Intelligent Interactive Displays in Vehicles with Intent Prediction: A Bayesian Framework

Bashar I. Ahmad, James K. Murphy, Simon J. Godsill, Patrick M. Langdon and Robert W. Hardy

Abstract

Using an in-vehicle interactive display, such as a touchscreen, typically entails undertaking a free hand pointing gesture and dedicating a considerable amount of attention, that can be otherwise available for driving, with potential safety implications. Due to road and driving conditions, the user input can also be subject to high levels of perturbations resulting in erroneous selections. In this article, we give an overview of the novel concept of an intelligent predictive display in vehicles. It can infer, notably early in the pointing task and with high confidence, the item the user intends to select on the display from the tracked free hand pointing gesture and possibly other available sensory data. Accordingly, it simplifies and expedites the target acquisition (pointing and selection), thereby substantially reducing the time and effort required to interact with an in-vehicle display. As well as briefly addressing the various signal processing and human factor challenges posed by predictive displays in the automotive environment, the fundamental problem of intent inference is discussed and a Bayesian formulation is introduced. Empirical evidence from data collected in instrumented cars is shown to demonstrate the usefulness and effectiveness of this solution.

I. Introduction

The complexity of in-vehicle infotainment systems (IVIS) has been steadily increasing to accommodate the growing additional services associated with the proliferation of smart technologies in modern vehicles. They aim to improve the driving experience and safety, for example advanced driver assistance, route guidance, driver inattention monitoring, and many others [1]. Consequently, minimising the effort and distraction of interacting with or controlling the IVIS is a key challenge [2]. This article introduces and presents an overview of the predictive in-vehicle display system, which utilises suitable statistical signal processing algorithms to enhance and simplify human machine interaction (HMI) in automotive applications, including IVIS-related interactions.

Lately, there has been a strong move towards replacing traditional static mechanical controls in vehicles, such as buttons, switches and gauges, with interactive displays, mainly touchscreens [2]. This is motivated by the evolution of the increasingly ubiquitous touchscreen technology and the ability

B. I. Ahmad, J. K. Murphy, and S. J. Godsill are with the Signal Processing and Communications Lab, Engineering Department, University of Cambridge, Trumpington Street, Cambridge, UK, CB2 1PZ. Emails:{bia23, jm362, sjg30}@cam.ac.uk.

P. M. Langdon is with the Engineering Design Centre, Engineering Department, University of Cambridge, Trumpington Street, Cambridge, UK, CB2 1PZ. Email:pml24@cam.ac.uk.

R. W. Hardy is with the HMI Research Group, Jaguar Land Rover, Coventry, UK. Email: rhardy@jaguarlandrover.com.

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the authors. The material includes two videos demonstrating the predictive display concept. Contact bia23@cam.ac.uk for further questions about this work.

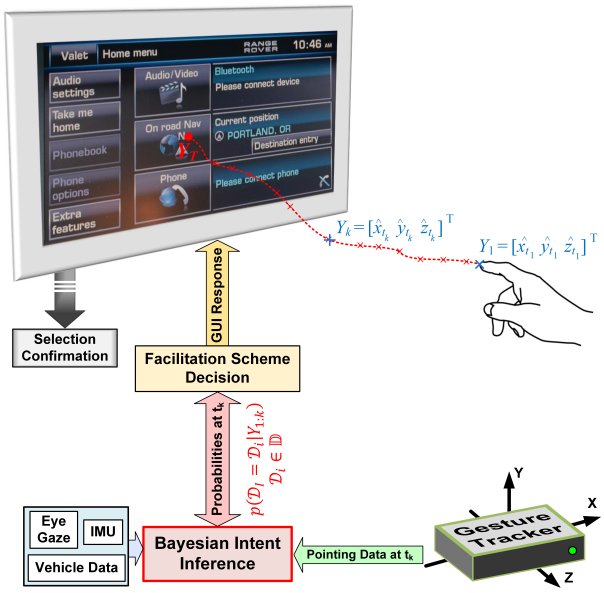


Figure 1: Block diagram of an in-vehicle predictive touchscreen system with a full 3D pointing finger-tip track, $t_k > t_1$. An infotainment menu of a Range Rover interface is displayed; the vehicle interior is not shown.

of these displays to: 1) effectively handle a multitude of functions by incorporating large quantities of information associated with in-vehicle infotainment systems, 2) promote intuitive interactions via free hand pointing gestures, especially for novice users, 3) offer design flexibility through a combined display-input–feedback module, 4) minimise clutter in the vehicle interior given their adaptability to the context of use, unlike mechanical controls. For example, the Tesla Model S car features a 17" touchscreen controlling most of the car functions [3]. Additionally, other types of displays, such as head-up displays (HUDs) and general 3D displays, have the potential of providing a more immersive driving experience and are becoming increasingly commonplace in vehicles [4], [5], for instance, the Jaguar Land Rover HUD windscreen incorporating laser holography [6]. However, such displays are often passive and users lack the means to easily interact with them in an automotive setting.

Interacting with an in-vehicle touchscreen typically involves undertaking a free hand pointing gesture to select an item on the display. This requires dedicating a considerable amount of visual, cognitive and manual attention, that is otherwise available for driving. The user input can also be subject to perturbations due to the road and driving conditions, resulting in incorrect on-screen selections [7], [8]. For example, the rate of successfully selecting an icon on the in-car display can be less than 50% when driving over a badly maintained road [8]. Rectifying an erroneous selection or adapting to the present noise will tie up more of the user's attention. This can render using interactive displays in vehicles effortful and distracting, with potential safety consequences [9]. Hence, there is a need for a solution that simplifies interaction with in-vehicle displays via intuitive free hand pointing gestures, or even enables it for emerging display technologies such as HUDs.

An intelligent in-vehicle predictive display[1], whose top-level block diagram is depicted in Figure

[1]The two attached videos show an early prototype of the predictive display with mid-air selection and prediction results during a few free hand pointing gestures; alternatively, follow the links: *https://youtu.be/Yco4v3N2QJk* and *https://youtu.be/Xf3W6nWeL4Y.*

1, employs a gesture tracker (and possibly other sensory data when available) in conjunction with a probabilistic prediction algorithm to determine the item the user intends to select on the display, remarkably early in the free hand pointing gesture [10]. It subsequently facilitates and expedites the target acquisition. Thus, the introduced intent-aware system can significantly improve the interactive display usability in vehicles and reduce the effort (attention) they require. Assuming that the prediction certainty meets a set criterion, the user need not touch the display surface to select the intended on-screen item, allowing mid-air selection. Therefore, this solution can also enable interacting with displays that do not have a physical surface, for example HUD and 3D displays or projections.

This article highlights and gives a unified treatment of the various signal processing (e.g., tracking-filtering, fusion, prediction, etc.) and human factors (e.g., feedback, prior experience, etc.) challenges posed by the in-vehicle intent-aware display concept, some of which were individually considered in previous publications (including those for non-automotive applications), such as [10]–[19]. In particular, the fundamental problem of intent inference within a Bayesian framework is addressed here, and suitable probabilistic prediction models are presented; they lead to a low-complexity implementation of the inference routine. Within this formulation, the task of smoothing perturbed pointing trajectories due to road and driving conditions via statistical filtering is discussed. The sensory requirements of the predictive system in the vehicle environment are also briefly outlined. Data collected in instrumented cars and results from a prototype predictive touchscreen system are shown to demonstrate the capabilities of this intelligent HMI solution.

