Impact of Haptic Warning Signal Reliability in a Time-and-Safety-Critical Task

Mario Enriquez & Karon E. MacLean
Dept. of Computer Science
University of British Columbia
E-mail: {enriquez, maclean}@cs.ubc.ca

Abstract

The bulk of current haptics human-factors research focuses on mapping basic human perceptual limits. However, many realistic applications demand a better understanding of how to construct more life-like but often less controllable experiment scenarios.

In this paper, we study this problem in the context of advanced automobile interfaces. We employ a throttle pedal with programmable force feedback to indicate potentially undesirable situations in the external environment and to gently but steadily guide the driver away from them. We have found evidence that within this scenario, errors in such a warning signal can have a negative effect on the behavior of the driver within the conditions studied.

These experiments required a complex protocol and necessarily permitted a variety of participant tactics. Post-experiment analysis revealed that very subtle variations in participant instruction produced large differences in tactics and consequent experiment outcome.

Keywords: Haptic force feedback, warning signal, false negative, false positive, driving performance, experiment design, participant instructions.

1. Introduction

Perceptual experiments usually fall into one of two categories: those where participants are asked to react to stimuli in some direct quantitative manner and those where the requested response is intended to capture more subtly perceived attributes of the stimuli. In research focused on determining basic perceptual limits, the former category is more common. However, as we integrate haptic feedback into sophisticated real applications, we need to better understand how to conduct more life-like – and often less controllable – experimental scenarios.

The latter type of experiment generally entails a realistic context and/or relatively complex tasks; and in the attempt to generate context, may invite involved or deliberately imprecise instructions. In such a situation, how can we get the participants to focus on the desired aspects of the experiment without giving away critical experiment information? How should we instruct participants so as to produce a desired performance tactic when tasks are complex and often cannot be clearly explained for experimental purposes? Finally, how can we draw strong conclusions from performance and response data collected in a deliberately uncontrolled environment?

In this paper, we examine these questions in the context of advanced automobile interfaces. Our paradigm employs programmable force feedback in the primary driving controls (here, the throttle pedal) to indicate potentially undesirable situations in the external environment and to gently but steadily guide the driver away from them.

For our experiments we consider a scenario that presupposes the existence of "drive by wire" automotive throttle control systems, whereby a pedal position sensor and electronic signal replace the traditional all-mechanical linkage from pedal to engine control module. These systems have begun to appear in the last several years for their virtue of improving fuel efficiency and throttle response. However, their existence incidentally affords a redefinition of how the primary controls feel to the driver and further allow the use of a newly-bidirectional channel to deliver new kinds of information in a new format. Given the critical nature of the driving task and in particular of the role played by the throttle and its feel, it is essential that such new interfaces be well designed. The experiments described here address one aspect of this larger problem: *driver behavior when information delivered through this new channel is not completely reliable.*

1.1. Remainder of This Paper

We begin by describing previous work in the areas of haptic constraints, warning signal response and signal reliability and introduce the concept of an Active Pedal (AP) using a virtual model of a physical system. The third section presents the implementation of a simple driving simulator used for our experiment. The fourth section provides a detailed account of the experiment design and implementation. Finally, we present the results obtained for three separate experiment series and discuss the subtle but forceful impact of instruction style on the outcome of this type of experiment.

2. Background

2.1. Haptic Constraints

The concept of using programmed force feedback to subtly inform and/or modify user behavior in a realtime manipulation task is not new. Rosenberg first proposed using haptic virtual fixtures to constrain user motion through a space in ways analogous to the use of a ruler or compass in mechanical drafting [7]. More recently, others have employed dynamically and sometimes automatically generated fixtures in applications such as surgical teleoperation; for example Payandeh & Stanisic demonstrated improvement in terms of performance, workload and task training time [6], and Okamura's group has been optimizing characteristics of the haptic signal itself [5]. Most relevantly, Steele & Gillespie looked at shared control of steering in a car where use of a haptic steering wheel improved tracking performance and reduced visual demand in a visual tracking task [9].

These and other studies consistently document the potential for appropriately displayed haptic feedback to provide information that enhances performance and reduces user effort in demanding realtime tasks. However, we have yet to consider characteristics of the information used to *generate* the informative signals; in particular its reliability, and how this may play out in the ability of the user to utilize that information.

2.2. Virtual Models of Physical Systems

Haptically portrayed models of familiar physical systems can make a haptic aid more intuitive [8]. We hypothesize that use of this approach in a driving situation can influence driver behavior towards a more conservative driving pattern in a subtle and non-irritating way, and potentially without the driver's explicit attention or awareness. However, we do not know how a user might respond to a haptic signal based on a virtual physical model when the signal cannot be guaranteed to be reliable.

