In *Proceedings of the 2018 IEEE International Conference: Robot and Human Interactive Communication* (RO-MAN 2018),
Nanjing, China, August 2018

A Study of Human-robot Copilot Systems for En-route Destination Changing

Yu-Sian Jiang[1], Garrett Warnell[2], Eduardo Munera[3], and Peter Stone[4]

***Abstract*— In this paper, we introduce the problem of en-route destination changing for a self-driving car, and we study the effectiveness of human-robot *copilot* systems as a solution. The copilot system is one in which the autonomous vehicle not only handles low-level vehicle control, but also continually monitors the intent of the human passenger in order to respond to dynamic changes in desired destination. We specifically consider a vehicle parking task, where the vehicle must respond to the user's intent to drive to and park next to a particular roadside sign board, and we study a copilot system that detects the passenger's intended destination based on gaze. We conduct a human study to investigate, in the context of our parking task, *(a)* if there is benefit in using a copilot system over manual driving, and *(b)* if copilot systems that use eye tracking to detect the intended destination have any benefit compared to those that use a more traditional, keyboard-based system. We find that the answers to both of these questions are affirmative: our copilot systems can complete the autonomous parking task more efficiently than human drivers can, and our copilot system that utilizes gaze information enjoys an increased success rate over one that utilizes typed input.**

I. Introduction

There has recently been a great deal of interest in developing systems that are able to automate most automobile driving tasks. In many cases, we seek to replace low-level human control of the vehicle with control provided by autonomous, or self-driving, systems. These autonomous vehicles, which would operate according to a common set of traffic and safety rules, are expected to provide safer transportation, reduce driver stress, and improve their riders' productivity.

A popular vision for self-driving cars is that of a passenger pre-specifying a destination and the car autonomously maneuvering to that destination. Under this paradigm, before the system may begin its maneuver, the autonomous agent needs to know the destination of its passenger as communicated using, e.g., a keyboard or spoken command. Given the specified destination, the agent may then begin autonomous driving.

Consider, however, what might happen if the passenger wishes to modify the destination during their trip (e.g., deciding to stop for lunch when the passenger happens to notice a restaurant through the vehicle's window). The passenger would need to either respecify the destination using the procedure described above, or disengage the autonomous driving agent to take over steering and manually drive there. Assuming we wish to avoid requiring the user to drive manually, if the system is not explicitly designed to accommodate this scenario, destination respecification may prove to be too difficult; if the human is not able to quickly instruct the vehicle, it may end up passing by the desired destination.

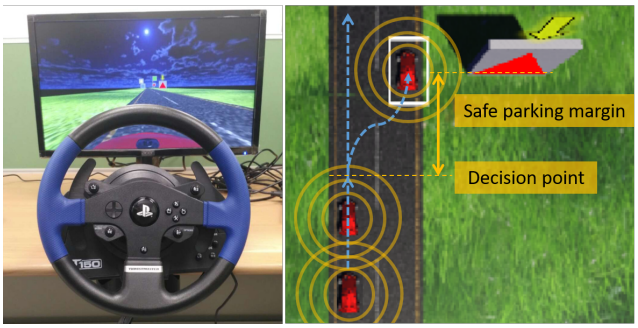

Fig. 1: **Depiction of autonomous parking task.** *Left:* driving simulator and controls available to users who performed manual driving. *Right:* top-down illustration of autonomous parking task for users who utilized a copilot system. The car proceeds straight down the road at a fixed speed and attempts to infer the user's intended destination at pre-specified decision points. As the car crosses the decision point, the system attempts to infer whether or not the user's intended destination is the parking spot corresponding to that decision point. If it is, the car autonomously parks there.

In this paper, we explicitly consider the above problem, i.e., *en-route destination changing*. Addressing en-route destination changing requires new techniques that can enable a higher speed of communication between the human and the vehicle. Here, we propose and study one candidate technique: a human-robot *copilot* system that not only handles low-level vehicle control, but also continually monitors the passenger in order to infer and respond to changes in destination in a timely fashion. We are particularly interested in whether or not the destination inference can be performed using information obtained by monitoring the passenger's eye gaze. Our interest is inspired by neurophsychology studies that have suggested that, by observing a partner's gaze, humans can infer their partner's intention or goal towards a particular

[1]Yu-Sian Jiang is with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712, USA sharonjiang@utexas.edu
[2]Garrett Warnell is with the Computational and Information Sciences Directorate, US Army Research Laboratory, Adelphi, MD 20783, USA garrett.a.warnell.civ@mail.mil
[3]Eduardo Munera is with Mindtronic AI, Taipei, Taiwan eduardo.munera@mindtronicai.com
[4]Peter Stone is on the Faculty of the Department of Computer Science, University of Texas at Austin, Austin, TX 78712, USA pstone@cs.utexas.edu

object [1]. Indeed, several examples in the literature (e.g., [2], [3], [4], [5], [6], [7]) have demonstrated that an autonomous agent utilizing human gaze cues can better interpret the human's intent and thus make for a better partner. Here, we posit that this information can also be useful in addressing en-route destination changing.

