

Evaluation of a Predictor Based Framework in High-Speed Teleoperated Military UGVs

Yingshi Zheng, Mark J. Brudnak, Paramsothy Jayakumar, Jeffrey L. Stein, Tulga Eersal*

Abstract—Mobility of teleoperated unmanned ground vehicles can be significantly compromised under large communication delays, if the delays are not compensated. This paper considers a recently developed delay compensation theory and presents its first empirical evaluation in improving mobility and drivability of a high-speed teleoperated vehicle under large delays. The said delay compensation theory is a predictor based framework. Two realizations of this framework are considered; a model-free realization that relies only on model-free predictors, and a blended realization where the heading predictions from the model-free predictor are blended with those from a steering-model based feedforward predictor for a more accurate prediction of the vehicle heading. A teleoperated track following task is designed in a human-in-the-loop simulation platform. This platform is used to compare the teleoperation performance with and without the predictor based framework under both constant and varying delays. Through repeated measurement analysis of variance, it is concluded that the predictor based framework is effective in achieving a higher vehicle speed, more accurate lateral control and better drivability as indicated by the three performance metrics of track completion time, track keeping error, and steering control effort, respectively. In addition, it is shown that the blended architecture can lead to further improvements in these metrics compared to using the model-free predictors alone. The analysis also shows that there is no statistically significant difference between constant and varying delay cases in the designed experiment, nor there is any direct relation between drivers' skill level and level of improvement in metrics.

Index Terms—Teleoperation, unmanned ground vehicles, delay compensation, predictor based framework, human-in-the-loop experiment

I. INTRODUCTION

A teleoperated unmanned ground vehicle (UGV) is a vehicle that is remotely driven by an off-board human operator. The control commands of the human operator and vehicle response such as vehicle states and on-board camera view are transmitted between the driver station and the vehicle via communication networks. One of the main challenges of teleoperated UGVs is that large communication delays in

the networks degrade vehicle mobility. This is due to delays causing noticeable asynchrony between the human operator's commands to the vehicle and observations of its response. This degradation becomes more pronounced at higher vehicle speeds. At high speeds, the vehicle dynamics become more critical and there is less time for the operator to respond to errors in vehicle control and sudden changes in the environment. For example, in a simulated driving task, in which human operators were asked to control the vehicle in a lane while maintaining a speed of 55 mph, driving performance was found to be significantly degraded with delays of 170 ms [1].

The literature presents many human-in-the-loop studies of teleoperated driving under delays; see [2, 3] for comprehensive reviews and see [4–6] for quantified teleoperated driving performance under delays in terms of vehicle mobility. Of particular interest for this work are those studies that consider delay compensation methods to improve performance under delays [1, 7–19]. Common practical delay compensation methods are passivity based methods and predictive displays. The former are mostly applied in those experimental setups where drivers operate arm-like manipulators to drive small mobile robots [8–11]. Passivity based methods have been shown to improve the teleoperation performance in terms of transparency. Transparency is defined as the degree of how well human operators feel the environment when teleoperating compared to when they directly interact with it [20]. Hence, transparency is mostly in the sense of haptics. However, driving teleoperated vehicles especially at high speeds relies more on vision than haptic feedback [21]. Therefore, it is unknown how improving haptic transparency helps with improving vehicle mobility in high-speed vehicle teleoperation with delays [3]. Furthermore, since mobile robots are usually operated at low speeds, vehicle mobility is less sensitive to delays as mentioned above.

The second category of methods, namely, predictive displays, have been proven effective to compensate delays and improve vehicle mobility in human-in-the-loop experiments [12–19]. However, only a few were conducted in the scope of high-speed teleoperated UGVs with large delays [16–19]. Predictive displays simulate the vehicle response that is likely to result from the current actions of the operator. Then they visually display to the operator either the predicted virtual vehicle in the delayed view [13] or directly the predicted view [13, 18]. Thus they reduce the asynchrony between the human's control actions and the subsequent vehicle response. A full vehicle model is required to predict the vehicle response, and the prediction accuracy depends on the accuracy of the vehicle model. Such high accuracy models, however, may not

This work was supported by the Automotive Research Center (ARC) in accordance with Cooperative Agreement W56HZV-14-2-0001 U.S. Army Ground Vehicle Systems Center (GVSC) Warren, MI. This material is based upon work supported by the National Science Foundation under Grant No. 1646019.

Y. Zheng, J. L. Stein and T. Eersal are with the Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI 48109, USA. Email: (zhengys@umich.edu, stein@umich.edu, tersal@umich.edu).

M. J. Brudnak and P. Jayakumar are with U.S. Army GVSC, Warren, MI 48092, USA. Email: (mark.j.brudnak.civ@mail.mil, paramsothy.jayakumar.civ@mail.mil).

Distribution A. Approved for public release. Distribution is unlimited. (OPSEC 1540)

* Corresponding author

always be available.

In light of the limitations of the state of the art identified above, and with a focus on high-speed teleoperated driving, here we pursue a delay compensation method that does not require knowledge of the full vehicle model to improve teleoperated driving performance. A predictor based framework has been developed previously in [22–25]. Within the framework, two model-free predictors are implemented to predict the control commands and vehicle states, respectively, without requiring to know the governing dynamics of the vehicles or human operators. The framework with two predictors has been tested in a pilot study in [26] and showed the potential of improving the teleoperated UGV mobility performance under constant round-trip delays of 0.9 s. This predictor based framework is developed further in [27] to include a blended architecture for a more accurate heading prediction. In particular, this architecture combines the feedback heading prediction of the model-free predictor with the feedforward heading prediction based on a simple steering model. Only partial information of the vehicle steering response is needed to generate the steering model in this blended architecture, which is in contrast to the full vehicle model used in predictive displays. The heading prediction accuracy has been evaluated in open loop in [27], indicating that adding the blended architecture helps predict the vehicle heading under delays more accurately compared to using the model-free predictor alone.

However, in previous work [22–27], the developed prediction methods have not been evaluated with a thorough human-in-the-loop user study. The analyses have been mostly theoretical, the only user study being the one in [26] with a very small number of subjects as a pilot study, without any blending, and for only constant delays. Performance of the predictors and the blended architecture under varying delays or driver skills has not been studied before. This paper addresses these gaps through the following original contributions:

- 1) Human-in-the-loop evaluation of the performance of the predictor based framework in a high-speed teleoperated UGV setting with large delays.
- 2) Human-in-the-loop evaluation of the performance of the predictor based framework without and with the blended architecture.
- 3) Analysis of factors of delay type (constant or varying delays) and different driver skills when teleoperating with the said predictor based framework without and with blended architecture.

In the experiments, 19 human drivers controlled a simulated vehicle in a virtual environment to complete a track following task under the following conditions: (1) without any delays as a baseline for teleoperated driving performance; (2) with either constant or varying round-trip delays, but without predictor based framework, to quantify performance degradation due to delays; and (3) with the same amount of delays and the predictor based framework without and with the blended architecture to evaluate the changes in performance caused by the framework. Three metrics, namely, track completion time, track keeping error and steering control effort, are used

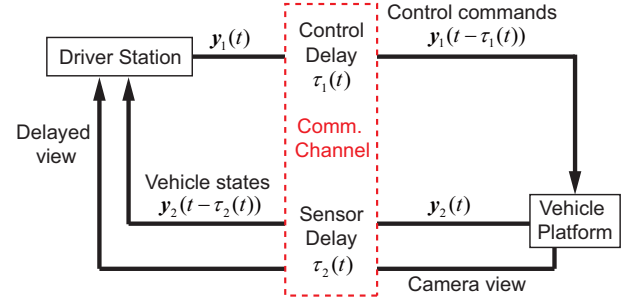


Fig. 1. Generic paradigm of teleoperated UGV systems. Bilateral delays cause asynchrony between operators’ control action at the driver station and the corresponding remote vehicle response.

to quantify vehicle mobility in terms of vehicle’s longitudinal performance, lateral performance and drivability, respectively. Through repeated measurement Analysis of Variance (RM-ANOVA) and t-tests, the metrics based on drivers’ data are analyzed and changes in vehicle mobility and drivability are tested for their statistical significance for all the conditions.

The rest of the paper is organized as follows. Sec. II summarizes the background of the predictor based framework applied to a teleoperated UGV system. The simulation platform developed for human-in-the-loop experiments to test the predictor based framework is described in Sec. III. Details of the design of the human-in-the-loop experiments, including the test task, test scenarios, selected design parameters and analysis methods are explained in Sec. IV. Sec. V discusses the results based on the experimental data and conclusions are given in Sec. VI.

