Predicting Human Operators' Detection of Time-Varying Changes in Controlled Element Dynamics

Barragan, M.; Pool, D.M.; van Paassen, M.M.; Mulder, Max

DOI
10.2514/6.2025-0974

Publication date
2025

Document Version
Final published version

Published in
Proceedings of the AIAA SCITECH 2025 Forum

AIAA SciTech Forum
6-10 January 2025, Orlando, FL
AIAA SCITECH 2025 Forum

10.2514/6.2025-0974

Predicting Human Operators' Detection of Time-Varying Changes in Controlled Element Dynamics

Martin Barragan*, Daan M. Pool†, Marinus M. van Paassen‡ and Max Mulder§
Delft University of Technology, Delft, Zuid-Holland, The Netherlands

While human control behavior is well-understood in continuous control tasks, little is still known about how human operators detect sudden changes in the controlled element dynamics. This paper focuses on modeling this detection phase for pursuit tracking tasks. Potential triggers for the human operator to detect changes in the controlled element dynamics were investigated via a time-varying computer simulation. Based on the results, hypotheses were generated and later tested in a single-axis pursuit tracking experiment with fifteen participants. Transitions from approximate single to approximate double integrator dynamics and vice versa were investigated, for which participants indicated if they detected the transition by pressing a button. Using the button push data, a model for each transition was developed and validated. The models work under the assumption that human operators use a threshold, a multiple of the steady-state standard deviation, on certain signals to detect transitions. The models developed for the transition from single to double integrator dynamics and vice versa are proposed to trigger on the tracking error and system output acceleration, respectively. They have an accuracy of 88.9% and 99.4%, respectively. However, a consistent underestimation of the detection lag remains a limitation of both models. Nonetheless, this research helped confirm the tracking error can be used in a model for the transition from single to double integrator dynamics, proposed a model for the opposite transition, and identified that the relationship between control inputs and the system's response as a crucial factor for the detection.

Nomenclature

$A_t[i]$	=	Amplitude of ith sine wave in forcing function (rad)
e	=	Error (rad)
$\dot{e}$	=	Error rate (rad/s)
$\ddot{e}$	=	Error acceleration (rad/s^2)
f_t	=	Forcing function (rad)
FN	=	False negative
FP	=	False positive
G	=	Maximum rate of change (s^{-1})
$H_c(s,t)$	=	Controlled element transfer function (-)
$H_{del}(s,t)$	=	Human operator delay transfer function (-)
$H_n(s,t)$	=	Remnant filter transfer function (-)
$H_p(s,t)$	=	Human operator transfer function (-)
i	=	Sine wave index (-)
IQR	=	Interquartile range (-)
k_c	=	Controlled element gain (-)
k_n	=	Remnant filter gain (rad)

*M.Sc. student (graduated), Control and Simulation section, Faculty of Aerospace Engineering, P.O. Box 5058, 2600GB Delft, The Netherlands; barraganmartin10@gmail.com.
†Assistant Professor, Control and Simulation section, Faculty of Aerospace Engineering, P.O. Box 5058, 2600GB Delft, The Netherlands; d.m.pool@tudelft.nl. Associate Fellow AIAA.
‡Associate Professor, Control and Simulation section, Faculty of Aerospace Engineering, P.O. Box 5058, 2600GB Delft, The Netherlands; m.m.vanpaassen@tudelft.nl. Member AIAA.
§Professor, Control and Simulation section, Faculty of Aerospace Engineering, P.O. Box 5058, 2600GB Delft, The Netherlands; m.mulder@tudelft.nl. Associate Fellow AIAA.

k_p	=	Human operator gain (-)
k	=	Multiple of base frequency (-)
M	=	Centered on time (s)
N_t	=	Number of sine waves used to construct forcing functions (-)
N	=	Number of samples (-)
n	=	Human operator remnant (rad)
P_1	=	Initial parameter value (-)
P_2	=	Final parameter value (-)
RMSE	=	Root-mean-square error (deg)
T_L	=	Human operator lead time constant (s)
T_m	=	Measurement time (s)
T_n	=	Remnant filter time constant (s)
t	=	Time (s)
TN	=	True negative
TP	=	True positive
u	=	Control input (rad)
$\dot{u}$	=	Control input rate (rad/s)
w	=	Gaussian white noise input (-)
y	=	System output (rad)
$\dot{y}$	=	System output rate (rad/s)
$\ddot{y}$	=	System output acceleration (rad/s^2)
ω_b	=	Controlled element break frequency (rad/s)
ω_m	=	Fundamental frequency (rad/s)
ω_{nm}	=	Neuromuscular system natural frequency (rad/s)
$\omega_t[i]$	=	Frequency of ith sine wave in forcing function (rad/s)
$\phi_t[i]$	=	Phase shift of ith sine wave in forcing function (rad)
ρ	=	Pearson correlation coefficient (-)
σ	=	Standard deviation (-)
τ_e	=	Effective time delay (s)
ζ_{nm}	=	Neuromuscular system damping ratio (-)

I. Introduction

Over the past decades, the role of automation has increased significantly in society. In the context of control tasks, this has led to human operators taking on more of a supervisory role. A situation where interventions made by the human operator are often in moments where adaptability, accuracy, and quick-thinking are required. Adaptability, in particular, is currently one of the main strengths humans have over automation. However, with the advancement of technology and development in how human operators interact with systems, the models used to describe their behavior have failed to keep up [1]. It is beneficial to understand and model the adaptive behavior of human operators for several reasons. First, it would help improve the design of current systems, since different design choices can be tested through simulation to identify potential problems. Additionally, adaptive models can be used to create better support systems for human operators. In the case of failures, for example, this can help reduce the number of loss of control incidents, which are still one of the largest worldwide contributors to fatal accidents [2]. Finally, it would allow for the development of safer and improved vehicle designs [3].

Young et al. [4] identified that the adaptive process of a human operator consists of three phases, namely 'detection', 'identification', and 'modification'. In the detection phase, the human operator realizes that something has changed and that adaptation is required. This triggers the identification phase to begin, in which the human operator identifies the nature of the change. Finally, in the modification phase, the human operator modifies their control behavior to suit the new situation. It can be argued that the identification and modification phases happen simultaneously, though the detection phase is certainly distinct. A good adaptive model would be capable of approximating the human operator's behavior throughout all three phases. However, the scope of this paper is limited to the detection phase, specifically the case where there is a change in the controlled element dynamics.

Significant research efforts have been dedicated to studying the adaptive nature of human operators and how they are able to detect changes to the controlled element dynamics [5–7]. Several models have been proposed, especially

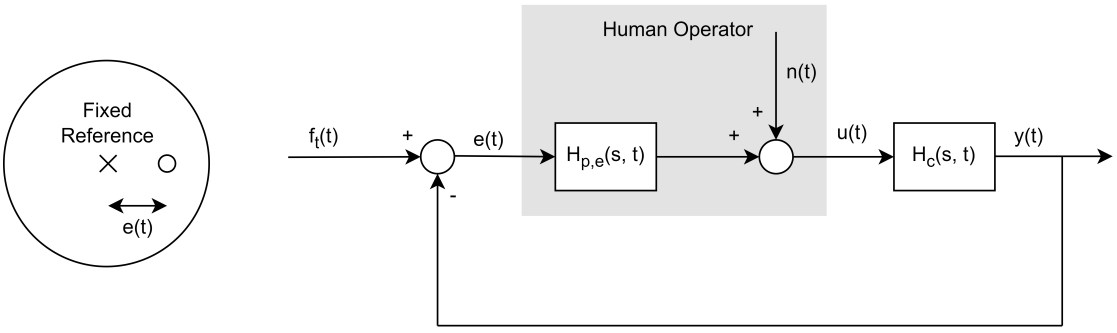
for compensatory tracking tasks in which the controlled element dynamics transition from more stable to less stable dynamics, that use a threshold on either the tracking error or error rate [8, 9]. The motivation behind these models is that well-trained human operators keep track of the statistical properties of those signals and are thus able to identify when either becomes abnormally large. Furthermore, human operators are able to predict the value of a particular signal in the (near) future via an internal model [10, 11] of the controlled element dynamics and knowledge of the control inputs. A significant mismatch between the observed and predicted value can also lead to the human operator detecting a change in controlled element dynamics. Building on this, Hess [12, 13] proposed to model the detection phase in pursuit tracking tasks based on a threshold of a signal proportional to the error between the observed system output rate and the (internally generated) desired system output rate.

Thus far, the majority of research efforts have focused on the transition from more stable to less stable dynamics (e.g., from approximate single to approximate double integrator dynamics). Modeling the human operator's detection of this transition based on outliers (i.e., extreme peaks) in the error or error rate has been successful [8, 14]. However, the same cannot be said for the inverse transition from double to single integrator, because the system does not become unstable if the current control strategy is kept, and thus the error does not grow abnormally large [15, 16]. This suggests that only part of the detection phase is properly understood. Furthermore, the majority of research efforts thus far have only studied the detection phase for tracking using compensatory displays.

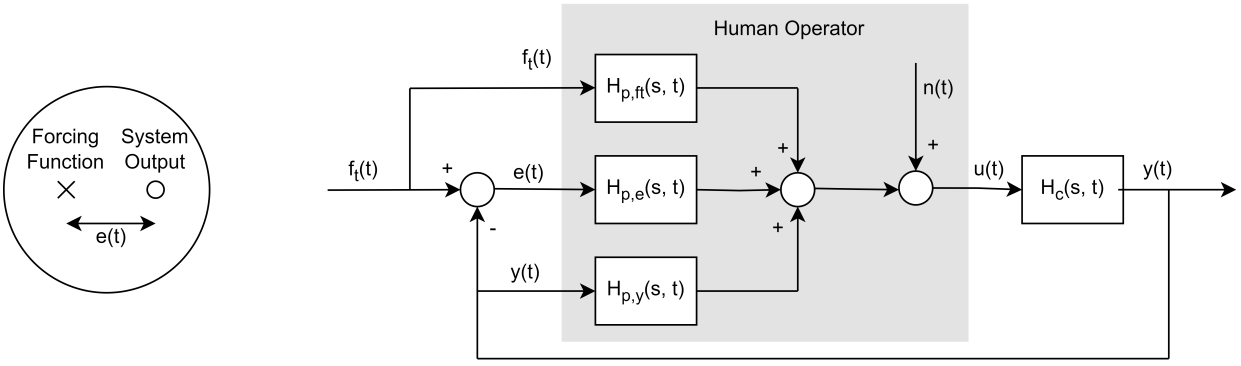
In real-life applications, compensatory displays are the exception rather than the norm. Studying the detection phase with a pursuit display, where more information is available to the human operator, can increase the applicability of the research. The presence of additional information also significantly impacts the control strategies available to the human operator [17] and can therefore also impact the way changes to the controlled element dynamics are detected. For example, the explicit knowledge of the relationship between the control inputs given by the human operator and the system's response can aid with the detection phase when using a pursuit display. The difference between compensatory tracking tasks and pursuit tracking tasks is illustrated in Figure 1. The same signals appear in both Figure 1a and Figure 1b, namely the forcing function $f_t(t)$, error $e(t)$, human operator remnant $n(t)$, control input $u(t)$, and the system output $y(t)$. However, the human operator is modeled differently in both as a result of only the error being displayed with a compensatory display, whereas the forcing function and system output are both displayed with a pursuit display. For any controlled element dynamics $H_c(s,t)$, the human operator can only respond to the error (modeled as $H_{p,e}(s,t)$) when using a compensatory display. With a pursuit display, however, the human operator can also respond to the forcing function (modeled as $H_{p,ft}(s,t)$), and the system output (modeled as $H_{p,y}(s,t)$) in addition to the error [17]. Thus, it is clear that the control behavior and the way changes to the controlled element dynamics are detected can be very different with a pursuit display compared to a compensatory display.

The goal of this paper is to investigate and provide insight into how human operators detect changes in the controlled element dynamics and consequently adapt their control behavior when using pursuit displays. Particular attention is paid to the transition from approximate double to approximate single integrator dynamics since this is less well understood. First, a simulation was developed to investigate what the trigger for a human operator to detect a change in controlled element dynamics might be. This involved analyzing six candidate signals (the system output y, the control input u, the tracking error e, and all of their derivatives), and evaluating potential models that are based on when the respective signal exceeds a certain threshold (a multiple of the steady-state standard deviation). Afterwards, a human-in-the-loop pursuit tracking task experiment was performed at the Faculty of Aerospace Engineering of TU Delft in which several transitions from approximate single integrator dynamics to approximate double integrator dynamics and vice versa were conducted. Participants indicated if and when they detected the change by pressing a button on a joystick. The goal was to find a link between the subjective button push data and some property in one or more of the signals that may have triggered participants to detect that the controlled element dynamics changed. Ultimately, two models are proposed and validated, one for each transition.

The structure of the paper is as follows. First, the design and results from a time-varying computer simulation will be presented in section II. Next, the experiment design and data analysis methodology will be presented in section III. The results of the experiment data are presented in section IV, followed by a discussion of the results in section V. Finally, the conclusions are presented in section VI.



(a) Compensatory display (left) and control diagram (right).



(b) Pursuit display (left) and control diagram (right). The control diagram is adapted from Mulder et al. [17].

Fig. 1 Difference between compensatory and pursuit displays, and the accompanying control diagrams.

II. Computer Simulation

A. Simulation Design

In preparation for the human-in-the-loop experiment and to formulate hypotheses, a time-varying simulation was implemented using MATLAB and Simulink to gain a better understanding of what might trigger a human operator (HO) to detect a change in the controlled element (CE) dynamics. To do this, two HO models were compared with each other following a sudden change in CE dynamics. One of the HO models is of an adaptive HO that immediately adapts their own dynamics to match the new CE dynamics after a transition, while the other is of a constant HO that never adapts their own dynamics and continues controlling the new CE in the same way as before the transition. The latter is representative of an HO who has not (yet) detected that there has been a change in CE dynamics. Comparison between the signals of the adaptive HO and constant HO were used to generate hypotheses on what triggers a HO to detect a transition.

The naming convention for the CE dynamics is as follows: the numbers are the order of the system and the sequence is the order in which the conditions occur [15]. For example, DYN12 is a trial where the CE dynamics start off as an approximate single integrator (first order) and transitioned to an approximate double integrator (second order) during the trial; DYN21 has the same conditions, but with the order reversed. A time-invariant trial only has one number. For example, DYN1 is a trial where the CE dynamics remained an approximate single integrator throughout the entire trial. The simulations were primarily used to investigate the time-varying conditions, DYN12 and DYN21. The time-invariant trials, DYN1 and DYN2, were only used to calculate steady-state reference values to be used in the analysis, as described in subsection II.B.

Despite the fact that a pursuit display was used in the experiment, the simulation is performed for compensatory tracking because there is a universally accepted model for it, i.e., the simplified precision model [18]. The block diagram for the compensatory tracking simulation is given in Figure 2. $H_p(s,t)$ is the linear part of the HO dynamics (including

the neuromuscular system dynamics but excluding the time delay), $H_{del}(s,t)$ is the effective HO time delay, $H_n(s,t)$ is the remnant filer dynamics used to model the HO remnant, and $H_c(s,t)$ is the CE dynamics.

