The TEORT Problem:
Finding a Path to a Solution for Modern In-Vehicle HMIs

Bruce Mehler*1, Linda Angell2, Bobbie Seppelt1, Sean Seaman2, and Bryan Reimer1

1 Massachusetts Institute of Technology AgeLab & New England Univ. Transportation Center, USA
(bmehler@mit.edu, bseppelt@mit.edu, reimer@mit.edu)
2 Touchstone Evaluations, Inc., USA
(langell@touchstoneevaluations.com, sseaman@touchstoneevaluations.com)

Keywords: Attention management; Driver distraction; Glance metrics; Guidelines

EXTENDED ABSTRACT

The National Highway Traffic Safety Administration (NHTSA) of the U.S. Dept. of Transportation issued a set of voluntary visual-manual distraction guidelines for in-vehicle electronic devices in 2013 [1]. They were developed in response to a growing concern regarding the effects of distraction on motor vehicle safety [1-3]. This initial phase of a broader set of planned guidelines applies to original equipment in-vehicle human-machine interfaces (HMIs) operated through visual-manual means (i.e. involve a driver looking at a device, manipulating a device-related control, and/or looking for visual feedback/content) to engage in activities secondary to the primary driving task (e.g., communication, entertainment, information gathering, navigation). The purpose of the guidelines is to promote safety by discouraging the introduction of overly distracting devices in vehicles.

Functionally, the guidelines quantify visual-manual demand in terms of objective visual behavior. One method of meeting the guidelines is to demonstrate that at least 21 out of 24 participants from a defined age and gender balanced sample must fall at or under defined thresholds for three off-road glance metrics (mean single glance duration (<2 seconds), percentage of glances greater than 2 seconds (≤15%), and total eyes off road time (TEORT) (≤12 seconds)) during use of the HMI in specified driving simulation conditions [1].

While NHTSA describes these guidelines as being built upon principles developed as part of the earlier Alliance of Automotive Manufacturers (Alliance) guidelines [5], some important differences stand out, particularly as they impact the TEORT metric. One has to do with the definition of “eyes off road”. Under the Alliance approach, metrics consider only glances to any controls or display associated with a device. In contrast, NHTSA’s “off-road” designation applies to any glances not directed on the forward roadway. Thus, driving task-relevant glances to the instrument cluster to check vehicle speed, or to mirrors or out a window to check surrounding traffic all count as off-road glances under this definition.

NHTSA provided a pragmatic argument for this approach [6]. NHTSA-supported efforts ([6] footnote 105) found that eye tracking systems then in use did not have “enough accuracy to reliably characterize whether eye glances are focused toward the device upon which the task is being performed or toward some other in-vehicle location”, but were practical to use when categorization was limited to on or off forward roadway regions. Recent analyses of field data [7-8] have found that during standard visual-manual HMI tasks such as manual radio tuning, drivers almost exclusively limited their glances to the forward roadway and to the HMI under test, thus making the off-the-forward-roadway simplification realistic for such assessments. Nonetheless, HMIs have evolved significantly in recent years and may include both visual-manual and auditory-vocal (i.e., multi-modal) interactions. The previously mentioned work [7-8] also shows that multi-modal HMIs frequently include glances to safety-relevant off-road locations (e.g. mirrors, etc.). The need to evaluate HMIs
that include substantive auditory-vocal (voice-based) components that seem to engender such driving relevant glances bring into question the use of glance metrics that monotonically assign “off-road” glances during the period of a secondary task interaction to the “minus” column of TEORT. A series of field studies [9-16] looking at actual production voice-based HMIs suggest that many multi-modal HMIs would have significant difficulty meeting the NHTSA TEORT metric if it were applied (Figure 1).

**Figure 1.** TEORT collected in the field across seven studies and six vehicle models during voice-based, multi-modal navigation address and point of interest entry [9-16]. Colored bars represent the sample mean, vertical lines the standard error of the mean, and colored dots the 85% point in the sample distribution for individual tasks. The horizontal line shows the threshold at or under which approximately 85% (21 out of 24 participants) (dots) should fall if the NHTSA metric were applied.