II. A TELEOPERATED UGV SYSTEM WITH PREDICTOR BASED FRAMEWORK

A generic teleoperated UGV system is shown in Fig. 1. Human operators are located at the Driver Station instead of being on-board the vehicle. They drive the remote Vehicle Platform by sending the control vector $\mathbf{y}_1(t)$ that includes steering δ , throttle Th , and brake Br commands, while monitoring the vehicle response based on the received camera view and vehicle state vector $\mathbf{y}_2(t)$ that includes heading ψ , speed u and location in X, Y . All these signals and camera view are transmitted over communication channels and there exist control delays $\tau_1(t)$ and sensor delays $\tau_2(t)$ during transmission. The round trip delays of $\tau_1(t) + \tau_2(t)$ cause asynchrony in time between operators’ control actions and their observation of the corresponding vehicle responses. This asynchrony can make teleoperated driving very challenging and significantly deteriorate the vehicle mobility performance as found in [1].

To reduce this asynchrony, a predictor based framework developed previously in [27] is integrated into the teleoperated UGV system to compensate delays and predict the undelayed signals, as shown in Fig. 2. One modification to the teleoperated UGV system in Fig. 1 is that in addition to the vectors $\mathbf{y}_1(t)$ and $\mathbf{y}_2(t)$, their derivatives, $\dot{\mathbf{y}}_1(t)$ and $\dot{\mathbf{y}}_2(t)$, as well as the send-time stamp need to be included in the communication packets. A Driver Predictor and Vehicle Predictor are placed at

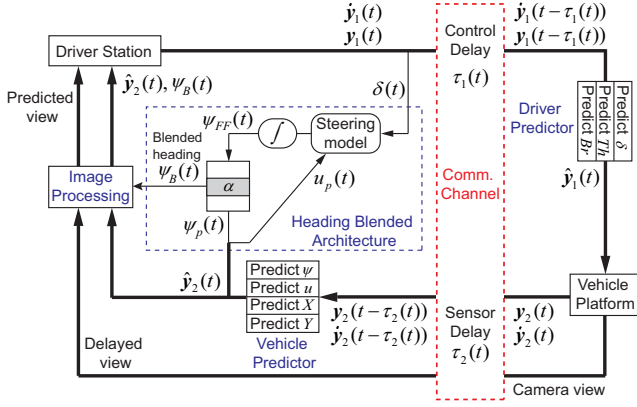


Fig. 2. A predictor based framework with two model-free predictors and a blended architecture for heading is applied to the teleoperated UGV system to predict and recover the undelayed signals that are transmitted through the communication channel.

opposite sides of the communication channel. These predictors receive the delayed vectors and their derivatives, and perform a separate prediction on each signal in the vectors. Thus they generate corresponding predicted outputs $\hat{\mathbf{y}}_1(t)$ and $\hat{\mathbf{y}}_2(t)$. Both predictors have the same dynamics:

$$\dot{\hat{\mathbf{y}}}_i(t) = \dot{\mathbf{y}}_i(t - \tau_i(t)) + \lambda_i[\mathbf{y}_i(t - \tau_i(t)) - \hat{\mathbf{y}}_i(t - \tau_i(t))] \quad (1)$$

for $i \in \{1, 2\}$, where $\hat{\mathbf{y}}_i(t)$ are the predictor states as well as the outputs, and $\hat{\mathbf{y}}_i(t - \tau_i(t))$ are the retarded states that are delayed by the same amount as the measured one-way delay $\tau_i(t)$. By synchronizing the clocks at the Driver Station and Vehicle Platform, $\tau_i(t)$ can be determined as the difference between the send-time stamp included in the packets and the local receive time. λ_i is a square matrix with non-zero terms λ_j ($j = 1, 2, \dots, n$) on the diagonal only, as individual predictor gains to predict the j^{th} signal in the n -dimensional vector $\mathbf{y}_i(t)$. Predictor stability and performance have been well studied in [25] and a design procedure about how to select the predictor gain has been developed in [27]. Note that both predictors are model-free in the sense that predictor dynamics in (1) does not involve any information about the governing equations or parameters of the remote system, where the delayed signals originate. Being model-free, predictors are robust to modeling errors and can be applied to general teleoperated UGV systems regardless of the vehicle platform or different operators with minimal parameter retuning. However, the prediction accuracy – i.e., how close the predicted $\hat{\mathbf{y}}_i(t)$ is to the undelayed signal $\mathbf{y}_i(t)$ – may be lower than model-based approaches if accurate models are available.

To balance robustness with prediction accuracy, the blended architecture that was developed in [27] is included to generate a blended heading output $\psi_B(t)$, which linearly combines the heading $\psi_p(t)$ in the Vehicle Predictor output $\hat{\mathbf{y}}_2(t)$ with a feedforward heading $\psi_{FF}(t)$. The blending law is

$$\psi_B(t) = (1 - \alpha)\psi_{FF}(t) + \alpha\psi_p(t) \quad (2)$$

with a blending weight α between 0 and 1. Note that a steering model is included in the blended architecture to convert the

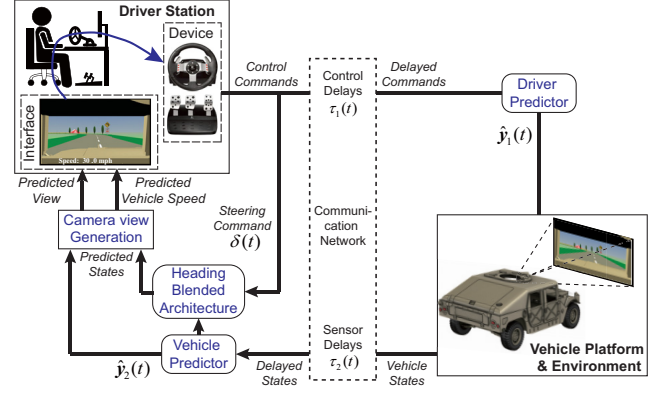


Fig. 3. Human-in-the-loop simulation platform for teleoperated UGV systems with predictor based framework.

operators' steering commands $\delta(t)$ to vehicle yaw rate and to generate $\psi_{FF}(t)$ by integrating the yaw rate. Since the steering gains are usually related to vehicle speed, gains are scheduled depending on the predicted speed signal $u_p(t)$ from $\hat{\mathbf{y}}_2(t)$. Compared to the predictive display methods where a full vehicle model is required to perform prediction, the blended architecture only requires identification and modeling of the steering subsystem of the vehicle platform. Thus, a minimum amount of information about the vehicle is used, benefiting robustness to modeling errors. Only the prediction of vehicle heading involves blending to achieve higher accuracy. This is due to the observation from the pilot study in [26] that, among all the predicted signals, human subjects are most sensitive to the prediction errors in heading.

With Driver Predictor, Vehicle Predictor, and the blended architecture, the final predicted vehicle response vector includes the blended heading $\psi_B(t)$ as well as the predicted speed and location in the Vehicle Predictor output $\hat{\mathbf{y}}_2(t)$.

Based on the predicted vehicle response, the delayed camera view can be processed in two different ways [13]. Specifically, either the predicted shadow vehicle can be overlaid onto the delayed camera view, or the view captured from the predicted vehicle position can be directly displayed. The latter approach is used in this simulation platform.

The effectiveness of this predictor based framework on vehicle mobility performance is evaluated through a driver-in-the-loop experiment in a simulation-based test platform introduced in the following sections.

III. SIMULATION PLATFORM

A real-time driver-in-the-loop simulation test platform is developed in MATLAB Simulink to emulate a teleoperated UGV system as shown in Fig. 3.

The driver station contains a set of Logitech G27 steering wheel and pedals to generate control commands, as well as a monitor to provide visual feedback. The physical steering wheel and pedals are connected as a joystick to the computer that runs the simulation platform. The encoder reading of the steering wheel along with the throttle and brake pedal positions are measured and scaled as the control commands. Also, due

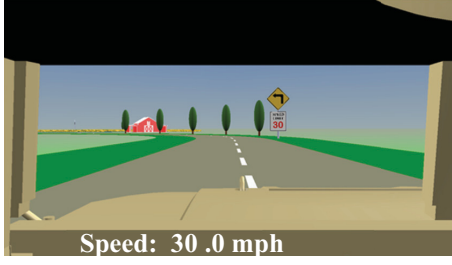


Fig. 4. The monitor displays the camera view and current vehicle speed. The hook on the vehicle's hood serves as a reference of the vehicle's center.

to the observation that the steering commands generated by a human do not exceed a frequency of 1 Hz in a majority of normal driving scenarios [28], a first-order low pass filter with cutoff frequency of 1 Hz is applied to the steering commands to reduce measurement noise. Throttle and brake signals are scaled between 0 and 1 to generate feasible commands for vehicle powertrain dynamics. The visual driving interface has a resolution of 1420x800 pixels (i.e., an aspect ratio of 16:9) and is updated at 20 frame per second (FPS), which is within the typical frame rate range used for teleoperated vehicles [29, 30].