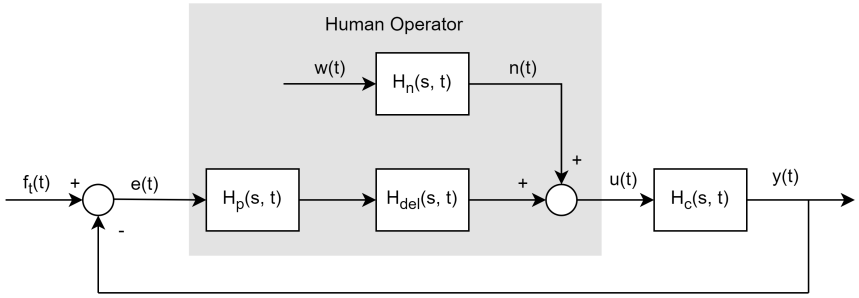


Fig. 2 Block diagram for the compensatory tracking system in the simulation.

The structure of $H_c(s,t)$ is given in Eq. (1), where k_c is the CE gain and ω_b is the CE break frequency, both of which are time-varying. The break frequency is used to change the order of the CE dynamics in the crossover region. For the DYN1 condition, ω_b is set to 20 rad/s, whereas for the DYN2 condition, it is set to 0.2 rad/s. The gain k_c is varied to keep the level of control activity approximately constant for both conditions and has a value of 15 in DYN1 and 2 in DYN2.

$$H_c(s,t) = \frac{k_c(t)}{s(s+\omega_b(t))} \tag{1}$$

The HO dynamics are given in Eq. (2) and have the form dictated by the simplified precision model plus a term for the neuromuscular system dynamics [18]. For the results presented here, the HO gain k_p was tuned after performing the experiment to match the mean DYN1 and mean DYN2 crossover frequencies from the experiment. This results in a k_p of 0.233 for DYN1 and 3.73 for DYN2. The lead time constant T_L is 0 s for DYN1 (since no lead generation is required) and 5 s for DYN2, to perfectly compensate for the second integrator that has a break frequency of 0.2 rad/s. The neuromuscular system parameters were kept constant for both conditions with a natural frequency ω_{nm} of 15 rad/s and a damping ratio ζ_{nm} of 0.7. $H_{del}(s,t)$ accounts for the HO effective time delay, $e^{-\tau_e(t)s}$, and was modeled using the 'variable time delay' block in Simulink, with a delay of 0.09 s in DYN1 and 0.23 s in DYN2.

$$H_p(s,t) = k_p(t)[1+T_L(t)s] \cdot \frac{\omega_{nm}^2}{s^2+2\zeta_{nm}\omega_{nm}+\omega_{nm}^2} \tag{2}$$

Finally, the remnant filter dynamics have the form given in Eq. (3), which Van Grootheest et al. [19] found to be the best structure for fitting HO remnant. The time constant T_n was kept constant at 0.06 s, and the gain k_n was chosen such that the ratio of the power in the control input due to the remnant to the total power in the control input was 0.2. This resulted in a value of k_n of 0.00787 for DYN1 and 0.00813 for DYN2. The signal $w(t)$ in Figure 2, which is Gaussian white noise with zero mean and unit variance, is the input to the remnant filter, and the HO remnant $n(t)$ is the output. All of the parameter values are summarized in Table 1.

$$H_n(s,t) = \frac{k_n(t)}{(T_n s+1)^2} \tag{3}$$

For the time-varying conditions, the transition from one parameter value to the other is done according to the sigmoid function defined in Eq. (4), where P_1 is the initial parameter value, P_2 is the final parameter value, G is the maximum rate of change (equal to 100 s^{-1}), and M is the sigmoid's midpoint time (equal to 75 s). These values of G and M apply to all of the time-varying parameters in both the CE dynamics and the adaptive HO model.

$$P(t) = P_1 + \frac{P_2 - P_1}{1+e^{-G(t-M)}} \tag{4}$$

The resulting crossover frequencies and phase margins are given in Table 2. In addition to the standard DYN1 and DYN2 conditions, two further conditions arise after the change in CE dynamics as a result of the constant HO keeping the same control strategy. These are shown in the bottom two rows of Table 2. 'DYN1 CE with DYN2 HO' occurs

Table 1 Values for all of the parameters used in the simulation.

CE Dynamics	k_c (-)	ω_b (rad/s)	k_p (-)	T_L (s)	τ_e (s)	ω_{nm} (rad/s)	ζ_{nm} (-)	k_n (-)	T_n (s)	G (s^{-1})	M (s)
DYN1	15	20	0.233	0	0.09	15	0.7	0.00787	0.06	100	75
DYN2	2	0.2	3.73	5	0.23	15	0.7	0.00813	0.06	100	75

post-transition in the DYN21 trials while 'DYN2 CE with DYN1 HO' occurs post-transition in the DYN12 trials. As can be seen, the system with the constant HO becomes unstable post-transition in DYN12, while the system with the constant HO remains stable post-transition in DYN21, but with a very low crossover frequency and large phase margin. The crossover frequencies in the standard DYN1 and DYN2 conditions are 2.77 rad/s and 2.33 rad/s, respectively, and the phase margins are 52.8 deg and 46.7 deg, respectively.

Table 2 Crossover frequencies and phase margins for the different conditions in the simulation.

Condition	ω_c (rad/s)	φ_m (deg)
DYN1	2.77	52.8
DYN2	2.33	46.7
DYN1 CE with DYN2 HO	0.359	143
DYN2 CE with DYN1 HO	2.73	-24.6

B. Simulation Analysis Methodology

The goal of the analysis of the simulation data is to identify differences in the properties of different signals throughout separate phases of the simulation, as this could be something that a HO can use to detect that there has been a change in CE dynamics. The candidate signals that will be analyzed are the system output y, the system output rate $\dot{y}$, the control input u, the control input rate $\dot{u}$, the error e, and the error rate $\dot{e}$. These were chosen because they are the signals that can either be explicitly seen or can be visually estimated by the HO during a trial.

Each trial in the simulation lasts 120 s. However, all of the forcing functions were designed to have a period of 30 s, meaning there are exactly four periods in each trial [14, 15]. This is useful to compare different phases of the simulation because, if a whole period is considered, there is no dependence on the local forcing function properties. As such, each trial is divided into four 30 s phases, the first of which (0 s < t < 30 s) is used as run-in time, the second (30 s < t < 60 s) is the pre-transition steady-state phase, the third (60 s < t < 90 s) is the transition region, and the fourth (90 s < t < 120 s) is the post-transition steady-state phase [14]. The transition region is further narrowed down to a phase called the "detection period", which is 2.2 s to 7.4 s after the transition, and represents the period of time where a detection is expected [14]. Since the transition always occurs at t = 75 s (as M = 75 s in Table 1), the detection period is 77.2 s < t < 82.4 s.

A total of nine forcing functions were analyzed with the simulation to ensure the conclusions are not specific to local forcing function properties. For each forcing function, fifteen remnant realizations were simulated, resulting in 135 trials in each condition (DYN1, DYN2, DYN12, and DYN21). The reasoning behind using specifically nine forcing functions and fifteen remnant realizations for each is explained in subsubsection III.A.3 and subsubsection III.A.5.

The analysis of the results took place in two steps. In the first step, the standard deviation (σ) of the six candidate signals was compared over the different phases of the simulation. Of particular interest is the comparison of the pre-transition steady-state standard deviations with the standard deviations of the constant HO in the detection period, since this is representative of a HO that has not yet detected the change in CE dynamics. Thus, the goal in this step is to identify the signals that have the largest differences in their properties before and immediately after the transition, because that makes it likely that they are important in the process of detecting a change in CE dynamics.

In the second step of the analysis, a model is made using the signals identified to be the most relevant in the first step, and the accuracy of the model is evaluated. An illustration of the method used to evaluate the accuracy of the model is given in Figure 3. The model triggers (i.e., predicts the HO would detect a change in CE dynamics) whenever the instantaneous value of a particular signal exceeds a particular threshold, of which a whole range is evaluated [14]. The threshold is a multiple of the steady-state standard deviation of the respective signal, which is calculated from the time-invariant trials in the simulation (this is the only point in the simulation analysis where the time-invariant trials are

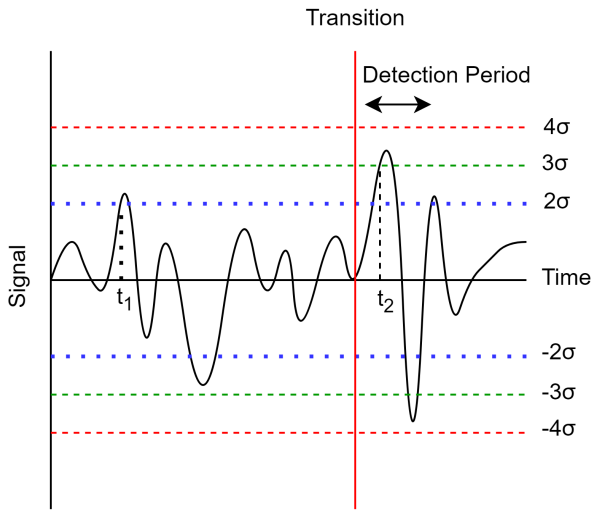


Fig. 3 Illustration showing how the second step of the simulation data analysis is done.

used) [14]. A total of 135 trials in DYN12 and 135 trials in DYN21 were evaluated. If, for a particular trial, the threshold was exceeded before the transition, as would be the case at time t_1 for a threshold of 2σ in Figure 3, it counts as a false positive (FP, "too early detection"). If the threshold is *not* exceeded before the transition and *is* exceeded in the detection period, as would be the case at time t_2 for a threshold of 3σ in Figure 3, it counts as a true positive (TP). Finally, if the threshold is never exceeded, as would be the case for a threshold of 4σ in Figure 3, it counts as a false negative (FN) [14]. The accuracy of the model can then be calculated using the equation below. Note that it is impossible to have a true negative (TN) in the simulation since only the time-varying trials were analyzed with the model, but it is included in Eq. (5) for completeness and because it will be needed when evaluating the experimental results in section IV.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{5}$$

C. Simulation Results

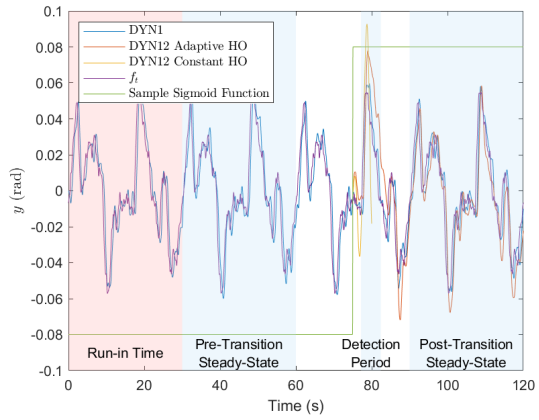
1. Sample Time Traces

To get a better feel for how the simulation results look in the time domain, sample time traces for each of the six candidate signals in each of the four CE dynamics conditions are presented in Figure 4-6 (for one forcing function and remnant realization). The different phases of the simulation introduced in subsection II.B can be seen by the shaded regions in the plots. Furthermore, a sample sigmoid function (Eq. (4)) is also shown in Figure 4a, where it can be seen that the transition is almost instantaneous. Note that there is only one line plotted in each of the figures before the transition because the adaptive HO and the constant HO are the same during that phase of the simulation. After the transition, the signals are shown for both HO models. In DYN12, the constant HO signals are only shown for 5 s following the transition because they become unstable and start congesting the figures.

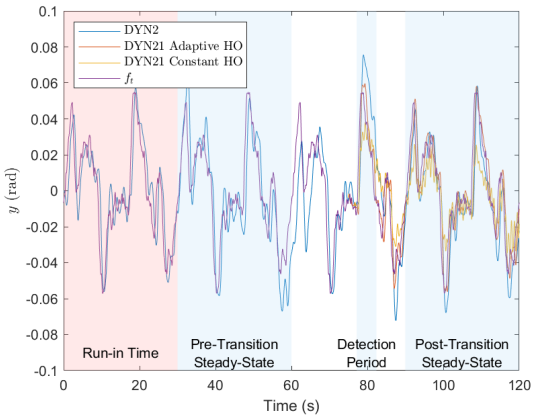
Several main observations can be made. First, it can be seen that the tracking performance is better in DYN1 than in DYN2, as expected due to the more difficult nature of the DYN2 condition. Second, the adaptive HO is able to quickly adapt to the CE dynamics post-transition and resume steady-state tracking. Finally, the constant HO post-transition in DYN21 gives very aggressive control inputs (Figure 5b), resulting in large (and quickly oscillating) values for $\dot{y}$ (Figure 4d) that lead to oscillations in y (Figure 4b). Thus, despite the the system with the constant HO not becoming unstable post-transition, the tracking performance is still significantly worse than that of the adaptive HO.

2. DYN12 Results

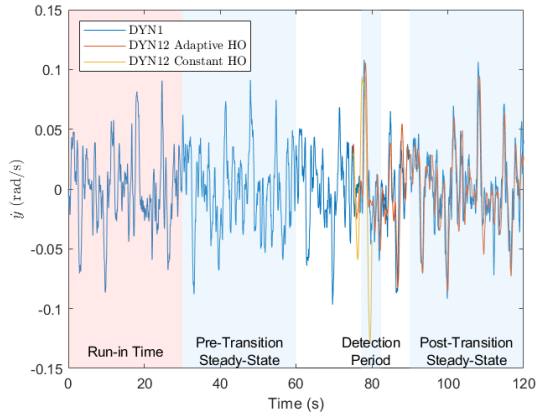
Next, the results for DYN12 are presented. In the first step of the analysis (as described in subsection II.B), a significant difference between the standard deviations in the pre-transition steady-state phase and the standard deviations of the constant HO in the detection period was found for all of the candidate signals since the system with the constant



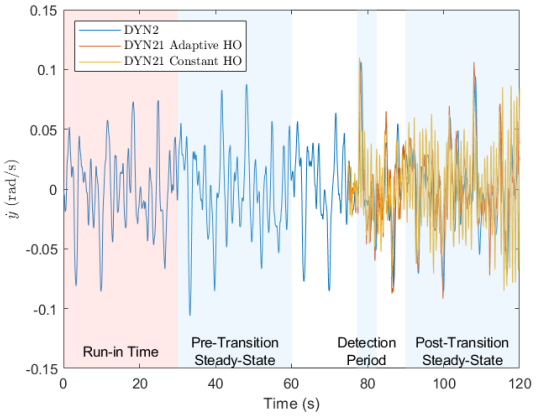
(a) y DYN1 and DYN12.



(b) y DYN2 and DYN21.

