Head-tracked off-axis perspective projection improves gaze readability of 3D virtual avatars

Tamas Bates
Honda Research Institute Europe
Offenbach, Germany
Technical University of Delft
Delft, The Netherlands
t.bates@tudelft.nl

Jens Kober
Technical University of Delft
Delft, The Netherlands
j.kober@tudelft.nl

Michael Gienger
Honda Research Institute Europe
Offenbach, Germany
michael.gienger@honda-ri.de

ABSTRACT

Virtual avatars have been employed in many contexts, from simple conversational agents to communicating the internal state and intentions of large robots when interacting with humans. Rarely, however, are they employed in scenarios which require non-verbal communication of spatial information or dynamic interaction from a variety of perspectives. When presented on a flat screen, many illusions and visual artifacts interfere with such applications, which leads to a strong preference for physically-actuated heads and faces.

By adjusting the perspective projection used to render 3D avatars to match a viewer's physical perspective, they could provide a useful middle ground between typical 2D/3D avatar representations, which are often ambiguous in their spatial relationships, and physically-actuated heads/faces, which can be difficult to construct or impractical to use in some environments. A user study was conducted to determine to what extent a head-tracked perspective projection scheme was able to mitigate the issues in readability of a 3D avatar's expression or gaze target compared to use of a standard perspective projection. To the authors' knowledge, this is the first user study to perform such a comparison, and the results show not only an overall improvement in viewers' accuracy when attempting to follow the avatar's gaze, but a reduction in spatial biases in predictions made from oblique viewing angles.

CCS CONCEPTS

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

KEYWORDS

Human-Computer Interaction, Virtual Reality, Augmented Reality, and Mixed Reality, Eye Gaze

ACM Reference Format:
Tamas Bates, Jens Kober, and Michael Gienger. 2018. Head-tracked off-axis perspective projection improves gaze readability of 3D virtual avatars. In *SIGGRAPH Asia 2018 Technical Briefs (SA '18 Technical Briefs), December 4–7, 2018, Tokyo, Japan.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3283254.3283271

SA '18 Technical Briefs , December 4–7, 2018, Tokyo, Japan

1 INTRODUCTION

Many robots in recent years have begun using displays mounted on the robot to present a virtual avatar to nearby observers [Kalegina et al. 2018]. The vast majority of such systems typically use a stylized 2D face (or even just eyes), although there are a few robots (such as the FURo-D[1]) currently employing 3D avatars. Virtual avatars can provide a cheap and effective means of providing emotional expression for a robot, or for displaying information about the robot's internal state. Due to many visual illusions which arise from displaying an avatar on a 2D screen it is difficult for such an avatar to give unambiguous cues or gestures to physical locations around it. In particular, eye contact and gazing behavior has been shown to be very important in a wide variety of human interaction scenarios [Kleinke 1986]. Without the ability to unambiguously convey these signals the usefulness of a virtual avatar will always be limited. As a result, physically-actuated heads are generally preferred for human-robot interaction, but depending on the intended environment these can be difficult or impractical to construct.

For purely social interactions during which it can be assumed that the user is standing directly in front of the screen, it may be sufficient simply to have an expressive avatar. Many practical applications, however, require that we be able to communicate specific, detailed, information, like the physical location of an object we need or the correct time to perform an action. To intuitively communicate such information, it must be possible to infer specific spatial relationships between the avatar, the environment, and the observer. When the avatar gazes at something or gestures towards it, human observers need to be able to quickly and intuitively understand which object the avatar is referring to (ideally with accuracy as close as possible to when observing human gazing and gestural behaviors). This is a difficult hurdle to overcome for anything projected onto a flat display, but by aligning the perspective used for projection with the observer's true physical viewpoint, as described by [Kooima 2008], it may be possible to arrive at a useful middle ground which does not require physically actuated facial features. This only requires tracking the viewer's head, and with recent developments in real-time human pose and face tracking [Zollhöfer et al. 2018] it would be easy to apply to almost any system.