An example of the visual driving interface is shown in Fig. 4. It contains the current vehicle longitudinal speed information and a first-person view that is captured from a camera placed above the driver's seat within the vehicle model in the 3D environment. The camera has a horizontal and vertical field of view (FOV) of 55 deg and 33 deg, respectively, and the environment viewed through the left-half of the vehicle windshield at the driver's side is captured. The FOVs of the camera are set within the common values used for teleoperated navigation tasks [31, 32].

On the vehicle side, a typical military truck, namely, a notional High Mobility Multipurpose Wheeled Vehicle (HMMWV), is chosen as the vehicle platform. The hook on this vehicle's hood serves as a reference of the vehicle's center, i.e., if the vehicle is driven on a dashed straight line and the hook aligns with the closest strip as shown in Fig. 4, the center of the vehicle is exactly on the line. The vehicle is simulated using the same model as in [27], including the 14 DoF vehicle dynamics and Pacejka tire model that were integrated in [33], and the powertrain model from [34] with a 6L V8 diesel engine map, drivetrain and automatic transmission. The vehicle is animated in the virtual environment using the MATLAB Simulink 3D Animation Toolbox.

The simulation model is run using the ODE4 solver, with a fixed time step of 0.01 s. The Simulink Execution Control block is included to slow down the simulation to real time. The signals transmitted between the Driver side and the Vehicle side are control commands (including steering, throttle and brake), vehicle states (including heading angle, global position and vehicle speed) and their derivatives as required by the predictors. While the derivative of all the vehicle states are directly available within the simulated vehicle model, numerical differentiations are performed on the control commands, because only the measurements of commands are available from the Logitech G27 gaming set. This simulation platform is

not set for distributed simulation. Instead, both the driver side and vehicle side are simulated in the same Simulink model on one computer and signals are transmitted internally without a physical communication network. Therefore, simulated delays, either constant or varying, are added optionally to simulate a teleoperated vehicle system with the delays of interest. This is further explained in detail in Sec. IV-B. Control delays $\tau_1(t)$ and sensor delays $\tau_2(t)$ can be specified independently, either as a predefined time sequence or as a random process. Also, the same one-way delay values for each packet are directly provided to the predictors instead of through measurement, thus zero measurement error on the delays is assumed.

The simulation platform also has the general option to activate or deactivate the predictor based framework without any structural change. When all the terms of the predictor parameters λ in the Driver Predictor and Vehicle Predictor are set to zero, there is no prediction of control commands and vehicle states using model-free predictors. As for the vehicle heading prediction, blended architecture is optional to use, as well. A transfer function based steering model is developed based on the lateral response of the same HMMWV vehicle platform as in [27] and implemented in the blended architecture. The blending weight α can be adjusted to generate different blendings of heading. When $\alpha = 1$, blended heading completely relies on the predictor outputs and thus the predictor based framework is model-free, emphasizing robustness to modeling error. When α is set between 0 and 1, blended heading is a linear combination of outputs from the Vehicle Predictor and feedforward heading, and can improve the prediction accuracy of heading even more than the former case of using the Vehicle Predictor only (i.e., $\alpha = 1$), as studied in [27].

Note that in Fig. 3, when the teleoperated UGV is simulated in the virtual environment, there exists one difference compared to Fig. 2: the camera view is not transmitted as video frames. Instead, it is directly available from the virtual environment by placing the vehicle to the required position. By moving the vehicle to the position that is specified by the predicted location and heading, the predicted view is readily displayed in the visual interface.

IV. HUMAN-IN-THE-LOOP EXPERIMENTS

Human-in-the-loop experiments were performed using the aforementioned test platform to evaluate the performance of teleoperated UGVs when using the predictor based framework under both constant and varying delays. The experiment design including the test task and scenarios, test setup, test procedures and analysis methods are described in detail in this section.

A. Experiment Design

A test track is generated in the virtual environment as shown in Fig. 5. The track (in gray) is 810 meters long and 10 meters wide with three left turns, three right turns, and a dashed, white centerline. Shoulders of 6 meters width (in dark green) are located on either side of the track. The safe speed at each turn is determined prior to the test as the maximum allowable speed when the vehicle can successfully be driven on the centerline without causing any tire lift-off. The safe speed is displayed

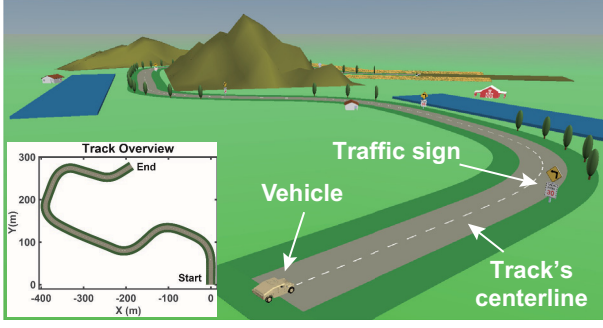


Fig. 5. Designated track, vehicle and landmarks in the virtual environment.

on a traffic sign that is placed several meters before each turn to warn test subjects about the upcoming turn. Trees at the turns and other landmarks such as mountains, houses, corn fields and water ponds serve as visual cues to aid the drivers on speed and distance assessment. In the training session of the experiment a reversed version of the track is used. During the testing session, the original track and its mirrored version are considered as two tracks for testing.

The task of the experiment is designed to be related to real-world military applications. Military vehicles may need to drive with minimum time-to-goal targets to minimize exposure time. Similarly, they may need to follow the tracks left by another vehicle to minimize visual signature. Hence, the task for test subjects is to operate the vehicle to complete the track as fast as possible while following the track's centerline as closely as possible.

Three performance metrics are considered as the dependent variables of the experiment: track completion time, track keeping error and steering control effort. Track completion time captures how fast the vehicle is driven along the track from start line to finish line. Track keeping error is calculated as the area between the trajectory of the vehicle and track's centerline, indicating how close the vehicle follows the centerline. These two metrics reflect the longitudinal and lateral mobility performance of the vehicle, respectively. Smaller metrics lead to better teleoperated driving performance and thus higher mobility. Steering effort is determined as the average magnitude of the steering wheel angle in degrees, aiming to capture the drivability of the vehicle. In a teleoperated driving task with delays, overactuation in the form of oversteering and repeated corrections is expected [35]. Smaller steering effort reflects that subjects feel easier and more comfortable to control the vehicle to complete the task. The above three metrics are referred to as *Time*, *Error* and *Effort*, respectively.

Two independent variables that may affect the dependent variables are studied: *Delay Type* and *Prediction Method*. *Delay Type* can be chosen from {**CD**, **VD**}, where **CD** or **VD** refer to the setups with non-zero bilateral constant delays or varying delays, respectively. *Prediction Method* is among {**NoPred**, **Pred**, **PredBlend**}, where **NoPred** represents the setup with non-zero delays, but without any prediction. The only difference between the method of **Pred** and **PredBlend** is the α value. For **Pred**, $\alpha = 1$ so that only the two model-free predictors are activated, whereas for **PredBlend**,

α is selected to be 0.5 and a resultant blended heading prediction is generated with both the predictors and blended architecture included in the setup. The parameter α is chosen using the optimization approach in [27]. Specifically, $\alpha = 0.5$ leads to the most accurate vehicle heading prediction in open loop under the configured delay values and predictors in this experiment. This open-loop optimization is employed only as a design guideline, as the open-loop optimal value may not be the optimal one for closed loop. Finding the latter is not critical for the purposes of this work and is thus left for future work. The detailed delay values and predictor parameters used in the test are presented in Section IV-B and Section IV-C, respectively.

In summary, the experiment follows a 2 (*Delay Type* with two levels) $\times$ 3 (*Prediction Method* with three levels) within-subjects factorial design. The goal of the experiment is to study how the *Delay Type* and *Prediction Method* affect the vehicle mobility and drivability in teleoperated UGVs at high speed. Seven scenarios with three repetitions in each scenario are tested. Seven scenarios include one **NoDelay** scenario with zero delays used as the baseline performance of human drivers to handle this track-following driving task, as well as six scenarios with non-zero delays and different combinations of *Delay Type* and *Prediction Method*. Thus, a total of 21 test runs need to be completed by each test subject. Each run has a different setup, combining one of the scenarios with one of the test tracks. The order of the runs is randomized in an evenly distributed manner to reduce the learning effects on a single scenario or one track. Specifically, no single scenario or the same track is tested twice in a row and one scenario is not tested more than the others.