2.3. Warning Signal Effect & Signal Reliability

Tipper [11] found a classic and robust warning signal effect [1, 4] in response time when subjects were given a haptic warning (a buzz on the hand) 100-1000 msec before receiving a visual stimulus to which they were to respond by pushing a computer key: response time improved in proportion to the advance warning given. Signal reliability has been shown to play a role in the way people process information contained within the signal [3, 10]. Based on this, Tipper proceeded to manipulate the reliability of the warning signal by corrupting it successively with 25% false negatives ("misses" or MI), 25% false positives ("false alarms" or FA) or a mix of these two types of errors. She found that the presence of FAs within a set of trials eliminated the warning signal improvement in response time even for those trials where the signal was present ("valid trials"); MI trials, on the other hand, had no such influence on the valid trials. Mixed errors produced the same negative effect as purely FA errors.

Figure 1: Setup. "Driver" at simulator with force feedback pedal.

We argue that the reason for this "bleeding" of a deleterious effect on subject behavior when a warning signal is subject to false positives is due to the subject's destroyed trust (whether conscious or not) in the reliability of the warning signal. This data suggests that false negatives do not similarly destroy trust. However, it was collected in a highly abstract context.

2.4. Sensor Reliability and Potential Impact on User Trust

It is generally very difficult to *guarantee* a technical system's perfect performance. In our situation of an intelligent system that warns a user of a critical situation, imperfect performance might occur when the system finds a critical situation when one does not exist (FA's), or fails to find one when it does exist (MI's). Further, a class of "perceptual" errors can occur through no fault of the technology: if the system finds and signals a warning for any situation that truly exists but which the user never perceives, the user may erroneously believe that system has delivered a false positive. In terms of impact on the user's trust of the system, this "perceptual" false alarm is indistinguishable from a "technical" false alarm. A user interface that takes input from sensors must therefore accommodate potential imperfections in the source input by understanding how the user will react to various amounts and types of sensor inconsistency or unreliability.

3. Driving Simulator

We wished to (a) establish whether use of a warning signal displayed as a haptic model of a familiar physical system can modify driving behavior in its perfect (reliable) form, and (b) explore how the same signal when unreliable might impact the driver's ability or willingness to make effective use of this information. We therefore developed a graphically simple driving simulator that reproduced several key aspects of a complex driving environment. A visual tracking task was executed via a force feedback pedal that superimposed an Active Pedal representation on the usual pedal spring force (Figure 1).

For this analysis, the physical system we modeled is that of a spring attached to the front of the car with a rest length equal to a nominal following gap behind the car ahead. When the driven car approaches the leading car, the driver feels the "compression" of this spring as an additional resistance through the throttle pedal: he must push a little harder to maintain the same gap. The smaller the gap between cars, the greater this extra push. This is, of course, only one of the possibilities for augmenting the information presented by the driving interface.

We implemented the simulator in Visual Basic on a 1 GHz P3 Windows 2000 computer, with an 18" LCD monitor.

3.1. Graphical Interface

The graphical interface (Figure 2) portrayed two cars on a road. The participant controlled the speed of the following car - which is stationary in the reference frame of the screen - using the pedal. The motion of the participant's car was conveyed by the rate at which road posts move toward the bottom of the screen. The speed of the (upper) lead car and ultimately its distance from the bottom car varied according to a pseudo-randomized control algorithm outlined below.

3.2. Workload Task: "Road Signs"

People often perform more than one task while driving; adjusting the radio, talking on the phone and using navigation systems absorb driver attention. In order to test the pedal force feedback in a multitasking environment, we provided an additional workload task: shapes (Figure 3) appeared at random locations and time intervals on the road margins and slowly faded away. Participants were asked to press the <ENTER> key when a particular shape (the triangle) was presented.

