Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, 26(6):1508–1518, Nov. 2010. doi: 10.1016/j.chb.2010.05.015

[20] M. A. Just and P. A. Carpenter. A theory of reading: From eye fixations to comprehension. *Psychological review*, 87(4):329, 1980.

[21] M. Kipp and P. Gebhard. Igaze: Studying reactive gaze behavior in semi-immersive human-avatar interactions. In *International Workshop on Intelligent Virtual Agents*, pp. 191–199. Springer, 2008.

[22] C. L. Kleinke. Gaze and eye contact: A research review. *Psychological Bulletin*, 100(1):78–100, 1986. doi: 10.1037/0033-2909.100.1.78

[23] B. H. Le, X. Ma, and Z. Deng. Live speech driven head-and-eye motion generators. *IEEE transactions on visualization and computer graphics*, 18(11):1902–1914, 2012.

[24] S. P. Lee, J. B. Badler, and N. I. Badler. Eyes alive. *ACM Transactions on Graphics*, 21(3), July 2002. doi: 10.1145/566654.566629

[25] X. Ma and Z. Deng. Natural eye motion synthesis by modeling gaze-head coupling. In *Virtual Reality Conference, 2009. VR 2009. IEEE*, pp. 143–150. IEEE, 2009.

[26] N. Murray, D. Roberts, A. Steed, P. Sharkey, P. Dickerson, and J. Rae. An assessment of eye-gaze potential within immersive virtual environments. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 3(4):8, 2007.

[27] A. Normoyle, J. B. Badler, T. Fan, N. I. Badler, V. J. Cassol, and S. R. Musse. Evaluating perceived trust from procedurally animated gaze. In *Proceedings of Motion on Games*, pp. 141–148. ACM, 2013.

[28] K. L. Nowak and F. Biocca. The Effect of the Agency and Anthropomorphism on Users' Sense of Telepresence, Copresence, and Social Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 12(5):481–494, Oct. 2003. doi: 10.1162/105474603322761289

[29] R. B. Queiroz, L. M. Barros, and S. R. Musse. Providing expressive gaze to virtual animated characters in interactive applications. *Computers in Entertainment (CIE)*, 6(3):41, 2008.

[30] S. Rogers, M. Lunsford, L. Strother, and M. Kubovy. The mona lisa effect: Perception of gaze direction in real and pictured faces. *Studies in Perception and Action VII*, pp. 19–24, 2003.

[31] D. Roth, C. Kleinbeck, T. Feigl, C. Mutschler, and M. E. Latoschik. Beyond replication: Augmenting social behaviors in multi-user virtual realities. In *Proceedings of the 25th IEEE Virtual Reality (VR) conference*, 2018.

[32] D. Roth, M. E. Latoschik, K. Vogeley, and G. Bente. Hybrid Avatar-Agent Technology A Conceptual Step Towards Mediated "Social" Virtual Reality and its Respective Challenges. *i-com*, 14(2), Jan. 2015. doi: 10.1515/icom-2015-0030

[33] D. Roth, K. Waldow, F. Stetter, G. Bente, M. E. Latoschik, and A. Fuhrmann. SIAM-C - A Socially Immersive Avatar Mediated Communication Platform. In *Proceedings of the International Conference on Artificial Reality and Telexistence / Eurographics Symposium on Virtual Environments*, 2016.

[34] K. Ruhland, S. Andrist, J. Badler, C. Peters, N. Badler, M. Gleicher, B. Mutlu, and R. Mcdonnell. Look me in the eyes: A survey of eye and gaze animation for virtual agents and artificial systems. In *Eurographics State-of-the-Art Report*, pp. 69–91, 2014.

[35] L. Tickle-Degnen and R. Rosenthal. The Nature of Rapport and Its Nonverbal Correlates. *Psychological Inquiry*, 1(4):285–293, Oct. 1990. doi: 10.1207/s15327965pli0104_1

[36] V. Vinayagamoorthy, M. Garau, A. Steed, and M. Slater. An Eye Gaze Model for Dyadic Interaction in an Immersive Virtual Environment: Practice and Experience. *Computer Graphics Forum*, 23(1):1–11, Mar. 2004. doi: 10.1111/j.1467-8659.2004.00001.x