B. Test Setup: Delays

In the setups for the scenarios except **NoDelay**, constant delays or varying delay sequences are generated based on real network data. This procedure is described in detail below.

Round-trip time (RTT) delays for teleoperated UGV systems generally vary from around 0.1 s to more than 1 s, depending on the types of communication connections (e.g., radio/WLAN, cellular network, satellite, etc.), communication distance and bandwidth. Longer distance and lower bandwidth generally lead to larger RTT. Some RTT values were reported in teleoperated UGV systems with physical communication networks [36–38]. In [36], an average of 0.121 s RTT was measured based on a teleoperated vehicle road test with 3G cellular network, while the peak delays were more than 1 s. In [37], RTT delays were measured as a summation of a one-way delay of 0.33 s through a low-bandwidth radio link and another one-way delay of 0.55 s through a high-bandwidth WLAN video link. In [38], RTT delays in a ViaSat satellite link were around 0.75 s, when the Driver Station and Vehicle Platform were located in Washington DC and California, respectively. This RTT range is also aligned with the RTT values used in most of the simulation based experiments in the literature [1, 17, 18, 39, 40].

To simulate a representative teleoperated UGV system, large round-trip delays of around 0.9 s are tested. Specifically, they are distributed in the two ways of communication as control

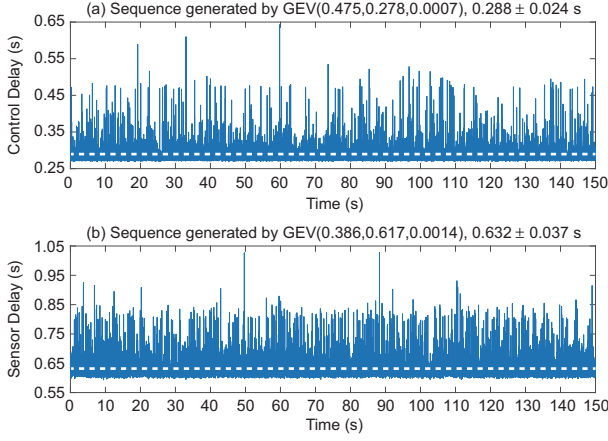


Fig. 6. Simulated varying delay sequences of (a) control delays and (b) sensor delays, generated by the corresponding fitted GEV models. Mean values are represented with white, dashed lines.

Table I. Simulated varying control and sensor delays generated by the fitted GEV models.

	Fitted GEV			Mean (s)	Standard deviation (s)
	ξ	μ	σ		
Control delays	0.475	0.278	0.0007	0.288	0.024
Sensor delays	0.386	0.617	0.0014	0.632	0.037

delays of around 0.3 s and sensor delays of around 0.6 s. This choice reflects the general knowledge that sensor delays are typically higher than control delays due to the larger packet size and conversions associated with video feeds. For each scenario involving **CD**, control commands and vehicle response are delayed by constant values of 0.3 s and 0.6 s, respectively. For scenarios with **VD**, delay sequences are generated based on data collected in [22] using an actual network between Ann Arbor, Michigan (MI) and Palo Alto, California (CA). This network is referred to as the MI-CA network in the rest of the paper.

A Generalized Extreme Value (GEV) distribution is used to represent the MI-CA network as a probability distribution delay model. Its probability density function is expressed as

$$f_{\xi \neq 0, \mu, \sigma} = \frac{1}{\sigma} (1 + \xi \frac{x - \mu}{\sigma})^{-1-1/\xi} \exp(-(1 + \xi \frac{x - \mu}{\sigma})^{-1/\xi})$$

$$f_{\xi = 0, \mu, \sigma} = \frac{1}{\sigma} e^{-\frac{x - \mu}{\sigma}} \exp(-e^{-\frac{x - \mu}{\sigma}}) \quad (3)$$

where x is the random variable of delay value and ξ, μ, σ are the shape, location and scale parameters, respectively. When $\xi > 0$, a heavy tailed distribution can be represented that has a lower bound $\mu - \frac{\sigma}{\xi}$ based on the extreme value theory. A GEV distribution model, denoted as $GEV(\xi, \mu, \sigma)$, with $\xi = 0.707$, $\mu = 0.0546$, $\sigma = 0.0012$ fits well with the delay distribution of MI-CA network, whose mean and standard deviation are 0.058 s and 0.006 s. 5 and 9 independent randomly generated sequences using $GEV(0.707, 0.0546, 0.0012)$ are summed respectively to simulate sequences of varying control and sensor delays that have mean values close to the corresponding constant values of 0.3 s and 0.6 s.

The histograms of summed sequences are fitted again using GEV distributions. Then new sets of sequences are randomly generated based on the fitted distributions for 150 s and are

used as the simulated control and sensor delays in the tests, as shown in Fig. 6. They meet the assumption that delays can be modeled using a two-time-scale Markov Chain and thus the predictor stability with varying delay can be established as in [25]. The fitted parameters of the resulting GEV distributions as well as means and standard deviations of the simulated delays are listed in Table I. The mean control and sensor delays are 0.288 s and 0.632 s, respectively, and thus for the varying delay case, the round-trip delays have a mean of 0.920 s.

No physical network is included in the simulator and factors like packet reordering and dropping are not considered based on the negligible values observed in the tested networks [22, 34]. Only the simulated communication delays are considered for compensation. Other than that, mechanical delays related to the vehicle such as the delays in the steering and brake systems are captured in the simulated 14 DoF vehicle dynamics. Based on the time constants of the vehicle model, mechanical delays are on the order of 0.1 s. Computational delay of applying the predictors with or without the blended architecture is negligible due to the simple structure of the predictors and blended architecture. Delays of rendering the virtual interface are also negligible in this simulation environment.

C. Test Setup: Predictor Based Framework Parameters

The design parameters in the predictor based framework include the λ terms in both the Driver Predictor and Vehicle Predictor to predict each of the transmitted signals.

Predictor parameters λ are chosen independently for each transmitted signal type and based on the design procedure developed in [27] considering both predictor stability and performance. The values are listed in Table II. Firstly, all λ are between 0 and the maximum allowable range to ensure predictor's asymptotic stability. The maximum values are denoted as $\lambda_{\max}(\tau)$ for constant delay τ and $\lambda_{\max}(\tau(t))$ for varying delay $\tau(t)$ and are calculated as follows [25]. For constant delays,

$$\lambda_{\max}(\tau) = \frac{\pi}{2\tau} \quad (4)$$

and for varying delays $\tau(t)$ that can be modeled using a two-time-scale Markov Chain,

$$\lambda_{\max}(\tau(t)) = \frac{3}{2\tau_{\text{avg}}(t)} \quad (5)$$

where $\tau_{\text{avg}}(t)$ is the average value of $\tau(t)$. Pilot test data collected in [26] under the same test task and scenario are analyzed to provide some insights into each transmitted signal $y(t)$ with constant measured delays τ_m . One relevant information is the dominant frequency components within the coupling error $c(t)$. $c(t)$ is defined as the difference between an undelayed and delayed signal, i.e.,

$$c(t) := y(t) - y(t - \tau_m) \quad (6)$$

To achieve good steady state prediction performance, it is recommended that the predictor bandwidth ω_p is greater than the coupling error bandwidth ω_c [27]. Here, the coupling error bandwidth ω_c is estimated as the frequency until which the signal power of $c(t)$ is accumulated to reach 90% of the total power within the Nyquist frequency of $c(t)$. The predictor

Table II. Predictor settings for individual signals of interest

Predicted Signal	Measured Delay τ_m [s]	ω_c [rad/s]	λ [s^{-1}]	Sat. & Reset.	ω_p [rad/s]
Throttle	0.3	11.5	$0.30\lambda_{\max}(0.3)$	No	1.65
Brake		19.4	$0.30\lambda_{\max}(0.3)$	No	1.65
Steering		5.37	$0.10\lambda_{\max}(0.3)$	Yes	0.94
Speed	0.6	3.35	$0.40\lambda_{\max}(0.3)$	Yes	1.92
Global X		1.88		No	
Global Y		1.30		No	
Heading		1.86		Yes	

bandwidth ω_p is related to the selection of λ and is the smallest positive value that satisfies

$$\lambda = 2\omega_p \sin(\tau_p \omega_p) \quad (7)$$

where τ_p is the amount of delay to be compensated by the predictor. Larger λ and smaller τ_p result in larger ω_p . To achieve good transient prediction performance, larger λ leads to faster transient, but causes greater magnitude of overshoot in $\hat{y}(t)$. Applying the saturation & resetting scheme developed in [27] could help reduce the overshoot when the transmitted signal $y(t)$ has change of direction. A detailed analysis of λ 's effect on the predictor's overall performance in terms of steady state and transient with respect to the constant delays has been addressed in [27]. The λ values for scenarios with varying delays are chosen to be the same as the λ values for constant delays.

