2021 IEEE International Conference on Robotics and Automation (ICRA 2021)
May 31 - June 4, 2021, Xi'an, China

Watch Where You're Going! Gaze and Head Orientation as Predictors for Social Robot Navigation

Blake Holman[1], Abrar Anwar[1], Akash Singh[1], Mauricio Tec[2], Justin Hart[1], and Peter Stone[1,3]

2021 IEEE International Conference on Robotics and Automation (ICRA) | 978-1-7281-9077-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICRA48506.2021.9561286

***Abstract*— Mobile robots deployed in human-populated environments must be able to safely and comfortably navigate in close proximity to people. Head orientation and gaze are both mechanisms which help people to interpret where other people intend to walk, which in turn enables them to coordinate their movement. Head orientation has previously been leveraged to develop classifiers which are able to predict the goal of a person's walking motion. Gaze is believed to generally precede head orientation, with a person quickly moving their eyes to a target and then following it with a turn of their head. This study leverages state-of-the-art virtual reality technology to place participants into a simulated environment in which their gaze and motion can be observed. The results of this study indicate that position, velocity, head orientation, and gaze can all be used as predictive features of the goal of a person's walking motion. The results also indicate that gaze both precedes head orientation and can be used to predict the goal of a person's walking motion at a higher level of accuracy earlier in their walking trajectory. These findings can be leveraged in the design of social navigation systems for mobile robots.**

I. Introduction

Social navigation is the task of people and robots navigating in shared spaces [1]. As robotic systems move out of carefully-controlled environments such as factories and warehouses and into places designed for people, it is important to develop systems that can gracefully co-exist with people; including safely and comfortably navigating in human presence. People are able to infer each other's intended motion trajectories based on observing each other's gaze [2]. Gaze is also an important social cue involved in the coordination of passing behavior, when people walk past each other in areas such as hallways [3]. Unhelkar et al. [4] demonstrated head orientation to be predictive of human walking motion and developed a classifier that leverages head orientation to make discrete predictions of the target of a person's walking trajectory. Head orientation, however, is sometimes used as a proxy for gaze, due to the difficulty of measuring or expressing gaze in many scenarios [5], [6]. While gaze is difficult to detect on mobile robot platforms using current technology, deep learning methods for gaze detection that may be suitable to the task are currently under development [7]. In prior work, Hart et al. [3] have demonstrated that a robot using a gaze-based head-turn cue can signal its intention to a person passing it in a hallway, enabling the person to react to the cue and move out of the way. The present work takes an important step towards performing the inverse of that task — having the robot react to a person's gaze in order to get out of the way — by making predictions of human walking motions based on gaze.

Both head orientation and gaze increase in their predictive power of a person's motion trajectory as a person approaches the target of their walking motion. We claim that gaze orientation gives better predictions earlier in this trajectory than previously explored features such as position, velocity, and head orientation. While all of these approaches reach 100% accuracy by the time the person reaches their navigational goal, our results show that gaze achieves this accuracy earlier than competing measures. Using high-fidelity tracking in virtual reality, this paper presents a study based on Unhelkar et al. [4] to measure gaze orientation as a predictor of the target of a person's walking motion. Our tests confirm these claims using a linear mixed effects model in combination with a Tukey test. We also provide statistically significant experimental evidence supporting the generally-accepted, but unproven, hypothesis that gaze precedes head orientation. This study represents important progress towards the long-term goal of leveraging the predictive power of head and gaze orientation to develop systems wherein robots navigate comfortably and safely in close proximity to people in human-populated environments.

II. Related Work

Head orientation has been established as an anticipatory signal to the direction of movement of a pedestrian [2], [4], [8]–[11]. This work attempts to establish the predictive power of head orientation and gaze so they may be leveraged towards the ends of what has been called "social navigation" [3] or "socially-aware navigation" [1], in which humans and robots navigate in close proximity in human-populated spaces. Though the present work concentrates on gaze and head orientation, the predictive power of head orientation, body posture, and gaze has been investigated from a variety of perspectives [2], [8]–[11].