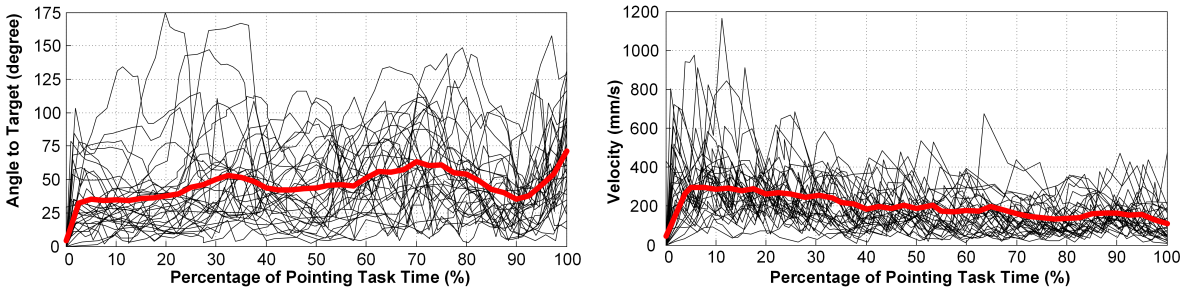
II. Background

According to the renowned human movement model Fitts' law [20], the index of difficulty (ID) and total time (T) of acquiring an interface icon (i.e., pointing and selection) are given by

$$\begin{aligned} \text{ID} &= \log_2\left(1+\ell/W\right), \\ T &= a + b\log_2\left(1+\ell/W\right), \end{aligned} \tag{1}$$

where W and ℓ are the the width of the target item and its distance from the starting position of the pointing object (mouse cursor or pointing finger), respectively [12]; a and b are empirically estimated. As intuitively expected, the selection task can be simplified and expedited by applying a *pointing facilitation scheme*, such as increasing the item size (larger W) or moving it closer to the cursor (smaller ℓ). Since a typical Graphical User Interface (GUI) contains several selectable items, any assistive pointing strategy should be preceded by a predictor to identify the intended on-screen icon [12]. Hence, the endpoint prediction problem has received notable attention in the Human Computer Interaction (HCI) area, for example [11]–[14] (see [10] and [14] for a brief overview).

The majority of existing HCI studies focus on pointing in 2D via a mouse or mechanical device on a computer screen to acquire GUI icons. They often use deterministic pointing kinematics models for endpoint prediction assuming: 1) the pointing object (cursor) velocity has a consistent profile and



(a) Angle to GUI icon $\theta_k \triangleq \angle(Y_k - Y_{k-1}, d_I)$. (b) Pointing finger-tip velocity $\|Y_k - Y_{k-1}\|_2$.

Figure 2: Angle to on-display icon and velocity profile for 30 in-car pointing tasks; thick red line is the mean.

is zero at arrival at destination, and 2) the cursor heads at a nearly constant angle towards its endpoint. Both premises make intuitive sense for mouse pointing in 2D, however, they do not necessarily hold for free hand pointing gestures in 3D [10]. For example, Figure 2a shows that the pointing finger-tip heading angle to an on-screen icon drastically changes throughout a sample of free hand pointing gestures recorded in an instrumented car; d_I is the location of the intended on-screen destination in 3D and Y_k is the 3D Cartesian coordinates of the pointing finger-tip at the time instant t_k.

Data driven prediction techniques, such as in [13] and [19], can be applied to infer the intended destination of a pointing task. They often utilise a pointing motion model learnt from *a priori* recorded interactions, necessitating the availability of a complete data set of training examples of pointing trajectories. This requirement is particularly stringent for free hand gestures approaching a display in 3D to select icons on GUIs of various possible layouts, due to the very large number of possible paths. More so in an automotive HMI context, where a user might be expected to undertake a few pointing gestures, for instance, to set up the IVIS preferences, during his or her first system use. On the other hand, the predictive display system discussed here employs known motion as well as sensor models, and thus can use a state-space-modelling approach, albeit with a few unknown parameters. It requires minimal training and is computationally efficient.

In the area of object tracking, for example in surveillance applications, knowing the destination of a tracked object not only leads to more accurate tracking results, but also offers vital information on intent, revealing potential conflict or threat [16], [21], [22]. Destination prediction can be viewed as a means to assist planning and decision making at a system level higher than that of established conventional sensor-level tracking algorithms, whose objective is to infer the current value of the latent *state* X_t (e.g., the tracked object position, velocity, etc.) [22]. For example, destination-aware trackers that include an additional mechanism to determine the object endpoint are proposed in [16]. These methods discretise the state space area into predefined regions and the object can only pass through a finite number of these zones; such a discretisation can be a burdensome task for free hand pointing gestures in 3D. On the contrary, the predictive display solution presented in this article uses continuous state space motion models that do not impose any restrictions on the path the pointing finger has to follow to reach its intended on-display endpoint and can easily handle noisy as well as

asynchronous observations. Nevertheless, other conforming destination-aware tracking methods can be applicable.

A related scenario in which there is a growing interest is the user input on a smartphone, perturbed due to situational impairment, for example walking [17]. Typically in such cases, the GUI is dynamically adapted to compensate for the measured noise. For an in-vehicle display, the pointing time and distance is notably longer than that for a hand-held device and the correlation between the pointing hand movements and the experienced in-car accelerations or vibrations can be ambiguous [10]. This is attributed to the complexity of the human motor system and its response to noise as well as the seat position, cushioning, reaching style or distance, etc. Thus, compensating for the measured in-vehicle noise can have limited effects on improving the display usability. Here, perturbed user input is tackled within the statistical inference framework of a predictive display.

III. An In-Vehicle Predictive Display System

Below, we describe the various modules that compliment the present in-vehicle interactive display, e.g., a touchscreen, to realise the intelligent predictive display system in Figure 1.

A. Gesture Tracker

Motivated by extending HCI beyond traditional keyboard input and mouse pointing, new 3D vision sensory devices have emerged that can track, at high rates, hand gestures, including pointing finger(s)-tip(s), for example Microsoft Kinect, Leap Motion (LM) and SoftKinetic DepthSense. However, operating in a mobile vehicle enviroment can be challenging to these trackers due to dynamically changing light conditions, in-car vibrations-accelerations, occlusion with limited in-car mounting positions, large coverage area (e.g., steering wheel or armrest to display and the front passenger) and others. Fortunately, the current interest in gesture-based HCI in cars (e.g., current BMW 7 Series cars have a gesture control for some features) is driving the development of automotive-grade gesture trackers [15]. In Figure 1, a tracker provides, in real-time, the pointing hand/finger(s) locations, $Y_{1:k} \triangleq \{Y_1, Y_2, ..., Y_k\}$, at the discrete time instants $t_1, t_2, ..., t_k$. For instance, $Y_n = [\hat{x}_{t_n} \;\; \hat{y}_{t_n} \;\; \hat{z}_{t_n}]'$ is the 3D Cartesian coordinates of the pointing finger-tip at t_n. In general, the predictive display demands reliable pointing finger tracking at a rate exceeding 30Hz, as majority of in-vehicle pointing tasks can have durations in the range of $0.2s \leqslant T \leqslant 4s$ [8]. Figure 3 depicts three complete 3D pointing trajectories, $Y_{1:T}$, collected in a car using a LM controller under three conditions, which visibly affect the pointing gesture.

B. Bayesian Intent Inference

Let $\mathbb{D} = \{\mathcal{D}_i : i = 1, 2, ..N\}$ be the set of N selectable items on the interactive display. Whilst no assumptions are made about the layout of the icons in $\mathbb{D}$, each item is modelled as a distribution representing the extended regions in space of various shapes and sizes occupied by the corresponding

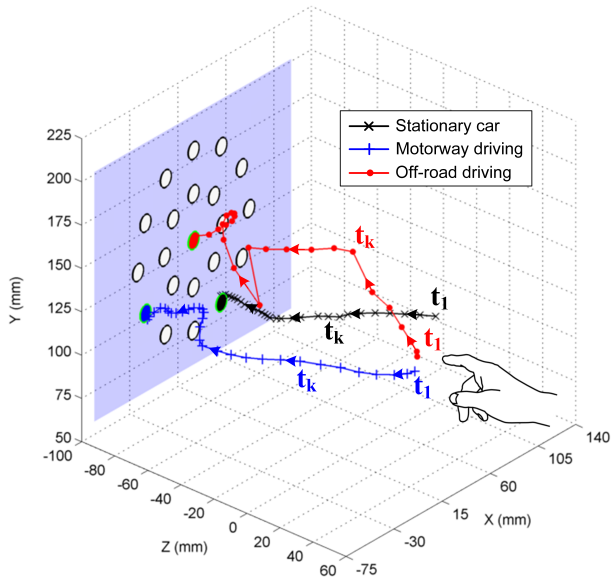


Figure 3: Recorded 3D pointing finger-tip tracks to select icons (circles) on an in-car touchscreen (blue surface) under various driving/road conditions [10]; $t_1 < t_k$.