We specifically consider a simulated parking task (illustrated in Figure 1) in which en-route destination changes correspond to having the vehicle park next to a particular roadside sign board. In the context of this task, we conduct a human study and investigate the following questions:

1) *Is there benefit in using a copilot system over manual driving?*
2) *For a copilot system, is there any benefit in using gaze tracking to detect a destination change versus a system that instead uses a more traditional, keyboard-based method of input?*

We study these questions by having humans use a manual driving system and two proposed copilot systems. Our first copilot system, *CopilotKey*, is a more traditional, keyboard-based system for detecting a destination change in which the system gets the user's intended destination from the keyboard and plans a new path upon receiving a new destination. The second copilot system, *CopilotGaze*, is a gaze-based system for detecting destination changes in which the system utilizes the recently-proposed dynamic interest point detection (DIPD) technique [8] to recognize the user's intended destination based on gaze information. Once the intended destination has been inferred, both copilot systems plan a new path to the new location.

By analyzing the results of our user study, we find that there is benefit to using the proposed copilot methods over manual driving: the time to arrive at the intended destination is deterministic, and the vehicle speed undergoes less fluctuation than in a manual driving counterpart. We also find that there is benefit to using gaze information in our copilot system compared to one that uses the keyboard: participants using *CopilotGaze* tend to be more successful at completing the parking task than those using *CopilotKey*, especially when the vehicle is driving at a high speed.

The rest of this paper is organized as follows. In Section II, we review previous works that relates to shared autonomy and intention inference. Then, we describe our copilot system in more detail in Section III, followed by the setup and results for user experiments as well as the analysis of the experiment data in Section IV. Finally, we discuss our observations and conclude our work in Section V and Section VI, respectively.

II. Related Work

Autonomous en-route destination changing is related to work in both *shared autonomy* and *intention inference*. Our problem is related to work in shared autonomy because it involves balancing the autonomous system's low-level control of the vehicle with the human's high-level control of which destination the vehicle should go towards. The problem is related to work in intent inference because it requires the autonomous system to actively monitor cues from the human and use these cues in order to infer the human's intended destination. In this section, we provide a review of the literature in each of these areas. Importantly, to the best of our knowledge, our work is one of the first that considers these problems in the context of autonomous vehicles; we are aware of only one other work which proposes the idea of using an autonomous co-driver with a behavioral architecture that enables co-driving [9].

A. Shared Autonomy

Work in shared autonomy focuses on designing techniques that allow humans and autonomous systems to collaborate in order to achieve a goal. One example where such systems are desirable is robot teleoperation: requiring a human to control every detail of the robot's motion can be tedious. Moreover, the user's control may be noisy, or the user may not have enough information to determine how the robot should be controlled in order to achieve their goal. Work in shared autonomy aims to address these issues by allowing the user to cooperate with an autonomous agent so that the overall goals can be more easily achieved.

Much of the existing work in shared autonomy uses a predict-then-blend approach [10], [11], [12], [13]. Predict-then-blend approaches can be viewed as ones that arbitrate between the user's policy and a fully-autonomous policy [13], and the hope is that resulting policy helps the user to achieve the goal more efficiently. These approaches first predict the most likely goal from the user's input, use this goal to build a predicted user policy, and then blend this policy with a fully-autonomous one. The policy blending is done according to an arbitration function which is typically a weighted summation, though selecting the specific function often requires user studies.

Another approach described in the shared autonomy literature is to formulate the problem as one of optimizing a partially observable Markov decision process (POMDP) over the user's goal [14], [15], [16], [17]. These approaches assume that the user's behavior is approximately optimal in the context of a certain Markov decision process (MDP) that contains information about their intended goal. The shared autonomy agent, which does not know the intended goal, models the system as a POMDP that includes observations of the human, and uses techniques from this field in order to try to execute optimal assistive behaviors.

The approach we study in this paper is closely related to predict-then-blend approaches. In the copilot systems we design, the autonomous agent seeks to infer the intended goal of the user and effectively blends this information with its current policy as a means by which to steer the vehicle toward the intended goal.

B. Intention Inference

Work in the area of human intention inference focuses on using state and/or action observations of the human in order to reason about the human's plans and goals [18], [19], [20], [21], [22]. In contrast to human intention inference, another kind of work studies how to present robot intentions

