Deriving Extended Keystroke Level Model Resumability Operators: An Occlusion Study

Heather Wilson, David R. Large*, Gary Burnett

Human Factors Research Group, University of Nottingham, Nottingham. UK
*David.R.Large@nottingham.ac.uk

Abstract: Theoretical techniques to model and predict drivers' visual behaviour during the execution of secondary in-vehicle tasks, such as the extended keystroke level model (eKLM), are predicated on perfect task resumability during the interleaving of attention, i.e., it is assumed that the secondary task can resume without penalty – irrespective of task characteristics – as soon as attention is redirected to it. In practice, this is unlikely to be the case. Moreover, it is reasonable to opine that resumability may improve over increasing numbers of glances. A formative occlusion study was devised to explore search-and-touch HMI interactions in which the number of glances and task complexity varied, with the aim of deriving new eKLM resumability operators.

1. Introduction

Predictive modelling is a valuable tool that can be used to determine the expected level of visual demand associated with secondary tasks in a vehicle without extensive user testing (see: Large et al., 2018). However, modelling techniques often fail to account for any changes in performance due to the interleaving of primary and secondary tasks, notably when the secondary task requires multiple off-road glances. In theory, predictive models often assume that the secondary task can resume without interruption as soon as the driver re-engages their visual attention and that each subsequent glance has the same net effect on resumability. However, it is reasonable to opine that, in practice, a driver may be better prepared to resume the secondary task in each subsequent glance due to cumulative increases in task familiarity and mental preparation.

2. Background

2.1 Keystroke Level Model

The Keystroke Level Model (KLM) (Card, Moran, Newell, 1980) aims to predict how long it will take an expert user to accomplish a routine task without errors using an interactive computer system.

KLM reduces the elements of interacting with an interface or system to individual operators, each representing a unique activity, such as mental preparation, moving, pointing, swiping, etc. Each operator has a defined time allowance and thus, complete task-time can be calculated by summing individual times in the sequence they must occur.

In an automotive context, KLM task-time predictions can be used to estimate the visual demand (i.e. total duration of off-road glances) associated with an in-vehicle human-machine interface (IV-HMI). However, this may be distributed over several glances, with the driver redirecting their attention to the road during the intervening time.

KLM predictions fail to adequately articulate this interleaved attention. Indeed, directly equating task-time with off-road glance time fails to consider activities that may be possible (or may continue or conclude) without vision. It also fails to recognise that a driver will need to refamiliarise themselves with the secondary task in each subsequent glance.

The extended-KLM (eKLM) (Pettitt, Burnett & Stevens, 2007) aims to address the first point by incorporating the underlying principles of the visual occlusion protocol in its predictions (ISO, 2017). The approach is predicated on three assumptions:

1. During periods of vision, the operator sequence can progress without interruption;
2. An operator that begins during a period of vision can continue into an occluded period *providing* it is not reliant on the provision of vision (e.g., moving hand towards HMI);
3. An operator can *only* begin in an occluded period when vision is not required at any point in its duration (e.g., a keystroke operation in which the finger has already been moved to its target).

The authors demonstrate significant correlations between observed and predicted results in a validation study (Pettitt, Burnett & Stevens, 2007). Nevertheless, the technique assumes that secondary task operators requiring vision can resume without penalty in each new glance (i.e. perfect resumability), and therefore fails to account for potential changes in behaviour and performance over multiple glances.

3. Method

3.1 Overview

An occlusion study was devised to explore point-and-touch HMI interactions, in which the number of glances was enforced (1, 2 or 3). Participants were always required to select the target in the final glance. For example, if provided with two glances, they were instructed to use the first glance to visually acquire the target and mentally prepare, and the second glance to complete the task; participants were only allowed to move their hand/arm during the final glance, and were told to do so only when they were confident that they could make the correct selection (thereby effectively segregating searching and pointing time). Task complexity (based on the structuring of the interface, number of elements and Fitts' index of difficulty (Fitts, 1954)) varied.

3.2 *Participants and Experimental Setup*

Twelve experienced drivers took part. Participants sat in the University of Nottingham Human Factors driving simulator and were presented an in-vehicle touchscreen HMI located in the centre console of the vehicle.

Wearing occlusion glasses, participants were provided with either one, two or three "glances" for each task (and were told which before each task). Each glance provided 1.5s of vision (equivalent to a 2.0s off-road glance, in line with the occlusion protocol (ISO, 2017)). Thus, the total shutter open times (TSOT) were 1.5, 3.0, or 4.5s, respectively.

Targets were either 6mm (ID=6.0) or 12mm (ID=5.0) wide, based on relevant literature (Jin, Plocher & Kiff, 2007; Sesto, Irwin, Chen, Chourasia & Wiegmann, 2012), and presented as structured or unstructured arrays of 1, 4, 9 or 25 similar targets, arranged in a uniform square. Indices of difficulty (IDs) were calculated based on the width of the target and the distance from the participant's hand to the target when their hand was placed on the steering wheel at the "10 o'clock" position, in line with Fitts (1954).

3.3 *Procedure*

For each condition (defined by the permissible number of glances – one, two or three), participants were required to find and select the target containing a designated number, which was spoken aloud to them. Participants completed 12 tasks for each condition, and therefore completed 36 tasks in total.