The effort required for this task was adjusted during pilot experiments by varying the size, number, frequency and distinctiveness of shapes until pilot participants felt the workload task was "reasonably challenging" and we felt it was competing substantially for attention with their primary driving task. In these experiments, the workload task appeared about 2-16 times per minute

3.3. Speed Control: FF Pedal and Brake

The simulator included a force feedback pedal with a position sensor interfaced through an IO board for force display and throttle input; the participants also used the keyboard space bar as a "brake". The pedal position input determined the acceleration of the participant-controlled car

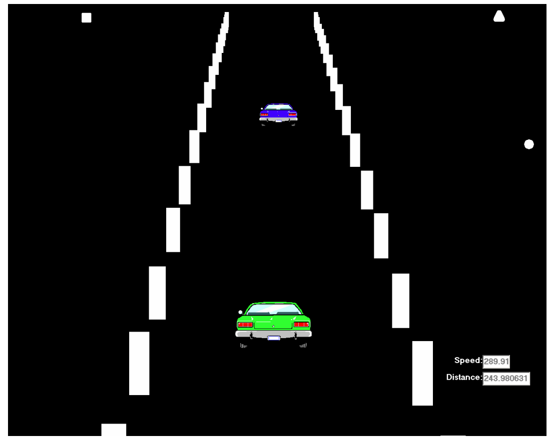

Figure 2: The main screen. Participant controls the speed of the green car (bottom) using the FF pedal in tandem with a "brake".

Figure 3: Workload shapes which appear at random locations on the "road" margins. Their size relative to other graphical features can be seen in the previous figure.

in the simulation. In the following, "lead" refers to leading vehicle, "car" refers to the participant's vehicle, "TTC" is Time to Contact (time until a collision should current relative velocities be maintained) and "THW" is Time Headway between the two. We employ a desired THW of 2 seconds, i.e. the car crosses a point on the road 2 seconds after the lead. *Xcar, Xlead and Xrel* refer to the position of the participant's car, the lead car and the distance between them. In a similar manner, *Vcar*, *Vlead* and *Vrel* refer to the cars' velocity.

$X_{rel} = X_{lead} - X_{car}$ An increasing X_{rel} is good.

$V_{rel} = V_{lead} - V_{car}$ An increasing V_{rel} is good.

$$THW = \frac{X_{rel}}{V_{car}} \qquad TTC_{control} = \frac{X_{rel}}{-V_{rel}}$$

$THW_{desired} = 2$ Desired THW.

$$OutputForcetoPedal = \left[a\left(\frac{1}{THW} - \frac{1}{THW_{desired}} \right) + \frac{b}{TTC_{control}} \right] C$$

The variables *a* and *b* are constants that define the displayed force profile given the distance between the cars and their velocity. We used values that maintain the relation b=-a/15 based on simulated results. *C* is a constant gain used to keep the total pedal force within a comfortable and comparable to a mechanical pedal system range.

3.4. Following-Car Dynamics

The position of the throttle pedal is used to calculate the "force" applied to the participant's car; the car's actual acceleration profile also depends on a wind and road-drag model component as well as on the vehicle's mass and internal friction. If the brake is not pressed:

$$Acceleration = \frac{F_{throttle} - F_{drag}}{M_{car}}$$

where

$F_{throttle}$ = Force generated by motor = f(pedal position)

F_{drag} = Drag force proportional to car's speed

M_{car} = Mass of car

If $F_{throttle} < F_{drag}$, the car gradually slows down. If the brake is pressed:

$$Acceleration = \frac{-F_{brake}}{M_{car}}$$

where F_{brake} is a constant brake force.

3.5. Lead-Car Dynamics

In order to observe the participant's response to Active Pedal activations in a finite amount of time, we needed the participant to interact with the AP fairly often. We aimed for 3 activations per minute as an acceptable facsimile of driving on a busy highway. This was achieved by adaptively adjusting the erraticity of the lead car velocity. Our algorithm randomly changed the lead car's virtual accelerator pedal and brake positions, and its consequently computed velocity, at discrete, randomly determined intervals ranging from 5 to 18 seconds. At all other times, the lead car maintained a constant accelerator and brake setting. Erraticity could be set at four levels from "steady" to "abrupt". The program evaluated the rate of Active Pedal activations every 30 seconds and adjusted erraticity level as needed.

4. Experiment Design

Experiment Units: All **trials** shared the same overall structure: the participant "drove" through 20 AP activations or "events", where each event is delineated by an activation (triggered when THW dropped below 2 seconds). The trial ended after 20 activations with an approximate duration of 6.7 minutes (3 events / minute). In post processing we segmented each trial into these 20 observations, and computed performance metrics independently on each segment. We used seven **trial types** representing 4 variables: workload task (W) present / absent; Active Pedal (force feedback) present or absent (P); and False Alarms (F) and Misses (M) committed by the AP at various frequencies. Table 1 identifies 4-letter trial labels.