Patla et al. [8] provide a detailed account of how the body rotates toward a new motion direction. Their work indicates that foot placement and trunk rotation follow head motion in reorienting the body along a new path of motion. Prevost et al. [10] present a series of experiments in which they posit that head turns precede full body turns by 500ms. Individuals walking around corners systematically direct their gaze to the

Department of [1]Computer Science, [2]Statistics and Data Science, The University of Texas at Austin, 2317 Speedway, Stop D9500, Austin, Texas, 78712-1757. [3]Sony AI. {blake.holman, abraranwar, akashsingh, mauriciogtec}@utexas.edu, {hart, pstone}@cs.utexas.edu

end point of their future trajectory, and they accompany this movement with a head turn in the same direction [9]. There is an intuition behind the idea that gaze precedes turning one's walking stride in a new direction, in that pedestrians must look to the terrain in order to safely walk in the new direction. Another study by Grasso et al. [12], however, demonstrates that gaze predicts the direction of 90 degree turns even in the dark, implying that gaze changes are deeply rooted in human behavior beyond the conscious need to inspect the walking path for obstacles. Unhelkar et al. [4] leverage motion capture data from a Vicon motion capture system to measure head orientation to predict the targets of walking trajectories of study participants.

While the overall belief is that gaze anticipates head motion, current literature fails to provide definitive statistical evidence of the degree to which this relationship may hold [11]. Studies on head orientation often fail to take gaze into account, or they infer the orientation of the eyes by interpolating from head orientation; thus they cannot fully disentangle the effects of head and gaze orientation [12], [13]. The present work follows up on the work of Unhelkar et al. [4], performing a similar experiment, but separately testing eye gaze and head yaw as predictors of the target of a pedestrian's walking motion. Our results show that gaze orientation both precedes head orientation and achieves a desired level of goal prediction accuracy at a statistically significant level earlier.

Gaze has been studied heavily in the field of human-robot interaction as a communication modality that occurs implicitly [6]. That is to say, even when one is simply using their eyes to observe their environment, others can look to their eyes to observe where they are looking. In the context of social navigation, studies have leveraged gaze as an implicit communicative cue to convey where an agent intends to walk in an interaction. Nummenmaa et al. [2] present an interaction with a virtual agent in which the agent shifts its gaze toward the direction that it intends to walk, in a simulation of the agent walking towards the study participant. Participants decide which side they walk toward in order to to pass the agent, revealing that they understand the gaze cue as indicating where the agent intends to walk. In related experiments, Hart et al. [3] validate the importance of gaze in coordinating passing behavior in hallways. In a human field study, research confederates look either in the direction that they intend to walk or counter to the direction that they intend to walk, bumping into more people when looking counter to their walking direction [3]. In their follow-on human-robot interaction study, participants passing a robot in a hallway more often move to pass the robot on the side opposite to the direction of a robot's gaze when it uses a head-turn, improving performance over an LED turn signal [3]. Part of the inspiration for this study is the goal of leveraging a robot's interpretation of human gaze, head orientation, and other body language to coordinate its behavior when passing people in a hallway.

Studies have leveraged virtual reality with embedded gaze tracking to provide a more realistic experiment in crowd simulations, and to study the avoidance of perpendicular collisions [14], [15]. Gandrud and Interrante [16] classify the destination of a human in virtual reality-based on head orientation and gaze direction, considering a binary goal (the present work considers five goals). Zank and Kunz [17] present a series of experiments in a simulated multi-story building in which participants are instructed to reach the top floor contrasting gaze bias against approaches based on the participant's position, showing that gaze provides an earlier prediction. As virtual reality hardware becomes more commonplace, we expect a broad adoption of these types of studies.

III. Study Design

Recently, inexpensive, consumer-grade virtual reality systems have entered into the mainstream. To produce an appealing immersive visual experience, these systems feature accurate, high frequency, low-latency tracking systems. Kreylos [18] reports an expected tracking accuracy of around 2mm for the HTC Vive system. The HTC Vive Pro Eye [19] features an integrated Tobii [20] eye tracker. This work leverages this virtual reality hardware to simplify software development and data collection for an integrated gaze-tracking and motion capture study.

In this study, participants walk through a simulated room. A schematic drawing of the room is presented in Figure 1, and a screenshot from inside the virtual reality simulation is presented in Figure 2a. For each trial, participants are instructed to walk along a straight path towards position "A," a target 1m directly in front of their starting position. Upon reaching position "A," participants proceed to one of five goals placed 4m in front of their starting position, labeled 1–5, and placed 1m horizontally apart from each other. The purpose of navigating to position "A" before Goals 1–5 is to avoid conflating the effects of beginning to walk with the measured effects of the study. Participants choose their own path when walking toward the goal.

This study design closely mimics that presented by Unhelkar et al. [4], with a few notable exceptions. The participant navigates by walking through a virtual environment, rather than the real world. The instrumentation used in this study is an HTC Vive Pro Eye head-mounted display, a Vive Tracker, and a Vive Controller, rather than a Vicon [21] motion capture system. This study captures eye tracking data using a Tobii eye tracker built into the headset in addition to motion capture data, whereas Unhelkar et al. [4] leverage only motion capture data.