GUI elements. For simplicity and computational efficiency, Gaussian distributions can be considered and the i^{th} item is modelled as $\mathcal{N}(\mu_i, C_i)$. The mean μ_i and covariance matrix C_i capture the 3D location and the extent-orientation of $\mathcal{D}_i$, respectively. At the time instant t_k, the inference module in Figure 1 calculates the posterior probabilities for the N destinations,

$$\mathcal{P}(t_k) = \{p(\mathcal{D}_I = \mathcal{D}_i | Y_{1:k}) : i = 1, 2, ..., N\}, \tag{2}$$

which represent the likelihood of each of the icons in $\mathbb{D}$ being the *unknown* intended on-display endpoint $\mathcal{D}_I \in \mathbb{D}$. This uses the gesture-tracker measurements $Y_{1:k}$ (and possibly other sensory data), i.e., the available partial pointing finger track at t_k whose extraction might require simple data sorting and associating routines. Each observation Y_k is assumed to be derived from an underlying pointing finger true (perturbation-free) latent state X_{t_k}, that can include its position, velocity, etc.

Within a Bayesian framework, we have

$$p(\mathcal{D}_I = \mathcal{D}_i | Y_{1:k}) \propto p(Y_{1:k} | \mathcal{D}_I = \mathcal{D}_i) p(\mathcal{D}_I = \mathcal{D}_i), \tag{3}$$

where the prior $p(\mathcal{D}_I = \mathcal{D}_i)$ on the selectable items (independent of $Y_{1:k}$ or the current pointing task) can be attained from relevant semantic and contextual information, such as selection frequency, GUI design, user profile, etc. This makes the adopted formulation particularly appealing as additional information, when available, can be easily incorporated. For example, the priors in (3) can be gradually and dynamically learnt as the system is being used, starting from uninformative ones. Therefore, it is an adaptable probabilistic (belief-based) approach.

A prediction is performed at the arrival of each (or a few) new sensor observations. The inference module can use a number of low complexity, computationally efficient, probabilistic endpoint predictors that are amenable to real-time implementation, given the limited computing resources

and training data available in vehicles. The linear models discussed in the next section lead to a Kalman-filter-type implementation, combining endpoint prediction and filtering out of noise induced by road/driving conditions. For severe perturbations, a separate statistical filter can be employed to remove the highly nonlinear gesture motion arising from perturbations.

C. Facilitation Scheme and Decision

To assist the selection task, the displayed interface may be modified at t_k, for example icons can be expanded/shrunk, colored/faded or other [11], [12], [14], as per their probabilities $\mathcal{P}(t_k)$ in (2). Such facilitation strategies can require major modifications to legacy in-vehicle GUI designs and possibly the related software-hardware architectures. Their impact on the user experience in a split attention scenario (driving and interacting) is non-trivial and can be advised by experimental studies. For instance, unlike mouse pointing on a computer screen, constantly changing the in-car interface can increase visual demand to monitor the ongoing changes. A promising pointing facilitation scheme is mid-air selection, where the system auto-selects the predicted intended on-screen item on behalf of the user, who does not need to physically touch the display surface. Whilst mid-air selection can reduce the free hand pointing gesture duration and thus effort (visual, cognitive and manual), its implementation entails only sending/reading a select signal to/by the existing interface software module with minimal display overheads.

After inferring $p(\mathcal{D}_I = \mathcal{D}_i \mid Y_{1:k})$ at time t_k, the endpoint $\hat{\mathcal{D}}_I(t_k) \in \mathbb{D}$ of a free hand pointing gesture can be estimated (if needed) by minimising the expected value of a cost function over all of the possible destinations in $\mathbb{D}$. This can be expressed by

$$\hat{\mathcal{D}}_I(t_k) = \underset{\mathcal{D}^* \in \mathbb{D}}{\arg\min} \sum_{i=1}^{N} \mathcal{C}(\mathcal{D}^*, \mathcal{D}_I) p\left(\mathcal{D}_I = \mathcal{D}_i | \, Y_{1:k}\right), \tag{4}$$

where $\mathcal{C}(\mathcal{D}^*, \mathcal{D}_I)$ is the cost of deciding $\mathcal{D}^*$ as the destination given that $\mathcal{D}_I$ is the true intended on-display icon. If the binary decision criterion $\mathcal{C}(\mathcal{D}^*, \mathcal{D}_I) = 1$ if $\mathcal{D}^* \neq \mathcal{D}_I$ and $\mathcal{C}(\mathcal{D}^*, \mathcal{D}_I) = 0$ otherwise, is used, it can be easily seen that (4) leads to the Maximum *a Posteriori* (MAP) estimate; it implies that the most probable endpoint is deemed to be the intended on-display selectable icon. Within the Bayesian framework, more elaborate cost functions can be applied [23]; groups $\hat{\mathbb{D}}_q \subset \mathbb{D}$ rather than an individual icons may also be considered for expansion or fading purposes.

Whilst the intuitive MAP estimate can be used to assess the suitability of the prediction model, it can produce fast fluctuating decisions during the pointing task. This can be detrimental to mid-air selection due to the resultant false positives. In such cases, a simple decision rule can stipulate that the probability of an icon $p\left(\mathcal{D}_I = \mathcal{D}_i | \, Y_{1:k}\right)$, namely the one delivered by the MAP classifier, should exceed a certain threshold for a given duration of time before triggering an auto-selection action.

D. Adaptable GUI and Selection Confirmation

The displayed interface implements seamlessly, in real-time, the applied pointing facilitation scheme. If an on-screen item is selected or auto-selected, the user can substantially benefit (i.e., in terms of reducing the visual workload) from a feedback confirming the selection action, for example an audible or haptic signal. For a predictive display with mid-air selection, the emerging ultrasonic mid-air haptic technology [24] presents itself as a suitable equivalent to the conventional on-screen haptic feedback, which is used in standard smartphone devices, with established benefits.

E. Additional Sensory Data

The availability of additional vehicle sensory data, such as suspensions travel data via the controller area network (CAN) bus or an on-board inertia measurement unit (IMU), can enable the intelligent predictive display system to establish the operating conditions, for instance, allowing it to determine whether the user input is perturbed, or even estimate the level of noise present. It can then modify the applied statistical model, by adapting its parameters or performing pre-processing prior to intent inference. Eye-gaze measurements can also offer valuable information on areas of interest on the display and can be used as an input modality in HCI, e.g., in [25]. Eye-gaze trackers are primarily utilised to examine the human performance behaviour in a controlled setting, such as simulators, and a corpus of literature exists [26]. Obtaining accurate data from such a tracker, that is not head-mounted, in a mobile vehicle can be challenging given the currently available commercial sensors. However, the fusion or simultaneous use of eye-gaze and pointing gesture data for an in-vehicle predictive display is a promising research area. In summary, if any additional information becomes available, it can be easily incorporated into the Bayesian framework via the priors $p\left(\mathcal{D}_I = \mathcal{D}_i\right), \mathcal{D}_i \in \mathbb{D}$, or alternatively treated as a part of the measurements vector $Y_{1:k}$.

IV. Bayesian Endpoint Prediction

Given the available measurements $Y_{1:k}$ at t_k, determining the probability of each of the endpoints in $\mathbb{D}$ being the intended destination requires calculating the observation likelihood $p(Y_{1:k} \mid \mathcal{D}_I = \mathcal{D}_i)$ conditioned on each endpoint, as stated in (3). The prior $p(\mathcal{D}_I = D_i)$, which is independent of the current pointing task, is presumed to be available; here, for simplicity, all icons are assumed to be equally probable with $Pr(\mathcal{D}_I = D_i) = 1/N, i = 1, 2, ..., N$. The key problem in the intent prediction procedure is therefore that of evaluating the observation likelihood, i.e., the probability of having made a series of observations, under the assumption that the tracked object is ultimately heading to a given destination. This can be tackled by adopting an underlying motion model of the pointing finger, describing its trajectory on its journey towards the intended endpoint and including an element of randomness in the followed track. This capitalises on the premise that the motion of the pointing finger in 3D is dictated by the intended icon on the display. Since the true destination $\mathcal{D}_I$ is unknown *a priori*, N such models for each $\mathcal{D}_i \in \mathbb{D}$ are postulated, and the objective becomes calculating the

likelihood of the observed partial pointing trajectory being drawn from a particular endpoint-driven model. In other words, the destination that leads to a model that best explains $Y_{1:k}$ is assigned the highest probability of being $\mathcal{D}_I$, and vice versa.