2 RELATED WORK

Studies have shown that physical embodiment is generally preferred over the use of a virtual avatar [Li 2015]. However many small-scale features, like eyebrow movement or pupil dilation, are difficult and

[1] http://www.myfuro.com/furo-d/service-feature/

expensive to construct, and many environments in which we want to take advantage of human-robot interaction may not be suitable for complex or delicate actuators (an industrial factory, for example, may be hazardous for actuated eyes and ears). On the other hand, developing a virtual avatar which is capable of communicating effectively through gestures or eye gazing cues is also difficult. Some robots have presented avatars using back-projected masks [Kuratate et al. 2011], which has been shown to produce more effective and less ambiguous results compared avatars rendered on a 2D screen [Al Moubayed et al. 2012], particularly when viewed at an oblique angle. It has also been shown that when viewed from the correct position even photographs of humans can convey gaze targets nearly as accurately as physical humans [Bock et al. 2008] (but the readability of a photo's gaze target when viewed from an angle was not evaluated). By borrowing some ideas from the Virtual Reality field, it may be possible to render a 3D avatar on a screen with a perspective matching the viewer's physical viewpoint, which could allow for a reasonable level of practical usefulness in the avatar's gaze, gestures, etc.

3 3D AVATAR IMPLEMENTATION

While our robot (see Fig 1) is vaguely humanoid in shape and the motion planner produces minimum-jerk velocity profiles which resemble those of humans, it lacks a number of features (e.g. head, eyes, etc) which humans often rely on to communicate with each other when working together [Admoni and Scassellati 2017][Basili et al. 2012]. To help compensate for this, a 3D virtual avatar (shown in Fig. 1) was created to display on a screen attached to the robot. This provides an additional avenue for non-verbal communication, as long as the screen is in view and the content is clear.

In order to avoid uncanny valley issues, the avatar was designed not to have a humanoid appearance, but is rendered in a somewhat realistic manner as this has been shown to be important for the perception of a virtual avatar's behavior [Garau et al. 2003]. Most importantly, the avatar's eyes resemble human eyes in structure and move independently of the head. Head orientation alone has been shown to be insufficient for accurately conveying the target of a gaze [Kennedy et al. 2015], so when a new gaze target is selected the eyes focus first and the head follows. The eyes of the avatar were also exaggerated in size in order to make them easy to read.

Eye saccades were implemented in a simple (not biologically correct) manner. Eye movement follows a sinusoidal curve for its initial acceleration, then approximates a damped spring model for the latter part of the movement, which results in reasonably-natural-looking motion (no study participants complained). Minor secondary behavior, like blinking, was also implemented to increase user comfort (several users noted that even the deer avatar appeared "creepy" if it did not blink). The avatar blinks at random every 2-4 seconds, which is similar to human blinking behavior when engaged in active conversation [Bentivoglio et al. 1997].

In order to make the avatar's gaze as readable as possible, the user's head position is tracked using VICON markers attached to a hat and a generalized off-axis projection is used to render the avatar from the user's physical viewpoint, as described by Kooima [Kooima 2008]. This was motivated by earlier results which have shown that even 3D faces projected onto a flat viewing surface often produce ambiguous and unreliable gazes from an outside observer's perspective [Al Moubayed et al. 2012]. Projecting the view from the observer's actual viewpoint could reduce the ambiguity of the avatar's gaze when the user views the screen at an oblique angle, and easily allows the avatar to look directly at the user. Fig 2 shows an example of the off-axis projection vs. standard projection.

4 USER STUDY

A small user study (14 participants; 9 male, 5 female) was carried out to evaluate how well people could interpret the target of the avatar's gaze with a standard perspective projection and with an off-axis projection centered at the user's physical viewpoint. Our hypothesis was that the off-axis projection would improve readability of the avatar's gaze enough to allow an avatar to provide useful spatial cues to users through gazing and gestures. All participants had normal or corrected-to-normal vision.