Tasks were presented in a randomised order to counteract learning effects. Participants were observed via cameras located inside the driving simulator. Videos were subsequently analysed using Behavioural Observation Research Interactive Software (BORIS) (Friard & Gamba, 2016) to determine task-times and accuracy (i.e., task performance). In addition, *resumability time* was defined as the time it took participants to resume the task in the final glance (i.e., the time between the start of the final glance and the moment the participant's hand left the steering wheel).

3.4 *Hypotheses*

It was predicted that *resumability time* would reduce with increasing number of glances and increase with increasing task complexity. In addition, task performance (i.e., accuracy of item selection) was expected to improve with increasing number of glances and decrease with increasing task complexity.

4. Results

4.1 *Task Resumability*

Results show that the resumability time decreased with increasing number of glances, as expected. A Pearson's Correlation analysis showed that the resumability time for each task was negatively correlated to the number of glances, such that as the number of glances increased, participants took significantly less time to resume the task ($r=0.792$, $p<0.01$). A Linear regression equation was derived showing this relationship: $y = 0.77 – 0.24x$. There was no significant correlation between ID (i.e., index of difficulty) and resumability time. However, complexity in terms of array size (i.e., number of potential targets) was positively correlated with resumability time ($r=0.318$, $p<0.05$), indicating that it took longer to resume the task when there were more potential targets to select.

4.2 *Task Performance*

Results show that task performance improved with increasing number of glances, as expected (Pearson's correlation: $r=0.385$, $p<0.05$). In addition, a negative correlation was found between ID (i.e., task complexity) and accuracy, such that as ID increased, accuracy decreased ($r=-0.550$, $p<0.01$).

4.3 *Derived Resumability Time Operators*

NHTSA guidelines (NHTSA, 2013) stipulate that an IV-HMI should only permit tasks expected to have an average success rate of 85% or higher. Resumability time was therefore derived using only data where accuracy was 85% or higher. In practice, this equated to tasks with an ID of 5.0 or less (i.e., 12mm targets) (which also suggests that smaller targets would not be permissible on an IV-HMI).

Resumability time ('operators') differed based on secondary task characteristics (i.e., structuring and number of targets.) Mean values are summarised in Table 1. These should be applied in accordance with existing eKLM assumptions.

Table 1 Derived eKLM Resumability Operators

Number of Glances	Resumability Operator
1	.44s
2	.29s
3	.27s

5. Discussion and Conclusion

The results indicate a cumulative effect, i.e., the time to resume the task reduced in each successive glance. This suggests that drivers built up and retained residual task familiarity and could therefore respond more quickly in later glances. In addition, the results suggest that complex search-and-touch IV-HMI interactions could potentially be completed with greater efficacy if searching and pointing were segregated, i.e., by encouraging drivers to use an initial glance/s to mentally prepare/visually acquire their target and an additional, final glance to make the selection.

It is noted that to extract the required data, the number of glances was enforced, and there was no provision to execute the task sooner, even if the participant was ready; we also imposed a strict >85% accuracy criterion when deriving operators. This may have artificially extended task-time in some situations, e.g., for less-complex tasks undertaken over multiple glances, and limited formative data.

Using the newly derived *resumability* operators, predictions made using eKLM are likely to be more accurate, although findings will need to be validated with on-road/simulator data in further work.

References

Card, S. K., Moran, T. P., & Newell, A. (1980). The keystroke-level model for user performance time with interactive systems. Communications of the ACM, 23(7), 396-410.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology, 47(6), 381–391. https://doi.org/10.1037/h0055392

Friard, O., & Gamba, M. (2016). BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. Methods in ecology and evolution, 7(11), 1325-1330.

Hick, W. E. (1952). On the rate of gain of information. Quarterly Journal of Experimental Psychology, 4(1), 11–26. https://doi.org/10.1080/17470215208416600

ISO 16673:2017 (2017). Road vehicles — Ergonomic aspects of transport information and control systems — Occlusion method to assess visual demand due to the use of in-vehicle systems. Geneva, Switzerland: International Organization for Standardization (ISO).

Jin, Z. X., Plocher, T., & Kiff, L. (2007, July). Touch screen user interfaces for older adults: button size and spacing. In International Conference on Universal Access in Human-Computer Interaction (pp. 933-941). Springer, Berlin, Heidelberg.

Large, D. R., Burnett, G., Crundall, E., van Loon, E., Eren, A. L., & Skrypchuk, L. (2018). Developing predictive equations to model the visual demand of in-vehicle touchscreen HMIs. International Journal of Human–Computer Interaction, 34(1), 1-14.

NHTSA. (2013). Visual-Manual NHTSA Driver Distraction Guidelines For In-Vehicle Electronic Devices. NHTSA-2010-005

Pettitt, M., Burnett, G., & Stevens, A. (2007). An extended keystroke level model (KLM) for predicting the visual demand of in-vehicle information systems. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1515-1524).

Sesto, M. E., Irwin, C. B., Chen, K. B., Chourasia, A. O., & Wiegmann, D. A. (2012). Effect of touch screen button size and spacing on touch characteristics of users with and without disabilities. Human Factors, 54(3), 425-436.