Taking into account the trade-off in λ between steady state and transient performance, λ for each transmitted signal is designed as shown in Table II. Note that for control commands of throttle, brake and steering, it is unlikely that the predictor bandwidth ω_p can cover all the frequencies up to ω_c . Thus, smaller λ values are chosen to minimize the potentially significant degradation of steady state performance. On the contrary, ω_c for the vehicle states of speed, position and heading are relatively small. Therefore, partial delays of 0.3 s (i.e., $\tau_p = 0.3$) out of 0.6 s measured (i.e., $\tau_m = 0.6$) are compensated with the selected λ values to satisfy the expectation of $\omega_p > \omega_c$ as much as possible. Since frequent direction changes in steering, vehicle speed and heading are observed, saturation and resetting scheme is applied when predicting these three signals to reduce the effect of overshoot on transient performance. The predictor outputs for throttle and brake commands are saturated between 0 and 1, as well.

D. Test Procedure

Prior to the test, subjects were asked to fill out an informed consent form and answered some basic background questions about their age, proficiency of driving a vehicle in the real world, and proficiency of driving in a virtual environment with a gaming steering wheel and pedals.

The whole test process lasted no more than 2 hours and was divided into a training session (1 hour) and a testing session (45 minutes). The training session was necessary to help subjects adapt to driving in a simulation based teleoperation setup under large delays. Subjects were verbally instructed on the test details, including the track-following driving task, the performance goals, as well as the seven scenarios to test. They were informed that driving as fast and as close to the

centerline as possible were equally important. However, the subjects were not told how the performance metrics were specifically measured and were allowed to adopt a driving behavior suitable for themselves to meet the performance goals according to their own priorities.

The remaining time in the training session was left for subjects to adapt to the teleoperated driving setup and practice all seven test scenarios one by one. The current scenario to practice was visually displayed on the visual driving interface and the three performance metrics were calculated each time the subjects completed the driving task to monitor their performance under training. A scenario would be practiced until metrics between the successive trials indicated consistent performance and subjects felt confident about driving in such a scenario. Subjects also could ask for more time to be distributed on the scenarios with which they were not familiar.

In the testing session, each subject was asked to complete a total of 21 valid runs in randomized order, with three repetitions in each scenario. The subjects were aware of the scenario in each run before testing. The run was considered valid if none of the following conditions were observed:

- Vehicle was off track for more than 5 s or did not pass the finish line of the track.
- At least two tires lifted off in the vehicle model.
- Average vehicle speed was less than 25 mph.

A warning was displayed on the visual driving interface when one of the tires lifted off the ground. The 25 mph speed bound was determined based on the driving data of the beta test. In the beta test, a human driver with rich experience of performing teleoperated driving under delays went through the training session and became well-trained to generate replicates of test run data with consistent performance. In the most challenging scenarios with varying delay and without prediction (i.e., **VD** $\times$ **NoPred**), the driver was able to complete the task with an average vehicle speed of 30 mph. Up to 5 mph reduction in speed compared to the beta test was then allowed in the experiments, resulting in the 25 mph minimum bound on average speed. A run with average speed lower than 25 mph meant that the subject did not push the vehicle dynamics to the limit to complete the driving task in this run. An average speed of more than 25 mph is considered as high speed in this experiment, given that other vehicles or ground robot platforms are usually tested in speeds of only several m/s [41]. After all test runs were completed, subjects were thanked for their participation and received a compensation.

E. Analysis Methods

2-way repeated measures analysis of variance (RM-ANOVA) was used to study the effect of the two independent variables (i.e., *Delay Type* and *Prediction Method*) on each of the dependent variables (i.e., performance metrics *Time*, *Error* and *Effort*). Applying RM-ANOVA requires that data meet two assumptions: normality (whether metrics follow normal distribution) and sphericity (whether variance of metrics within each group are equal). For each metric, normality was checked using the Anderson-Darling test and sphericity was checked by comparing variances across the six test conditions of *Delay Type* $\times$ *Prediction Method* based on Levene's test. Processing

such as performing log scale transformation on data was done when needed to satisfy the assumption of normality.

Two null hypotheses for each metric were tested using an F-test based on Type III sums of squares with 95% confidence level. These null hypotheses were:

- 1) There does not exist a significant difference in performance metrics when different prediction methods are applied to compensate delays.
- 2) There does not exist a significant difference in performance metrics between the constant and varying delays used for testing.

If the F-test indicated that at least one mean is different from the rest (i.e., $P < 0.05$), Fisher's least significant difference (LSD) method was used to identify the groups with pairwise significant differences in the means.

Another dimension to study was whether there was any correlation between drivers' skill level and performance in this teleoperated driving task with large delays. Metrics of *Time* and *Error* were normalized to their mean values among all drivers' data. An overall performance metric P_{comb} combining normalized *Time* and *Error* was defined:

$$P_{\text{comb}} := 0.5 \frac{\text{Time}}{\text{Time}_{\text{avg}}} + 0.5 \frac{\text{Error}}{\text{Error}_{\text{avg}}} \quad (8)$$

The smaller P_{comb} was, the better a driver's overall performance was. The two metrics were combined with equal weights, because the drivers were informed before the test that driving fast and accurately would be evaluated equally. P_{comb} in scenarios (**CD** $\times$ **NoPred**) and (**VD** $\times$ **NoPred**) was used as the criterion to distinguish driver skill level. All drivers were divided into 2 skill groups, **SKILLED** and **NOVICE**, via 2-means clustering.

The following null hypotheses were tested using two-sample t-tests with 95% confidence level:

- 1) Means of metrics for **SKILLED** drivers are equal to those of **NOVICE** drivers under the same scenario.
- 2) Means of level of improvement in metrics for **SKILLED** drivers and **NOVICE** drivers after applying the predictor based framework are equal.

V. RESULTS AND DISCUSSION

The simulation was run in a Dell Z210 Desktop computer with CPU core frequency of 3.4 GHz. The experiments were approved by the University of Michigan Health Sciences and Behavioral Sciences Institutional Review Board (UM IRB #HUM00112376). Drivers at all skill levels were welcomed and were recruited publicly through email announcements and flyers. A total of 22 test subjects participated in the experiments. 2 of them were observed to still struggle with completing the driving task especially in the delayed conditions without any prediction after 1 hour of training session. The background information collected revealed that they reported to have only 1-6 months of experience on driving vehicles in the real world and no experience of driving virtual vehicles at all. Another subject did not push the vehicle to its limits as instructed. Therefore, these three subjects could not complete the experiments successfully and their data are not included in the analysis.

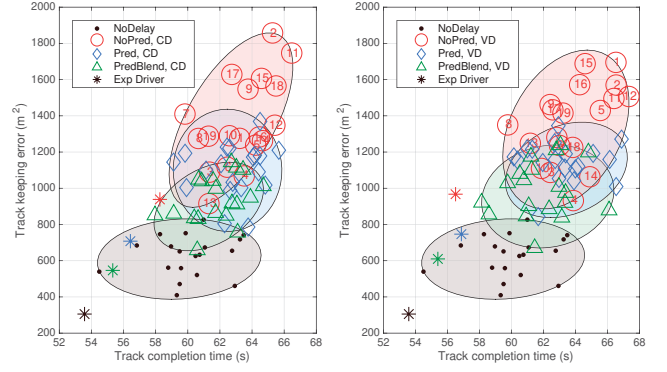


Fig. 7. Distribution of performance metrics of track completion time and track following error for various scenarios among all test subjects. Exp Driver denotes the expert teleoperator in the beta test.

The remaining 19 subjects had an average age of 22.7 years old with standard deviation of 2.6 years. Majority of them were familiar with doing daily driving in the real world: 18 subjects reported to own driver licenses and 15 of them had more than 2 years of driving experience. However, only 6 subjects were familiar with driving in a virtual environment such as playing racing games. In terms of the situation most similar to this experiment, where humans drive the vehicle in simulated environment using steering wheel and pedals, only 3 of them had prior experience. Given the general lack of experience among the test subjects on driving in the used test setup, the importance of training session is emphasized again to help them adapt to the test setup and become familiar with the scenarios especially when large round trip delays of around 0.9 s were added into the system.