According to the chain rule of probability,

$$p(Y_{1:k} \mid \mathcal{D}_I = \mathcal{D}_i) = p(Y_k \mid Y_{1:k-1}, \mathcal{D}_I = \mathcal{D}_i) p(Y_{1:k-1} \mid \mathcal{D}_I = \mathcal{D}_i), \tag{5}$$

where $p(Y_{1:k-1} \mid \mathcal{D}_I = \mathcal{D}_i)$ is the likelihood estimated at the previous time instant t_{k-1}. Thus, the observation likelihood in (5) can be calculated sequentially, i.e., with the arrival of each new sensor measurement of the pointing gesture, and determining the prediction error decomposition (PED), $p(Y_k \mid Y_{1:k-1}, \mathcal{D}_I = \mathcal{D}_i)$, at t_k suffices. Below, we outline simple destination-driven models, including the bridging distributions approach introduced in [27], [28], and show how sequential calculation of the PED can be performed, permitting the posterior probability distribution over intended endpoints in (2) to be calculated at each stage.

A. Modelling Pointing Movement

The pointing gesture movement towards an on-screen item is not deterministic. The person making the pointing gesture is capable of autonomous action and is in control of a complex motor system with numerous physical constraints, and is likely to also be subjected to external motion, jolting, rolling, acceleration and braking in a moving vehicle. Hence, models of the pointing finger movements, albeit driven by intent, are uncertain, and this can be captured by adopting stochastic models. This implies that the predictions of the tracked object motion are not single deterministic paths, but are rather probabilistic processes, with the pointing finger position at a future time expressed as a probability distribution in space. By adequately incorporating this uncertainty, relatively simple models of pointing finger motion can be used successfully to evaluate the corresponding observation likelihoods and the probabilities of $\mathcal{P}(t_k)$ in (2). It is emphasised here that the in-vehicle predictive display system objective is to infer the intent of the hand movement and *not* to accurately model the complex human motor system. Thus, an approximate motion model that enables reliably determining the destination of a free hand pointing gesture is sufficient.

Calculating the transition density of a stochastic model, for example between two successive observation times t_{k-1} and t_k, is required to condition the tracked pointing finger state X_t (e.g., position, velocity, etc.) on a nominal endpoint $\mathcal{D}_i$. Continuous-time motion models are a natural choice, where the tracked object's dynamics are represented by a continuous-time stochastic differential equation (SDE). This SDE can be integrated to obtain a transition density over any time interval. Although numerous models for object tracking exist, the class of Gaussian linear time invariant (LTI) models for the evolution of X_t is utilised by the in-vehicle predictive display, as they lead to a low-complexity inference procedure (unlike non-linear and/or non-Gaussian models). This class includes many models used widely in tracking applications, for example the (near) constant

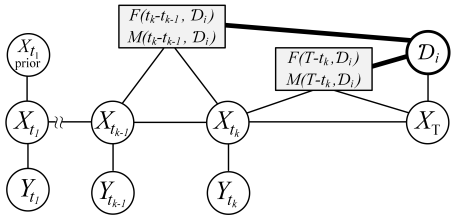


Figure 4: The system graphical structure; endpoint $\mathcal{D}_i$ acts as a prior and affect the state transition.

velocity (CV) and linear destination reverting (LDR) models highlighted below, in addition to other Gaussian LTI models that can describe higher order kinematics (acceleration, jerk, etc.) [22].

Whilst the system governing the pointing finger dynamics is assumed not to change over time, it does depend here on the intended endpoint $\mathcal{D}_I \in \mathbb{D}$, which intrinsically drives the pointing motion. Conditioned on knowing this endpoint, e.g., the i^{th} GUI icon $\mathcal{D}_i$, and integrating the Gaussian LTI model, the relationship between the system state at times t and $t+h$ can be written as

$$X_{i,t+h} = F(h, \mathcal{D}_i) X_{i,t} + M(h, \mathcal{D}_i) + \varepsilon_t, \tag{6}$$

with $\varepsilon_t \sim \mathcal{N}(0, Q(h, \mathcal{D}_i))$ is the dynamic noise embodying the randomness in the motion model. The matrices F and Q as well as the vector M, which together define the state transition from one time to another, are functions of the time step h and, notably, the destination $\mathcal{D}_i \in \mathbb{D}$. Thereby, N such models are constructed to establish the endpoint of the pointing gesture.

The k^{th} observation, for example, the pointing finger position as provided by the gesture-tracking device, is also modelled as a linear function of the time t_k state perturbed by additive Gaussian noise,

$$Y_k = G X_{i,t_k} + \nu_k, \tag{7}$$

where G is a matrix mapping from the hidden state to the observed measurement and $\nu_n \sim \mathcal{N}(0, V_n)$. For instance, if the gesture-tracker provides the pointing finger positions directly and the system state includes only position, then G is a 3×3 identity matrix. The noise covariance can be utilised to set the level of noise in each of the x, y and z axes as per the gesture-tacker specifications, for example a time-of-flight based tracker such as the SoftKinetic DepthSense camera exhibits higher inaccuracies in observations along the depth axis. It is noted that no assumption is made about the observation arrival times t_k and irregularly spaced, asynchronous measurements can naturally be addressed within this formulation. The system structure, for each nominal endpoint $\mathcal{D}_i$, is depicted graphically in Figure 4, where the destination $\mathcal{D}_i$ influences the endpoint-driven state at all times.

Amongst linear Gaussian models, linear destination reverting models, such as the Mean Reverting Diffusion (MRD) and Equilibrium Reverting Velocity (ERV) models, make particularly suitable candidates for the pointing finger motion in (6), as discussed in [10]. Their state evolution explicitly incorporates the destination information. For example, the governing SDE for the mean reverting diffusion model is given by: $dX_{i,t} = \Lambda\left(d_i - X_{i,t}\right) dt + \sigma dw_t$. It indicates an attraction of the motion

towards the location of destination d_i (e.g., the mean of the Gaussian distribution representing $\mathcal{D}_i$), with Λ (a design parameter) capturing the strength of this reversion for each axis in 3D, and w_t is a Wiener process. Whilst the MRD is based on a multivariate Ornstein-Uhlenbeck process [29] and the system state only includes the position information in 3D, the state of the ERV model proposed in [10] additionally includes the velocity of the pointing-finger, in 3D, driven by the endpoint. This facilitates modeling pointing velocity profiles like those shown in Figure 2b. Integrating the SDE of the MRD and ERV results in (6), each with specific F, M and Q matrices.

During a pointing task, the path of the pointing finger, albeit random, must end at the intended destination at time T (i.e., the pointing finger reaches its endpoint on the display). This can be modelled by an *artificial* prior probability distribution for X_T corresponding to the geometry of the destination; alternatively, it can be treated as a *pseudo-observation* at T. In order to maintain the linear Gaussian structure of the system in (6) and (7), this distribution is assumed to be Gaussian, such that $p(X_T \mid \mathcal{D}_I = D_i) = \mathcal{N}(X_T; a_i, \Sigma_i)$; see [28] for a discussion on this construct. The mean vector a_i specifies the constrained system state at the destination, whereas Σ_i is a covariance matrix of the appropriate dimension. For instance, for the MRD model, in which only pointing finger position is considered, $a_i = \mu_i = d_i$ representing the location-centre of the destination in 3D. In the case of the ERV model, defining the final state distribution also involves specifying a distribution of the pointing finger velocity at endpoint. A large-scale prior covariance can be used to model the uncertainty in this, however certain properties might be assumed, e.g., relatively high velocity in the direction towards the screen.

Exploiting the artificial prior on the distribution of X_T requires that the state of the motion models in (6) to be conditioned not only on $\mathcal{D}_i \in \mathbb{D}$, but also on the arrival time T. Including this permits the posterior of the system state at time t_k to be expressed as $p(X_{t_k} \mid Y_{1:k}, T, \mathcal{D}_I = \mathcal{D}_i)$, and the sought observation likelihood in (5) is subsequently given by $p(Y_{1:k} \mid T, \mathcal{D}_I = \mathcal{D}_i)$ after k measurements. The inclusion of the prior on X_T in the motion model changes the system dynamics (even for MRD and ERV models), where the predictive distribution of the pointing finger state changes from a fully random walk to a bridging distribution, terminating at the endpoint. This encapsulates the long term dependencies in the pointing finger trajectory due to premeditated actions guided by intent. Since the intended destination is not known, N such bridges are constructed, one per nominal endpoint. Consequently, all Gaussian linear models, including the non-destination reverting ones, whose dynamic models are not dependent on $\mathcal{D}_i$ like Brownian motion (BM) and CV, can be utilised for destination prediction within the presented Bayesian framework. This technique of conditioning on the endpoint is dubbed bridging distributions (BD) based inference.