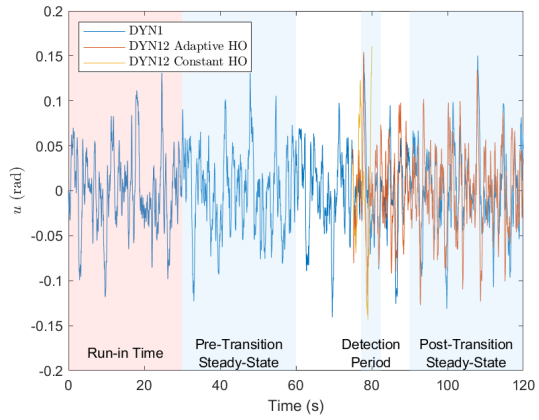


(c) $\dot{y}$ DYN1 and DYN12.

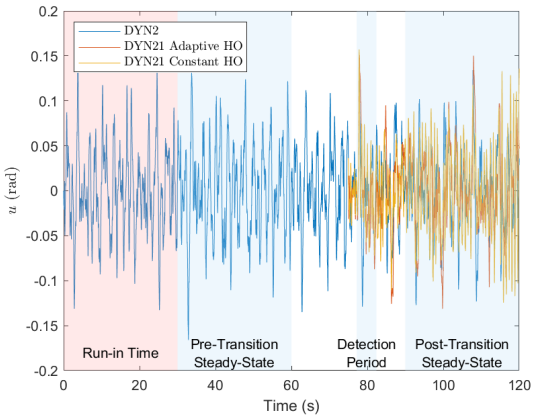


(d) $\dot{y}$ DYN2 and DYN21.

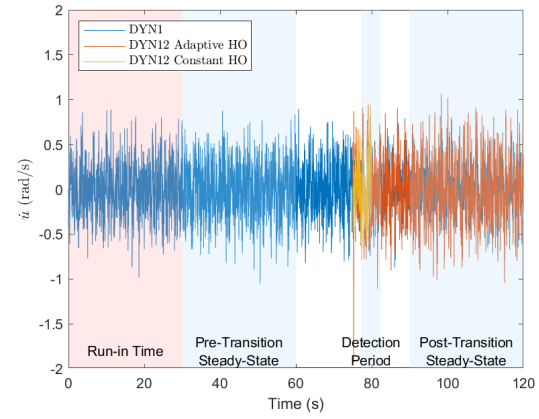
Fig. 4 Comparison of sample time traces for y and $\dot{y}$ in the four CE dynamics conditions, and between the constant HO and the adaptive HO post-transition.



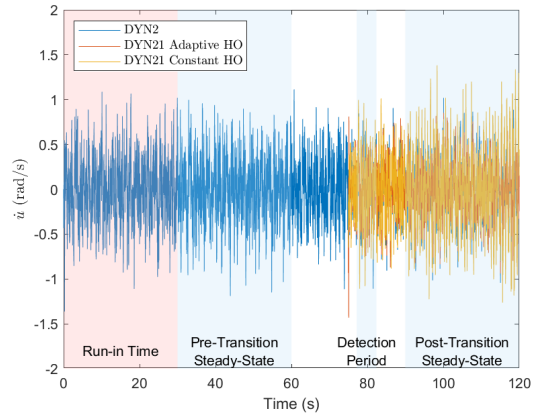
(a) u DYN1 and DYN12.



(b) u DYN2 and DYN21.

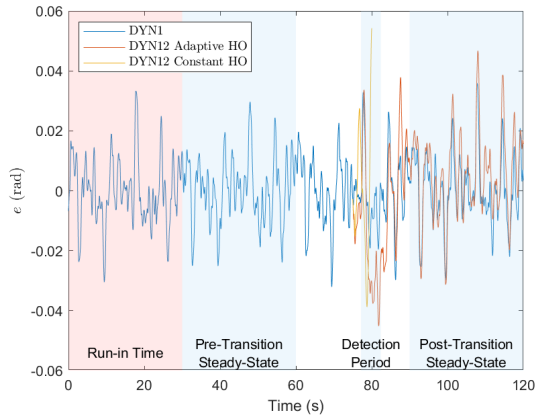


(c) $\dot{u}$ DYN1 and DYN12.

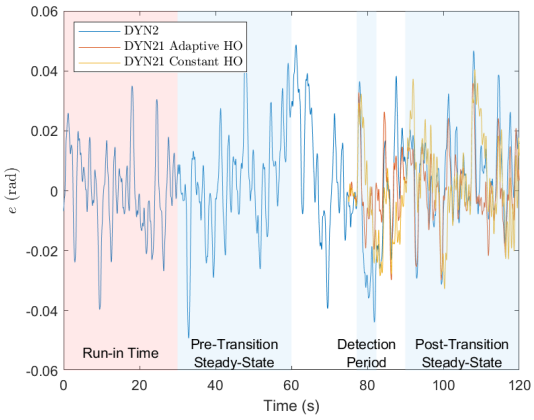


(d) $\dot{u}$ DYN2 and DYN21.

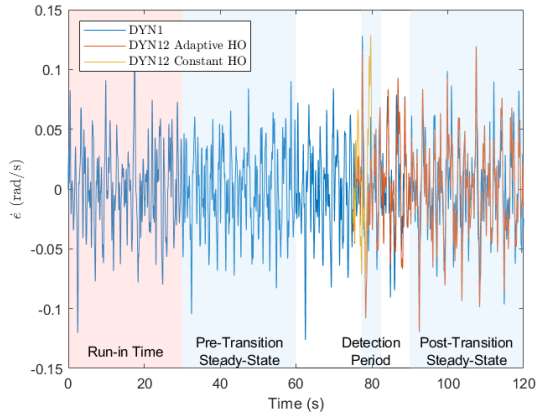
Fig. 5 Comparison of sample time traces for u and $\dot{u}$ in the four CE dynamics conditions, and between the constant HO and the adaptive HO post-transition.



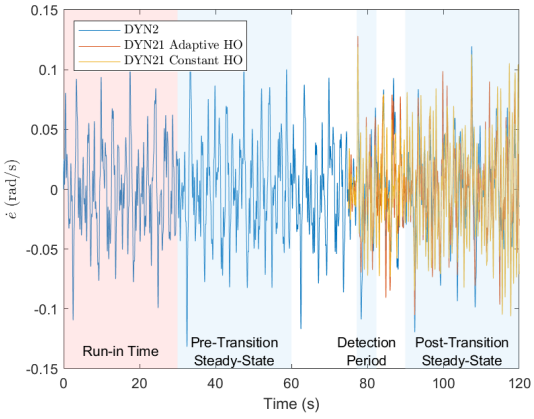
(a) e DYN1 and DYN12.



(b) e DYN2 and DYN21.

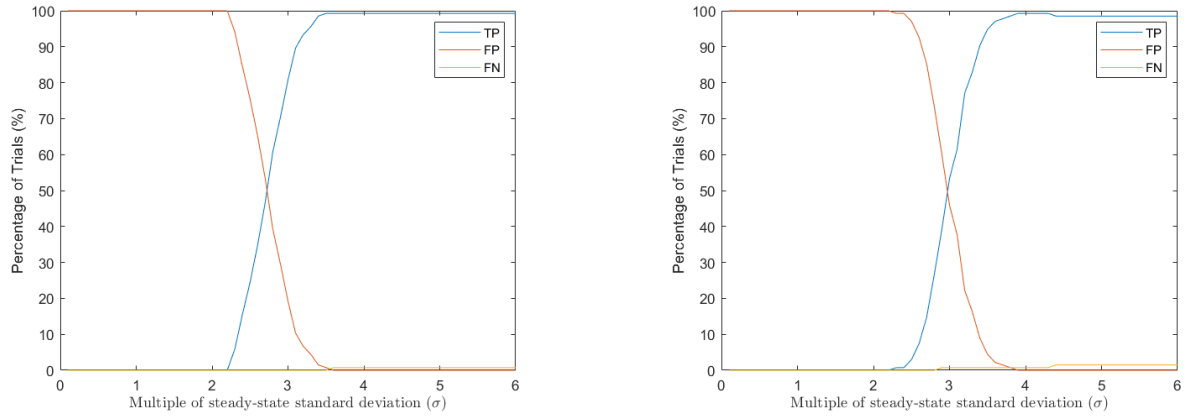


(c) $\dot{e}$ DYN1 and DYN12.



(d) $\dot{e}$ DYN2 and DYN21.

Fig. 6 Comparison of sample time traces for e and $\dot{e}$ in the four CE dynamics conditions, and between the constant HO and the adaptive HO post-transition



(a) Error (e). **(b) Error rate ($\dot{e}$).**

Fig. 7 Number of TPs, FPs, and FNs for the constant HO over various thresholds in the DYN12 condition.

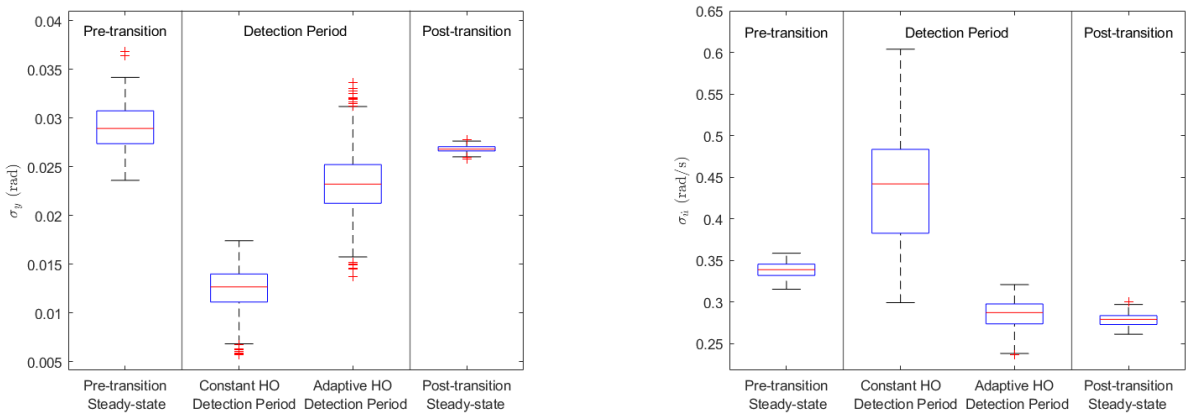
HO becomes unstable after the transition. For this reason, and due to space constraints, the figures from the first step of the analysis are not presented here, but can be found in [20].

Since it is impossible to tell which of the six candidate signals is the first to diverge, e and $\dot{e}$ will be used in the second step of the analysis. This is because previous research efforts (analyzing compensatory displays) have been successful in modeling the DYN12 transition using e and $\dot{e}$ [8, 9], and large increases to either is something that a HO can notice and use as a trigger to detect a change in CE dynamics. This is particularly true when using a compensatory display since they are the *only* two signals the HO can see (or estimate in the case of $\dot{e}$). The number of TPs, FPs, and FNs over different thresholds (i.e., different multiples of the steady-state standard deviation) are given in Figure 7. As can be seen, the number of FPs drops to zero for thresholds larger than 3.6σ for e and 3.9σ for $\dot{e}$ (the same threshold on $\dot{e}$ proposed by Van Ham [21]), while there is a TP in more than 99% of trials for a threshold larger than 3.5σ for e and more than 98.5% of trials for a threshold larger than 3.9σ for $\dot{e}$. Thus, above the threshold of 3.5σ and 3.9σ for e and $\dot{e}$, respectively, an accuracy greater than 98.5% can be achieved. This suggests that e and $\dot{e}$ could be good signals off which to base a model for the DYN12 transition.

3. DYN21 Results

Finally, the results for DYN21 are presented. In the first step of the analysis it was found that the biggest difference in standard deviations between the steady-state pre-transition phase and the constant HO in the detection period occurs for y and $\dot{u}$, as shown in Figure 8. The number of samples $N = 135$ in each box plot. As can be seen, there is a 56.2% decrease in the median standard deviation of y (Figure 8a) and a 30.4% increase in the median standard deviation of $\dot{u}$ (Figure 8b). However, there is still overlap with the standard deviations of the pre-transition steady-state phase in the case of $\dot{u}$, which makes developing a model more difficult. It can also be seen that there is a large difference between the constant HO in the detection period, and the adaptive HO in the detection period and post-transition steady-state phase, suggesting that adaptation is still beneficial in the DYN21 transition, despite the system with the constant HO not becoming unstable post-transition.

In the second step of the analysis, the accuracy of a model over a range of thresholds was evaluated (according to Eq. (5)), with the results shown in Figure 9. As can be seen, there is no threshold for any of the six candidate signals that results in an accuracy greater than 31%. The highest accuracy is achieved for $\dot{u}$, as predicted from the first step of the analysis, but it is very low in comparison to the model for DYN12 based on e or $\dot{e}$. As alluded to earlier, the overlap between the pre-transition steady-state standard deviations and the standard deviations in the detection period for the constant HO contributes significantly to the low accuracy. Regarding y, since the magnitude of the oscillations for the constant HO becomes smaller post-transition (Figure 4b), it is impossible to achieve an accuracy greater than 0% as there are no TPs for any threshold. Either the threshold is too low and results in a FP, or the threshold is too high and results in a FN. However, by introducing the error acceleration $\ddot{e}$ (which is related to $\dot{u}$ in DYN2) into the analysis, a model with an accuracy of 73.0% can be achieved with a threshold of 3.3σ. Visually estimating accelerations is difficult for humans [22], however, especially of e when using a pursuit display since the error itself has to be estimated. Thus, these results



(a) Output (y). **(b) Control input rate ($\dot{u}$).**

Fig. 8 Comparison of the standard deviations of y and $\dot{u}$ in the steady-state pre-transition phase with the constant HO in the detection period, the adaptive HO in the detection period, and the adaptive HO in the steady-state post-transition phase for the DYN21 condition.

suggest that detecting the DYN21 transition will be significantly harder than detecting the DYN12 transition, assuming HOs detect the transition based on a threshold for a particular signal, which is simply an assumption at this stage.

4. Simulation Analysis Conclusions

Overall, several conclusions can be made from the simulation results. First, it can be seen that in both conditions (DYN12 and DYN21) there are significant differences between the constant HO and the adaptive HO's control behavior. In DYN12, the system with the constant HO becomes unstable, while in DYN21 both HO models remain stable but the tracking performance of the adaptive HO is significantly better. Thus, detecting the transition and adapting to the new CE dynamics is crucial to maintaining good tracking performance. Second, for the DYN12 transition, a model based on e or $\dot{e}$ was able to achieve a high accuracy (>98.5%) suggesting these are good signals off which to base a model, which is consistent with previous research. [9, 14]. On the other hand, detecting the DYN21 transition may be more difficult because there are no significant differences in any of the six candidate signals. Only for a model based on $\ddot{e}$ could a high accuracy be achieved, but visually perceiving accelerations of a signal that has to be estimated itself is difficult for humans [22].

III. Method

A. Experiment Data

For developing and validating the model presented in this paper, a human-in-the-loop experiment was conducted at the Faculty of Aerospace Engineering of TU Delft. Details of the experiment will be provided here.