The Alliance guidance [5] from 2006 specifically recognized that the then just-emerging voice-based HMIs might require future modifications in guidelines. The subsequent NHTSA Phase I guidelines [1] specifically state that they are not “currently” applicable to the auditory-vocal portions of HMI devices. This begs the question of how “portions” are functionally segregated in a multi-modal HMI where varying degrees of supporting information may be displayed or remain on a display during “speaking” or “listening” portions of a task. It is critical to consider the totality of glance behavior during a multi-modal HMI interaction in order to more fully take into account potential impacts on functional distraction. Voice-based interfaces should not be given a “pass” simply because auditory-vocal components are present (see also [17]).

Recent efforts have explored the utility of reconceptualized and modified versions of Kircher and Ahlström’s AttenD algorithm [18] to more broadly consider how attention is distributed across space and time, and to better understand how various features of resulting “attention buffer” metrics are associated with actual crash risk in naturalistic data [19-21]. Adjusted buffer metrics highlight substantive differences in patterns of glance behavior characterizing voice-based vs. primary visual-manual tasks across multiple HMIs (Figure 2). Reanalysis of existing field data indicates that further enhancements to the attention buffer model are sensitive to varying levels of cognitive load associated with auditory-vocal task engagement of working memory [22].

In brief, it can be argued that the TEORT metric has logical safety relevance in the context of classic visual-manual interfaces [23] that do not provide inherent task pacing. However, TEORT as used within NHTSA guidelines penalizes driving relevant glances to mirrors, the instrument cluster, etc. that are often present during longer multi-modal interface tasks. Further, it does not take into account the significance of on-road glance characteristics and the interleaving of off-road and on-road glances in impacting overall
situation awareness [19]. The TEORT 12 second metric developed around the model of standard visual-manual radio tuning does not logically have the same applicability if appropriate-length glances off-road are interspaced with sufficient on-road glance time, which, as threaded in sequence, reflect attention management more similar to baseline driving. The modified attention buffer takes these factors into account and has been demonstrated to have safety relevance to actual crash risk [19-21].

Figure 2. Attention buffer values for 56 HMI tasks across multiple vehicles in on-road studies.

This paper is intended to stimulate discussion and further efforts along a possible path toward a solution to the TEORT problem. Current NHTSA guidelines provide two options for assessing visual demands of HMIs: 1) Occlusion method, and 2) off-road Glance Metrics (mean single glance duration, percentage of long duration glances, and TEORT). Rather than replacing either of the existing options, this approach proposes adding a third. Under Option 3, the mean single off-road glance duration and percentage of long duration glance metrics would still apply (as both capture safety-relevant aspects of off-road glance behavior); however, TEORT (which can unfairly penalize modern, multi-step, multi-modal tasks) would be replaced by an attention buffer metric (which considers the strategic nature of how glances are distributed both off and on-road over the course of an HMI interaction). The rational for such an attention buffer metric would include, in part, a demonstrated link to safety in existing or to be developed naturalistic data. It is emphasized that this proposal does not call for removing or modifying the existing options, but to add an option that should allow for more appropriate assessment of the visual demand associated with multi-step, multi-modal tasks such as voice-based entry of addresses into a navigation system.

This proposal is referred to as a “path to a solution” as there is still work to be done on advancing a refined attention buffer model, developing guidance around a buffer metric, and developing basic science to support suggesting a threshold value for a new assessment option. While mean attention buffer values have been shown to differ significantly between crash and near-crash events [19-21], efforts are actively underway exploring a further hybrid measures that may provide more sensitivity in discriminating meaningful, safety-relevant HMI demand differences. Further collaborative investment by industry, academic partners, safety advocates, and governmental bodies are likely to enhance our shared understanding around how to reduce distraction and enhance supportive attention management in the vehicle.
References