B. Intent Inference: Sequential Likelihood Evaluation

We recall that the primary objective of the intent inference routine is to determine the observation likelihoods $p(Y_{1:k} \mid \mathcal{D}_I = \mathcal{D}_i), \mathcal{D}_i \in \mathbb{D}$, at t_k, rather than the posterior distribution of the system

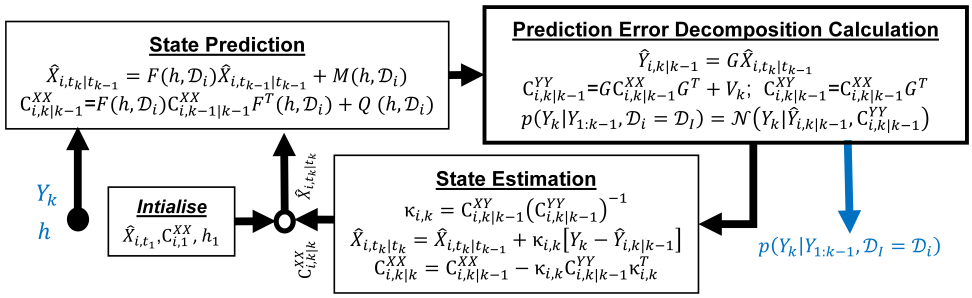


Figure 5: Kalman filter for sequentially evaluating the PED for endpoint $\mathcal{D}_i$ at the arrival of observation Y_k; state prediction at t_k relies on the state estimation results, including covariance $C^{XX}_{i,k-1|k-1}$, from the previous time step and $h = t_k - t_{k-1}$.

state X_{t_k}, as in traditional tracking applications [22]. Nonetheless, the latent state estimation, which might be relevant in certain scenarios, is addressed below. Based on (6) and (7), a classical Kalman filter can be employed to sequentially calculate the prediction error decomposition in (5) as depicted in Figure 5 and, thereby the sought observation likelihood for the current set of measurements $Y_{1:k}$ conditioned on $\mathcal{D}_i$. The computationally efficient Kalman filter is particularly desirable since running, concurrently, multiple Kalman filters for all $\mathcal{D}_i \in \mathbb{D}$ is plausible in real-time, even in settings where limited computing power is available. This solution is also amenable to parallelisation.

For the bridging approach, it is shown in [27] and [28] how the PED and observation likelihood in (5) from each constructed bridge, i.e., conditioned on T and $\mathcal{D}_i$, can be estimated using a modified Kalman filter. As the true arrival time T is unknown *a priori* in practice, approximating

$$p(Y_{1:K} \mid \mathcal{D}_I = \mathcal{D}_i) = \int_{T\in\mathcal{T}} p(Y_{1:k} \mid T, \mathcal{D}_I = \mathcal{D}_i)p(T \mid \mathcal{D}_I = \mathcal{D}_i)dT, \tag{8}$$

is necessary, where $p(T \mid D)$ is the a prior distribution of arrival times at destination $\mathcal{D}_i$ and $\mathcal{T}$ is the time interval of possible arrival times T. In the simplest case, arrivals might be assumed at some specific future time. This is a crude approximation, nevertheless is often quite effective [28]. To improve inference accuracy (and possibly also to learn about expected arrival time), arrivals can be modelled as having a prior distribution, such as being expected uniformly within some time period $[t_a, t_b]$, giving $p(T \mid D) = \mathcal{U}(t_a, t_b)$. In this case, numerical quadrature, for example via Simpson's rule, can be applied. Although bridging-distribution based intent inference involves running multiple Kalman filters, and hence is more computationally demanding, it can significantly improve the endpoint inference capability of a predictive display and leads to a more robust performance.

In summary, the introduced modelling approach to infer the item the user intended to select on the display as early as possible in the free hand pointing gesture is generic and offers considerable flexibility in terms of catering for various sensing technology specifications (for example, observation error) as well as adaptability in terms of adjusting the motion model parameters. It is simple and relatively computationally efficient, which makes it suitable for the requirements of an automotive environment. In the developed predictive display prototype (an optimised C# implementation of the system in Figure 1 on a typical automotive computing platform), prediction with Kalman filtering

was tested with up to $N = 64$ destinations and an observations data rate $\geqslant 30$Hz without any noticeable delays in the system response in terms of the pointing facilitation routine.

C. Handling Perturbed Pointing Trajectories

When the user input is perturbed in a moving vehicle due to the road and driving conditions, the predictive display system can handle noisy free hand pointing gestures by setting the noise covariance in the motion model in (6) relative to the measured (experienced) in-vehicle vibrations-accelerations. This conforms with the modelling assumptions and a higher covariance corresponds to having less certainty in the inferred endpoint-driven latent state $X_{i,t}$, i.e., pointing finger position, velocity, etc. This technique is suitable for low to medium perturbation levels that can be represented by Gaussian noise, for instance, driving on smooth to moderately bumpy-paved roads. The output of the filters, calculating the posterior of each nominal destinations $p(\mathcal{D}_I = \mathcal{D}_i|Y_{1:k})$ at t_k, can be used to estimate the posterior probability of the system latent state X_{t_k}, including the perturbation-free pointing finger position. This is given by the the Gaussian mixture

$$p(X_{t_k}|Y_{1:k}) = \sum_{i=1}^{N} p\left(X_{i,t_k}\,|Y_{1:k}\right) p\left(\mathcal{D}_I = \mathcal{D}_i\,|Y_{1:k}\right), \tag{9}$$

where $p\left(X_{i,t_k}\,|Y_{1:k}\right)$ pertains to the i^{th} destination and is also calculated by the Kalman filter.

The assumption of Gaussian noise in a motion model can be overly restrictive in a highly perturbed environments, e.g., driving on rough terrain or a badly maintained road, since the pointing hand/finger can move in a highly erratic manner. It can exhibit sudden unintentional noise-related movements or jolts, as can be seen in Figure 3 for off-road driving. In such scenarios, the perturbations present can be treated as an additional nonlinear random jump process, denoted by P_t in the motion model, causing sudden large changes in the pointing finger position and velocity. For example, this can be modelled by the mean-reverting jump-diffusion velocity process whose SDE is given by

$$d\dot{P}_t = \sigma_p dW_{2,t} + \sigma_J dJ_t - \lambda_1 \dot{P}_t dt, \tag{10}$$

such that dJ_t is the instantaneous change in the jump process $J_t = \sum_{i=1}^{\tau_t} \rho_i$, with $\rho_i \sim \mathcal{N}(0,1)$, τ_t is the number of jumps in $[0,t]$ governed by a Poisson distribution, and the next jump time τ is set by an exponential distribution [18]. Likelihood estimation for such motion models relies on sequential Monte Carlo (SMC), particle, filtering [30], which is computationally costly and approximate compared to the original models in (6) with Kalman filtering, even in the efficient Rao-Blackwellized form [22], [30]. A practical alternative to applying this expensive inference procedure N times, one per destination, is to apply the SMC filtering once as a pre-processing stage prior to the destination prediction routine. The pre-processing objective is to remove the most severe effects of large jolts from the gesture-tracker observations $Y_{1:k}$ at t_k and allow the utilisation of the original linear motion models for intent inference [10], [28]. This approach represents a compromise

between the better filtering results of the jump model in a high-perturbations environment, and the computational efficiency of the original models.