1. Control Task

The experiment consisted of 30 runs, excluding training runs, of a single-axis (horizontal) pursuit tracking task. An annotated representation of the pursuit display can be seen in Figure 10. The forcing function value was shown by the white square and the system output was shown by the green square, meaning participants controlled the green square and had to keep it as close to the white square as possible. The error $e(t)$ could be visually estimated by participants as the distance between the two squares. Participants were told that their primary goal was to keep the error as low as possible for the duration of the trial. If, at any moment, they believed the CE dynamics had changed, they were instructed to immediately press the trigger button on a Logitech joystick, separate from the one used to control the system, which they held in their off hand. The decision to use a separate joystick was based on a recommendation from a previous experiment [14] in which it was found that reaction times for detecting a transition could increase if both the

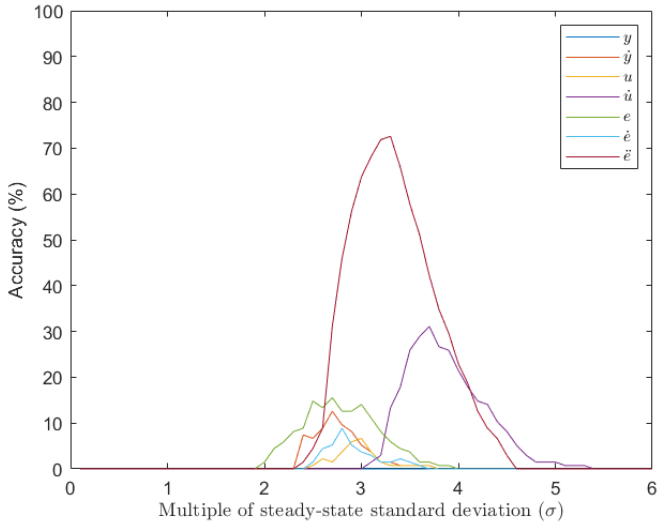


Fig. 9 Accuracy of a model based on each of the six candidate signals and the error acceleration $\ddot{e}$ over various thresholds for the DYN21 transition.

primary task of keeping the error as low as possible and the secondary task of pressing a button were done with the same hand on the same side-stick.

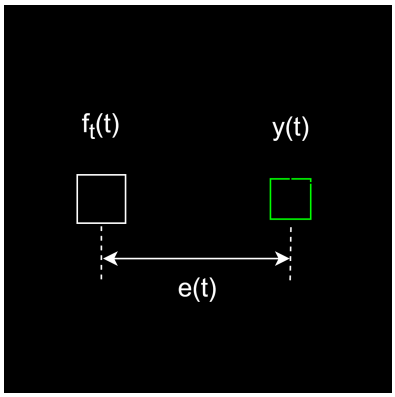


Fig. 10 Diagram of the pursuit display used in the experiment.

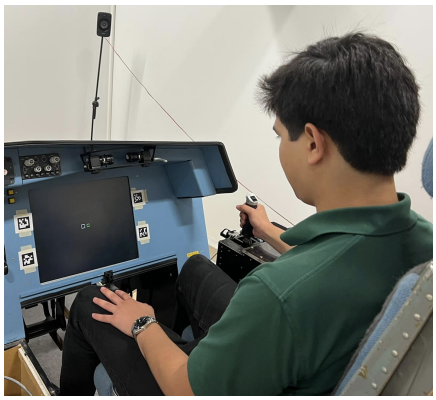


Fig. 11 Set-up during the experiment, with the exception of the joystick used for the button push.

2. Apparatus

The experiment was conducted in the fixed-base simulator in the Human-Machine-Interaction Laboratory at the Faculty of Aerospace Engineering of TU Delft. The set-up during the experiment can be seen in Figure 11. Participants sat in the right seat and gave control inputs using their right hand. The side-stick could only be moved left and right during the experiment, which consequently made the CE output move to the left or to the right, respectively. The maximum stick deflection was ±15 deg. The center of the stick is 9 cm above the axis of rotation. It had a stiffness of 3.54 N m rad^{-1}, a damping coefficient of 0.22 N m s rad^{-1}, an inertia of 0.01 kg m^2, and a breakout moment of 0 N m. The joystick used to indicate they noticed a change in CE dynamics (not pictured) was positioned on their left leg and help with their left hand.

3. Independent Variables

Two independent variables characterized each trial, namely the CE dynamics and the forcing function realization. The CE dynamics were varied to introduce time-varying behavior and allow for the investigation into the detection lags, and several forcing function realizations were used to investigate the effect of the forcing function properties around the transition moment, prevent predictability in the forcing function, and improve the generalization of the model [14].

Like in the simulation, there were a total of four CE dynamics conditions tested in the experiment, namely DYN1, DYN2, DYN12, and DYN21. The structure of the CE dynamics is the same as in the simulation and given in Eq. (1), and the beginning and end values of the gain k_c and break frequency ω_b are given in Table 3. Transitions between the beginning and end values of the two time-varying parameters were done according to the same sigmoid function as used in the simulation and given in Eq. (4). The value of G is 100 s^{-1} and M is 45 s (the middle of the measurement time).

Table 3 Parameters of the CE dynamics and sigmoid function for the four dynamics cases.

CE Dynamics	k_{c_1} (-)	k_{c_2} (-)	ω_{b_1} (rad/s)	ω_{b_2} (rad/s)	G (s^{-1})	M (s)
DYN1	15	15	20	20	-	-
DYN2	2	2	0.2	0.2	-	-
DYN12	15	2	20	0.2	100	45
DYN21	2	15	0.2	20	100	45

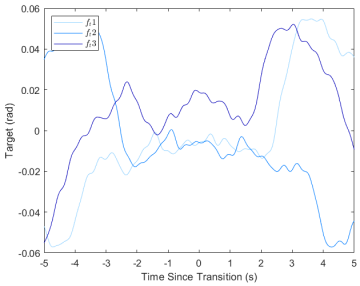
The forcing functions used in the experiment all follow the structure given in Eq. (6), where N_t is the number of sine waves used to construct the forcing function (ten in this experiment), and $A_t[i]$, $\omega_t[i]$, and $\phi_t[i]$ are the amplitude, frequency, and phase shift of the i^{th} sine wave, respectively. The amplitudes and frequencies used for each of the ten sine waves were kept the same for all forcing function realizations, with the phase shifts being the only difference. The amplitudes and frequencies are given in Table 4. It should be noted that, in order to avoid spectral leakage, the frequency of all of the sine waves must be an integer multiple k of the fundamental frequency, defined as $\omega_m = \frac{2\pi}{T_m}$, where T_m is the measurement time. For this purpose, the measurement time is taken to be 30 s.

$$f_t(t) = \sum_{i=1}^{N_t} A_t[i] \sin(\omega_t[i]t + \phi_t[i]) \tag{6}$$

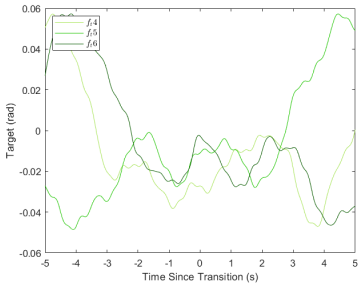
Table 4 Parameters (excluding the phase shifts) of the ten sine waves used to construct the forcing functions.

n (-)	k (-)	ω_t (rad/s)	A_t (rad)
1	2	0.419	$2.905 \cdot 10^{-2}$
2	5	1.047	$1.916 \cdot 10^{-2}$
3	9	1.885	$1.020 \cdot 10^{-2}$
4	13	2.723	$6.032 \cdot 10^{-3}$
5	19	3.979	$3.356 \cdot 10^{-3}$
6	27	5.655	$1.983 \cdot 10^{-3}$
7	39	8.168	$1.230 \cdot 10^{-3}$
8	51	10.681	$9.331 \cdot 10^{-4}$
9	67	14.032	$7.541 \cdot 10^{-4}$
10	83	17.383	$6.674 \cdot 10^{-4}$

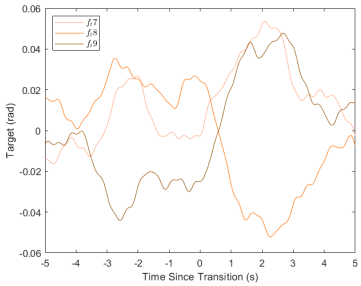
One of the reasons different forcing function realizations were used in the experiment is that previous experiments have found that the gradient of the forcing function around the transition moment can have a large impact on the detection lags [14, 16]. Specifically, larger gradients are expected to lead to lower detection lags. To quantify how large the gradients were around the transition moment, the power of the forcing functions was calculated for the two-second window centered around the transition. Consequently, three forcing functions were chosen that had low power, three that had medium power, and three that had high power, for a total of nine forcing function realizations used in the experiment (the same nine forcing functions were used in the simulation, hence the choice of specifically nine forcing functions in subsection II.B). What counts as "low", "medium", and "high" power was determined by looking at the distribution of power over an arbitrary two-second window in 1,000 forcing functions that had an average crest factor (calculated from



(a) Low power forcing functions (numbers 1, 2, and 3).



(b) Medium power forcing functions (numbers 4, 5, and 6).



(c) High power forcing functions (numbers 7, 8, and 9).

Fig. 12 All nine forcing functions used in the experiment, grouped by power level, in the 10 s window centered around the transition.

100,000 randomly generated forcing functions). From this, it was concluded that a power between 0 rad^2 and $2.5{\cdot}10^{-5}$ rad^2 can be considered low power, between $2.5{\cdot}10^{-5}$ rad^2 and $7.5{\cdot}10^{-5}$ rad^2 can be considered medium power, and above $7.5{\cdot}10^{-5}$ rad^2 can be considered high power. The forcing functions, numbered 1-9, are ordered in terms of power in the two-second window centered around the transition, with one being the lowest power and nine being the highest power. Time traces of the forcing functions, grouped by power level (low, medium, or high), in the transition region can be seen in Figure 12. The shades of the colors in each subfigure correspond to the amount of power in the two-second window centered around the transition, with the lightest shade being the least power and the darkest shade being the most power. Three of the nine forcing functions were used to collect data exclusively for validation, namely the middle one in each power level (forcing function numbers 2, 5, and 8). Thus, the validation forcing function for each power level is the line that is neither the lightest nor the darkest shade in each subfigure. The phase shifts and power in the two-second window centered around the transition for each forcing function are given in Table 5.

Table 5 Phase shifts and power in the 2-second window centered on the transition for the nine forcing functions. The forcing function numbers indicated with "(V)" are used to collect validation data.

	Forcing Function Number								
	1	**2 (V)**	**3**	**4**	**5 (V)**	**6**	**7**	**8 (V)**	**9**
ϕ_1, rad	6.065	3.454	0.885	4.315	5.608	3.671	1.040	3.219	6.172
ϕ_2, rad	0.223	3.073	1.120	3.308	6.027	2.757	1.958	5.615	2.662
ϕ_3, rad	4.232	5.440	4.806	0.988	3.106	5.220	2.037	4.049	0.903
ϕ_4, rad	1.880	1.296	4.251	3.925	3.109	4.162	4.768	4.820	2.290
ϕ_5, rad	4.757	4.062	1.260	6.259	5.506	3.685	1.522	4.160	6.112
ϕ_6, rad	1.072	3.464	4.764	6.236	5.560	5.188	2.820	0.440	2.162
ϕ_7, rad	2.754	0.153	5.692	0.609	5.668	5.261	1.520	2.525	4.821
ϕ_8, rad	1.593	1.678	3.003	4.782	2.511	5.699	1.577	3.368	0.101
ϕ_9, rad	5.404	4.585	0.116	2.776	1.259	5.566	0.114	0.213	6.090
ϕ_{10}, rad	4.101	0.583	1.338	4.126	2.843	0.776	0.755	1.461	1.853
Power Level	Low	Low	Low	Medium	Medium	Medium	High	High	High
Power, rad^2	$6.2 \cdot 10^{-6}$	$1.2 \cdot 10^{-5}$	$1.6 \cdot 10^{-5}$	$4.4 \cdot 10^{-5}$	$5.8{\cdot}^{-5}$	$6.3 \cdot 10^{-5}$	$1.3 \cdot 10^{-4}$	$2.2 \cdot 10^{-4}$	$2.3 \cdot 10^{-4}$

4. Experiment Procedures

Each participant performed a total of 30 measurement trials (i.e., excluding training trials) in the experiment, each with a unique combination of CE dynamics and forcing function. For each of the three validation forcing functions, one trial was performed in each of the two time-varying conditions (DYN12 and DYN21), for a total of six trials. For the other six forcing functions, one trial was performed in each of the four conditions (DYN1, DYN2, DYN12, and DYN21), for a total of 24 trials. The experiment was split into two halves by combining the fifteen DYN1 and DYN12 trials into

one half and the fifteen DYN2 and DYN21 trials into the other half. Within each half, the order of the conditions was determined via a random incomplete Latin square, so that participants could not predict whether a transition would occur in a particular trial.

The experiment started with several training runs in which participants could get familiar with the four CE dynamics conditions. Depending on the participant's performance, 1-5 training runs were done in each condition, usually two or three. If participants showed good tracking and detection performance on the first trial, only one additional training trial would be done in that condition, provided the performance remained stable. If the performance improved or was not at the level expected, an additional training run would be done in that condition until performance stabilized at a good level. All training runs used forcing function number 1.

Each trial lasted between 95 s and 105 s. The measurement time was always 90 s, but a random run-in time of either 5 s, 10 s, or 15 s was used to avoid participants anticipating the moment a transition would occur (since, if it was a time-varying condition, the transition would always occur 45 s into the measurement time) and to minimize the chance a participant could identify which forcing function was being used based on the starting position of the target. Despite the measurement time being 90 s, the period of all of the forcing functions was 30 s (hence why T_m in subsubsection III.A.3 is 30 s), which means there are exactly three periods of the forcing function in the measurement time. The first of those could be used to look at steady-state tracking in the first condition (e.g., DYN1 in DYN12), the second could be used to consider the transition period, and the third could be used to look at steady-state tracking in the second condition (e.g., DYN2 in DYN12).

If a participant made a large control error in any of the trials (due to a lack of concentration, for example) that trial would be re-done at the end of that half of the experiment. However, this only happened with one participant, where they had a FP in the first trial and two FPs in the second trial. For the remainder of that half of the experiment, the participant did not have any FPs or FNs and they stated the reason for the three FPs was a lack of concentration and sufficient knowledge of the CE dynamics (only one training run per time-varying condition was done with that participant).

5. Participants and Instructions

A total of sixteen participants took part in the experiment. One participant (Participant 8) had particularly poor tracking and detection performance and was thus excluded from the data presented in this paper. Thus, the data from fifteen participants are used to generate the results presented here, hence also fifteen remnant realizations were simulated per forcing function in subsection II.B. Each of the fifteen participants performed the experiment according to the order given in a different row of the Latin square to balance out order effects. As a result of the odd number of participants, and the inability to increase the number of participants to 30 due to time constraints, eight participants performed the DYN1 and DYN12 conditions first, while seven performed the DYN2 and DYN21 conditions first. All fifteen participants were university students or staff. They were chosen either based on the fact that they had experience with tracking tasks and had shown good performance previously, or had hobbies that generally translate to good performance in tracking tasks, such as playing video games that require quick reaction times (e.g., first-person shooting games) or simulator race car drivers.