Each participant completes 5 trials walking to each goal (25 total) in a randomized order. At the start of each trial, the participant moves to the start position and initiates the trial by pulling the trigger of the Vive Controller. A pre-recorded voice dictates — "Go to Goal G" — where G denotes the goal that the participant should walk to. Similarly, upon reaching the goal and finishing the trial, the participant is asked to return to the start position by the same pre-recorded voice — "Go back to the start position."

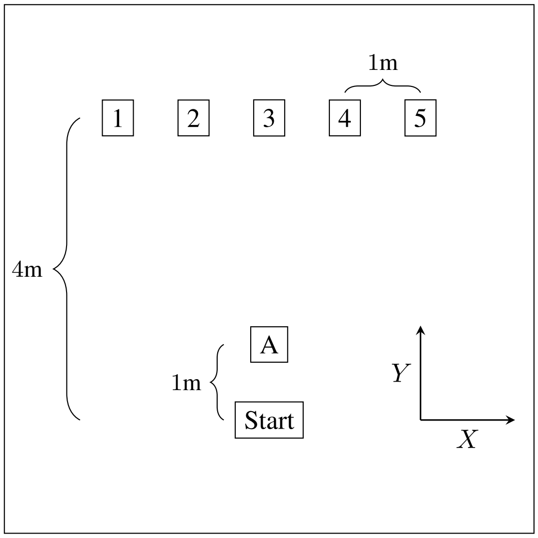


Fig. 1: Diagram of the scene in which the study takes place, with Goals 1–5. The participant begins at "Start" and passes through "A" before walking along their chosen path to the specified goal.

The virtual reality environment is implemented in Unity [22] version 2019.4.8f1 and utilizes the SteamVR [23] plugin to interface with the HTC Vive Pro Eye headset through an attached WiGig wireless kit. The virtual reality system samples position information at a rate of 60 Hz. The area through which the participant can interact inside the virtual reality environment requires external tracking and is referred to as the "play area." The tracked area in these experiments leverages two SteamVR Base Station 2.0's positioned approximately 7 meters diagonally from each other. Head orientation is measured using the tracking built into the virtual reality headset and gaze tracking data is sampled using the Tobii SDK. The participant's positional data is measured with respect to a Vive Tracker mounted to the participant's waist with a belt as seen in Figure 2b.

The experiment is laid out on a plane, in which the X axis spans horizontally across goals 1–5 whereas the Y axis spans vertically from the start position to the goals. The HTC Vive headset was used to determine the head yaw θ_{head}, gaze yaw θ_{gaze} and focus point of the participant (the point where the participant focuses their eyes). The x coordinate of the focus point is denoted γ_x. An example of a participant's head yaw and gaze yaw throughout the experiment is shown in Figures 4a and 4b, respectively. All of these paths are shown in Figure 3. Height interacts with walking speed, so the velocity v_x is approximated using the discrete forward-derivative of the x position and is normalized by the participant's height. Height is approximated by taking the average height of the Vive headset over each trial.

The collected data are used to analyze two hypotheses:

H1 Gaze tracking data provide better predictions earlier in a participant's walking trajectory than previously-demonstrated measurements, particularly head yaw.

H2 Gaze orientation precedes head orientation.

A total of 7 participants (6 male, 1 female) ranging in age from 19–31 (mean 22.7) participated in this study. All participants have normal or corrected vision. Due to the COVID-19 pandemic, regular Institutional Review Board (IRB)-regulated human studies were halted. Though naïve human study participants would be preferable, conducting this study during the pandemic required that we perform this study on ourselves. All participants in this study are members of our research lab who were acting in accordance with the COVID-19 restrictions in place at the University of Texas at Austin; four of whom are co-authors on this paper.

IV. Methodology

This study has two basic goals: to perform an analysis to detect the earliest point at which gaze, head orientation, and other factors are predictive of the goal of the participants' walking motion; and to develop a simple classifier to identify the target of the participant's walking motion.