4.1 Experimental Setup

A within-subjects experiment was carried out to determine the benefit (if any) of applying a head-tracked perspective projection to

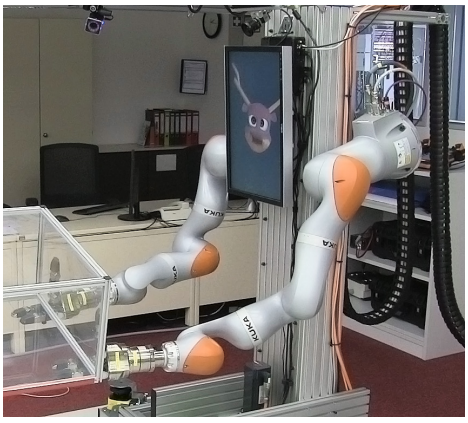

Figure 1: Virtual Avatar used to convey robot's internal state. 3D model adapted from CC0 'Stagley' mesh available at: https://www.blendswap.com/blends/view/4506

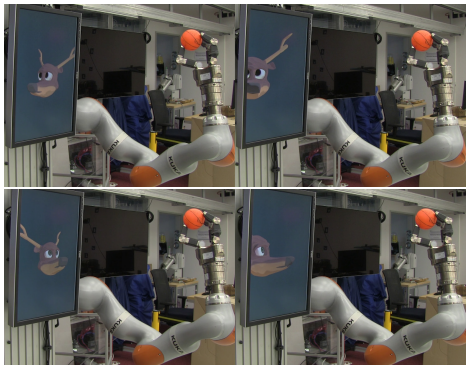

Figure 2: Avatar gazing at the camera (top) and a ball (bottom), without (left) and with (right) head-tracked perspective projection. View from ~45°angle.

the rendering of a 3D avatar vs. using a standard screen-centered projection. The goal was specifically to evaluate the readability of the avatar's gaze target when the screen is viewed from an oblique angle, and to compare this to the readability of the gaze target when a standard perspective projection is used.

The experimental setup is shown in Fig 3. A table with 19 numbered cups was placed between the participant and the screen. The cups were arranged in a hexagonal pattern so adjacent cups could be placed at a fixed distance from each other (20cm). The avatar was programmed to focus its gaze on different cups at regular intervals, and the participant was instructed to call out which cup they believed the avatar was gazing at. The avatar gazed at each cup for 7 seconds. In between gaze targets, the avatar looked at the participant for 3 seconds. This allowed a total time of 10 seconds within which the participant could give their answer. After 15 trials, the participant was asked to move to a new location and the process was repeated. The positions participants stood at were all 2m from the center of the screen, and varied in viewing angle from 0 to 45 degrees in 15-degree intervals. A short calibration round in which the avatar gazed only at the outer corner cups was given to each participant before the experiment began so they had a reference for the most extreme orientations of the avatar's head. Each participant repeated the experiment twice, first with a standard perspective projection, then with a head-tracked off-axis projection.

5 RESULTS

Tables 1 and 2 show simple accuracy scores for different viewing angles, averaged across all participants. **Correct** indicates the percentage of correct answers participants gave, while **N/A Answers** indicates the percentage of cases in which a participant was unable to give any answer (e.g. the avatar did not appear to be looking at the table at all, or the participant could not give an answer within 10 seconds). Even a small increase in the viewing angle leads to a large drop off in accuracy without applying the head-tracked projection, and, while accuracy still decreases with increasing angles when the head-tracked projection is applied participants were almost always certain that the avatar was at least looking at a cup on the table.

Fig 4 shows the average success rate for each cup from different viewing angles without applying the head-tracked projection. At all angles, cups closer to the screen exhibit higher success rates. This is quite intuitive since the closer a pair of targets are to the avatar, the greater the difference in the avatar's head angle when looking at them. At angles beyond 0° a very clear bias appears, showing that only cups on the same side of the screen as the user can be reliably guessed. Fig 5 shows the same data when the head-tracked projection is applied. The general pattern of cups closer to the screen being easier to guess is reproduced, but at higher angles there is much less bias in which cups can be accurately identified. To further explore this, the average distance between the correct answer and the guessed answer was plotted in figures 6 and 7, which again shows not just an overall increase in accuracy but a clear reduction in the bias of participants' answers.

Fig 8 shows the distribution of answers given (aggregated across all trials) when the avatar gazed at the center-most cup. Using a standard perspective projection, a very clear bias in the direction of the viewer's horizontal offset from the screen can be seen. This bias increases with the viewing angle and was observed across all cups on the table. When the head-tracked projection is used, this bias disappears. And while overall accuracy at larger viewing angles is reduced, the errors are still clustered around the correct answer, and in this case participants' answers were almost never more than one cup away from the correct answer.