Before starting the experiment, participants were briefed on the experiment procedure. Special attention was given to highlighting that their primary task throughout the experiment should be to keep the tracking error as low as possible, while the button push was a secondary task. However, in the event that they detected a change in CE dynamics, the button push could momentarily be elevated to the primary task. Participants were also told to push the button when they "think" the CE dynamics changed. After the training runs, no feedback was given to participants on their performance other than the root-mean-square error (RMSE) score that was shown on the display.

This experiment was reviewed and approved by the Human Research Ethics Committee at TU Delft under application number 3307.

6. Hypotheses

A total of four hypotheses were developed for the experiment, as listed below.

- **H1:** In both time-varying conditions all participants will adapt their control strategy to avoid the degradation in tracking performance observed in the computer simulations (i.e., their behavior will be similar to that of the adaptive HO).
- **H2:** In both time-varying conditions participants will be quicker to detect transitions in the CE dynamics when the gradients in the forcing function around the moment of transition are larger (quantified by the amount of power in the two-second window centered around the transition).

H3: In DYN12, a potential model with a high accuracy for predicting the HO's detection of a change in CE dynamics can be based on significant increases in the error and error rate signals.

H4: In DYN21, a potential model with a high accuracy for predicting the HO's detection of a change in CE dynamics cannot be based on any of the six candidate signals (y, $\dot{y}$, u, $\dot{u}$, e, $\dot{e}$).

H1, **H3**, and **H4** were formulated based on the conclusions from the simulation analysis (subsubsection II.C.4), while **H2** is based on Van Ham et al.'s findings [14].

B. Data Analysis

1. Performance Metrics

There are three primary performance metrics analyzed from the experiment data, namely the RMSE, crossover frequencies, and detection lags. The first two relate purely to tracking performance, while the latter has to do with detection performance.

The RMSE is useful for two reasons, one of which is to verify the quality of the experiment data and ensure it is representative of skilled HO behavior. For this purpose, the RMSE was calculated over the entire measurement time for all of the trials in DYN1 and DYN2. The other reason is that the RMSE in the final 30 s of the measurement time in the time-varying trials can be used to accept or reject **H1**. If participants always adapt their control behavior following a transition, the distribution of RMSE values when looking at the final 30 s of a time-varying trial (such that it covers one period of the forcing function) should be similar to the distribution of the RMSE over the six time-invariant trials in the same condition.

The crossover frequencies are also useful to verify the quality of the experiment data. For each trial, they were calculated using either the first or last 30 s of the measurement time, again to cover one period of the forcing function. Therefore, DYN12 trials were considered DYN1 trials when using the first 30 s of the measurement time, and DYN2 trials when considering the last 30 s of the measurement. The same can be said for the DYN21 trials but reversed. The crossover frequencies were then found by calculating the crossover frequency of the HO and CE frequency response. The HO frequency response was found by dividing the frequency response of the control input by the frequency response of the error *at the frequencies* of the ten sine waves which make up the forcing function ν_j, as given by Eq. (7). The corresponding phase margins were also calculated along with the crossover frequencies.

$$\hat{Y}_p(j\omega;\nu_j) = \frac{U(j\omega;\nu_j)}{E(j\omega;\nu_j)} \tag{7}$$

Finally, the detection lags were calculated by looking at how long it took between the transition (45 s into the measurement time) and the moment the button push occurred. This information is useful both to have a reference as to where to look in the time traces to determine what may have triggered the detection, as well as to be able to compare with the model predictions to determine the quality of the model.

2. Model with Multiple of Steady-State Standard Deviation as Threshold

The primary goal of the data analysis is to find a relationship between a certain property in one of the six candidate signals and the moment participants pressed the button to indicate they detected the change in CE dynamics. To do this a very similar method to what was used to develop the models for the simulation data (outlined in subsection II.B) will be followed. The model for the experiment data also uses a threshold based on the steady-state standard deviation of the respective signal and the moment it is exceeded is considered the moment the model predicts the HO would detect the change in CE dynamics.

Referring to Figure 3 and the accompanying explanation of it in subsection II.B, there are only a few differences with the method that will be followed here. First, for the analysis of the experiment data, the "detection period" is no longer considered, meaning a TP can occur at *any* moment after a transition and before the end of the trial. Second, it is possible to have TNs here, since some of the experiment trials were time-invariant. For these cases, a TN occurs if the threshold is *never* exceeded during the trial, and a FP occurs if the threshold is exceeded at any moment during the trial. In the time-varying trials, TPs, FPs, and FNs are calculated in the same way as was explained in subsection II.B (with the exception of TPs only being able to occur in the detection period, as just described). Finally, the only other difference with respect to the simulation data analysis is that for the experiment data, the reference steady-state standard deviation that is used to determine the threshold is calculated *per participant*, as opposed to using every participant's data. The accuracy of the model is then calculated using Eq. (5). As a second measure of the quality of the model, the

(time) difference between the model detections (when the threshold is exceeded) and the participant detections (when participants pressed the button) was analyzed.

IV. Results

A. RMSE

To verify the quality of the experiment data, the RMSE sorted by participant over the six time-invariant trials of DYN1 is shown in Figure 13a, alongside the RMSE for the six time-invariant trials of DYN2 in Figure 13b. As a reference, the right-most box plot in each subfigure is the data for *all* participants combined. Therefore $N = 6$ for each box plot except the right-most in each subfigure, where $N = 90$. It is clear that there is a difference in tracking performance between participants. However, different skill levels are to be expected between different people in any task, and the differences in our dataset are not large enough to consider a specific (group of) participant(s) as outliers or unskilled. This is supported by the fact that there are no data points considered outliers in the box plot for all of the participant's data (right-most box plot) in Figure 13a and only three outliers in Figure 13b. Thus, these results suggest that the quality of the experiment data is good, and all participants can be considered skilled HOs.

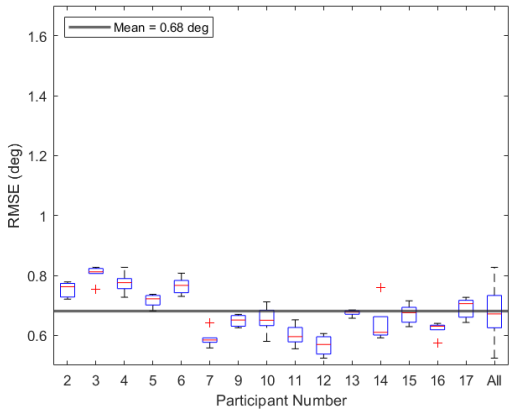
With regards to using the RMSE as a metric to determine if participants always adapted their control behavior to the new CE dynamics (as hypothesized by **H1**), the RMSE is sorted by participant for the last 30 s of the DYN21 trials in Figure 13c and the last 30 s of the DYN12 trials in Figure 13d. The former can be compared to the DYN1 trials (Figure 13a), while the latter can be compared to the DYN2 trials (Figure 13b). As can be seen, the distribution in RMSE for each participant is very similar between the two pairs, indicating that the control behavior was similar and thus that adaptation took place. One noticeable difference is that the spread in RMSE, quantified by the interquartile range (IQR), is somewhat larger in Figure 13c and Figure 13d than in Figure 13a and Figure 13b, respectively. However, this is to be expected given that the former two are calculated from only one period of the forcing function (30 s) while the latter two are calculated from three periods of the forcing function (90 s). Thus, mistakes, or lack thereof, have a bigger impact on the RMSE when only looking at a 30 s period.

As a final point, sorting the RMSE by forcing function and the order in which the conditions were performed was found not to impact the RMSE, see [20] for more details.

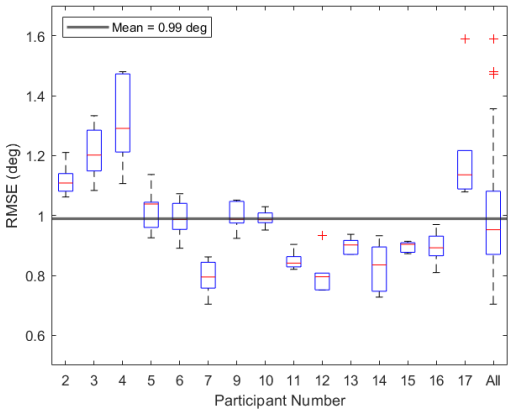
B. Crossover Frequencies

The crossover frequencies and phase margins sorted by participant, calculated using the first 30 s of the measurement time according to the method described in subsubsection III.B.1, are presented in Figure 14. The crossover frequencies and phase margins from the six DYN1 trials and six DYN12 trials that did not use the validation forcing functions are shown in Figure 14a and Figure 14c, respectively, while the equivalent results for the DYN2 and DYN21 trials are shown in Figure 14b and Figure 14d, respectively. Therefore, $N = 12$ in all of the box plots, except the right-most in each subfigure where $N = 180$ since it shows all of the participant's data combined. The scattered points overlayed are the results of the three DYN12 trials (or DYN21 trials for Figure 14b and Figure 14d) that used the validation forcing functions. The mean crossover frequency in DYN1 is 2.77 rad/s, while in DYN2 it is 2.33 rad/s. The phase margins are also good, with a mean of 55.6 deg in DYN1 and 41.1 deg in DYN2. In the context of verifying the quality of the experiment data, these results show that all participants had good crossover frequencies and phase margins, and no (group of) participant(s) performed significantly worse than the others. The latter can be seen by the fact that there is only one outlier in the box plot with all of the participant's data in Figure 14a, two in the box plot with all of the participant's data in Figure 14c and Figure 14d, and none in Figure 14b. Thus, like with the RMSE in subsection IV.A, these results suggest the experiment data is of good quality. Given the relationship between crossover frequency and RMSE, the results from Figure 14 are consistent with the results from Figure 13 in that participants who had low RMSE scores generally had high crossover frequencies. For the time-invariant DYN1 trials, the correlation coefficient ρ between the RMSE and crossover frequency is -0.81, and for the time-invariant DYN2 trials, ρ is -0.52, both of which can be considered high [23].

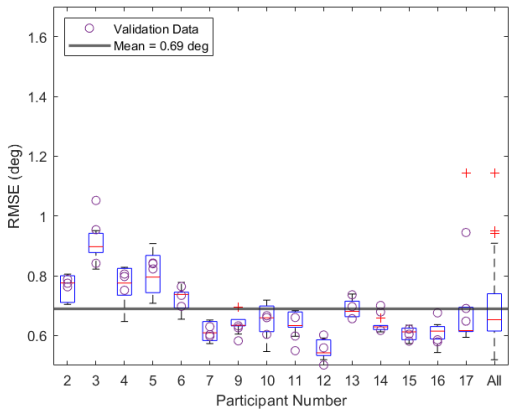
Figure 15 shows the crossover frequencies and phase margins calculated using the final 30 s of the measurement time in the six time-varying trials that did not use the validation forcing functions. Thus, $N = 6$ in each of the box plots, except the right-most one in each subfigure where $N = 90$ since it shows all of the participant's data combined. The scattered points overlayed are the results of the three time-varying trials that used the validation forcing functions. It should be noted that only the time-varying trials are included in Figure 15 because the goal of this figure is to compare it with



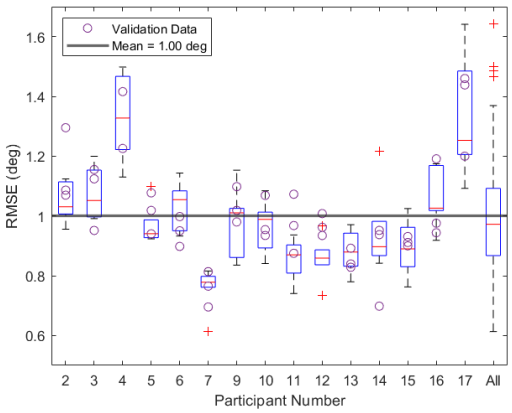
(a) Time-invariant DYN1 trials.



(b) Time-invariant DYN2 trials.

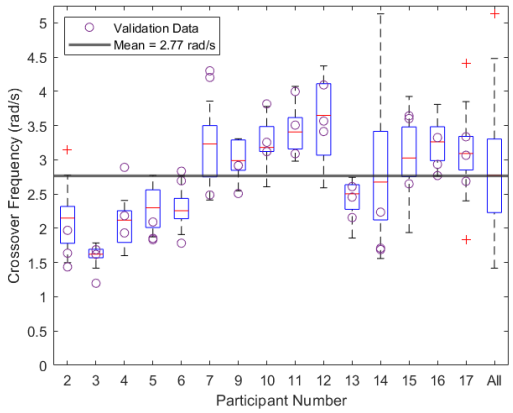


(c) Last 30 s of the DYN21 trials, so representative of DYN1.

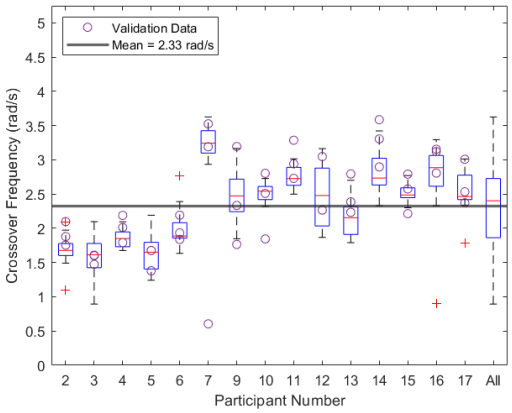


(d) Last 30 s of the DYN12 trials, so representative of DYN2.

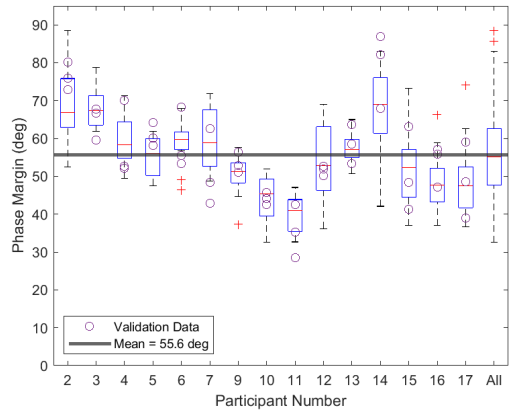
Fig. 13 RMSE by participant in different conditions and phases of the trials.



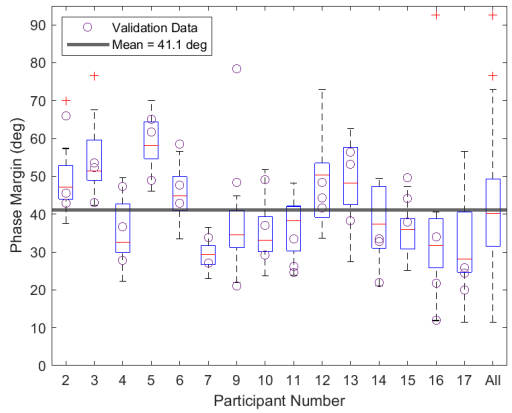
(a) DYN1 crossover frequencies.



(b) DYN2 crossover frequencies.