The three performance metrics are calculated based on 19 subjects' data. Averaging the three repeated runs within the same scenario, metrics of track completion time (*Time*) and track following error (*Error*) from each driver's data are plotted in Fig. 7. **NoPred**, **Pred** and **PredBlend** represent the scenarios with different prediction methods. Delay type **CD** and **VD** stands for the constant and varying bilateral delays. Each ellipse encloses all the data points under the same scenario with minimum volume and represents the distribution of subjects' driving performance. The numbers marked in scenario (**NoPred** $\times$ **CD** or **VD**) represent the n^{th} driver's metric. As mentioned in Sec. IV-E, P_{comb} for each driver is calculated. Drivers 3, 4, 6, 7, 8, 10, 13 and 14 are classified into the group of **SKILLED** and the remaining drivers comprise the **NOVICE** skill group. The performance of the aforementioned experienced driver in the beta test is marked in asterisk. Despite the performance variation among drivers, metrics of *Time* and *Error* are reduced in the scenarios of **Pred** and **PredBlend** compared to the scenario of **NoPred**, which indicates better performance. Similar trends are seen when tested with either constant delays or varying delays.

The means and standard deviations of 19 subjects' data are reported in Table III. For each metric, the first row with parentheses and the second row with brackets represent the results of **SKILLED** and **NOVICE** driver group, respectively. **NoDelay** is considered as the baseline of performance metrics.

Table III. Comparison of means and standard deviations of metrics under different scenarios. Numbers in parentheses and brackets are for the **SKILLED** and **NOVICE** group, respectively.

Means and Stds NoDelay	<i>Delay Type</i>	CD			VD		
	<i>Prediction Method</i>	NoPred	Pred	PredBlend	NoPred	Pred	PredBlend
(58.9 ± 2.34) [60.9 ± 2.60]	<i>Time</i> [s]	(62.1 ± 2.03) [64.3 ± 2.00]	(61.3 ± 2.08) [63.9 ± 1.59]	(60.4 ± 2.13) [62.8 ± 1.93]	(62.5 ± 2.35) [64.8 ± 2.36]	(61.6 ± 2.07) [64.4 ± 2.12]	(60.5 ± 2.33) [62.9 ± 2.12]
(578 ± 180) [662 ± 165]	<i>Error</i> [m ²]	(1178 ± 210) [1490 ± 260]	(1019 ± 205) [1155 ± 159]	(850 ± 180) [998 ± 161]	(1163 ± 270) [1499 ± 211]	(1141 ± 197) [1148 ± 152]	(910 ± 158) [1045 ± 183]
(36.7 ± 1.11) [36.5 ± 1.15]	<i>Effort</i> [deg]	(43.9 ± 2.97) [45.3 ± 3.41]	(39.2 ± 1.86) [39.9 ± 1.91]	(38.4 ± 1.22) [39.5 ± 1.86]	(43.8 ± 2.95) [45.3 ± 3.03]	(39.3 ± 1.78) [40.4 ± 2.43]	(39.1 ± 1.76) [39.3 ± 2.21]

Table IV. RM-ANOVA table for the metric *Time*.

Factor	DF	Adjusted SS	Adjusted MS	<i>F</i>	<i>P</i>
Delay	1	0.0019	0.0019	2.71	0.101
Method	2	0.0456	0.0228	31.71	0.000
Delay × Method	2	0.0007	0.0007	0.51	0.601
Skill group	1	0.1277	0.1277	18.99	0.000
Method × Skill group	2	0.0005	0.0003	0.36	0.697
Subject	17	0.1143	0.0067	4.75	0.000
Repetition	38	0.0538	0.0014	1.97	0.001
Error	278	0.2000	0.0007		
Total	341	0.5452			

Table V. RM-ANOVA table for the metric $\log(\text{Error})$.

Factor	DF	Adjusted SS	Adjusted MS	<i>F</i>	<i>P</i>
Delay	1	0.0021	0.0021	3.40	0.066
Method	2	0.1263	0.0631	103.33	0.000
Delay × Method	2	0.0013	0.0006	1.04	0.354
Skill group	1	0.0445	0.0445	25.19	0.000
Method × Skill group	2	0.0083	0.0041	6.79	0.001
Subject	17	0.0300	0.0018	2.57	0.008
Repetition	38	0.0261	0.0007	1.13	0.290
Error	278	0.1699	0.0006		
Total	341	0.4171			

It is observed that the metrics in scenario **NoPred** without any prediction are much worse than the baseline **NoDelay**. Applying the predictor based framework either without or with the blended architecture (**Pred** or **PredBlend**) helps with reducing the means of all three metrics compared to **NoPred**. Thus, the predictor framework helps improve vehicle mobility and drivability as quantified by the three metrics. Additionally, the **SKILLED** group outperforms the **NOVICE** group in almost all the scenarios and metrics. The detailed results related to the effect of prediction methods, skill groups, and delay types are presented in Section V-A, Section V-B and Section V-C, respectively.

A. Relation between Prediction Methods and Performance

To study whether the improvement is statistically significant, RM-ANOVA is applied to the data within the six scenarios with delays only, since the metrics in scenario **NoPred** are significantly different than those in other scenarios.

ANOVA is based on the assumption of normality. Hence, normality of the metrics are checked first using Anderson-Darling Test and only the calculated metrics of *Time* follow normal distribution with 95% confidence. For the remaining two metrics of *Error* and *Effort*, box-cox transformation with $\lambda = 0$ in Minitab 17 is applied, which is equivalent to taking the natural logarithm of the metrics. Normality then holds

Table VI. RM-ANOVA table for the metric $\log(\text{Effort})$.

Factor	DF	Adjusted SS	Adjusted MS	<i>F</i>	<i>P</i>
Delay	1	0.0001	0.0001	0.39	0.531
Method	2	0.0811	0.0406	264.31	0.000
Delay × Method	2	0.0001	0.0001	0.20	0.818
Skill group	1	0.0034	0.0034	2.40	0.140
Method × Skill group	2	0.0002	0.0001	0.73	0.482
Subject	17	0.0244	0.0014	8.88	0.000
Repetition	38	0.0061	0.0002	1.05	0.394
Error	278	0.0427	0.0002		
Total	341	0.1615			

for the transformed metrics $\log(\text{Error})$ and $\log(\text{Effort})$ based on the Anderson-Darling Test with 95% confidence. Through Levene's test, the null hypothesis of equal variances of metrics *Time*, $\log(\text{Error})$ and $\log(\text{Effort})$ fails to be rejected with 95% confidence and therefore the sphericity condition also holds for the metrics.

Two-way RM-ANOVA on each metric of *Time*, $\log(\text{Error})$ and $\log(\text{Effort})$ is performed individually in Minitab 17 as a general linear model. The two factors of interest are *Delay Type* and *Prediction Method*. The factor of skill group and its nested factor of subject and repetition are also included in the model, but set as random or uncontrolled to set up the requirement of repeated measures.

The detailed RM-ANOVA tables are shown in Table IV–VI. With significance level of 0.05, *P* values of the F-test with respect to the factor of *Prediction Method* are close to 0 and much smaller than 0.05 in all three RM-ANOVA tables, indicating a significant difference in all three metrics among different prediction methods. The effect sizes (partial eta squared) of the factor *Prediction Method* for the metrics *Time*, $\log(\text{Error})$ and $\log(\text{Effort})$ are 0.085, 0.329 and 0.522, respectively. Pairwise comparison results are shown in Fig. 8, where the asterisk represents that there exists significant difference in mean pairwise. In terms of the effect of *Delay Type*, the effect sizes are very close to 0 and the *P* value for the metrics of *Time*, *Error*, *Effort* in the RM-ANOVA results are 0.101, 0.066 and 0.531, respectively. All *P* values are greater than 0.05 and thus the hypothesis of no significant difference between constant and varying delays tested cannot be rejected with 95% confidence level.

Define the level of improvement LoI as:

$$\text{LoI} := \frac{|r_p - r_d|}{|r_{nd} - r_d|} \quad (9)$$

where r is the mean of a given transformed metric, with subscript *nd* representing **NoDelay**, subscript *p* and *d* representing the scenarios with prediction (**Pred** or **PredBlend**)

Table VII. The level of improvement in performance metrics with **Pred** or **PredBlend**, compared to **NoPred**.