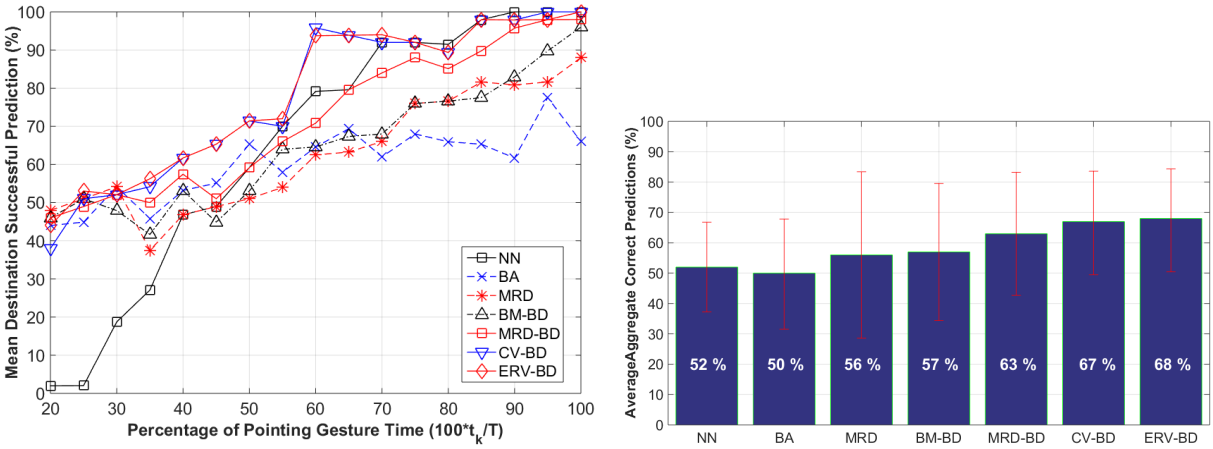
Applying a pre-processing SMC filter or dynamically adjusting the motion model covariance can be guided by additional sensory data, such as changes in the suspension height (by probing the vehicle CAN bus), IMU accelerometer, front-facing cameras, etc. These can reliably measure the level of accelerations and vibrations experienced in the vehicle. Additionally, the filtered free-hand pointing gesture can be used not only for pointing, but also for general gesture-based interactions.

V. Performance Analysis: Empirical Results

The performance of the intelligent predictive display concept is assessed here using data collected in two cars (Jaguar XK and Range Rover) instrumented with the system in Figure 1 under various road and driving conditions, namely when the vehicle is: 1) stationary, 2) driven over a well maintained road (i.e., motorway) at varying speeds, 30-70 mph, and 3) driven on a badly maintained roads with rutted and potholed surfaces with random patches and manhole covers raised-sunken where mild to severe in-car perturbations are experienced. A Leap Motion sensor is used to track, in real-time, the free hand pointing gestures (namely pointing finger-tips) and an experimental GUI is displayed on an 11.5" touchscreen mounted to the car dashboard; the attached videos show an early prototype of a predictive touchscreen system. The interface has $N = 21$ selectable circular icons, each of width $W \leqslant 2$cm that are approximately 2 cm apart in a circular formation, identical to that in Figure 3; the detailed set-up is described in [8]. Similar to the Fitts' law task in HCI, one randomly chosen GUI item is highlighted at a time and the user is expected to select it via a free hand pointing gesture. To maintain an objective testing procedure, all possible endpoints in $\mathbb{D}$ are assumed to be equally probable, $Pr(\mathcal{D}_I = \mathcal{D}_i) = 1/N, i = 1, 2, ..., N$. Maximisation of the likelihood function $\prod_{j=1}^{J} p(Y_{1:n}^{j} \mid \mathcal{D}_I = \mathcal{D}_i, \Omega)$ for a sample of J typical full pointing finger trajectories is used to set the motion model parameters Ω, and thus, constitutes training for the system. Next, the performance results of several Bayesian predictors and an in-car prototype system are examined. It is emphasised that predictors have no knowledge of the user intent in any of the experiments below.

A. Endpoint Prediction Performance

To examine the prediction accuracy throughout the pointing task, from its start at t_1 until touching the display surface at time T, 50 *a priori* recorded in-car full pointing gestures are used; no pointing facilitation routine is applied. The inference performance is evaluated in terms of: 1) the ability to determine the intended on-screen icon via a MAP estimate $\hat{\mathcal{D}}(t_k) = \arg\max_{\mathcal{D}_i \in \mathbb{D}} p(\mathcal{D}_I = \mathcal{D}_i \mid Y_{1:k})$, i.e., how early the predictor assigns the highest probability to true endpoint $\mathcal{D}^+$ and 2) the aggregate inference success, i.e., proportion of the total pointing gesture (in time) for which the predictor correctly inferred $\mathcal{D}^+$. The success is defined by $S(t_n) = 1$ if $\hat{\mathcal{D}}(t_n) = \mathcal{D}^+$ and $\mathcal{S}(t_n) = 0$ otherwise, for observations at times $t_n \in \{t_1, t_2, ..., T\}$. Whilst $J = 5$ pointing trajectories are used for training, the prior on the distribution of the durations of typical in-car pointing tasks, $p(T \mid \mathcal{D}_I = \mathcal{D}_i)$, for



(a) Mean successful destination inference. (b) Gesture portion (in time) with successful predictions.

Figure 6: Endpoint inference performance with a MAP estimate as a function of the percentage of pointing time [28].

the bridging-distribution predictors is obtained from the experimental study in [8]. It is noted that utilising 10% of the available tracks to set the model parameters is aimed at demonstrating the low training requirement of the applied state-space-modelling-based inference approach. This feature is highly desired in an automotive context as discussed in Section II. However, as the driver/passenger uses the predictive display, the system can refine the applied model parameters from the larger available data set(s). This can result in a more accurate modelling and prediction procedure.

In Figure 6, the linear destination reverting, Brownian motion and constant velocity models with the bridging prior, notated by MRD-BD, ERV-BD, BM-BD and CV-BD, are assessed. A mean reverting diffusion model without bridging, MRD, is also examined. The figure also depicts the outcome of the probabilistic Nearest Neighbour (NN), which assigns the highest probability to GUI item closest to the current position of the pointing finger-tip as per $p\left(Y_n|\mathcal{D}_I=\mathcal{D}_i\right)=\mathcal{N}\left(Y_n;d_i,C_{\text{NN}}\right)$, and Bearing Angle (BA) where $p\left(Y_n|Y_{n-1},\mathcal{D}_I=\mathcal{D}_i\right)=\mathcal{N}\left(\theta_{i,n};0,\sigma^2_{\text{BA}}\right)$ [10]. The latter assumes a minimal cumulative angle to the destination located in 3D at d_i; C_{NN} is the covariance of the multivariate Gaussian distribution and $\theta_{i,n}\triangleq\angle\left(Y_n-Y_{n-1},d_i\right)$ is the angle to $\mathcal{D}_i\in\mathbb{D}$.

Figure 6 illustrates that the bridging-distributions-based inference models CV-BD and ERV-BD, achieve the earliest successful predictions, since they capture the importance of the velocity component. This is particularly visible in the first 70% of the pointing task in Figure 6a, where a pointing facilitation scheme can be most effective. Destination prediction towards the end of the pointing gesture can have limited impact, since by that stage the user would have already dedicated the necessary attention-effort to execute the selection task . The performance of all depicted predictors generally improves as the pointing finger is closer to the display. This is particulary visible for the NN model, which is built on the premise that the pointing finger is closest to the intended endpoint. An exception is the BA model, since the reliability of θ_n as a intent measure declines as $t_n\rightarrow T$. Overall, this figure shows that probabilistic predictors can successfully infer the intended destination on the display remarkably early in the free pointing gesture. For example, in 60% of cases, the

bridged ERV model, ERV-BD, can infer the true intent only 40% into the pointing gesture (with overall correct decision exceeding 65%), thus, it can reduce pointing time-effort by over 60%.

The gains of combining the MRD motion model with the bridging method are noticeable in Figure 6a. This is due to to the ability of bridging technique (the prior on X_T) to reduce the sensitivity of LDR models to variability in the processed tracks; it tapers the system sensitivity to parameter estimates and the parameter training requirements.

B. Real-time Results from a Prototype System

Here, results from a pilot user study with 20 participants are presented. Whilst none of the participants have used an intent-aware display before, the study employs a prototype in-car intelligent predictive touchscreen system that performs intent inference in real-time and seamlessly implements the *mid-air selection* facilitation scheme as discussed in Section III-C; see the attached videos for a demonstration. An audible cue, i.e., a short ping sound signal, is produced by the predictive display to confirm to the user that an interface icon has been auto-selected. The *subjective* workload of interacting with an in-vehicle touchscreen with and without the predictive functionality is recorded using the NASA TLX test [31], which is widely utilised in HMI-HCI studies. It requires the participant to complete a questionnaire to rate and weight the mental, physical and temporal demand as well as performance, effort and frustration experienced when carrying out the in-vehicle pointing tasks. The durations T of accomplishing selection tasks in the trials are also assessed. This can be viewed as an objective measure of the effort involved.