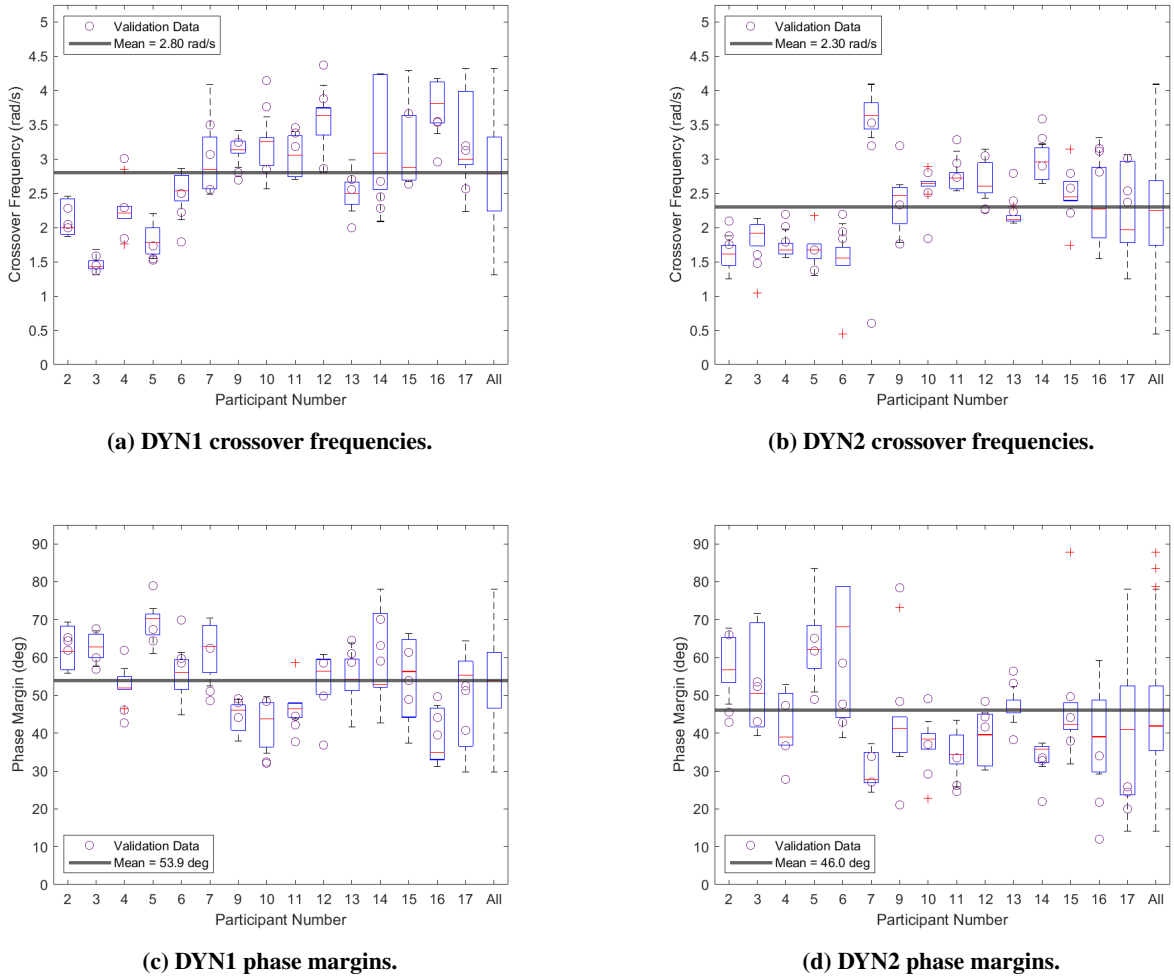


(c) DYN1 phase margins.



(d) DYN2 phase margins.

Fig. 14 Crossover frequencies and phase margins by participant calculated using the first 30 s of the measurement time in the six time-invariant trials and the six time-varying trials that did not use the validation forcing functions. The scattered points are the results of the three time-varying trials that used the validation forcing functions.



(a) DYN1 crossover frequencies.

(b) DYN2 crossover frequencies.

(c) DYN1 phase margins.

(d) DYN2 phase margins.

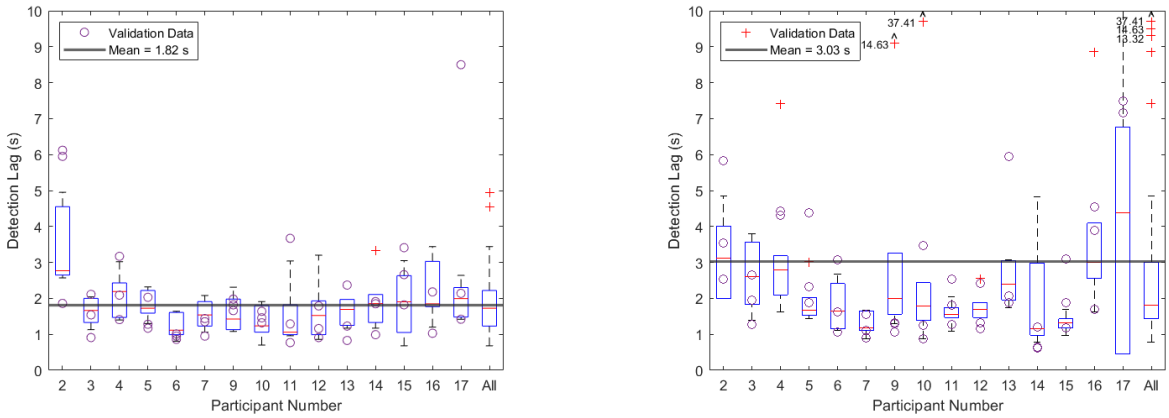
Fig. 15 Crossover frequencies and phase margins by participant calculated using the last 30 s of the measurement time in the six time-varying trials that did not use the validation forcing functions. The scattered points are the results of the three time-varying trials that used the validation forcing function.

Figure 14 to see whether adaptation to the new CE dynamics took place. As can be seen, the means and distributions are similar. Thus it can be concluded, like with the RMSE, that participants adapted their control behavior.

In comparison to similar previous experiments [14, 15], the crossover frequencies found here are significantly higher. Possible reasons for this are that the display used in this experiment is different (not only the fact that it is a pursuit display but also the symbols on the display are different), the CE dynamics were modified, and the horizontal axis was used in this experiment as opposed to the vertical (pitch) axis. Like with the RMSE, the forcing function and order of conditions were found to have minimal impact on the crossover frequencies.

C. Detection Lags and Detection Performance

The detection lags sorted by participant for the DYN12 transition are shown in Figure 16a, and the detection lags for the DYN21 transition are shown in Figure 16b. Again, the right-most box plot shows the data for all participants combined. Participant 2 had a FP in one of the DYN21 trials, thus $N = 5$ for that box plot, and Participant 17 had one FP and one FN in the DYN21 trials so $N = 4$ for that box plot. For the remaining box plots $N = 6$, except the right-most in each subfigure where $N = 90$ and $N = 87$ in Figure 16a and Figure 16b, respectively. The scattered points overlayed are the detection lags for the trials that used the validation forcing functions. Note that only TPs are shown and some



(a) DYN12. **(b) DYN21.**

Fig. 16 Detection lags by participant in the time-varying trials.

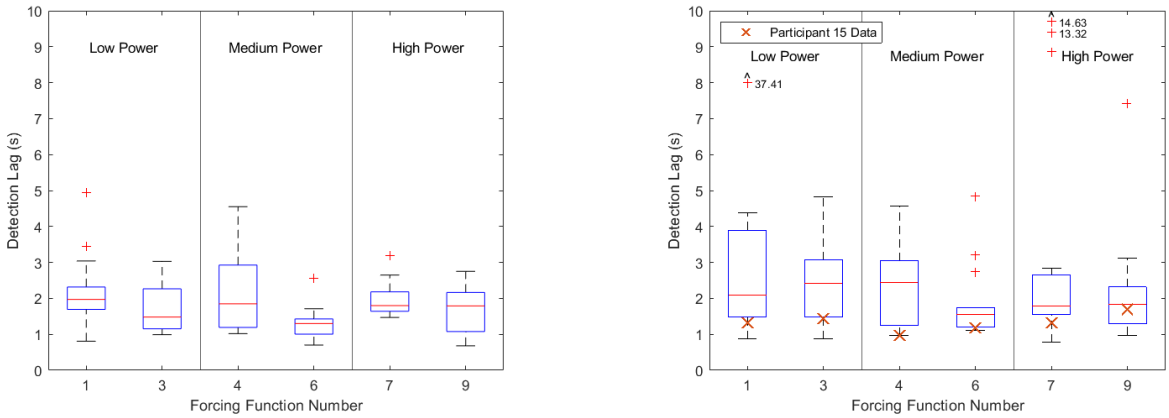
(outliers in the) detection lags in Figure 16b are outside of the y-axis range so are indicated by an upwards arrow along with the corresponding detection lag.

Despite the results from section II suggesting the DYN21 transition would be much harder to detect, it is interesting to note that the detection lags for the two conditions are not vastly different. The mean for the DYN12 transition is 1.82 s compared to 3.03 s for the DYN21 transition. However, the medians are very similar, with a median detection lag of 1.73 s in DYN12 and 1.83 s in DYN21. The large difference between the mean and median in DYN21 is evident from the box plot with all of the participant's data (right-most box plot) in Figure 16b, where it can be seen that the median is much lower than the mean, and it is the third quartile line that is approximately equal to the mean. This indicates that the mean is being dragged up by a few slow detections. It can also be seen in Figure 16 that the spread per participant is generally larger for the DYN21 transition than for the same participant in the DYN12 transition. The larger spread can be seen by the bigger IQRs in Figure 16b compared to Figure 16a. For DYN12, the mean IQR over all participants is 0.93 s, while for DYN21 it is 1.35 s. The larger variability suggests the DYN21 transition was more difficult to detect.

In comparison to similar previous experiments, the detection lags in this experiment were much lower. Van Ham et al.[14] and Plaetinck et al. [24] both performed compensatory tracking experiments investigating the detection of the DYN12 transition where participants were also instructed to press a button when they detected a transition. The median detection lag for Van Ham's experiment [14] was 6.6 s and for Plaetinck's experiment [24] it was 5.2 s (both significantly higher than the median for DYN12 *and* DYN21 in this experiment). Two factors likely contributed to this. The first is that the break frequency ω_b of the CE dynamics in DYN1 was increased from 6 rad/s to 20 rad/s for this experiment, so the system responded more like pure a single integrator in the crossover region and made the two conditions more distinct. The second is the fact that a pursuit display was used in this experiment, which will be further discussed in section V.

The detection lags sorted by (non-validation) forcing function can be seen in Figure 17a for DYN12 and in Figure 17b for DYN21. Again, the outliers that exceed the y-axis range in Figure 17b are shown by an upwards arrow along with the corresponding detection lag. As can be seen, there is no clear downward trend in either condition, so the impact of the forcing function was not as large as expected by **H2**. For DYN21, it can be seen that the spread decreases for the latter forcing functions, suggesting the transition is slightly easier to detect when the gradients in the forcing function are large. However, the fact that the lower bound is similar across all forcing functions suggests that the forcing function does not have a very big influence for a well-trained HO. For example, looking at the data for Participant 15 (the scattered crosses in Figure 17b), who arguably had the best detection performance in DYN21 (and good tracking performance in both DYN1 and DYN2), it is clear that increasing gradients in the forcing function did not lead to smaller detection lags.

For the DYN12 detection lags, a repeated-measures ANOVA test revealed no significant differences in the means depending on the power in the forcing function in the transition region $F(2, 58) = 1.71$, $p = 0.19$. For the DYN21 detection lags, Mauchly's test indicated that the sphericity assumption had been violated $\chi^2(2) = 26.2$, $p = <0.001$. As a result, the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = 0.61$). Again, the results revealed no significant difference in the means depending on the power in the forcing function in the transition



(a) DYN12. **(b) DYN21.**

Fig. 17 Detection lags sorted by forcing function.

region $F(1.22, 33.0)$, $p = 0.353$. Note that the two (non-validation) forcing functions in each power level were combined for the ANOVA test, such that the comparison was between low power, medium power, and high power.

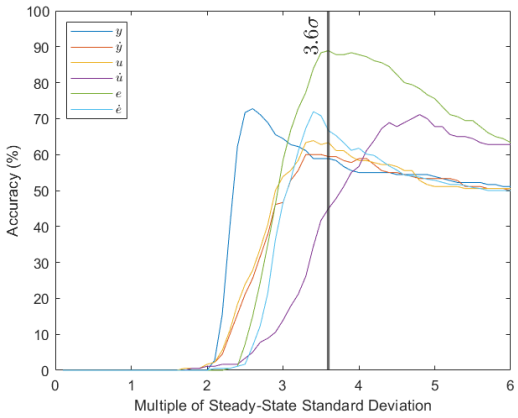
Overall, there were a total of twelve FPs, split over six participants, and one FN. Thus, the accuracy over all of the participants, calculated according to Eq. (5), is 97.1%. Of the twelve FPs, seven were in DYN2, two were in each of DYN21 ("too early detections") and DYN1, and one was in DYN12 (also a "too early detection"). The FN was in DYN21. Therefore, the participant accuracy in DYN1 and DYN12 was 98.7%, and in DYN2 and DYN21 it was 95.6%, again suggesting the DYN21 transition was (slightly) more difficult to detect. Six of the twelve FPs occurred during one of the first three trials in the experiment (excluding training trials), and a further three occurred in one of the first three trials of the second half of the experiment, where the set of conditions being tested was different from the first half. This suggests that either there was insufficient training, or the level of confidence participants used to press the button was too low, and after realizing it resulted in too many false button pushes, they adjusted and waited a bit longer before pressing the button.

D. Model for DYN12 Transition Detection

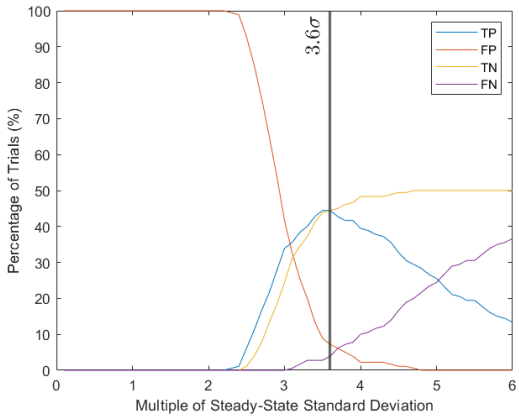
Using the data reserved for developing the model and the methodology given in subsubsection III.B.2, a model based on each of the six candidate signals analyzed can be made and evaluated. In Figure 18a, the accuracy of a model based on the different signals over various thresholds can be seen. As expected based on **H3**, e is the signal that results in the highest accuracy, with a maximum accuracy of 88.9% at a threshold of 3.6σ. For the validation data set, the accuracy at this threshold is 84.4%. The proposed 3.6σ threshold is very similar to Van Ham [14], who proposed a threshold of 3.9σ on the error. In Figure 18b, the breakdown of the number of TPs, FPs, TNs, and FNs over the same range of thresholds can be seen. At the threshold of 3.6σ, there are 80 TPs (out of 90 possible), 13 FPs, 80 TNs, and 7 FNs. Since 50% of the trials considered here (i.e., excluding the trials used to collect validation data) were time-varying, a 100% accuracy would be achieved if there was a TP in 50% of the trials and a TN in 50% of the trials.