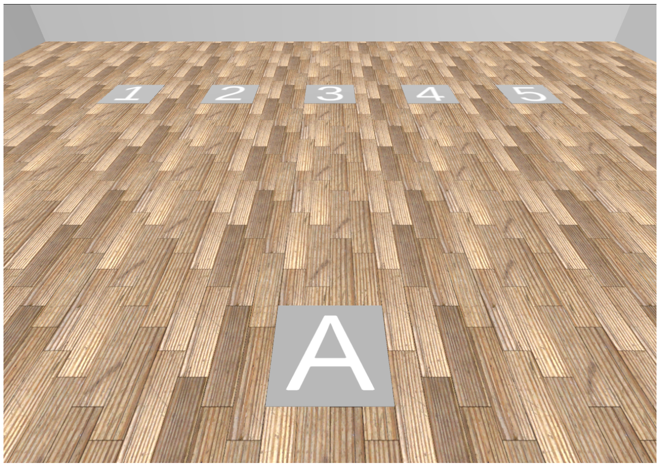
A. Repeated Measures ANOVA

A signal (x, v_x, θ_{head}, θ_{gaze}, or γ_x) is predictive of a participant's motion goal if its mean value corresponding to each goal is distinguishable before the participant's turning motion (x). This prediction can be measured with respect to the temporal domain or the spatial domain. The temporal domain consists of trajectories normalized by their total duration to represent percent completion. The spatial domain measures the distance from start using the Y coordinate, ranging from -0.5–3.5 meters.

To find the mean value for each signal conditioned on its goal, the statistical analysis fits a linear mixed effects model and contrast design at each of 10 discretized points in either the spatial or temporal domain. This analysis is adapted from [4]. More specifically, at each of one of the 10 splits resulting from discretization, the full procedure consists of the following:

- A Repeated Measures ANOVA, over each of the signals (x, v_x, θ_{head}, $\theta_{\text{gaze}}, \gamma_x$), consisting of a linear mixed effects model where the fixed-effect predictors are the intended goals and the random-effect predictors are participant indicators with participant-goal interactions.
- A contrast design and Tukey test for all possible pairwise comparisons of fixed-effect coefficients, adjusting for multiple testing at a family-wise error rate of 1%.
- A block bootstrap procedure which resamples the data 64 times stratified by participant and refits the linear mixed effects model to obtain standard deviations for the model coefficients.

The fixed effects of the model capture the mean value of each signal on a discretized range conditioned on its goal, controlling uncertainty estimation for the natural randomness among individuals. Tukey tests reveal the first time out of the 10 discretized points at which all possible comparisons of coefficients are significant, as well as the first time at which at least one significant difference emerges. Since



(a) Participants see this view of the virtual environment in which the study takes place. They are instructed to first walk to "A," then to proceed to one of Goals "1"–"5."

(b) Participants wear an HTC Vive Pro Eye head-mounted display with a WiGig wireless adapter and an HTC Vive Tracker on a belt mounted to the waist. They interact with controls in the experiment using the Vive Controller.

Fig. 2: View from inside and outside the virtual environment used in this experiment.

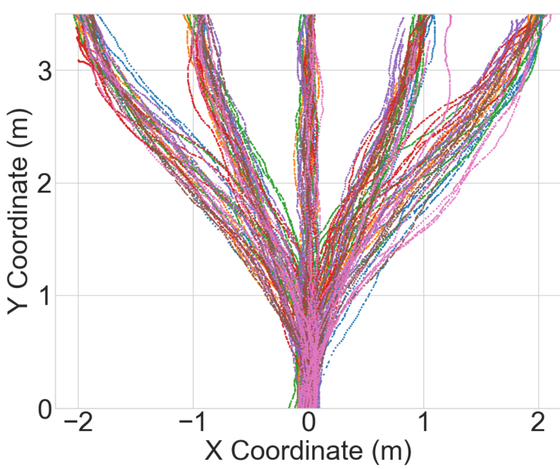


Fig. 3: Participants' trajectories to the goals. Each color represents a participant.

this is a discrete measurement the results are averaged over all bootstrap resamples, obtaining a continuous value. Bootstrapping also provides confidence bounds around the model coefficients.

B. Prediction Algorithm

A multivariate Gaussian time series prediction algorithm [4], [24] is used to predict the goal that each participant is heading toward. The results demonstrate that various combinations of input variables produce similar performance to Unhelkar et al. [4], and that gaze improves the predictive power the model.

Each of the variables in the problem formulation (x, y, v_x, θ_{head}, θ_{gaze}, and γ_x) are interpolated using a cubic spline, allowing their values to be sampled at uniform time-steps. At 60 Hz, each participant takes an average of 308 time steps (5.1 seconds) to reach the goal (from completion of the automated voice's playback). To account for the differences of trial duration, walking motion trajectories are normalized so that each trajectory is represented by 300 equally spaced points per trial. This differs from the implementation by Unhelkar et al. [4] which instead uses Dynamic Time-Warping [25].

The algorithm uses a vector of predictors Z_t at each timestep t modeled with a multivariate Gaussian distribution conditional on a goal G

$$Z_t \mid G = g \sim \mathcal{N}(\mu_t^g, \Sigma_t^g). \tag{1}$$

A trajectory with observed predictors $z_{1:t}$ from a trial is classified into a specific goal g by maximizing the *a posteriori* probability in Bayes' rule.