A **Session** consisted of a practice trial followed by five experiment trials; each of a different type. Every participant completed trial types 0000, W000, 0P00 and WP00 (WL and reliable AP present/absent) in random order, followed by one of type WPF0, WP0M or WPFM (either False Alarms, Misses or both plaguing the AP signal in the presence of the workload task).

Table 1: Trial types

Trial Description				Label
Work Load	Active **P**edal	**F**alse Alarms	**M**isses	
NO	NO	NO	NO	0000
YES	NO	NO	NO	W000
NO	YES	NO	NO	0P00
YES	YES	NO	NO	WP00
YES	YES	25%	NO	WPF0
YES	YES	NO	25%	WP0M
YES	YES	12.5%	12.5%	WPFM

Training: All participants performed a practice trial where the different combinations of parameters that would be presented in the following 5 trials (AP+- and WL+-) were experienced. However, False Alarms and Misses were not experienced, nor was their possibility mentioned.

Participants: We used 36 participants in three runs of the experiment (12 per run). The participants were between 18-40 years of age, 14 female and 22 male, all with valid driver's licenses and normal vision and motor capability.

Instructions: For all three runs of the experiment, participants were told they were competing in a virtual "driving rally" with scoring on race time, safety errors and performance in the workload task. In the first two runs, instructions were read from a script, by a different individual for each run. For the third run, instructions were conveyed by a video recording of an experimenter relating the same script. At the time, we considered the instructions for all sessions and runs to be effectively identical.

5. Analysis

Analysis was conducted via Matlab scripts and Visual Basic code created for data segmentation, computation of performance metrics, collation of segment results, statistical comparisons and graphical display. This section describes how several analysis issues were handled.

Data Segmentation: To delineate the 20 activation events in each session, we defined a segment to begin as the participant *leaves* the critical THW zone from the previous segment and continue through to the end of the next critical zone penetration.

Performance Metrics: We used three performance metrics to examine the impact of warning signal reliability on driving behavior. These metrics were computed for every segment of every non-practice trial for each participant. For all three, more positive values indicate worse performance (see Figure 4).

P_{crit} (Critical Zone Penalty): weighted integral of time spent inside the critical region (THW < its nominal 2-second value) where the AP signal is activated. The closer the driver is to the lead car, the higher the penalty:

$$mP_{crit} = \sum_{k=nEnterZone}^{nLeaveZone} \frac{1}{THW_k},$$

where *nEnterZone* and *nLeaveZone* refer to the time steps during which the critical zone was entered and departed respectively. THW_k is the Time-Headway at that time step; its inverse is larger when the driver is closer to the lead car.

Brake: # of samples in the segment where the "brake" was pressed, multiplied times sample period.

Crashes: # of crashes during a segment (THW = 0).

Statistical Comparisons: To determine relative driving performance among the different experiment conditions, we compared *distributions* of segment performance metrics rather than trial and/or segment mean values. Mean values of metrics like amount of braking or number of crashes exhibit large variance by their nature, and even statistically significant differences may not be very meaningful. Distributions, on the other hand, retain information related to frequency and likelihood of these kinds of events occurring under the different conditions studied.

A **Kolmogorov-Smirnov (KS) test** statistically evaluates the difference between data distributions. The response distributions include all observations of a particular metric for a given set of conditions: the KS test then provides the likelihood that two such distributions are different [2]. Specifically, KS uses as a test statistic the maximum difference over all *x* values of the cumulative distributions of the two data sets X_1 and X_2. Mathematically, this can be written as:

$$\text{KS test statistic} = \max(|F_1(x) - F_2(x)|),$$

where $F_1(x)$ is the proportion of X_1 values <= *x* and $F_2(x)$ is the proportion of X_2 values <= *x*.

It should be noted that the KS test does not distinguish between differences due to distribution means, shapes or variances; this is acceptable for our purposes since all of these are relevant, and in general the differences we found appeared due to a combination of these factors.

Participant blocking: Because we observed substantial between-participant variation and we were most interested in the effect on individuals of varying experimental conditions, we blocked on participants by computing the mean of all observations in a given metric for each participant and then removing that mean from the participant observations before comparison. (This meant that negative values were possible for the metrics).

Four Tests: We performed four different statistical comparisons on the described performance metrics, each based on a KS test between two distributions. In order to utilize a 2-distribution test to compare 3 distributions, we therefore had to carry out three pair-wise comparisons.

1. Effect of Reliable AP and of Workload

Does AP feedback help when reliable? What effect does our workload model have? To measure the effect of AP, trial were lumped as (0000+W000) and (0P00+WP00), then compared in a 2-way test. For workload, the same trials were lumped as (0000+0P00) and (W000+WP00). Each test utilized 36 participants x 4 trials x 20 segments = 2880 observations.