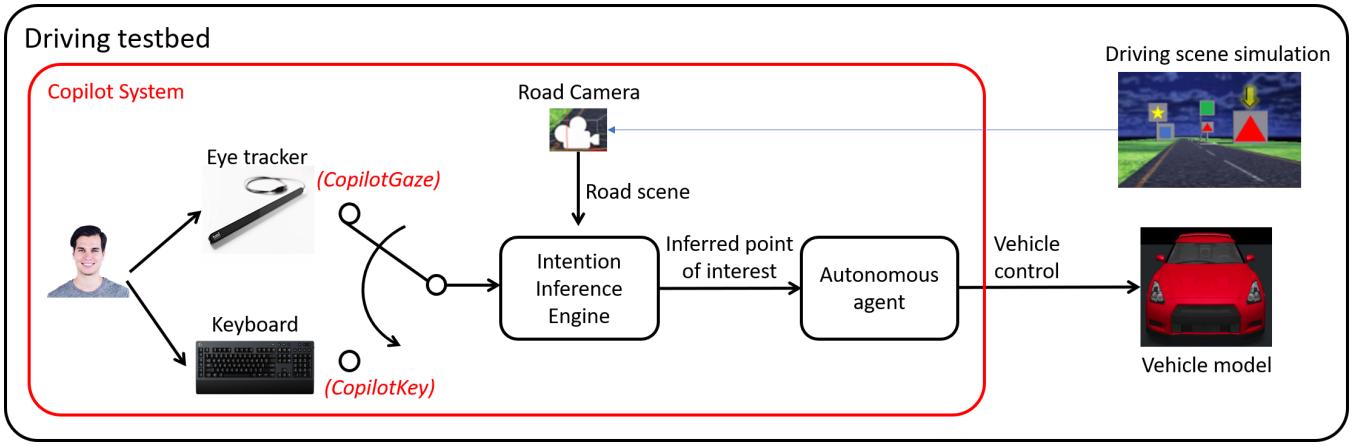

Fig. 2: **A diagram of a copilot system in a driving testbed.** *CopilotKey* interfaces with the user via a keyboard, while *CopilotGaze* monitors the user using an eye tracker. Each copilot system attempts to infer the user's intended destination from the corresponding input, and then the autonomous agent modifies the low-level vehicle behavior in response.

to humans (e.g., [23]). Our work is more related to techniques about human intention inference, i.e., those that seek to provide autonomous systems with the ability to determine the intent of humans. Since the techniques we study here utilize bio-sensing data to infer a human's intended destination, we now review the literature related to bio-sensing data-driven approache s.

Humans' physical status (e.g., pose, action, and other physiological signals) and their interaction with the surrounding environment can sometimes reveal their intent. Therefore, intention inference can be partially achieved by analyzing one or more of these physical statuses. For example, plan and intention information may be inferred from human speech [24], [25]. Some work has shown that modeling the relationship between human poses and objects in an image can be used to infer the person's next activity [26], [27]. In a driving application, head motion has been used as an important cue for predicting a driver's intent to change lanes [28]. Further, employing multi-modal data including GPS, speed, street maps, and driver's head movement can allow an ADAS (advanced driver assistance system) to anticipate the driver's future maneuvers [29].

Gaze cues, which implicitly include head pose information, can be particularly useful when attempting to infer human intent as it pertains to finer-grained points of interest (e.g., shop signs far away from a driver). In human-agent collaboration tasks, it is sometimes assumed that the human intends to interact with particular objects on the table [30]. In this paper, we assume that the human's intent is associated with a particular spatial location and, therefore, that the intent inference problem is simply that of inferring that location. With this goal in mind, a deep learning based method was recently proposed that combines gaze and saliency maps in order to form a predicted gaze direction [31]. The method was shown to be useful in both surveillance and human-robot teaming as a means by which to understand a person's intention from a third party perspective. In cases where the person's face and gaze targets were captured by different cameras, one needs to correlate the gaze tracking data from the face camera with the objects from the scene camera. Prior work on DAS has shown how to correlate a diver's gaze with road signs in the environment [2]. The system calculates the disparity between the scene camera and gaze angles for the sign, and then uses this disparity to determine whether or not the driver sees the road sign. Another approach is to divide the scene into several regions and train a classifier on a dataset which contains the face images with annotated regions to predict the region of user attention. For example, nine gaze zones in the vehicle such as driver's front, rear view mirror, passenger's front, etc., were defined and a CNN classifier was trained to categorize the face images into the predefined fixed nine gaze zones so as to recognize the point of driver's attention [3]. In other application areas such as hand-eye coordination tasks and player-adaptive digital games, machine learning-based methods (e.g., SVM, kNN, LSTM, ...) have been shown to be effective in predicting user intent from gaze observations [6], [7].

Because the en-route destination changing problem happens in a highly dynamic environment, the methods we study infer intent either through keyboard input or through gaze input via the recently-proposed Dynamic Interest Point Detection (DIPD) technique [8]. DIPD is a technique for inferring the point of interest corresponding to the human's intent from eye-tracking data and an environment video. The technique correlates the scene shown in the environment video with the human's gaze point to infer the human's point of interest and deals with various sources of errors such as eye blinks, high-speed tracking misalignment, and shaking video content. These advantages make DIPD useful for vehicle applications and we use it for our *CopilotGaze* copilot system.

III. Human-Robot Copilot System

As a means by which to study the questions set out in Section I, we propose here a simple human-robot copilot system that uses human intent recognition to influence the

behavior of the autonomous agent. Figure 2 shows the block diagram of the proposed system. Our copilot system includes both an intention inference engine and an autonomous agent. The goal of the intention inference engine is to detect the user's intended destination in the context of the surrounding environment. Based on that detection, the autonomous agent then performs path-planning and vehicle control in order to move toward a new destination.