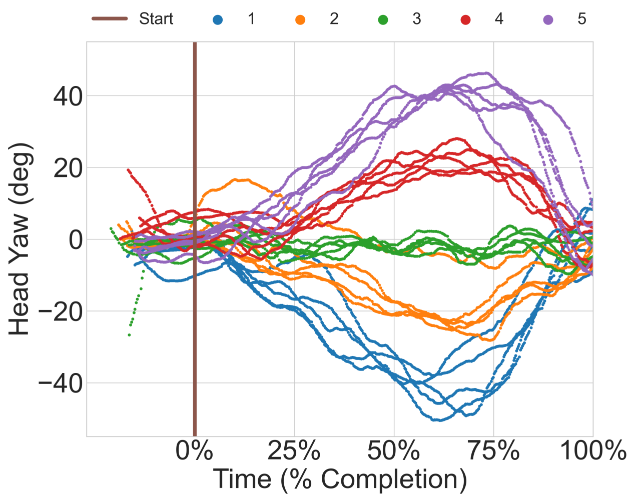
$$P(g \mid Z_{1:t} = z_{1:t}) \propto P(t) \cdot P(z_{1:t} \mid G = g), \tag{2}$$

$$P(z_{1:t} \mid G = g) = \prod_{i=1}^{t} \mathcal{N}(z_i \mid \mu_i^g, \Sigma_i^g). \tag{3}$$

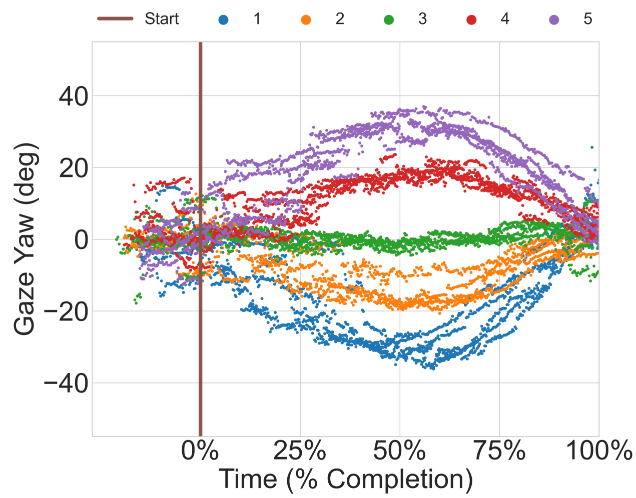
While the assumption of sequential independence is definitely not true, in practice, the likelihood term above is sufficient to build a good predictive model [4]. By choosing the goal with the highest likelihood, a reliable classification is made for the goal given a partial trajectory. Using this multivariate Gaussian formulation, it would also be easy to add new features to increase the predictive capability of the model.

V. Results

These results provide statistically significant evidence that gaze predicts — at the same level of accuracy as head orientation and earlier in a person's walking trajectory — the intended goal of a person's walking motion. These results also provide supporting evidence that gaze precedes head orientation.



(a) The variable θ_{head} versus percent completion over time.



(b) The variable θ_{gaze} versus percent completion over time.

Fig. 4: The variables θ_{gaze} and θ_{head} versus percent completion over time, with the colors representing the goals 1–5. The vertical line represents the moment in time when the recorded instructions finish dictating which goal to walk towards. The plots show trials for single participant, with 5 trials for each goal.

A. Results from the ANOVA

As can be seen in Figures 5a and 5b, in both the spatial and temporal domains, the signals v_x, θ_{head}, θ_{gaze} and γ_x all anticipate the physical turn x. It can be seen from the solid horizontal line (results of the Tukey test for the first moment at which the trajectories of all goals are statistically different) and the dashed line (the first emergence of a pairwise difference) that gaze discriminates between goals earlier than head orientation, and that θ_{gaze} is the earliest anticipatory signal.

All-pairwise differences for θ_{head} and θ_{gaze} emerge at 1.33m and 1.1m, respectively, in the spatial domain. This suggests an anticipation that is earlier by 0.22m when using θ_{gaze} versus θ_{head}. The analogous measurements are 4.75% versus 4.09% in the temporal domain. Equivalently, estimates using θ_{gaze} precede those of θ_{head} by 6.6% in the temporal domain.

The time of first emergence of differences occurs earlier in the temporal domain for θ_{gaze} relative to θ_{head}, but the difference is smaller in the spatial domain. This illustrates that participants first look at their target and then start moving, at which point they align their head yaw.