In Figure 19, a comparison can be made between the model detections (i.e., when the threshold is exceeded), and participant detections (i.e., when participants pressed the button). The ordering of the trials on the x-axis is arbitrary, as the intention is merely to show a comparison of the model detections with the participant detections for each trial. However, the trials are numbered according to the following (arbitrary) structure. First, the DYN1 trials are grouped (trials 1-90), and the DYN12 trials are grouped (trials 91-180). Within each half, the trials are ordered first by participant and then by forcing function number (in ascending order). Therefore, trials 1-6 are the six DYN1 trials for Participant 2 in order of ascending forcing function number, trials 7-12 are the six DYN1 trials for Participant 3 in order of ascending forcing function number, etc. The same applies to trials 91-180 but with the DYN12 trials instead of DYN1. Therefore, if a detection occurs in one of the areas shaded in red, it is a FP (the darker red is a FP in DYN1 and light red is a "too early detection" in DYN12), while if a detection occurs in the area shaded in green, it is a TP.

The first observation that can be made is that the model has more FPs than the participants, which is the source of the lower model accuracy compared to the accuracy of the participants in the DYN1 and DYN12 trials. Focusing on the



(a) Accuracy of a model based on each of the signals over a range of thresholds.



(b) Breakdown of the number of TPs, FPs, TNs, and FNs for a model based on e.

Fig. 18 Analysis of which signal results in the best accuracy for modeling the DYN12 transition.

TPs, it can also be seen that there are a considerable amount of trials where the model detections are slow. The slowest participant detection is 4.95 s, whereas there are thirteen trials where the model detections are slower than 4.95 s.

Overall, there are 80 trials (out of 90 DYN12 trials) where both the model and the participant have a TP. For these 80 trials, the mean difference between the model's detection lags and the participant's detection lags is 3.46 s. However, it is interesting to consider the cases where the model detections are quicker and the cases where the participant detections are quicker separately. Out of the 80 trials where both the model and the participant have a TP, the model detections were quicker in 42 trials and the participant detections were quicker in the remaining 38 trials. In the 42 trials where the model detections were quicker, the mean difference between the model detections and the participant detections is 0.61 s, which is close to a human reaction time [25]. In the 38 trials where the participant detections were quicker, the mean difference between the model detections and participant detections is 6.62 s. This suggests that in approximately half of the trials (the 42 trials where the model detections were quicker), the model detections are a good estimate of the actual participant detection lags. However, in the other half of the trials (the 38 trials where the participant detections were quicker), the model detections are not representative of the actual participant detection lags. A possible reason for this is that there is a different, or multiple, mechanisms at play in the detection phase. For example, it could be that the participant detections were triggered by a filtered version of e. Another possibility is that the detection was based on the proposed threshold for e in some cases (e.g., the 42 trials where the model detections were quicker), but on a threshold for $\dot{e}$ or the observed relationship between the control inputs and the system's response in other cases.

Overall, it can be concluded that the model proposed based on a threshold for e is capable of achieving a high accuracy in detecting *if* a transition occurred, but not as good at estimating the participants' detection lags.

E. Model for DYN21 Transition Detection

Following the same procedure that was just described for the DYN12 transition, a model can be defined for the DYN21 transition. The accuracy for all of the signals over a range of thresholds can be seen in Figure 20a. Note that an accuracy greater than 0.5 cannot be achieved using the six candidate signals, as hypothesized by **H4**. As a result, it was decided to include the output acceleration $\ddot{y}$ in the analysis as well. Further discussion on the implications of this choice of signal will be given in section V. However, with a model based on $\ddot{y}$ it is possible to achieve an accuracy as high as 99.4% at a threshold of 4σ. This results in a total of 90 TPs, 1 FP, 89 TNs, and 0 FNs. The accuracy of the validation data set at a threshold of 4σ is 100%.

In Figure 21, the same comparison that was shown in Figure 19 for the DYN12 model is presented for the DYN21 model. Again, the ordering of the trials on the x-axis is arbitrary, but is done in the same way as in Figure 19. Contrary to the model for DYN12, there are more participant FPs in this case, with only one model FP.

Out of the 90 DYN21 trials, there are 87 trials in which both the model and the participant had a TP. For these 87 trials, the mean difference between the model's detection lags and the participant's detection lags is 3.12 s. Again, it is interesting to consider the cases where the model detections were quicker and the cases where the participant detections

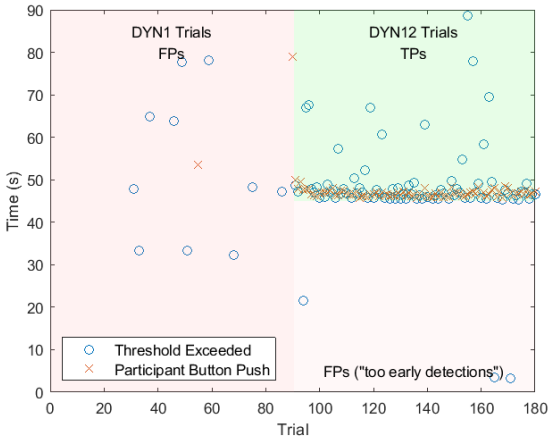
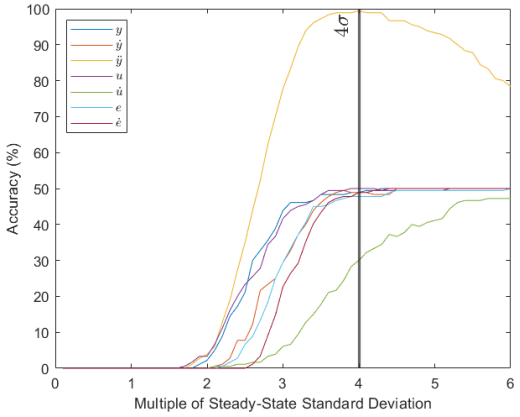
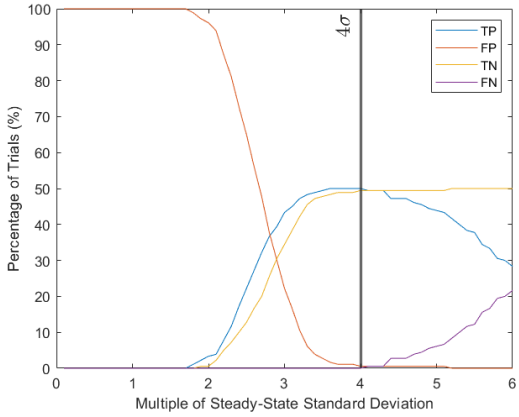


Fig. 19 Comparison of the moment the threshold is exceeded in the model and when participants pressed the button. FPs appear on the left (DYN1) and bottom half of the right (DYN12 but before the transition) while TPs appear on the top half of the right side.



(a) Accuracy of a model based on each of the signals over a range of thresholds.



(b) Breakdown of the number of TPs, FPs, TNs, and FNs for a model based on $\ddot{y}$.

Fig. 20 Analysis of which signal results in the best accuracy for modeling the DYN21 transition.

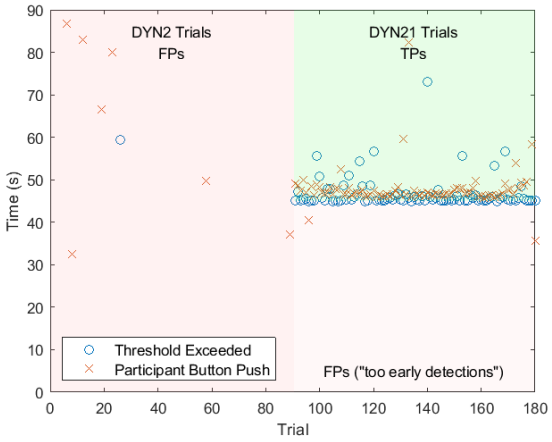


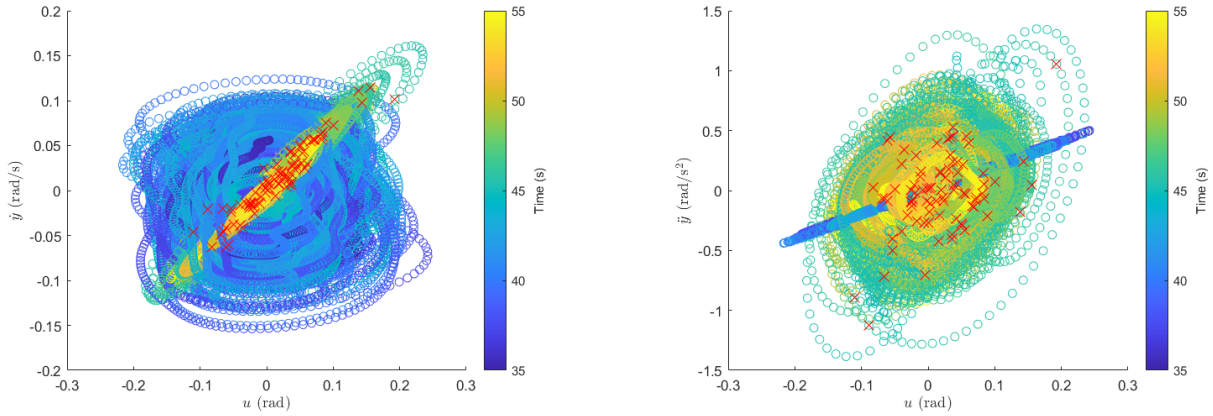
Fig. 21 Comparison of the moment the threshold is exceeded in the model and when participants pressed the button. FPs appear on the left (DYN2) and bottom half of the right (DYN21 but before the transition) while TPs appear on the top half of the right side.

were quicker separately. Here, there is a big difference in the number of cases of each, with the model detections being quicker in 72 of the 87 trials, and the participant detections being quicker in the remaining 15 trials. When the model detections are quicker, the mean difference between the model's detection lags and the participant's detection lags is 2.59 s. Over the same trials, the mean participant detection lag is 3.08 s, which indicates that the model detections are extremely quick in these cases. In fact, in 29 of the 72 trials where the model detections are quicker, the model detection occurs before $t = 45.03$ s, so before the transition is even complete. A discussion on the implications of this will be provided in section V. In the 15 trials where the participant detections were quicker, the mean difference in detection lags between the model and the participant is 5.65 s. Thus, the model's detection lags are not representative of the actual participant detection lags in neither the case where the model detections were quicker nor the case where the participant detections were quicker.

Overall, it can be concluded that the DYN21 model is able to achieve a near-perfect accuracy in detecting the transitions, but often the model's detection lags are extremely small and are therefore not representative of the actual participant detection lags.

F. Relationship Between u-$\dot{y}$ and u-$\ddot{y}$ in DYN21

To illustrate one of the benefits of using a pursuit display over a compensatory display, the relationship between u and $\dot{y}$, as well as between u and $\ddot{y}$, is presented here. This represents the relationship between what participants do (i.e., give control inputs) and what they see on the screen in response when using a pursuit display. The relationship between u and $\dot{y}$ for the ten-second window centered around the transition in every DYN21 trial is shown in Figure 22a. The relationship between u and $\ddot{y}$ for all of the DYN21 trials over the same period is given in Figure 22b. The red crosses are the moment participants pressed the button. Prior to the transition, the DYN2 system is being controlled, which means control inputs are proportional to $\ddot{y}$. Thus, it is logical that the relationship is linear, as in Figure 22b. After the transition, the control inputs are proportional to $\dot{y}$, and thus a linear relationship emerges in Figure 22a. As a result of this, all of the button pushes occur in the linear portion of Figure 22a (some of the button pushes towards the bottom left of Figure 22a appear to be outside the linear portion but this is not the case as the linear portion is relatively wide in DYN1). In Figure 22b, the button pushes do not exclusively occur in the linear portion, and when they do, the trajectory taken to get there always goes outside of the linear portion first. Therefore, in addition to a threshold on a particular signal (e.g., $\ddot{y}$ as proposed in subsection IV.E), deviations from the expected (linear) relationship between u and $\dot{y}$ or u and $\ddot{y}$ can also act as a trigger for participants to detect a transition.



(a) u-$\dot{y}$ relationship. **(b) u-$\ddot{y}$ relationship.**

Fig. 22 Relationship between u-$\dot{y}$ and u-$\ddot{y}$ in the ten-second window centered around the DYN21 transition for all participants. The button pushes are indicated by the red crosses.

V. Discussion

The goal of this research was to gain a better understanding of what triggers a HO to detect a change in CE dynamics. A human-in-the-loop single-axis pursuit tracking experiment was performed in which the CE dynamics sometimes transitioned from approximate single to approximate double integrator dynamics or vice versa. If and when participants detected the change in CE dynamics, they pressed a button on a joystick to indicate that they detected it. A total of fifteen participants took part, each performing a total of six trials in DYN1, nine trials in DYN12, six trials in DYN2, and nine trials in DYN21. The data from all of the DYN1 and DYN2 trials, as well as six of the trials from each of DYN12 and DYN21 were used to propose a model that predicts when participants would press the button. It was found that the error signal e was best for modeling HOs' detection of the DYN12 transition, where an accuracy of 88.9% could be achieved for a threshold of 3.6σ. For DYN21, none of the six candidate signals (y, $\dot{y}$, u, $\dot{u}$, e, or $\dot{e}$) resulted in an accuracy greater than 50%, so instead the model is based on the output acceleration $\ddot{y}$, which resulted in an accuracy of 99.4% for a threshold of 4σ.

The experiment was successful in collecting high-quality data for developing the models. Throughout the entire experiment, the participant detection accuracy was 97.1%, meaning in the large majority of cases they were able to detect when the CE dynamics changed. The median detection lags in this experiment were also small compared to previous experiments [14, 24], with a median of 1.73 s in the DYN12 transition and 1.83 s in the DYN21 transition. The mean crossover frequencies were high as well, with a mean of 2.77 rad/s in DYN1 and 2.33 rad/s in DYN2. A similar experiment was previously conducted [15], where the mean participant crossover frequency was only 0.97 rad/s in DYN1 and 1.67 rad/s in DYN2. Such crossover frequencies make it difficult to develop a good, representative model of HO control behavior because they do not really represent skilled HO control behavior. Thus, the results presented here seem to be more generalizable to skilled HO behavior, which is ultimately where the applications of this research lie.