2. No AP vs. Reliable AP vs. 25% Misses AP

What is the impact of Misses on performance? This 3-way test compared trial types W000, WP00 and **WP0M** for the 12 participants who performed WP0M (**4 from each experiment run**). Each of the three component 2-way tests utilized 12 x 2 trials x 20 = 480 observations; 12x3x20=720 observations in all were involved in the three comparisons.

3. No AP vs. Reliable AP vs. 25% False Alarms AP

What is the impact of False Alarms? This 3-way test also utilized 720 observations and compared trial types W000, WP00 and **WPF0** for a second subset of 12 participants.

4. No AP vs. Reliable AP vs. 12.5% M + 12.5% F

What is the impact of mixed False Alarms and Misses? This 3-way test also utilized 720 observations, and compared trial types W000, WP00 and **WPFM** for the final subset of 12 participants.

6. Results

In Figure 4, we see the baseline effects of the reliable AP signal (top half), and of workload (bottom). The Active Pedal signal (as implemented in our simulator) reduced the magnitude of all selected metrics, proving to be a significant aid over the no-AP case. However, our mechanism for imposing workload demonstrated mixed results, hurting performance for one of the metrics (braking), less significantly improving performance for another (P_{crit}) and having no significant effect on the third metric (Crashes) within the conditions studied.

Using a similar convention, Figure 5 shows the results of the 3-way KS comparisons (composed of three 2-way tests) of the trials that used no AP signal, a reliable AP signal, or a particular type of unreliable signal. In the top graph in Figure 5, the signal for this twelve-participant subset was corrupted by Misses. The Brake and Crash metrics do not show a significant alteration in driving behavior when a reliable AP signal was employed (No AP vs. Rel AP). However, P_{crit} shows a significant *increase* in time spent in the critical zone for reliable-AP trials (large positive blue bar), countering the 36-participant result shown in Figure 4 (negative blue bar), and the two other 12-participant results for this comparison for False Alarms and Mixed Error participant subsets below. In the comparison of reliable with Miss-prone AP trials (Rel AP vs. Misses), P_{crit} and Brakes indicate that for these 12

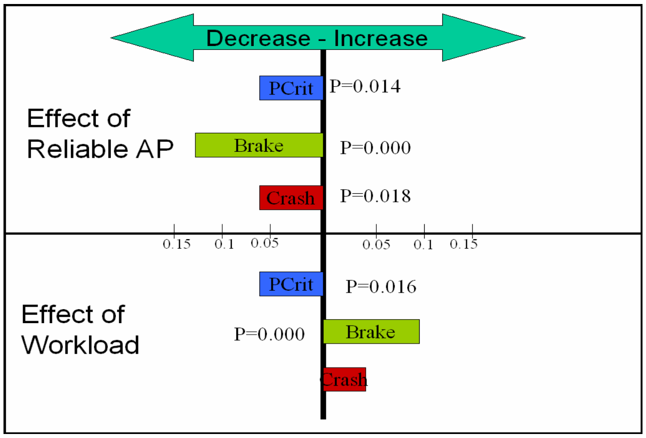

Figure 4: KS results for Test 1 show the effect of reliable AP and of workload for the three metrics (all 36 participants). Data from the same 2880 observations have been compared after lumping by presence/absence of AP signal (top) or workload task (bottom). The P-value indicates the statistical significance of the noted difference according to the KS test, whenever P≤0.050. The *x*-value is the dimensionless KS test statistic, i.e. the maximum difference between the two cumulative distribution functions. The *x*-direction of the arrows indicates whether the change in the performance metric denotes an increment or decrement in the related metric. **For these metrics, a more negative value indicates a more conservative driving pattern**.

participants, driving style was ***more extreme when the warning signal was reliable than when it was subject to misses***. The final row in this graph (No AP vs. Misses) is consistent with the first two for its only significant metric (Brakes): a miss-prone signal is better than none.

Proceeding in this manner through the remaining two graphs of Figure 5, we see in summary that a reliable AP signal usually results in a performance improvement over no AP signal (P_{crit} for the first group is the only exception), and that this result is often significant. False Alarms results in a performance most similar to that of no signal at all, i.e. the presence of false alarms appears to "wipe out" the benefits of the reliable signal for our specific setup and experiment design (Brakes metric). The presence of Mixed Errors results in a behavior intermediate between a reliable signal and none, for all metrics. The twelve FA participants seem to have been less reactive than the other 24, showing little diversity in performance for any metric except Brake (which followed the results of Mixed Errors). Unsurprisingly, Crashes shows the least consistent results among the three metrics. It represents the most extreme error, and the one most likely to be influenced by variations in the participant's accustomed driving style and the current driving mindset.