B. Results from the Predictive Model

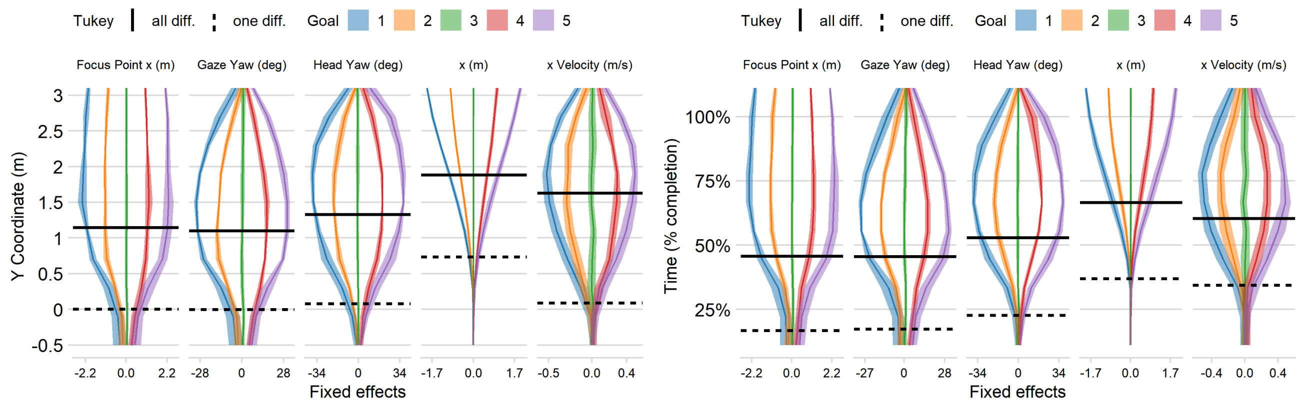
Due to the sparse number of participants, as a result of restrictions in place due to the COVID-19 pandemic, performance of the predictive model is estimated using leave-one-out cross-validation. These tests assess the robustness of gaze as a predictor of human motion.

Figure 6 shows that at $t = 0$, when using only the coordinate information or velocity added to it, the accuracy of the model is 20%; equivalent to random chance. By simply adding gaze yaw, performance increases to about 40% accuracy. Further, the gaze yaw model achieves 95% accuracy at 34% of trajectory completion, whereas θ_{head} achieves the same accuracy at 41.3% completion. The other predictors do so much later. The gain in anticipation from θ_{gaze} to θ_{head} is 7.13%; closely aligned to the observed gain of 6.6% using the Tukey test in the previous section. The uncertainty bounds in Figure 6 for θ_{gaze} and θ_{head} show little overlap near one third of completion. A paired two-sided t-test fixing time at one third to completion shows that θ_{gaze} and θ_{head} achieve $94.86\% \pm 1.67$ and $88.57\% \pm 2.4$ accuracy respectively, and this difference is statistically significant ($t = 2.716, p = 0.007$).

The presented results demonstrate that using θ_{gaze} as an anticipatory indicator of human motion increases the predictive power of the model. Compared to other parameters, gaze yaw predicts motion trajectories with higher accuracy on average at every point in time.

VI. Conclusion

As robots move into human-populated spaces, they need to be able to smoothly, safely, and naturally navigate in close proximity to people. This warrants an investigation into the anticipatory behaviors expressed by people, which implicitly communicate information that is used to successfully navigate such crowded environments. Prior work leveraging these anticipatory behaviors shows that gaze and head orientation are used to coordinate human motion [3], and various predictors have been built based on the premise that people will look into the direction that they are about to walk [4], [12], [17]. The virtual reality study presented in this work quantifies the anticipatory power of observing these behaviors. This study provides evidence that gaze orientation outperforms previously used features for anticipating the endpoint of human trajectories and supports the claim that



(a) Spatial domain

(b) Temporal domain

Fig. 5: Fixed-effect estimates of each signal (x, v_x, θ_{head}, θ_{gaze}, and γ_x) showing the average value conditioned on its goal. The horizontal solid line represents the time at which the effects of each goal are statistically different across all pairwise comparisons using the Tukey test at a 1% family-wise error rate adjusted for multiple comparisons. The horizontal dashed line shows the moment at which the first difference emerges. It can be seen that θ_{gaze} and γ_x predict all other measurements. The shaded regions represent two standard deviations obtained with bootstrap resampling.