For our study, we develop two kinds of copilot systems. The first is *CopilotGaze*, where the user's intended destination is inferred based on gaze observations obtained by an eye tracker. The second is *CopilotKey*, where the user's intention is obtained from keyboard input. For *CopilotGaze*, the intention inference engine utilizes the Dynamic Interest Point Detection (DIPD) algorithm [8] in order to infer the point of interest corresponding to the user's intended destination using gaze tracking data and a dynamic Markov Random Field (MRF) model. For *CopilotKey*, the intention inference engine is simply a text-matching procedure that assumes the intended destination is the closest sign board with the same color and shape that was indicated using the keyboard. In both systems, while the car is autonomously following a predefined path, the intent inference engines will report one (or none) of the sign boards in the driving scene as the user's intended destination for each simulation time step. [1]

Both of the proposed copilot systems utilize the same autonomous agent for vehicle control. First, we define a safe parking margin ahead of each parking spot, and we call the rear boundary of this margin the *decision point*. There is a unique decision point for every possible parking spot. When the car reaches a decision point, the copilot system needs to decide whether to autonomously park the car in the corresponding parking spot or not. If the inferred point of interest at a decision point is the same as the corresponding sign board, the autonomous agent slows the vehicle and follows a pre-defined feasible trajectory to park the car in the parking spot. Otherwise, the autonomous agent continues forward. The procedure is shown in Algorithm 1 and illustrated on the right side of Figure 1. The safe parking margin (and hence the location of the decision point) should be a function of the vehicle speed, car model, and horizontal distance to the parking spot, not a constant value.

Algorithm 1 Autonomous parking procedure in the proposed human-robot copilot system.

```
1:  procedure CopilotPark
2:      isParking = false
3:      inferredPOI = null
4:      parkGoal = null
5:      brakeDistance = safe parking margin
6:      while autonomous parking is enabled do
7:          carPosition = current car position
8:          inferredPOI = DIPD(gaze point, scene)
9:          if not isParking then
10:             if distance(inferredPOI, carPosition) > brakeDistance then
11:                 parkGoal = inferredPOI
12:             if parkGoal != null and distance(parkGoal's decision point, carPosition) <= brakeDistance then
13:                 isParking = true
14:                 Park the car based on a pre-defined trajectory while slowing down the car.
15:             else
16:                 Continue forward.
```

IV. Experiments

In this section, we describe the user study that we performed in order to answer the questions stated in Section I. That is, we seek to answer experimentally whether there is benefit to the proposed copilot system versus manual driving, and whether inference from gaze helps our copilot system. To do this, we recruited 15 human participants living in Taiwan between the ages of 23 and 46 comprised of 6 females and 9 males. Each participant had a driver's license, and five of them had no or very little scientific background. Participants were asked to test a manual driving system, the *CopilotKey* system, and the *CopilotGaze* system (as shown in Fig. 3) using a driving simulator. The vehicle information and user input data recorded by the driving simulator in each task are used to evaluate the manual driving system and the copilot systems in terms of vehicle trajectory, task completion time (i.e. arrival time to the parking destination), number of user actions, and success rate of completing the parking task.

A. Experiment Setup

In order to compare the performance of the three systems studied in our work (i.e., manual driving, *CopilotKey*, and *CopilotGaze*), all of our human participants were asked to perform three tasks:

1) Task 1 users controlled a non-autonomous car using a steering wheel and pedals. The steering wheel simulates turning left and turning right, and the pedals simulate accelerating and braking. The users needed to manually drive the vehicle to the parking spot as if they were driving a real car.
2) Task 2 users operated the *CopilotKey* copilot system that uses a keyboard to communicate the intended destination. The users indicated to the autonomous vehicle where they would like to park by typing on a keyboard the color and shape of the sign board next to the intended parking spot.
3) Task 3 users operated the *CopilotGaze* copilot system that uses a gaze-based intention inference algorithm [8]. The users indicated to the autonomous vehicle where they would like to park simply by looking at the sign board next to the parking spot.

For each task, we had the users sit in front of a computer monitor attached to a computer that ran our driving simulator.

[1] A sample video of the system is available at `http://www.cs.utexas.edu/~larg/index.php/Gaze_and_Intent`.

(a) Manual driving system (b) *CopilotKey* copilot system (c) *CopilotGaze* copilot system

Fig. 3: **Experimental setup for manual driving system and each copilot system.** *Left:* the manual driving system includes a steering wheel, pedals, and a desktop screen to display the driving scene. *Middle:* the *CopilotKey* copilot system includes a traditional keyboard and a desktop screen. *Right:* the *CopilotGaze* copilot system includes a desktop screen with an eye tracker attached to its bottom rim.

The simulator utilizes a car model that simulates vehicle physics under the control of steering wheel, accelerator, and brake. The driving simulator was designed using Unity3D [32]. The screen shows the road scene which, for our parking task, includes multiple sign boards with a parking spot next to each. Each sign board depicts a colored shape, where the specific color and shape were randomly generated for each trial such that users could not anticipate what they would be. Each shape was one of the set {square, triangle, star, circle}. Colors belonged to the set {red, yellow, blue, green, and orange}. Hence, in total, 20 different signs are possible. Additionally, in each trial, we ensured that at least two of the signs were the same. This was done introduce ambiguity similar to that of real-world driving scenarios where a user may not be able to clearly differentiate between similar signs or know what to say but does know where to look. A depiction of the systems used by each of the three tasks is shown in Figure 3.