7. Impact of Instructions

Considering the complex nature of the experiment and some initial non-intuitive observations, we were prompted to examine our data in greater detail. A close examinationd

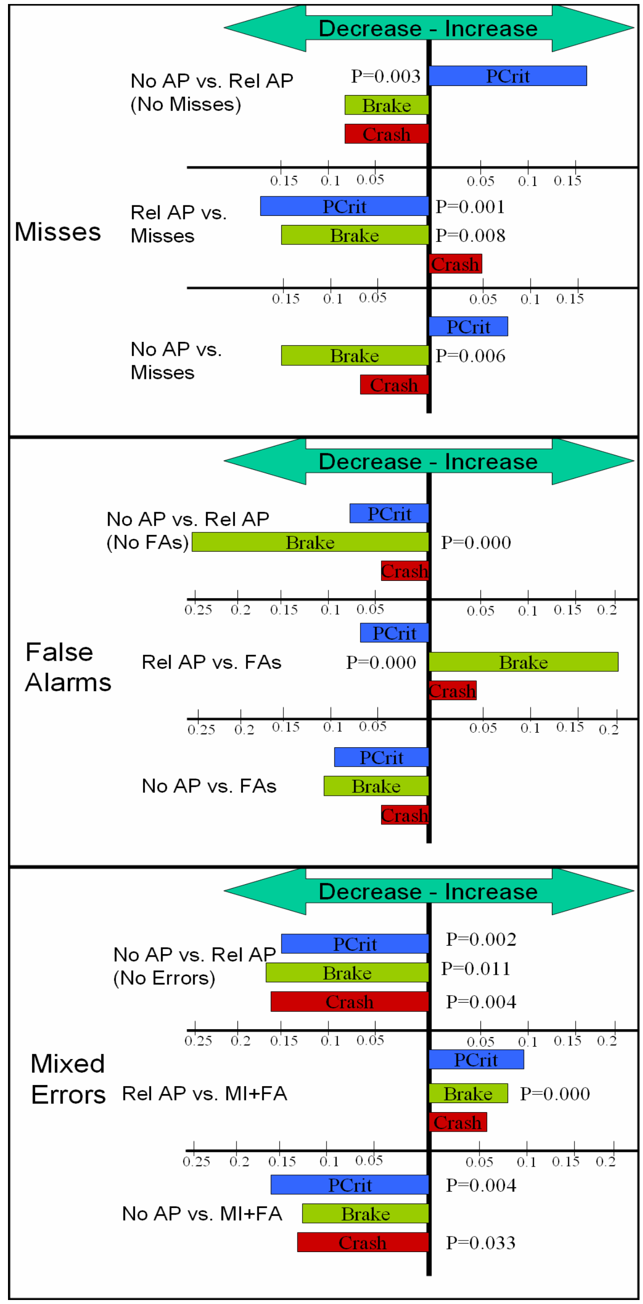

Figure 5: KS results are compared two at a time for (in each of the three graphs) three cases: No AP signal, a reliable AP signal, and an unreliable AP signal corrupted by one of three categories of errors (Tests 2-4). The KS test P-value is noted when significant, and the *x*-value is the dimensionless KS test statistic.

of the individual results revealed that each participant's overall driving behavior correlated with that of others in the same run.

Figure 6 shows the average values obtained for the Critical Zone Penalty (P_{crit}) for the 3 separate runs of the experiment (12 participants each; and each run included

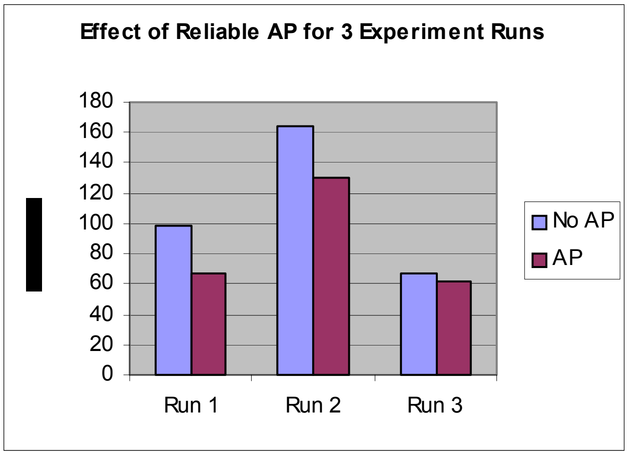

Figure 6: Average values for P_{crit} metric for the 3 experiment runs (12 participants each) after grouping by presence/absence of the AP signal. Data from all reliable trials (0000+W000 vs. 0P00+WP00) are compared. A more negative value indicates improved performance. Reliable AP always improves performance over no AP, but overall performance varies substantially between the three runs.

participants tested with all types of signal error).