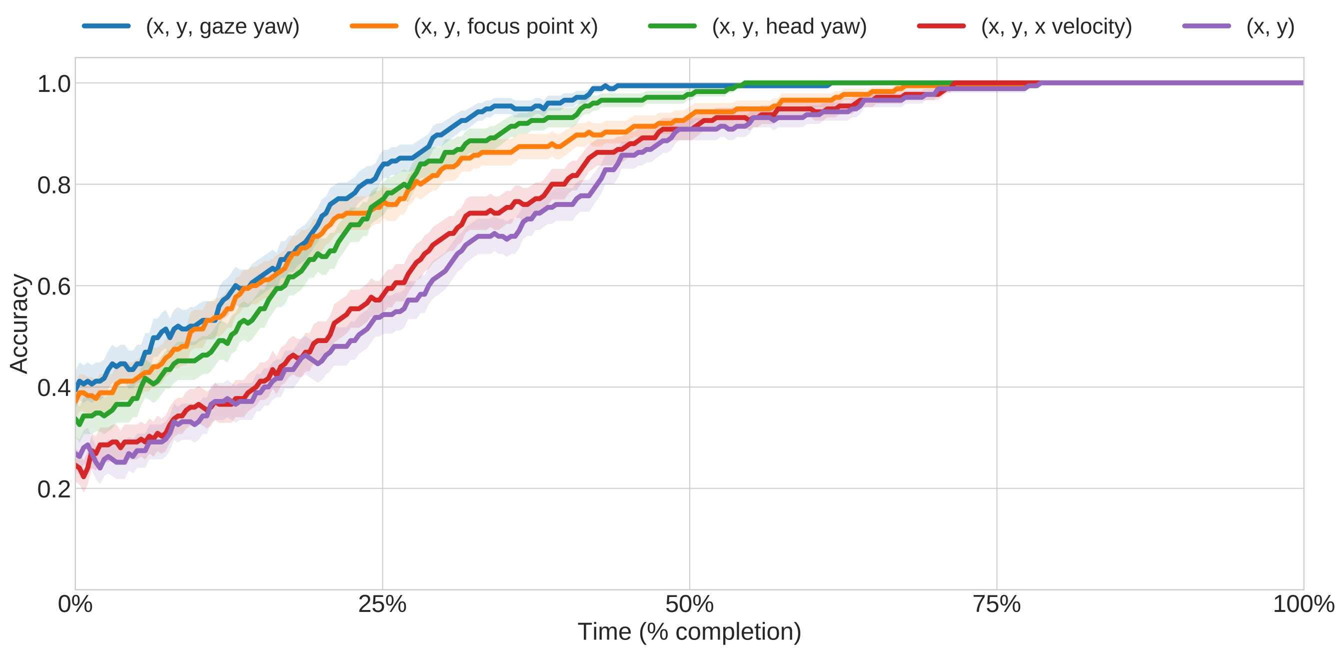


Fig. 6: Cross-validated accuracy of the multivariate Gaussian time series model over percent completion in time. Cross validation is computed with respect to a single participant over a model trained over all other participants, then computed as the mean when this procedure is repeated for all participants. The shaded region represents one standard deviation from the mean cross-validated accuracy.

gaze orientation precedes head orientation as an anticipatory predictor. These results can be leveraged in the construction of socially-aware navigational systems for mobile robots.

Acknowledgment

This work has taken place in the Learning Agents Research Group (LARG) at UT Austin. LARG research is supported in part by NSF (CPS-1739964, IIS-1724157, NRI-1925082), ONR (N00014-18-2243), FLI (RFP2-000), ARO (W911NF-19-2-0333), DARPA, Lockheed Martin, GM, and Bosch. Peter Stone serves as the Executive Director of Sony AI America and receives financial compensation for this work. The terms of this arrangement have been reviewed and approved by the University of Texas at Austin in accordance with its policy on objectivity in research.

References

[1] J. Rios-Martinez, A. Spalanzani, and C. Laugier, “From proxemics theory to socially-aware navigation: A survey,” *International Journal of Social Robotics*, vol. 7, no. 2, pp. 137–153, April 2015.

[2] L. Nummenmaa, J. Hyönä, and J. K. Hietanen, “I’ll walk this way: Eyes reveal the direction of locomotion and make passersby look and go the other way,” *Psychological Science*, vol. 20, no. 12, pp. 1454–1458, December 2009.

[3] J. Hart, R. Mirsky, X. Xiao, S. Tejeda, B. Mahajan, J. Goo, K. Baldauf, S. Owen, and P. Stone, “Using human-inspired signals to disambiguate navigational intentions,” in *Proceedings of the International Conference on Social Robotics (ICSR)*, Nanjing, China, August 2020, pp. 320–331.