Delay Type Prediction Method	CD		VD	
	Pred	PredBlend	Pred	PredBlend
<i>Time</i>	15%	47%	17%	52%
$\log(\text{Error})$	26%	47%	20%	40%
$\log(\text{Effort})$	60%	68%	57%	66%

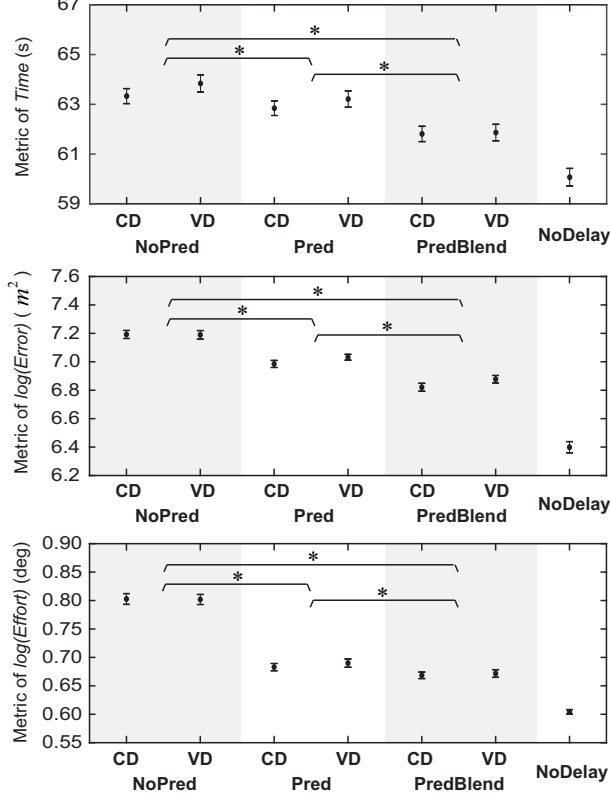


Fig. 8. Pairwise ANOVA comparison results on the three performance metrics *Time*, $\log(\text{Error})$, $\log(\text{Effort})$ with different delay types and prediction methods. There exists significant difference in mean among **NoPred**, **Pred**, **PredBlend**, indicated by the asterisks.

and without prediction (**NoPred**), respectively, with the same type of delays of either **CD** or **VD**. The levels of improvement for the three metrics are shown in Table VII. The larger LoI is, the more improvement in performance subjects benefit from the predictor based framework without and with the blended architecture.

Consider the constant delay case first. Compared to **NoPred**, metrics of *Time*, $\log(\text{Error})$ and $\log(\text{Effort})$ are improved significantly by 15%, 26% and 60%, respectively, in scenario **Pred**. With **PredBlend**, they are all improved significantly by 47%, 47% and 68%. Track completion time and track keeping error are improved more using **PredBlend** than **Pred** by 32% and 21%, respectively, and steering control effort is improved more by 8%. Similar level of improvement can also be observed in the varying delay case.

B. Relation between Driver Skill and Performance

In Table III, the **SKILLED** driver group (with means μ_1) outperforms **NOVICE** group (with means μ_2) in almost all the

scenarios and metrics. 2-sample one-tailed t-tests with 95% confidence are performed to study whether the difference in mean of performance between the two groups is statistically significant. The null hypothesis H_0 and its alternative hypothesis H_{a1} and H_{a2} are defined as:

$$H_0 : \mu_1 = \mu_2, \quad H_{a1} : \mu_1 < \mu_2, \quad H_{a2} : \mu_1 > \mu_2 \quad (10)$$

Error and *Effort* are transformed to $\log(\text{Error})$ and $\log(\text{Effort})$ to meet the normality assumption of t-tests. For each scenario and each metric *Time*, $\log(\text{Error})$ and $\log(\text{Effort})$, pooled standard deviation of two groups are calculated and t-values are listed in Table VIII. The sample sizes of the **SKILLED** group and **NOVICE** group are $N_1 = 24$ and $N_2 = 33$, respectively. The critical value is $t_{0.05,55} = -1.673$. Each cell colored in green indicates that H_0 can be rejected with 95% confidence and H_{a1} when $t < t_{0.05,55}$ is accepted. Cells not colored show no significant evidence for such scenario and metric. Overall, the **SKILLED** group has statistically better performance in terms of *Time* and $\log(\text{Error})$ than the **NOVICE** group.

Similarly, level of improvements LoI for *Time*, $\log(\text{Error})$ and $\log(\text{Effort})$ within each subject are calculated based on (9). Pooled standard deviation of two groups is calculated and t-values of LoI for different prediction methods and delay types are listed in Table IX. Same hypotheses as in (10) are tested and the critical value is $t_{0.05,17} = -1.740$. Cells colored indicate that H_0 can be rejected with 95% confidence. Green cells mean H_{a1} when $t < t_{0.05,17}$ is accepted and the **NOVICE** group improves its performance more than the **SKILLED** group when predictors are used in the varying delay scenario. Blue cells mean H_{a2} when $t > t_{0.95,17}$ is accepted and the **SKILLED** group improves its performance more (larger LoI) when applying predictors under varying delays as well as predictors with blended architecture under constant delays. The remaining LoIs are not significantly different due to large variations within each group. There is no straight conclusion that relatively novice drivers, who do not perform well in teleoperated driving with delays, will have more performance improvement than skilled drivers after applying the predictor based framework.

C. Relation between Delay Types and Performance

Note that means and standard deviations of the metrics with **VD** are slightly larger than those with **CD** in Table III, but the difference is negligible and not significant. The level of improvement in metrics between **VD** and **CD** shown in Table VII are also similar with the predictor based framework, as the same λ values in the predictors are used for scenarios with both constant and varying delays. In the literature, Refs. [17, 40] reported detectable performance difference between constant and varying delays (with a delay variation ratio of 0.3 to 0.67 between the standard deviation and mean of delays), while no significant difference between them was observed in [7], with a ratio of 0.5. In contrast, the said ratio is less than 0.1 in our experiment and it is more unlikely to lead to significant performance difference. However, our delay sequences are generated based on the measurements of an actual network and are therefore more realistic to represent an actual network connection than the aforementioned experiments in the literature.

Table VIII. t-values of the 2-sample one-tailed t-tests to compare two skill groups' mean performance, with significance level of 0.05. Green cells are those for which the **SKILLED** group has statistically better performance than the **NOVICE** group.

t-values NoDelay	Delay Type	CD			VD		
	Prediction Method	NoPred	Pred	PredBlend	NoPred	Pred	PredBlend
-3.005	Time	-4.115	-5.329	-4.381	-3.734	-4.798	-4.077
-1.991	log(Error)	-4.852	-2.899	-3.306	-5.382	-0.296	-2.807
0.726	log(Effort)	-1.653	-1.426	-2.601	-1.901	-1.978	-0.338

Table IX. t-values of the 2-sample one-tailed t-tests to compare the level of improvement (LoI) in performance metrics between two skill groups. Green (blue) cells are those for which the LoI in the **NOVICE** (**SKILLED**) group is statistically significantly larger than that in the **SKILLED** (**NOVICE**) group.

Delay Type	CD		VD	
Prediction Method	Pred	PredBlend	Pred	PredBlend
Time	0.823	-0.191	0.013	-0.474
log(Error)	-0.920	-0.222	-2.933	-0.464
log(Effort)	0.326	2.317	1.896	-0.683

As shown in Fig. 6, delay values are mostly distributed around their means and there exist sudden spikes in the sequences. However, these spikes do not last for a long duration and, in this closed-loop simulation based experiment, they do not have a notable effect on the performance. Further experiments could be conducted to study the effect of variations in the delay on performance when tested with a network that has a larger variation of delays.

VI. CONCLUSIONS

This paper presents the first human-in-the-loop evaluation of the predictor framework of [22–27] in teleoperated UGVs in terms of its ability to improve vehicle mobility and drivability in a track following task. Two realizations of the framework are considered. In the first realization all coupling signals are predicted in a model-free manner, whereas in the second one a blended prediction of the heading signal is considered by linearly combining the model-free prediction with a steering model based one. The performances of both realizations of the framework under both constant and varying delays are evaluated with human-in-the-loop simulations, where performance is measured using three metrics.

Based on the results, four conclusions are drawn: (1) The predictive framework significantly improves all metrics, thereby improving vehicle mobility and drivability. (2) The blended architecture offers more improvement in all three metrics compared to the purely model-free realization. (3) There is no direct relation between driver skill level and level of improvement in metrics when prediction methods are applied, even though the skilled driver group outperforms the novice driver group in all the metrics and all the scenarios. (4) There is no significant difference in any of the metrics under any conditions between constant and varying delays. These conclusions encourage further development of the predictor framework.

Future work includes applying the predictor based framework to actual vehicle platforms, which involves transmitting packets with actual communication networks, measuring one-way delays and developing image processing methods to convert the delayed camera view to the predicted view based on delayed and predicted vehicle states.