Here we can see that for all 3 runs, the presence of an Active Pedal improves performance overall (by reducing the average P_{crit}). Also evident from this figure is the difference in overall magnitude for the 3 separate runs of the experiment, regardless of AP signal presence or absence.

The only variable we have been able to identify that could explain this phenomenon is the delivery of participant instructions for each run. A different experimenter administered each of the first two runs of the experiment, reading the same script; becoming suspicious, we instructed 3rd-run participants with the aid of a video recording, using the voice of a third experimenter. We conjecture that intonation, expression and verbal emphasis might have varied enough between the different experimenters to encourage different degrees of driving conservatism among each set of participants.

8. Discussion

In summary, we can observe that for this simulator and the tested combinations of experiment conditions:

(*i*) AP forces (always vs. never present) had a significant impact on all of the three metrics considered (p=0.014, 0.000, 0.018 respectively; 36 participants and 2880 observations). This is a strong result. Our workload task, on the other hand, did not appear to have a consistent effect on these metrics for the conditions tested. Further work will investigate the effect of workload more directly.

(*ii*) The improvements observed when using a reliable AP are lost when the signal is plagued with False Alarms (25%) in this particular experimental context. Performance with FA's is never significantly different from that with *no* signal (e.g. a Drive-By-Wire system with no additional force feedback). However, this pattern of degradation is significant only for the Brake metric for this participant subset. Mixed errors (FA+Misses), produced behavior similar to that of just FA's, for those results that are significant, although with a smaller magnitude. This effect of false alarms suggests that this type of error may undermine the improvements gained with the uncorrupted AP and are consistent with those found by Tipper [11], but to our knowledge this is the first time they have been documented in a semi-realistic driving context (i.e. continuous task subject to additional workload tasks).

(*iii*) Within the conditions studied, the presence of 25% Misses significantly ***improves*** performance over the cases of both a reliable and an absent AP signal, for some metrics (in Figure 5, compare the Rel AP vs. Error AP for Misses relative to those for FA and Mixed errors; the pattern is markedly different). This is perhaps the most surprising result among Tests 2-4.

Why might the presence of missed events (Misses) in the warning signal improve performance over the case of a perfectly reliable AP signal – under these conditions and measured by these metrics? It is believed that a warning signal of any type places individuals in a state of heightened alert and thus decreases reaction times [1]. However, it may also be the case that when individuals come to fully trust a warning signal, they may not feel the need to attend so closely to the task, particularly when a second task is competing for that attention – and this may result in decreased performance, despite the warning signal. Conversely, we theorize that our participants seem to make good use of a signal that is always trustworthy when it does trigger, but cannot be depended on to trigger for every valid target, *without* abdicating responsibility for finding those other events. This result may not appear in the case of false positive signals because the user may then feel that the signal is *never* trustworthy. If so, determination of the cognitive or perceptual level at which this distinction is made will require further investigation.

(*iv*) There is a noticeable difference in participant behavior (and thus performance according to our measures) for the 3 separate runs of the experiment. This can be clearly seen in Figure 6, where there is an evident difference between the overall values for the P_{crit} metric for the three separate runs of the experiment. The same trend was observed in the two other metrics (not shown due to length restrictions).

We theorize that these differences are a result of variation in the participants' understanding of their assigned task. The three experiment runs were administered by different individuals. A careful postmortem suggested that these individuals inadvertently placed a subtly different emphasis on differ rent aspects of the instructions for each run, thus creating three different driving mindsets that could explain the evidence seen in Figure 6: *Slow/Calm* (Experiment 3), *Fast/Aggressive* (Experiment 2) and *Somewhere In Between* (Experiment 1).

The instructions were designed to situate the participants in a “drive conservatively but quickly” mindset. This gave the participant the responsibility of enacting a compromise between two often-conflicting goals, as most of us do in real-life driving on a daily basis. However, a simulator is not a real car and brings no real consequences to aggressive driving. If the experimenter read the instructions with a greater emphasis on “conservative” as opposed to “quickly”, the participant’s behavior might be different for that particular run of the experiment. This is what seems to have occurred.