We asked the participants to do their best to get the car to park in a particular parking spot. To communicate to participants which spot they should try to park in, we used a brief priming phase before the vehicle started moving or the participant could control the car. During the priming phase, a yellow arrow specified the goal parking spot. Participants were then allowed to press a start button on the screen to begin the trial, after which the yellow arrow disappeared and the car began to move (either by manual driving or by autonomous driving agent). After the start for the copilot systems, the car moves straight ahead at a fixed initial speed and only parks at the destination if the system identifies it at the decision point through the intent inference engine. After the start for the manual condition, the car responded to the participant's pedal and steering wheel control. These experiments are repeated for goals specified at different distances from the start (23.1 meters, 77.2 meters, and 131.2 meters) and, in the case of the copilot trials, different speeds (15 km/hr, 30 km/hr, and 45 km/hr).

In order to analyze the performance of each system, we collected vehicle information, user commands, car trajectory, and task success rate. In particular, we timestamped and logged the (x, y) location of the car, the human driving commands (i.e., left/right/accelerate/brake/none), and whether or not the task had been completed successfully for every frame during each trial. These data were then post-processed to calculate the 2D trajectory, arrival time to the parking destination, number of user operations, and vehicle speed. Users in each task were allowed to practice driving in the simulation testbed for 3-4 times to get familiar with the system before the experiments actually started. Here, practice means controlling the car using a steering wheel and pedals (for Task 1 users), communicating the intended destination via a keyboard (for Task 2 users), or communicating the intended destination using gaze (for Task 3 users), depending on the specific task.

For each of the questions of interest, we have a separate hypothesis. First, due to an autonomous system's ability to efficiently and reliably perform low-level vehicle control, we hypothesize that we will observe benefit in using copilot systems versus manual driving for our task. Second, due to the speed with which gaze information can be communicated and

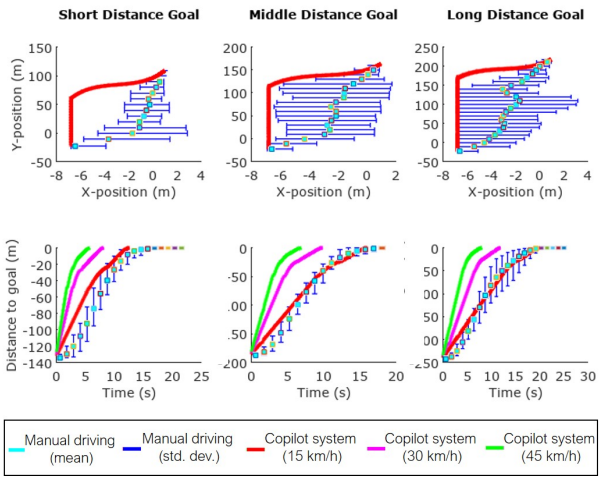

Fig. 4: **Comparison of manual driving mode and copilot modes when the intended destination is correctly inferred.** Each column of graphs represents the measures for a different set of trials, where each column represents trials conducted where the intended goal was at a different distance from the vehicle's starting position. Each row corresponds to a different measure. *Top:* Vehicle's coordinates (x, y) along the trajectory. All copilot data overlaps and so only the 15 Km/h line is visible. *Bottom:* Distance to goal versus time. User average and user standard deviation refer to data recorded in the manual driving condition. This figure is best viewed in color.

the usefulness of this information in helping to resolve spatial ambiguity, we hypothesize that our *CopilotGaze* system will be more successful than our *CopilotKey* system.

B. Experimental Results

To help answer our first question, Figure 4 compares data from users that used the manual driving system with those that used one of the copilot systems in trials with successful destination inference. From the top row, it is clear that the vehicle trajectories generated by the copilot system are both smoother and lower-variance. From the bottom row, it can be seen that, especially when the autonomous system was moving at higher speeds, the copilot system allowed the users to reach their intended goal more quickly than when they drove there manually.

Table I reports the mean and standard error of the number of user actions required for a participant to reach their intended goal for both the manual driving system and the *CopilotKey* system. We do not report any action information for the *CopilotGaze* system because the users can reach the goals just by gazing at the sign board next to the intended parking destination. In the manual driving system, actions occur when users turn the steering wheel to the left or right, or press or release the accelerator or brake pedals. In the *CopilotKey* system, actions occur when users press a key, regardless of whether or not the users achieve the correct goal. Since the number of actions for the *CopilotKey* system is related to the number of characters that describe

TABLE I: **The number of user actions required for reaching an intended goal.** The table shows the mean and standard error (in parentheses) of the number of user actions that were performed by our study's participants. There is no number reported for the *CopilotGaze* system since the users control the system using only their eye gaze. We see that manual driving requires more user activity than either of the copilot systems we study.