Table 1: Standard Projection Results

View angle	Correct	Var	Std Dev.	N/A Answers
0°	38.57%	159.86	12.64	2.86%
15°	8.57%	40.82	6.39	6.19%
30°	6.67%	57.14	7.56	16.19%
45°	1.90%	21.77	4.67	31.90%

Table 2: Head-Tracked Projection Results

View angle	Correct	Var	Std Dev.	N/A Answers
0°	41.90%	243.99	15.62	0.48%
15°	32.86%	288.66	16.99	0.00%
30°	27.62%	176.87	13.30	0.48%
45°	18.57%	210.66	14.51	0.95%

6 CONCLUSION

Applying an off-axis perspective projection aligned to an observer's physical viewpoint when rendering a 3D avatar has a substantial

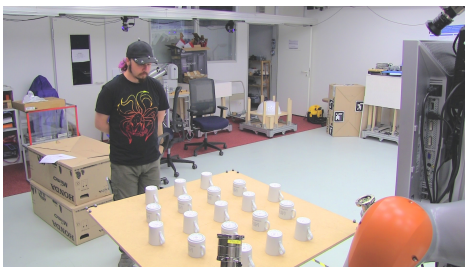

Figure 3: Experimental setup for evaluating readability of avatar gaze targets. The participant stands 2m from the screen at varying angles. Adjacent cups on the table are 20cm from center to center. All cups are 7.5cm in diameter.

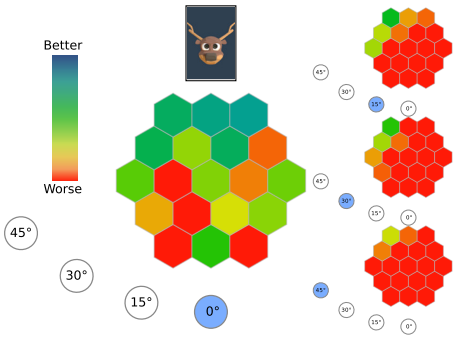

Figure 4: Average success rate for each cup from different viewing angles without head-tracked projection.

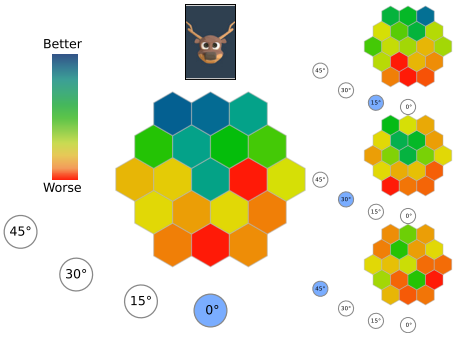

Figure 5: Average success rate for each cup from different viewing angles with head-tracked projection.

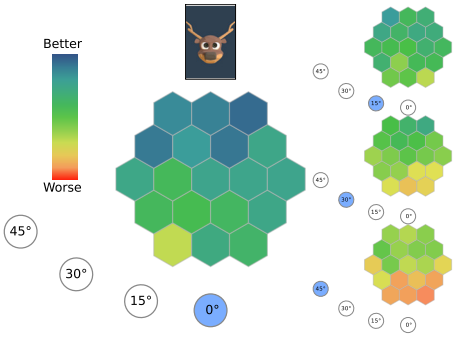

Figure 6: Average distance of participants' answers to ground truth without head-tracked projection.

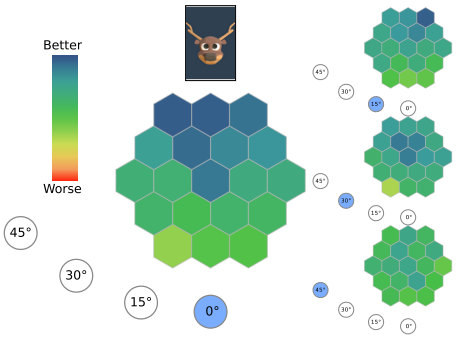

Figure 7: Average distance of participants' answers to ground truth with head-tracked projection.