9. Conclusions

The experiment described here confirms previous evidence of a deleterious effect of interspersed false positives (in contrast to the neutral effect of false negatives) on the ability to use a binary haptic warning signal. This work extends these findings to a substantially more sophisticated scenario involving a semi-realistic driving simulation with a pedal-controlled tracking task in the presence of additional workload, with intuitively generated continuous force feedback delivered through the pedal, and for a set of metrics which evaluate “conservative driving”.

This experiment has also introduced new possibilities regarding the potentially positive performance impact of interspersed false negatives in a warning signal for our specific context.

We conclude that participant instruction can strongly influence their attitude when immersed in complex scenarios such as the Active Pedal driving simulator. A post-experiment analysis of the results leads us to conclude that our instructions inadvertently created 3 different kinds of driver mindsets (Slow, Moderate, Fast). Specifically, we believe that the three experimenters tended to encourage the participants to drive more or less aggressively through both vocal emphasis in reading written instructions, and ad-hoc clarifications. The level of impact of Active Pedal force feedback (AP) varies given these different driver mindsets. At least within the conditions studied, the AP seems to have a stronger influence in moderating driving behavior for people who are driving aggressively.

In general, the strong sensitivity of this type of highly contextualized, stakes-based experiment to experimenter-influenced participant strategy underscores the need for care in experiment design and protocol as well as careful analysis of results to better understand the gathered data.

As implemented by us, the presence / absence of a workload task made no measurable difference in the impact of AP on driving performance. Possible causes for this are:

a) Our WL task was not hard enough to impact on the "automaticity" of the driver's mental state.

b) The principal response to the WL task was to drive less aggressively in general, a condition in which the AP had less effect. Thus WL (in this case) may have changed participant behavior, but independently of the AP.

We emphasize that the conclusions presented here apply only to our proposed haptic feedback model (Active Pedal) and the additional information it might provide to the driver and not to the general Drive-By Wire case.

In future work, we plan to further investigate the subtleties of warning signal reliability for complex scenarios such of that described here, and to innovate on mechanisms for reliable experimentation in these situations.

Acknowledgments

We are grateful to Nissan Motor Co, Ltd. for their sponsorship of this project.

REFERENCES

[1] Bertelson, P., “The time course of preparation,” *Quarterly Journal of Experimental Psychology*, vol. 19, pp. 272-279, 1967.

[2] Chakravarti, Laha*, et al.*, *Handbook of Methods of Applied Statistics,* vol. 1: John Wiley and Sons, 1967.

[3] Green, D. M. and Swets, J. A., *Signal Detection Theory and Psychophysics*. New York: Wiley, 1966.

[4] Mowrer, O. H., “Preparatory set (Expectancy) - Some methods of measurements,” *Psychological Review Monograph*, vol. 52:(Whole No. 233), 1940.

[5] Nolin, J., Stemniski, P. *and A. M. Okamura*, “Activation Cues and Force Scaling Methods for Virtual Fixtures,” in Proc. of 11th Int'l Symp on Haptic Interfaces for Virtual Environment and Teleoperator Systems, Los Angeles, CA, 2003.

[6] Payandeh, S. and Stanisic, Z., “On Application of Virtual Fixtures as an Aid for Telemanipulation and Training,” in Proc. of IEEE VR2002, Haptics Symposium, 2002.

[7] Rosenberg, L., *Virtual Fixtures: perceptual overlays enhance operator performance in telepresence tasks,* Ph.D. Thesis, Stanford University, 1994.

[8] Snibbe, S. S., MacLean, K. E.*, et al.*, “Haptic Metaphors for Digital Media,” in Proc. of ACM Symp. on User Interface Software & Technology (UIST 2001), Orlando, FL, 2001.

[9] Steele, M. and Gillespie, B., “Shared Control Between Human and Machine,” in Proc. of Human Factors and Ergonomics Society, Minneapolis, MN, 2001.

[10] Tanner, W. P., Jr. and Swets, J. A., “A decision-making theory of visual detection,” *Psychological Review Monograph*, vol. 61, pp. 401-409, 1954.

[11] Tipper, C. M., *Is Inhibition of Return Truly a Reflexive Phenomenon? An Investigation of the Role of Temporal Predictability in a Classic Inhibition of Return Paradigm.,* M.A. Thesis, University of British Columbia, 2003.