The first hypothesis investigated, **H1**, predicted that all participants would adapt their control behavior after a transition to approximate the behavior of the adaptive HO in the simulation (subsection II.C). In every DYN12 trial of the experiment, participants were able to stabilize the system after the transition and continue tracking the forcing function with the new CE dynamics while keeping the system output symbol on the display until the end of the trial (even if it momentarily left the display immediately after the transition). On the other hand, it was found that the system with the constant HO in the simulation became unstable after the transition, so this shows that the participants all adapted their control behavior in the DYN12 trials. In DYN21, the system with the constant HO did not become unstable in the simulation, but did suffer from worse tracking performance than the adaptive HO after the transition. It can be seen in Figure 13c and Figure 15 that this was not the case for participants in the experiment, as the distribution of RMSE and crossover frequencies calculated using the last 30 s of the DYN21 trials was equivalent to that of steady-state tracking in DYN1. Thus, **H1 is accepted.**

The second hypothesis, **H2**, predicted that larger gradients in the forcing function in the transition region would lead to smaller detection lags. To quantify how large the gradients in the transition region were, the power in the forcing

functions in the two-second window centered around the transition was used. Thus, the higher the forcing function number, the larger the gradients in the transition region. From Figure 17 and the repeated-measures ANOVA tests (subsection IV.C) it can be concluded that larger gradients in the transition region did *not* lead to smaller detection lags, and thus **H2 is rejected.** Despite the fact that Van Ham et al.'s experiment [14] has previously found that large gradients in the forcing function *did* generally lead to smaller detection lags, there is one crucial difference with this experiment that likely explains it, namely the type of display. Van Ham et al.[14] used a compensatory display, whereas a pursuit display was used in this experiment. With the pursuit display, participants can see the relationship between their control inputs and the system's response, whereas this relationship is impacted in an unknown way (to the participant) by the forcing function when using a compensatory display.

Knowledge of this relationship is less important when the gradients in the forcing function are large because large control inputs are required to track the large gradients. This causes the magnitude of the majority of the signals in the control loop to increase. Therefore, if exceeding a threshold (e.g., on e for DYN12 and on $\ddot{y}$ for DYN21) is a trigger for the human operator to detect a change in CE dynamics, the large control inputs will make it more likely that this threshold is exceeded.

On the other hand, knowledge of the relationship between the control inputs and the system's response can be more useful when the gradients in the forcing function are small. This is because the magnitude of the relevant signal may not be large enough to exceed the threshold. In this case, participants can use the fact that, in DYN1, u is proportional to $\dot{y}$, while in DYN2, u is proportional to $\ddot{y}$. Thus, in DYN1, moving the side-stick to the left or right causes the system output to *immediately* move in that direction. On the other hand, in DYN2, it is necessary to give a control input in the opposite direction to the direction of $\dot{y}$ to *first* slow it down, and *only then* it will start moving in the direction of the control input. Therefore, if participants notice that the relationship between the control inputs and the output has changed, namely that the sign of $\dot{y}$ does not immediately change when the sign of u changes for the DYN12 transition, or that the sign of $\dot{y}$ *does* immediately change when the sign of u changes for the DYN21 transition, that can also be used as a trigger to detect the transition. This relationship is something that cannot be seen by the HO with a compensatory display, so it is hypothesized here that this is why the gradients in the forcing function did not have a big impact on the detection lags.

The final two hypotheses, **H3** and **H4**, have to do with which signal is best to base a model on. For the DYN12 transition, e is the only one of the six candidate signals that results in a high accuracy. **H3** predicted the model could be based on e or $\dot{e}$ and is therefore **rejected**, even though the first half of the hypothesis is true. This outcome is consistent with previous research that found that a model with a threshold based on e performs well in detecting (specifically) the DYN12 transition [8, 9, 14]. The accuracy of the model with a threshold of 3.6σ on e is high, though the estimation of the detection lags can still be improved. In approximately half of the DYN12 trials, the model's detection lags were representative of the actual participant detection lags, but in the other half, there was a big difference in detection lags.

For the DYN21 transition, it can be seen in Figure 20a that none of the six candidate signals could be used to make a model with an accuracy greater than 50%. Thus, **H4 is accepted.** Instead, a model based on $\ddot{y}$ is proposed. Despite the fact that humans are not good at visually perceiving accelerations [22], there is still merit to the model because it is not argued that participants need to be able to estimate the value of $\ddot{y}$. Rather, they only need to be aware of significant changes to its properties, which, post-transition in DYN21, can be perceived as a system that is more aggressive, more responsive to control inputs, and less sluggish, or in the words of some of the participants, less "*oversteery*". $\ddot{y}$ also captures the the relationship between u and $\dot{y}$, and between u and $\ddot{y}$ described just above. This is because, in order for $\dot{y}$ to be proportional to u (as is the case in DYN1), it is necessary for $\ddot{y}$ to be larger in DYN1 than DYN2 so that the sign of $\dot{y}$ changes quicker after a change in the sign of u, and this can be used by participants to detect the transition.

A limitation of this model, however, is that the model detections (i.e., the moment the threshold is exceeded) are extremely quick, often occurring before $t = 45.03$ s for a transition at $t = 45$ s. This suggests the model could be fit to a property of the CE dynamics rather than something participants use to detect the transition, and thus may not be generalizable to other CE dynamics. On the other hand, it can be argued that this information *is* what participants use to detect the transition, but the fact that it is an acceleration and requires more precise knowledge of the expected response means it takes longer for the participant to notice.

The next step towards using this research in real-life applications is adjusting the models to more closely estimate the real HO detection lags. One complication in estimating the detection lags, however, is that it is not a binary decision to press the button, such as if someone was instructed to press the button when a picture pops up on the screen. Thus, there is a statistical element to the data [14]. Furthermore, the button push data are subjective, and different people have different thresholds for when they think they should press the button, in addition to cognitive decision time [25]. For this reason, it is recommended to introduce lags into the detection process. This can be done, for example, by passing the relevant signal through a low pass filter as done by Hess [13] or adding an explicit delay to model reaction times.

Another recommendation is to include u in the model. For this research, an attempt was made to include u in the threshold though it was unsuccessful in improving the estimation of the detection lags. Nonetheless, the results presented in this paper suggest that the relationship between u and the system's response is important, so additional investigation into how to include this in the model is recommended. One possibility is to develop a model that generates expectations of the behavior of u given the current y, $\dot{y}$, and f_t. Significant mismatch between the expectation and reality can then also be used as a trigger, or can be incorporated into the threshold (e.g., the higher the mismatch, the lower the threshold on $\ddot{y}$ for the DYN21 model). This is similar to Phatak and Bekey [9] and Van Ham et al. [14] who proposed independent thresholds on e and $\dot{e}$ but with a (possible) difference. The recommendation here is to try and include u in a *single* model such that it uses (at least) u and $\ddot{y}$ to determine a threshold, as opposed to a model that can trigger based on individually looking at different signals. To further understand the role that the relationship between the control inputs and system response plays, a similar experiment to the one performed for this research could be conducted but with several disturbance signals, each with differing amounts of power. If the relationship between the control inputs and the system's response is indeed critical, it should be found that increasing the power in the disturbance signal leads to larger detection lags because the relationship is lost.

Overall, however, this research provides the first steps to understanding how HOs detect a change in CE dynamics, particularly the DYN21 transition, using pursuit displays. A model for the DYN21 transition based on $\ddot{y}$ is proposed and recommendations for future improvements to the model as well as future investigations are made.

VI. Conclusion

This paper aimed to gain an understanding of what triggers a HO to detect a change in controlled element (CE) dynamics and see if this can be predicted with a model. A human-in-the-loop pursuit tracking experiment was performed to collect data to develop and validate the proposed models. Fifteen participants took part and had a combined accuracy of 97.1% in detecting the CE transitions. The mean detection lag was 1.82 s for the DYN12 transition and 3.03 s for the DYN21 transition. The observed crossover frequencies were high, with a mean of 2.77 rad/s for DYN1 and 2.33 rad/s for DYN2, indicating that skilled participants were selected. Two models were developed, one for DYN12 and one for DYN21. The model for DYN12 is based on e and has an accuracy of 88.9% at a threshold of 3.6σ. The model for DYN21 is based on $\ddot{y}$ and has an accuracy of 99.4% at a threshold of 4σ. The estimation of detection lags in both models remains one of their limitations, particularly the model for DYN21 where the model's detection lags are often extremely small, and, in the majority of cases, well before the participants pushed the button. Overall, however, this research helped confirm previous findings that a high-accuracy model can be developed based on e for the DYN12 transition, proposed a model with a high accuracy for the DYN21 transition, and identified that the relationship between the control inputs and system response is likely key in the detection process.

References

[1] Mulder, M., Pool, D. M., Abbink, D. A., Boer, E. R., Zaal, P. M., Drop, F. M., Van Der El, K., and Van Paassen, M. M., "Manual Control Cybernetics: State-of-the-Art and Current Trends," , 10 2018. https://doi.org/10.1109/THMS.2017.2761342.

[2] "Loss of Control In-flight (LOC-I)," , 2023. URL https://www.iata.org/en/programs/safety/operational-safety/loss-of-control-inflight/#:~:text=LOC%2DI%20refers%20to%20accidents,failures%2C%20icing%2C%20or%20stalls.

[3] Drop, F., "Control-Theoretic Models of Feedforward in Manual Control," 2016. https://doi.org/10.4233/uuid:7c1f62db-9a5a-4e02-8f11-488d6a299500, URL https://doi.org/10.4233/uuid:7c1f62db-9a5a-4e02-8f11-488d6a299500.

[4] Young, L. R., Green, D. M., Elkind, J. I., and Kelly, J. A., "Adaptive Dynamic Response Characteristics of the Human Operator in Simple Manual Control," *IEEE Transactions on Human Factors in Electronics*, Vol. HFE-5, No. 1, 1964, pp. 6–13. https://doi.org/10.1109/THFE.1964.231648.

[5] Young, L. R., and Stark, L., "Biological control system - a critical review and evaluation, developments in manual control," Tech. rep., In collab. with NASA, 3 1965. URL https://archive.org/details/nasa_techdoc_19650009660/page/n107/mode/2up.

[6] Young, L. R., "On Adaptive Manual Control," *Ergonomics*, Vol. 12, No. 4, 1969, pp. 635–674. https://doi.org/10.1080/00140136908931083.

[7] Niemela, R. J., and Krendel, E. S., "Detection of a Change in Plant Dynamics in a Man-Machine System," *Correspondence*, 1975.

[8] Elkind, J. I., and Miller, D. C., "Process of Adaptation by the Human Controller," *Second Annual NASA University Conference on Manual Control*, 1966, pp. 47–63.

[9] Phatak, A. V., and Bekey, G. A., "Model of the Adaptive Behavior of the Human Operator in Response to a Sudden Change in the Control Situation," *IEEE Transactions on Man-Machine Systems*, Vol. 10, No. 3, 1969, pp. 72–80. https://doi.org/10.1109/TMMS.1969.299886.

[10] Bastian, A. J., "Moving, sensing and learning with cerebellar damage," *Current Opinion in Neurobiology*, Vol. 21, No. 4, 2011, pp. 596–601. https://doi.org/10.1016/j.conb.2011.06.007.

[11] Imamizu, H., Miyauchi, S., Tamada, T., Sasaki, Y., Takino, R., Pütz, B., Yoshioka, T., and Kawato, M., "Human cerebellar activity reflecting an acquired internal model of new tool," *Nature*, Vol. 403, 2000, pp. 192–195.

[12] Hess, R. A., "Modeling pilot control behavior with sudden changes in vehicle dynamics," *Journal of Aircraft*, Vol. 46, No. 5, 2009, pp. 1584–1592. https://doi.org/10.2514/1.41215.

[13] Hess, R. A., "Modeling human pilot adaptation to flight control anomalies and changing task demands," *Journal of Guidance, Control, and Dynamics*, Vol. 39, No. 3, 2016, pp. 655–666. https://doi.org/10.2514/1.G001303.

[14] van Ham, J. M., Pool, D. M., and Mulder, M., "Predicting Human Control Adaptation from Statistical Variations in Tracking Error and Error Rate," *IFAC-PapersOnLine*, Vol. 55, No. 29, 2022, pp. 166–171. https://doi.org/10.1016/j.ifacol.2022.10.250.

[15] Terenzi, L., Zaal, P. M., Pool, D. M., and Mulder, M., "Adaptive Manual Control: a Predictive Coding Approach," *AIAA Science and Technology Forum and Exposition, AIAA SciTech Forum 2022*, American Institute of Aeronautics and Astronautics Inc, AIAA, 2022. https://doi.org/10.2514/6.2022-2448.

[16] Jakimovska, N., Pool, D. M., van Paassen, M. M., and Mulder, M., "Using the Hess Adaptive Pilot Model for Modeling Human Operator's Control Adaptations in Pursuit Tracking," American Institute of Aeronautics and Astronautics (AIAA), 2023. https://doi.org/10.2514/6.2023-0541.

[17] Mulder, M., Pool, D. M., van der El, K., Drop, F. M., and Van Paassen, M. M., "Manual Control with Pursuit Displays: New Insights, New Models, New Issues," *IFAC-PapersOnLine*, Vol. 52, No. 19, 2019, pp. 139–144. https://doi.org/10.1016/j.ifacol.2019.12.125.

[18] McRuer, D. T., and Jex, H. R., "A Review of Quasi-Linear Pilot Models," *IEEE Transactions on Human Factors in Electronics*, Vol. HFE-3, No. 3, 1967, pp. 231–249. https://doi.org/10.1109/THFE.1967.234304.

[19] Van Grootheest, H. A., Pool, D. M., Van Paassen, M. M., and Mulder, M., "Identification of time-varying manual-control adaptations with recursive ARX models," *AIAA Modeling and Simulation Technologies Conference*, American Institute of Aeronautics and Astronautics Inc, AIAA, 2018. https://doi.org/10.2514/6.2018-0118.

[20] Barragan, M., "Modeling the Human Operator's Detection of a Change in Controlled Element Dynamics," Master's thesis, Delft University of Technology, Faculty of Aerospace Engineering, 2023. URL http://resolver.tudelft.nl/uuid:5d9f79a5-6686-4767-a01d-01dc341fa990.

[21] Van Ham, J. M., *Adaptive Manual Control - The Human Response to Sudden Changes in Controlled Element Dynamics (MSc thesis)*, 2021. URL https://surfdrive.surf.nl/files/index.php/s/p3sY6Ur0IFjsjlX.

[22] Brenner, E., Rodriguez, I. A., Muñoz, V. E., Schootemeijer, S., Mahieu, Y., Veerkamp, K., Zandbergen, M., Van der Zee, T., and Smeets, J. B., "How can people be so good at intercepting accelerating objects if they are so poor at visually judging acceleration?" *i-Perception*, Vol. 7, No. 1, 2016. https://doi.org/10.1177/2041669515624317.

[23] Gignac, G. E., and Szodorai, E. T., "Effect size guidelines for individual differences researchers," *Personality and Individual Differences*, Vol. 102, 2016, pp. 74–78. https://doi.org/10.1016/j.paid.2016.06.069.

[24] Plaetinck, W., Pool, D. M., van Paassen, M. M., and Mulder, M., "Online Identification of Pilot Adaptation to Sudden Degradations in Vehicle Stability," *IFAC-PapersOnLine*, Vol. 51, No. 34, 2019, pp. 347–352. https://doi.org/10.1016/j.ifacol.2019.01.020.

[25] Thorpe, S., Fize, D., and Marlot, C., "Speed of processing in the human visual system," *Nature*, Vol. 381, 1996, pp. 520–522.