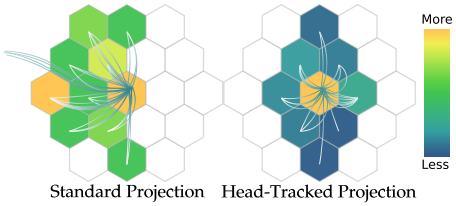

Figure 8: Distribution of answers given when the correct answer was Cup 10 (aggregated over all trials). Lines are drawn from the correct answer to the given answer; lines leaving the hex grid indicate trials without a given answer (e.g. participant thought avatar was not looking at any cup).

effect on the readability of the avatar's gaze when it is directed at physical objects outside of the screen. This will likely extend to any application in which a correct perspective is helpful or necessary, including gestures or pointing. Not only was a greater accuracy in predictions of the avatar's gaze target observed, also a distinct reduction in spatial biases of people's predictions. This implies that systems which employ gazing or gestural behaviors do not need to consider where an object is relative to the viewer to determine how ambiguous gazing or gesturing towards it will be. Rather, a simple relationship between the distance between the object and the avatar can be assumed, as the further an object is from the avatar the less the avatar's head/eyes will move to gaze at it relative to other objects at the same distance.

For any one-on-one interaction scenario, aligning the perspective projection of the avatar to the viewer's physical viewpoint could have a significant impact on usability and functionality. As many robots now include screens mounted on or near them, this could significantly reduce the cost and complexity of producing expressive features for robots and provides a solid argument in favor of developing more detailed 3D avatars rather than simple 2D avatars.

REFERENCES

Henny Admoni and Brian Scassellati. 2017. Social Eye Gaze in Human-Robot Interaction: A Review. *Journal of Human-Robot Interaction* 6, 1 (2017), 25–63.

Samer Al Moubayed, Jens Edlund, and Jonas Beskow. 2012. Taming Mona Lisa: communicating gaze faithfully in 2D and 3D facial projections. *ACM Transactions on Interactive Intelligent Systems* 1, 2 (2012), 25.

Patrizia Basili, Markus Huber, Omiros Kourakos, Tamara Lorenz, Thomas Brandt, Sandra Hirche, and Stefan Glasauer. 2012. Inferring the goal of an approaching agent: A human-robot study. In *IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 527–532. http://ieeexplore.ieee.org/document/6343805/

Anna Rita Bentivoglio, Susan B Bressman, Emanuele Cassetta, Donatella Carretta, Pietro Tonali, and Alberto Albanese. 1997. Analysis of blink rate patterns in normal subjects. *Movement Disorders* 12, 6 (1997), 1028–1034.

Simon W Bock, Peter Dicke, and Peter Thier. 2008. How precise is gaze following in humans? *Vision research* 48, 7 (2008), 946–957.

Maia Garau, Mel Slater, Vinoba Vinayagamoorthy, Andrea Brogni, Anthony Steed, and M Angela Sasse. 2003. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *SIGCHI conference on Human factors in computing systems*. 529–536.

Alisa Kalegina, Grace Schroeder, Aidan Allchin, Keara Berlin, and Maya Cakmak. 2018. Characterizing the Design Space of Rendered Robot Faces. In *ACM/IEEE International Conference on Human-Robot Interaction*. 96–104.

James Kennedy, Paul Baxter, and Tony Belpaeme. 2015. Head pose estimation is an inadequate replacement for eye gaze in child-robot interaction. In *ACM/IEEE international conference on human-robot interaction extended abstracts*. 35–36.

Chris L Kleinke. 1986. Gaze and eye contact: a research review. *Psychological bulletin* 100, 1 (1986), 78.

Robert Kooima. 2008. Generalized perspective projection. *School of Elect. Eng. and Computer Science* (2008), 1–7. http://csc.lsu.edu/~kooima/articles/genperspective/

Taakaki Kuratate, Yosuke Matsusaka, Brennand Pierce, and Gordon Cheng. 2011. "Mask-bot": A life-size robot head using talking head animation for human-robot communication. In *2011 11th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*. IEEE, 99–104.

Jamy Li. 2015. The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies* 77 (2015), 23–37.

Michael Zollhöfer, Justus Thies, Pablo Garrido, Derek Bradley, Thabo Beeler, Patrick Pérez, Marc Stamminger, Matthias Nießner, and Christian Theobalt. 2018. State of the Art on Monocular 3D Face Reconstruction, Tracking, and Applications. In *Computer Graphics Forum*, Vol. 37. 523–550.