	Manual driving	*CopilotKey*
Short Distance Goal	41.93 (2.41)	12.87 (1.60)
Middle Distance Goal	44.80 (4.21)	13.20 (3.24)
Long Distance Goal	44.60 (4.13)	13.18 (1.55)

the sign board, we only prompted users with sign boards of similar description length: "blue square" was used for the short and middle distance goals, and "orange circle" for the long distance goal. In other words, the minimum number of user actions required for reaching the short distance goal, middle distance goal, and long distance goal are 12, 12, and 14, respectively, each achieved using the *CopilotKey* system. Some users performed more actions because they needed to correct their wrong typing during each trial, while others performed fewer actions because they failed to type the entire description of the sign board before the car passed the sign. Overall, we see that manual driving required more user activity - in an order of fifty driving actions compared to an order of 10 character key-in (e.g. "blue square" and press the ENTER key) required by *CopilotKey* and gaze activity required by *CopilotGaze*.

Figure 5 depicts the experimental results related to our second question (i.e., *for a copilot system, is there any benefit in using gaze tracking to detect a destination change versus a system that instead uses a more traditional, keyboard-based method of input?*). It shows the average success rate (and standard error of this statistic) of users for each of our copilot systems, *CopilotKey* and *CopilotGaze* as the autonomous vehicle's speed was increased. We can see that, in trials where the vehicle moved more slowly, both systems do very well and have approximately the same performance. However, as the speed is increased, we see that the success rate of *CopilotGaze* only decreases slightly, while the success rate of *CopilotKey* decreases significantly. We ran a Student's t-test between the *CopilotKey* task and the *CopilotGaze* task, and found that *CopilotGaze* system leads to a significantly higher success rate for each vehicle speed shown (p = 0.019188 for 15km/h, p = 0.0021075 for 30km/h, and p = 5.0517e-006 for 45km/h).

V. Discussion

As passenger vehicles become more autonomous, the interaction between these vehicles and their drivers will change. For example, a driver choosing the remove their hands from the steering wheel of an autonomous vehicle may not necessarily mean that the driver intends to stop steering the vehicle. Therefore, these vehicles will require new methods of detecting and incorporating driver intent. As a limited example of this new paradigm, we studied the

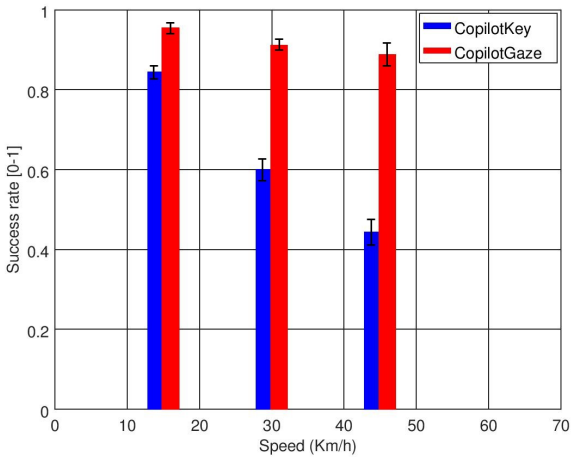

Fig. 5: **Success rate of *CopilotKey* and *CopilotGaze* systems.** The figure compares the average success rate of participants completing the autonomous parking task using either the *CopilotKey* or *CopilotGaze* system at different initial driving speeds. The height of the thick bars indicate the mean value of success rate averaged over participants, and the error bars indicate the standard error of this statistic.

benefit of a collaborative human-robot copilot system in the context of a vehicle parking task we designed in which the autonomous system is tasked with inferring en-route changes in the human driver's goal.

We first investigated if there is benefit in using a copilot system over manual driving in the context of our parking task. We hypothesized that there would be benefit, and the experimental data appears to support this hypothesis. In particular, we found benefit in that vehicle trajectories had lower variance in the copilot conditions, task completion was achieved more quickly in the copilot conditions, and the copilot conditions required fewer user actions. One factor that may have contributed to these results is the skill of human drivers: unskillful drivers may exhibit more oscillatory behavior as they try to determine which controls to apply in order to achieve the intended goal. Assuming that the intended destination was inferred correctly, autonomous vehicles, with a model of the vehicle dynamics and explicit knowledge of the goal, do not suffer from this behavior. Moreover, in our experiments, the copilot systems were actually able to achieve a faster vehicle speed than that achieved in the manual driving condition. These results imply copilot systems may be more time and energy-efficient, safer, and more comfortable than manual driving.

We also hypothesized that *CopilotGaze* would perform better than *CopilotKey*, and indeed the data support this hypothesis as well. We believe that a few factors could contribute to why utilizing gaze information leads to a higher rate of successfully-inferred destinations. First, *CopilotKey* users may not have enough time to key in their intended point of interest before the vehicle passes the decision point. This could be due to, for example, individual typing speed or sign description word length. *CopilotGaze* would not be affected by these temporal issues because of the speed at which the human can shift their gaze and the system can detect that shift. A second contributing factor might have been destination ambiguity. Because some of the sign board icons were identical in shape and color, *CopilotKey* users may have found it difficult to specify a particular board using only text input. For example, two sign boards may have both depicted a red triangle, and so typing the phrase "red triangle" would not be enough information for the system to know which sign board in particular was the user's intended destination. *CopilotGaze*, on the other hand, would not be affected by this form of ambiguity since each sign has a unique spatial position that corresponds to unique user gaze behavior.