[4] V. V. Unhelkar, C. Pérez-D’Arpino, L. Stirling, and J. A. Shah, “Human-robot co-navigation using anticipatory indicators of human walking motion,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Seattle, Washington, USA, May 2015, pp. 6183–6190.

[5] M. Slaney, A. Stolcke, and D. Hakkani-Tür, “The relation of eye gaze and face pose: Potential impact on speech recognition,” in *Proceedings of the International Conference on Multimodal Interaction (ICMI)*, Istanbul, Turkey, November 2014, p. 144–147.

[6] H. Admoni and B. Scassellati, “Social eye gaze in human-robot interaction: A review,” *Journal of Human-Robot Interaction*, vol. 6, no. 1, p. 25–63, May 2017.

[7] A. Saran, S. Majumdar, E. S. Short, A. Thomaz, and S. Niekum, “Human gaze following for human-robot interaction,” in *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)*, Madrid, Spain, October 2018, pp. 8615–8621.

[8] A. E. Patla and J. N. Vickers, “Where and when do we look as we approach and step over an obstacle in the travel path?” *Neuroreport*, vol. 8, no. 17, pp. 3661–3665, May 1997.

[9] M. A. Hollands, A. E. Patla, and J. N. Vickers, ““Look where you’re going!”: Gaze behaviour associated with maintaining and changing the direction of locomotion,” *Experimental Brain Research*, vol. 143, no. 2, pp. 221–230, March 2002.

[10] P. Prévost, Y. Ivanenko, R. Grasso, A. Berthoz, I. Yuri, G. Renato, and B. Alain, “Spatial invariance in anticipatory orienting behaviour during human navigation,” *Neuroscience Letters*, vol. 339, no. 3, pp. 243–7, March 2003.

[11] A. M. López, J. C. Alvarez, and D. Álvarez, “Walking turn prediction from upper body kinematics: A systematic review with implications for human-robot interaction,” *Applied Sciences*, vol. 9, no. 3, pp. 361–375, February 2019.

[12] R. Grasso, P. Prévost, Y. P. Ivanenko, and A. Berthoz, “Eye-head coordination for the steering of locomotion in humans: An anticipatory synergy,” *Neuroscience Letters*, vol. 253, no. 2, pp. 115–118, September 1998.

[13] T. Pozzo, Y. Levik, and A. Berthoz, “Head and trunk movements in the frontal plane during complex dynamic equilibrium tasks in humans,” *Experimental Brain Research*, vol. 106, pp. 327–38, February 1995.

[14] F. Berton, A.-H. Olivier, J. Bruneau, L. Hoyet, and J. Pettré, “Studying gaze behaviour during collision avoidance with a virtual walker: Influence of the virtual reality setup,” in *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, Osaka, Japan, March 2019, pp. 717–725.

[15] F. Berton, L. Hoyet, A.-H. Olivier, J. Bruneau, O. Le Meur, and J. Pettré, “Eye-gaze activity in crowds: Impact of virtual reality and density,” in *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, Atlanta, Georgia, USA, March 2020, pp. 322–331.

[16] J. Gandrud and V. Interrante, “Predicting destination using head orientation and gaze direction during locomotion in VR,” in *Proceedings of the ACM Symposium on Applied Perception (SAP)*, Anaheim, California, USA, July 2016, p. 31–38.

[17] M. Zank and A. Kunz, “Eye tracking for locomotion prediction in redirected walking,” in *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI)*, Greenville, South Carolina, USA, March 2016, pp. 49–58.

[18] O. Kreylos, *Lighthouse tracking examined*, May 25, 2016 (accessed October 30, 2020). [Online]. Available: http://doc-ok.org/?p=1478

[19] “Vive Pro Eye Series,” https://www.vive.com/eu/product/vive-pro-eye/overview/, Accessed: 2020-10-31.

[20] “Tobii,” https://www.tobii.com/, Accessed: 2020-10-31.

[21] “Vicon | Award winning motion capture systems,” https://www.vicon.com/, Accessed: 2020-10-31.

[22] “Unity real-time development platform,” https://www.unity.com/, accessed: 2020-10-31.

[23] “SteamVR - Valve Corporation,” https://www.steamvr.com/en/, Accessed: 2020-10-31.

[24] C. Pérez-D’Arpino and J. A. Shah, “Fast target prediction of human reaching motion for cooperative human-robot manipulation tasks using time series classification,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Seattle, Washington, USA, May 2015, pp. 6175–6182.

[25] J. Kruskal and M. Liberman, “The symmetric time-warping problem: From continuous to discrete,” in *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*, D. Sankoff, Ed. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc., 1983, ch. 4, pp. 125–161.