When the predictor is off, the trial is a classical experiment of interacting with a conventional touchscreen, where the user has to touch the display surface to select a GUI icon. Whereas, with the prediction and mid-air selection functionality on, the intent-aware predictive touchscreen often executes the selection action for the user. An auto-selection action is triggered at time $t_k \leqslant T$ once the calculated probability for a given GUI icon, as per the estimated $p(\mathcal{D}_I = \mathcal{D}_i|Y_{1:k})$, exceeds a set threshold for a predefined period of time (on average, threshold $\gamma = 0.55$ and its duration $T_S = 65$ ms are set empirically). When this prediction certainty requirement is not met or the pointing finger is not detected, the user can continue pointing until he/she touches (and selects) the intended interface icon. Since the system is not aware of the user intent, any erroneous selection of the unintended GUI icon will lead to a longer pointing time and higher subjective workload, e.g., a higher frustration score.

Figure 7 shows that the interactions subjective overall workload declines by over 47% when employing the predictive display system in Figure 1 with mid-air selection, which is a substantial reduction. Figure 8 depicts the normalised histogram of pointing tasks duration T for over 8,000 selection tasks for all 20 participants. This figure illustrates that T is reduced when the prediction-autoselection functionality is on. In particular, the histogram in Figure 8b is visibly shifted to the left with smaller durations being more frequent and high values (indicating lengthy effortful pointing

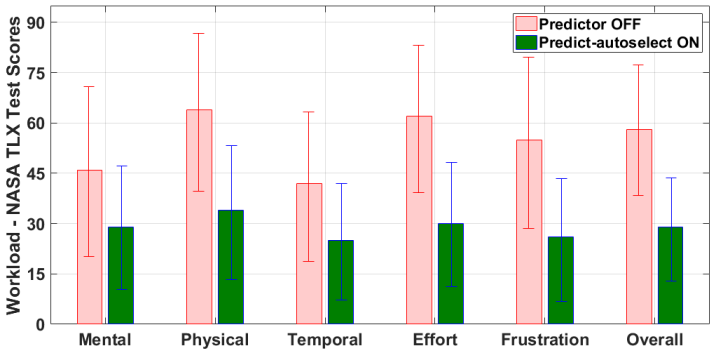
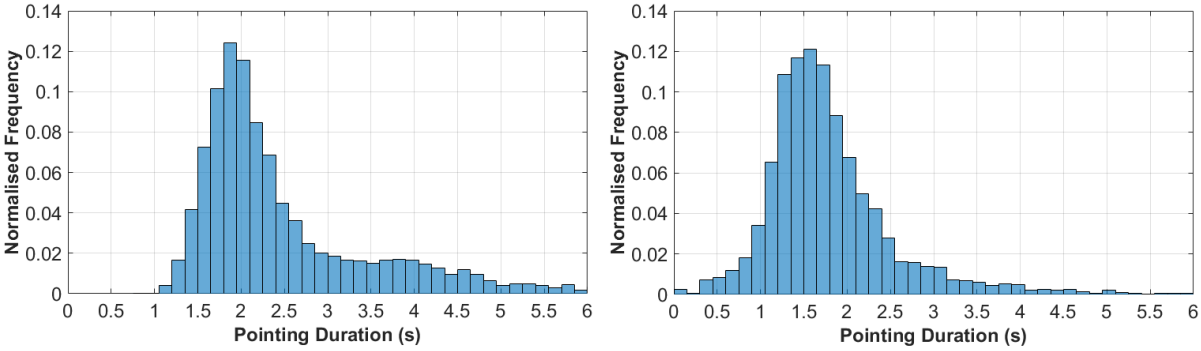


Figure 7: Mean workload of using an in-car display with and without the predictive functionality for 20 participants.



(a) Predict-autoselect off; $\mu = 2.63$s and $\sigma = 1.34$. (b) Predict-autoselect on; $\mu = 1.82$ s and $\sigma = 0.84$.

Figure 8: Pointing time with and without prediction-autoselection functionality for 20 participants.

gestures) are less recurring. On average, the introduced predictive solution reduces the duration of accomplishing an on-screen selection task via a free hand pointing gesture by approximately 30.75%. Higher reductions on the pointing time can be achieved, see Figure 6, by relaxing the requested prediction certainty (threshold or its duration) at the expense of, possibly, increasing the number of false auto-selections. This can have a negative impact on the user experience and system acceptance. It is a trade-off that has to be taken into account and the decision criterion can be adaptively changed based on the user requirements and the controlled IVIS functionality or the displayed GUI.

C. *Remarks on Results*

Since interactions with displays in modern vehicles are prevalent [2], small improvements in the pointing task efficiency, even reducing its duration by a few milliseconds, can have significant aggregate benefits on the user experience, notably for drivers. Therefore, the overviewed predictive solution can substantially reduce the effort and distraction of using in-vehicle interactive displays. However, further experimental evaluation is required for other pointing facilitation schemes, *in lieu* of mid-air selection that involves taking an action on behalf of the user. Additionally, devising a principled approach to setting the decision criterion for auto-selection according to the general cost minimisation problem in (4) is an open research question.

VI. Conclusions

Recent advances in sensing, data storage and communications technologies have led to the introduction of new smart vehicle functionalities and services aimed at offering personalised, more pleasant and safer driving experience. Nevertheless, little attention is often paid to the human machine interface aspect of these functionalities, for instance, interacting, controlling and customising them. Such interactions can be highly effortful and distracting, especially for drivers, with potential safety consequences. The reviewed concept of intelligent predictive displays in this article presents itself as a promising smart HMI technology. It can significantly reduce the effort and distractions associated with using an in-vehicle interactive display, which typically serves as a gateway to the available in-vehicle infotainment systems and services. This solution, whose cornerstone is suitable statistical signal processing algorithms, can also enable interacting with displays that do not have a physical surface, such as head-up displays for augmented reality and projections of 3D interfaces; such displays are poised to proliferate rapidly in the automotive environment in the near future.

Within the introduced general Bayesian framework, additional sensory or semantic data, if available, can be easily incorporated to enhance the prediction capabilities of the intelligent display and its handling of perturbed free hand pointing gestures due to road and driving conditions. The perturbations filtering aspect of this solution can be beneficial to general gesture-recognition-based interfaces in vehicles, not only pointing. Moreover, the predictive system can offer additional flexibilities in terms of the interface design and display placement in the vehicle interior as users might only need to reach (not necessarily touch) the display, with the mid-air selection scheme. This can be viewed in the context of inclusive design and ergonomics, where the display response or operation mode can be tailored to the user profile and motor abilities. Predicting the intended endpoint of a free hand gesture can extend, beyond the touchscreen, to various other items within the vehicle, such as the various mechanical controls.

Although a number of predictors that are based on Gaussian motion models were discussed here, several other probabilistic approaches can be employed within the presented Bayesian formulation, such as interacting multiple models [22], stochastic context-free grammars [16], and other destination-aware tracking algorithms. Whilst the presented empirical results testify to the efficacy of the intelligent predictive display system, this solution can benefit from future advancements in in-vehicle sensing technology, probabilistic intent inference algorithms, Bayesian decision strategies, fusion of multiple sensory data (not only gesture) and others. This article serves as an impetus for further research into using signal processing or machine learning techniques to alleviate the effort and attention required to interact with smart infotainment, connectivity and safety services in vehicles.

Acknowledgment

The authors would like to thank Jaguar Land Rover for funding this work under the the CAPE agreement and supporting the data collection on their test-track at the JLR Gaydon Centre, U.K.

REFERENCES

[1] R. Bishop, *Intelligent Vehicle Technology and Trends*. Artech House, Inc., 2005.

[2] C. Harvey and N. A. Stanton, *Usability Evaluation for In-vehicle Systems*. CRC Press, 2013.

[3] Tesla Motors, *An Evolution in Automobile Engineering (ModelS). Accessed on 18 /03/16: www.teslamotors.com/models*.

[4] K. Bark, C. Tran, K. Fujimura, and V. Ng-Thow-Hing, "Personal Navi: Benefits of an augmented reality navigational aid using a see-thru 3D volumetric HUD," in *Proc. of the Int. Conf. on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI)*. ACM, 2014, pp. 1–8.

[5] N. Broy, M. Guo, S. Schneegass, B. Pfleging, and F. Alt, "Introducing novel technologies in the car–conducting a real-world study to test 3D dashboards," in *Proc. of the 7th Int. Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI)*. ACM, 2015, pp. 179–186.