One may also envision alternative copilot systems that employ other modalities of human input, e.g., speech and touch. While designing and testing such systems are outside the scope of the current work, we argue here for the unique advantages of employing gaze instead of relying on only these modalities. A copilot system that uses speech input rather than keyboard would provide a slightly-faster method of communication, but such a system might still suffer from the ambiguity issue discussed above. A copilot system that utilizes a touch-based inference system may resolve this ambiguity issue, however, it requires drivers to shift their attention to the touch screen, which might lead to greater driver reaction time and therefor pose a safety concern.

Another point we address here is that of ambiguous gaze patterns. While it certainly warrants future study to analyze a diversity of possible human gaze patterns, we might consider, for example, a situation in the vehicle is closing in on a supermarket, and the driver happens to attend to a passing car near the supermarket sign. In such a situation, uur *CopilotGaze* system would not falsely infer the supermarket as the en-route destination because the supercar moves at different speed than the supermarket from the driver's view, and therefore the DIPD technique, which considers dynamic tracking and time consistency in its algorithm [8], would infer the driver is not "tracking" the supermarket.

In future work, we plan to investigate ways improve the gaze tracking accuracy and the intent inference algorithm - both of which we believe would improve the success rate of the *CopilotGaze* system. Another interesting direction would be to include other vehicles in the scene and have real-time traffic simulation.

VI. Conclusion

In this paper, we studied the effectiveness of human-robot copilot systems in addressing the challenge of autonomous en-route destination changing situation. We analyzed such systems in the context of a parking task, where the passenger's intent is to go to and park next to a particular sign board. A human study was conducted to investigate whether there is benefit in using a copilot system over manual driving, and whether a gaze-based copilot system has benefit compared to a more traditional keyboard-based

system for detecting intent. Our experiment results showed that copilot systems could be operated in a more time-efficient manner, with less variability in vehicle trajectory. Further, we found that a copilot system using gaze for detecting user's intent can result in higher success rates since it enables fast communication between the human passenger and the machine and resolves ambiguity that may be present in language-based destination specification. This work verified that including gaze-based intention inference in a copilot system is worthwhile, and it paves the way to future research that aims to improve the effectiveness and efficiency of human-robot copilot systems.

ACKNOWLEDGMENTS

The author would like to thank Mike Huang and MingXi Lee from Mindtronic AI for their valuable inputs and help on setting up the autonomous vehicle parking system for user experiments. They would also like to thank Justin Hart for helpful discussions. This work has taken place in the Learning Agents Research Group (LARG) at the Artificial Intelligence Laboratory, The University of Texas at Austin. LARG research is supported in part by grants from the National Science Foundation (CNS-1305287, IIS-1637736, IIS-1651089, IIS-1724157), The Texas Department of Transportation, Intel, Raytheon, and Lockheed Martin. Peter Stone serves on the Board of Directors of Cogitai, Inc. 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] A. J. Calder, A. D. Lawrence, J. Keane, S. K. Scott, A. M. Owen, I. Christoffels, and A. W. Young, "Reading the mind from eye gaze," *Neuropsychologia*, vol. 40, no. 8, pp. 1129–1138, 2002.

[2] L. Fletcher, G. Loy, N. Barnes, and A. Zelinsky, "Correlating driver gaze with the road scene for driver assistance systems," *Robotics and Autonomous Systems*, vol. 52, no. 1, pp. 71–84, 2005.

[3] I.-H. Choi, S. K. Hong, and Y.-G. Kim, "Real-time categorization of driver's gaze zone using the deep learning techniques," in *Big Data and Smart Computing (BigComp), 2016 International Conference on*. IEEE, 2016, pp. 143–148.

[4] M. Tall, A. Alapetite, J. San Agustin, H. H. Skovsgaard, J. P. Hansen, D. W. Hansen, and E. Møllenbach, "Gaze-controlled driving," in *CHI'09 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2009, pp. 4387–4392.

[5] Y. Matsumotot, T. Ino, and T. Ogsawara, "Development of intelligent wheelchair system with face and gaze based interface," in *Robot and Human Interactive Communication, 2001. Proceedings. 10th IEEE International Workshop on*. IEEE, 2001, pp. 262–267.

[6] Y. Razin and K. M. Feigh, "Learning to predict intent from gaze during robotic hand-eye coordination." in *AAAI*, 2017, pp. 4596–4602.

[7] W. Min, B. Mott, J. Rowe, R. Taylor, E. Wiebe, K. E. Boyer, and J. Lester, "Multimodal goal recognition in open-world digital games," 2017.

[8] Y.-S. Jiang, G. Warnell, and P. Stone, "DIPD: Gaze-based intention inference in dynamic environments," 2018.