REFERENCES

- [1] L. Frank, J. Casali, and W. Wierwille, "Effects of visual display and motion system delays on operator performance and uneasiness in a driving simulator," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 30, no. 2, pp. 201–217, 1988.
- [2] T. B. Sheridan, "Space teleoperation through time delay: Review and prognosis," *IEEE Transactions on Robotics and Automation*, vol. 9, no. 5, pp. 592–606, 1993.
- [3] J. Y. Chen, E. C. Haas, and M. J. Barnes, "Human performance issues and user interface design for teleoperated robots," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1231–1245, 2007.
- [4] J. Storms and D. Tilbury, "Equating user performance among communication latency distributions and simulation fidelities for a teleoperated mobile robot," in *IEEE International Conference on Robotics and Automation*, 2015, pp. 4440–4445.
- [5] M. Cross, K. A. McIsaac, B. Dudley, and W. Choi, "Negotiating corners with teleoperated mobile robots with time delay," *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 6, pp. 682–690, 2018.
- [6] D. J. Gorsich, P. Jayakumar, M. P. Cole, C. M. Crean, A. Jain, and T. Ersal, "Evaluating mobility vs. latency in unmanned ground vehicles," *Journal of Terramechanics*, vol. 80, pp. 11–19, 2018.
- [7] J. P. Luck, P. L. McDermott, L. Allender, and D. C. Russell, "An investigation of real world control of robotic assets under communication latency," in *ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, 2006, pp. 202–209.
- [8] F. Janabi-Sharifi and I. Hassanzadeh, "Experimental analysis of mobile-robot teleoperation via shared impedance control," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 2, pp. 591–606, 2010.
- [9] F. Penizzotto, S. Garcia, E. Slawinski, and V. Mut, "Delayed bilateral teleoperation of wheeled robots including a command metric," *Mathematical Problems in Engineering*, vol. 2015, 2015.
- [10] Z. Zuo and D. Lee, "Haptic tele-driving of a wheeled mobile robot over the internet: a PSPM approach," in *IEEE Conference on Decision and Control*, 2010, pp. 3614–3619.
- [11] Y.-J. Pan, C. Canudas-de Wit, and O. Sename, "A new predictive approach for bilateral teleoperation with applications to drive-by-wire systems," *IEEE Transactions on Robotics*, vol. 22, no. 6, pp. 1146–1162, 2006.
- [12] A. K. Bejczy, W. S. Kim, and S. C. Venema, "The phantom robot: predictive displays for teleoperation with time delay," in *IEEE International Conference on Robotics and Automation*, 1990, pp. 546–551.
- [13] F. Chucholowski, "Evaluation of display methods for teleoperation of road vehicles," *Journal of Unmanned System Technology*, vol. 3, no. 3, pp. 80–85, 2016.
- [14] A. Hosseini, F. Richthammer, and M. Lienkamp, "Predictive haptic feedback for safe lateral control of teleoperated road vehicles in urban areas," in *IEEE Vehicular Technology Conference*, 2016, pp. 1–7.
- [15] S. Mathan, A. Hyndman, K. Fischer, J. Blatz, and D. Brams, "Efficacy of a predictive display, steering device, and vehicle body representation in the operation of a lunar vehicle," in *Conference Companion on Human Factors in Computing Systems*, 1996, pp. 71–72.
- [16] J. H. Hogema, "Compensation for delay in the visual display

- of a driving simulator,” *Simulation*, vol. 69, no. 1, pp. 27–34, 1997.
- [17] J. Davis, C. Smyth, and K. McDowell, “The effects of time lag on driving performance and a possible mitigation,” *IEEE Transactions on Robotics*, vol. 26, no. 3, pp. 590–593, 2010.
- [18] M. Brudnak, “Predictive Displays for High Latency Teleoperation,” in *NDIA Ground Vehicle Systems Engineering and Technology Symposium*, 2016, pp. 1–16.
- [19] A. Kelly, N. Chan, H. Herman, and R. Warner, “Experimental validation of operator aids for high speed vehicle teleoperation,” in *Experimental Robotics*, 2013, pp. 951–962.
- [20] D. A. Lawrence, “Stability and transparency in bilateral teleoperation,” *IEEE Transactions on Robotics and Automation*, vol. 9, no. 5, pp. 624–637, 1993.
- [21] S. Lichiardopol, “A survey on teleoperation,” *Technische Universitat Eindhoven, DCT Report Vol. 2007.155*, 2007.
- [22] A. Tandon, M. J. Brudnak, J. L. Stein, and T. Ersal, “An observer based framework to improve fidelity in internet-distributed hardware-in-the-loop simulations,” in *ASME Dynamic Systems and Control Conference*, p. V002T21A004.
- [23] X. Ge, M. J. Brudnak, P. Jayakumar, J. L. Stein, and T. Ersal, “A model-free predictor framework for tele-operated vehicles,” in *American Control Conference*, 2015, pp. 4573–4578.
- [24] X. Ge, Y. Zheng, M. J. Brudnak, P. Jayakumar, J. L. Stein, and T. Ersal, *Analysis of a Model-Free Predictor for Delay Compensation in Networked Systems*, ser. Advances in Delays and Dynamics. Springer, 2017, vol. 7.
- [25] Y. Zheng, M. J. Brudnak, P. Jayakumar, J. L. Stein, and T. Ersal, “A predictor based framework for delay compensation in networked closed-loop systems,” *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 5, pp. 2482–2493, 2018.
- [26] Y. Zheng, M. Brudnak, P. Jayakumar, J. L. Stein, and T. Ersal, “An experimental evaluation of a model-free predictor framework in teleoperated vehicles,” *IFAC-PapersOnLine*, vol. 49, no. 10, pp. 157–164, 2016.
- [27] Y. Zheng, M. J. Brudnak, P. Jayakumar, J. L. Stein, and T. Ersal, “A delay compensation framework for predicting heading in teleoperated ground vehicles,” *IEEE/ASME Transactions on Mechatronics*, vol. 24, no. 5, pp. 2365–2376, 2019.
- [28] D. T. Day and L. D. Metz, “The simulation of driver inputs using a vehicle driver model,” in *SAE Technical Paper*, no. 2000-01-1313, 2000.
- [29] P. L. McDermott and A. Fisher, “The tradeoff of frame rate and resolution in a route clearing task: Implications for human-robot interaction,” in *Human Factors and Ergonomics Society Annual Meeting*, vol. 53, no. 4, 2009, pp. 177–181.
- [30] O. Bodell and E. Gulliksson, “Teleoperation of autonomous vehicle with 360 degree camera feedback,” Masters Thesis, Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden, 2016.
- [31] S. Lee Pazuchanics, “The effects of camera perspective and field of view on performance in teleoperated navigation,” in *Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 16, 2006, pp. 1528–1532.
- [32] US Department of Defense, “Department of defense design criteria standard, human engineering,” *MIL-STD-1472F*, 1999.
- [33] J. Liu, P. Jayakumar, J. L. Stein, and T. Ersal, “A study on model fidelity for model predictive control based obstacle avoidance in high speed autonomous ground vehicles,” *Vehicle System Dynamics*, vol. 54, no. 11, pp. 1629–1650, 2016.
- [34] T. Ersal, M. Brudnak, A. Salvi, J. L. Stein, Z. Filipi, and H. K. Fathy, “Development and model-based transparency analysis of an internet-distributed hardware-in-the-loop simulation platform,” *Mechatronics*, vol. 21, no. 1, pp. 22–29, 2011.
- [35] M. Barnes, K. Cosenzo, D. Mitchell, and J. Chen, “Human robot teams as soldier augmentation in future battlefields: An overview,” in *International Conference on Human-Computer Interaction*, 2005.
- [36] S. Gnatzig, F. Chucholowski, T. Tang, and M. Lienkamp, “A system design for teleoperated road vehicles,” in *International Conference on Informatics in Control, Automation and Robotics*, vol. 2, 2013, pp. 231–238.
- [37] P. Appelqvist, J. Knuuttila, and J. Ahtiainen, “Development of an unmanned ground vehicle for task-oriented operation - Considerations on teleoperation and delay,” in *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, 2007, pp. 1–6.
- [38] National Robotics Engineering Center, “Long Distance Teleoperation - Avatar,” 2016. Accessed: 2018-8-15. [Online]. Available: <https://www.nrec.ri.cmu.edu/solutions/defense/other-projects/long-distance-teleoperation.html>
- [39] J. Lane, C. Carignan, B. Sullivan, D. Akin, T. Hunt, and R. Cohen, “Effects of time delay on telerobotic control of neutral buoyancy vehicles,” in *IEEE International Conference on Robotics & Automation*, vol. 3, 2002, pp. 2874–2879.
- [40] J. Storms, “Modeling and Improving Teleoperation Performance of Semi-Autonomous Wheeled Robots,” PhD Thesis, Department of Mechanical Engineering, University of Michigan, 2017.
- [41] D. Lee, O. Martinez-Palafox, and M. Spong, “Bilateral teleoperation of a wheeled mobile robot over delayed communication network,” in *IEEE International Conference on Robotics and Automation*, 2006, pp. 3298–3303.