[6] Jaguar Land Rover, *XE In-Car Technology. Accessed on 18/03/2016: http://www.jaguar.co.uk/jaguar-range/xe/features*.

[7] N. Goode, M. G. Lenné, and P. Salmon, "The impact of on-road motion on BMS touch screen device operation," *Ergonomics*, vol. 55, no. 9, pp. 986–996, 2012.

[8] B. I. Ahmad, P. M. Langdon, S. J. Godsill, R. Hardy, L. Skrypchuk, and R. Donkor, "Touchscreen usability and input performance in vehicles under different road conditions: an evaluative study," in *Proc. of the Int. Conf. on Automotive User Interfaces and Interactive Veh. Apps. (AutomotiveUI)*, 2015, pp. 47–54.

[9] S. G. Klauer, T. A. Dingus, V. L. Neale, J. D. Sudweeks, and D. J. Ramsey, "The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data," National Highway Traffic Safety Administration, DOT HS 810 5942006, 2006.

[10] B. I. Ahmad, J. K. Murphy, P. Langdon, S. Godsill, R. Hardy, and L. Skrypchuk, "Intent inference for pointing gesture based interactions in vehicles," *IEEE Trans. on Cybernetics*, vol. 46, pp. 878–889, 2016.

[11] E. Lank, Y.-C. N. Cheng, and J. Ruiz, "Endpoint prediction using motion kinematics," in *Proc. of the SIGCHI Conf. on Human factors in Computing Systems*, 2007, pp. 637–646.

[12] M. J. McGuffin and R. Balakrishnan, "Fitts' law and expanding targets: Experimental studies and designs for user interfaces," *ACM Trans. on Computer-Human Interaction*, vol. 12, no. 4, pp. 388–422, 2005.

[13] B. Ziebart, A. Dey, and J. A. Bagnell, "Probabilistic pointing target prediction via inverse optimal control," in *Proc. of the 2012 ACM Int. Conf. on Intelligent User Interfaces (IUI)*, 2012, pp. 1–10.

[14] P. T. Pasqual and J. O. Wobbrock, "Mouse pointing endpoint prediction using kinematic template matching," in *Proc. of the SIGCHI Conf. on Human Factors in Comp. Systems (CHI)*, 2014, pp. 743–752.

[15] E. Ohn-Bar and M. M. Trivedi, "Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations," *IEEE Trans. on Intelligent Transportation Systems*, vol. 15, pp. 2368–2377, 2014.

[16] M. Fanaswala and V. Krishnamurthy, "Detection of anomalous trajectory patterns in target tracking via stochastic context-free grammars and reciprocal process models," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 1, pp. 76–90, 2013.

[17] M. Goel, L. Findlater, and J. Wobbrock, "Walktype: using accelerometer data to accommodate situational impairments in mobile touch screen text entry," in *Proc. of the CHI*, 2012, pp. 2687–2696.

[18] B. I. Ahmad, J. K. Murphy, P. M. Langdon, and S. J. Godsill, "Filtering perturbed in-vehicle pointing gesture trajectories: Improving the reliability of intent inference," in *Proc. of IEEE International Workshop on Machine Learning for Signal Processing (MLSP '14)*, 2014.

[19] T. Bando, K. Takenaka, S. Nagasaka, and T. Taniguchi, "Unsupervised drive topic finding from driving behavioral data," in *IEEE Intelligent Vehicles Symposium (IV)*, 2013, pp. 177–182.

[20] P. M. Fitts and J. R. Peterson, “Information capacity of discrete motor responses.” *Journal of Experimental Psychology*, vol. 67, no. 2, p. 103, 1964.

[21] D. A. Castanon, B. C. Levy, and A. S. Willsky, “Algorithms for the incorporation of predictive information in surveillance theory,” *International Journal of Systems Science*, vol. 16, no. 3, pp. 367–382, 1985.

[22] Y. Bar-Shalom, P. Willett, and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*. YBS Publishing, 2011.

[23] J. Berger, *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media, 2013.

[24] T. Carter, S. A. Seah, B. Long, B. Drinkwater, and S. Subramanian, “Ultrahaptics: multi-point mid-air haptic feedback for touch surfaces,” in *Proc. of the 26th Annual ACM Sym. on User Interface Software and Technology (UIST)*, 2013, pp. 505–514.

[25] Y. Zhang, S. Stellmach, A. Sellen, and A. Blake, “The costs and benefits of combining gaze and hand gestures for remote interaction,” in *INTERACT 2015*. Springer, 2015, pp. 570–577.

[26] J. R. Wilson and S. Sharples, *Evaluation of human work*. CRC Press, 2015.

[27] B. I. Ahmad, J. K. Murphy, P. M. Langdon, S. J. Godsill, and R. Hardy, “Destination inference using bridging distributions,” in *Proc. of IEEE ICASSP*, Brisbane, 2015, pp. 5585–5589.

[28] B. I. Ahmad, J. K. Murphy, P. M. Langdon, and S. J. Godsill, “Bayesian intent prediction in object tracking using bridging distributions,” *IEEE Transactions on Cybernetics*, 2017.

[29] A. Meucci, “Review of statistical arbitrage, cointegration, and multivariate Ornstein-Uhlenbeck,” *SSRN preprint 1404905*, 2009.

[30] O. Cappé, S. J. Godsill, and E. Moulines, “An overview of existing methods and recent advances in sequential Monte Carlo,” *Proceedings of the IEEE*, vol. 95, no. 5, pp. 899–924, 2007.

Bashar I. Ahmad (bia23@cam.ac.uk) received the B.Eng. (Hons.) in Electronic Engineering and Ph.D. degrees from the University of Westminster, U.K. He is currently a Senior Research Associate in the Signal Processing and Communications Laboratory, Engineering Department, Cambridge University, U.K. and a fellow of Wolfson College, Cambridge. Prior to joining Cambridge University, he was a postdoctoral researcher in the Signal Processing and Communications group at Imperial College London, U.K. His research interests include statistical signal processing, multi-modal human computer interactions, sampling theory and cognitive radio.

James K. Murphy (jm362@cam.ac.uk) received the undergraduate and Ph.D. degrees from Cambridge University, U.K., and the M.Sc. degree in mathematical modeling from Oxford University, U.K. He is currently a Research Associate in the Signal Processing and Communications Laboratory, Engineering Department, Cambridge University. Before this, he worked as a Senior Analyst with a quantitative finance consulting firm. His research interests include Monte Carlo methods for inference in univariate and multivariate time series problems.

Simon J. Godsill (sjg30@cam.ac.uk) received the B.A. and Ph.D. degrees from the University of Cambridge, U.K., in 1989 and 1994, respectively. He is a Professor of statistical signal processing in the Engineering Department, Cambridge University, Cambridge, U.K. His research interests include Bayesian and statistical methods for signal processing, Monte Carlo algorithms for Bayesian problems, modeling and enhancement of audio and musical signals, tracking, and high-frequency financial data. He has published extensively in journals, books, and conferences. He was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the journal Bayesian Analysis, and as a Member of IEEE Signal Processing Theory and Methods Committee. He has coedited in 2002 a special issue of the IEEE TRANSACTIONS ON SIGNAL PROCESSING ON MONTE CARLO METHODS IN SIGNAL PROCESSING, and has organized many conference sessions on related themes.

Patrick M. Langdon (pml24@cam.ac.uk) is a Principal Research Associate for the Engineering Department, Cambridge University, U.K., and a lead researcher in Inclusive design within the Engineering Design Centre. He received the undergraduate and Ph.D. degrees from the University of Sheffield, U.K. He has originated numerous collaborative research project in design for inclusion and HMI since joining the department in 1997. He has published extensively in various formats, as author, issues editor and co-author. Currently, he is the PI or Co-I of 3 industrial projects, 1 commercial collaboration in automotive, and a Co-I of a 4 year EPSRC project on HMI for Autonomous, Smart and Connected Control.

Robert W. Hardy (rhardy@jaguarlandrover.com) joined Jaguar Land Rover, Coventry, U.K., in 2011 and is a lead research engineer within the Human Machine Interface research team. He received his undergraduate degree in computer science at Lancaster University, U.K., and went on to complete a Masters and Ph.D. in the field of mobile and ubiquitous systems. He worked as a Research Associate at Lancaster University on a 3 year industrial project with the Japanese phone operator, NTT Docomo, working in the area of near field communications. His research focus at Jaguar Land Rover to date covers primarily interface design, head up display, vision systems, and haptics.