[9] M. Da Lio, F. Biral, E. Bertolazzi, M. Galvani, P. Bosetti, D. Windridge, A. Saroldi, and F. Tango, "Artificial co-drivers as a universal enabling technology for future intelligent vehicles and transportation systems," *IEEE Transactions on intelligent transportation systems*, vol. 16, no. 1, pp. 244–263, 2015.

[10] A. Fagg, M. Rosenstein, R. Platt, and R. Grupen, "Extracting user intent in mixed initiative teleoperator control," in *AIAA 1st Intelligent Systems Technical Conference*, 2004, p. 6309.

[11] D. Aarno, S. Ekvall, and D. Kragic, "Adaptive virtual fixtures for machine-assisted teleoperation tasks," in *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*. IEEE, 2005, pp. 1139–1144.

[12] J. Kofman, X. Wu, T. J. Luu, and S. Verma, "Teleoperation of a robot manipulator using a vision-based human-robot interface," *IEEE transactions on industrial electronics*, vol. 52, no. 5, pp. 1206–1219, 2005.

[13] A. D. Dragan and S. S. Srinivasa, "A policy-blending formalism for shared control," *The International Journal of Robotics Research*, vol. 32, no. 7, pp. 790–805, 2013.

[14] O. Macindoe, L. P. Kaelbling, and T. Lozano-Pérez, "POMCoP: Belief space planning for sidekicks in cooperative games," in *AIIDE*, 2012.

[15] T.-H. D. Nguyen, D. Hsu, W. S. Lee, T.-Y. Leong, L. P. Kaelbling, T. Lozano-Perez, and A. H. Grant, "CAPIR: Collaborative action planning with intention recognition," in *AIIDE*, 2011.

[16] S. Javdani, H. Admoni, S. Pellegrinelli, S. S. Srinivasa, and J. A. Bagnell, "Shared autonomy via hindsight optimization for teleoperation and teaming," *arXiv preprint arXiv:1706.00155*, 2017.

[17] S. Javdani, J. A. Bagnell, and S. S. Srinivasa, "Minimizing user cost for shared autonomy," in *Human-Robot Interaction (HRI), 2016 11th ACM/IEEE International Conference on*. IEEE, 2016, pp. 621–622.

[18] K. Yordanova, S. Whitehouse, A. Paiement, M. Mirmehdi, T. Kirste, and I. Craddock, "What's cooking and why? behaviour recognition during unscripted cooking tasks for health monitoring," in *Pervasive Computing and Communications Workshops (PerCom Workshops), 2017 IEEE International Conference on*. IEEE, 2017, pp. 18–21.

[19] L. M. Hiatt, A. M. Harrison, and J. G. Trafton, "Accommodating human variability in human-robot teams through theory of mind," in *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, vol. 22, no. 3, 2011, p. 2066.

[20] M. Ramırez and H. Geffner, "Goal recognition over POMDPs: Inferring the intention of a POMDP agent," in *IJCAI*. IJCAI/AAAI, 2011, pp. 2009–2014.

[21] G. Kaminka, "Comparing plan recognition algorithms through standard libraries," 2018.

[22] R. G. Freedman, Y. R. Fung, R. Ganchin, and S. Zilberstein, "Towards quicker probabilistic recognition with multiple goal heuristic search," 2018.

[23] A. Watanabe, T. Ikeda, Y. Morales, K. Shinozawa, T. Miyashita, and N. Hagita, "Communicating robotic navigational intentions," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. IEEE, 2015, pp. 5763–5769.

[24] S. Carberry, *Plan recognition in natural language dialogue*. MIT press, 1990.

[25] J. Stephanick, R. Eyraud, D. J. Kay, P. van Meurs, E. Bradford, and M. R. Longe, "Method and apparatus utilizing voice input to resolve ambiguous manually entered text input," May 18 2010, uS Patent 7,720,682.

[26] H. Koppula and A. Saxena, "Learning spatio-temporal structure from rgb-d videos for human activity detection and anticipation," in *International Conference on Machine Learning*, 2013, pp. 792–800.

[27] V. Delaitre, J. Sivic, and I. Laptev, "Learning person-object interactions for action recognition in still images," in *Advances in neural information processing systems*, 2011, pp. 1503–1511.

[28] A. Doshi and M. Trivedi, "A comparative exploration of eye gaze and head motion cues for lane change intent prediction," in *Intelligent Vehicles Symposium, 2008 IEEE*. IEEE, 2008, pp. 49–54.

[29] A. Jain, H. S. Koppula, B. Raghavan, S. Soh, and A. Saxena, "Car that knows before you do: Anticipating maneuvers via learning temporal driving models," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3182–3190.

[30] H. Ravichandar, A. Kumar, and A. Dani, "Bayesian human intention inference through multiple model filtering with gaze-based priors," in *Information Fusion (FUSION), 2016 19th International Conference on*. IEEE, 2016, pp. 2296–2302.

[31] A. Recasens, A. Khosla, C. Vondrick, and A. Torralba, "Where are they looking?" in *Advances in Neural Information Processing Systems*, 2015, pp. 199–207.

[32] unity, *Unity official website*, https://unity3d.com/, 2017.