Effects of Speech Cursor on Visual Distraction in In-vehicle Interaction: Preliminary Results

Staffan Larsson
Simon Dobnik
Sebastian Berlin
University of Gothenburg
Box 200, SE-405 30 Gothenburg, Sweden
sl@ling.gu.se, simon.dobnik@gu.se, sebastian.berlin@gu.se

Abstract

This paper presents preliminary results on visual distraction tests concerning various multimodality solutions for in-vehicle dialogue systems in the SIMSI project. In particular, the Speech Cursor concept is evaluated in comparison to other solutions and is found to decrease visual distraction, especially for tasks involving list browsing.

1 Background

The goal of the SIMSI (Safe In-vehicle Multimodal Speech Interaction) project is threefold. Firstly, to integrate a dialogue system for menu-based dialogue with a GUI-driven in-vehicle infotainment system. Secondly, to further improve the integrated system with respect to driver distraction, thus making the system safer to use while driving. Thirdly, to verify that the resulting system decreases visual distraction and cognitive load during interaction. This demo paper describes the test environment designed to enable evaluation of the system, and the planned visual distraction tests.

Based on Larsson (2002) and later work, Talkamatic AB has developed the Talkamatic Dialogue Manager (TDM) with the goal of being the most competent and usable dialogue manager on the market, both from the perspective of the user and from the perspective of the HMI developer.

TDM supports multi-modal interaction where voice output and input (VUI) is combined with a traditional menu-based GUI with graphical output and haptic input. In cases where a GUI already exists, TDM can replace the GUI-internal interaction engine, thus adding speech while keeping the original GUI design. All system output is realized both verbally and graphically, and the user can switch freely between uni-modal (voice or screen/keys) and multi-modal interaction.

To facilitate the browsing of lists (a well known interaction problem for dialogue systems), Talkamatic has developed its Speech Cursor technology[1] (Larsson et al., 2011). By reading out the item currently in focus, it allows a user to browse a list in a multi-modal dialogue system without looking at a screen and without being exposed to large chunks of readout information. A crucial property of TDM's integrated multimodality is the fact that it enables the driver of a vehicle to carry out all interactions without ever looking at the screen, either by speaking to the system, by providing haptic input, or by combining the two. We are not aware of any current multimodal in-vehicle dialogue system offering this functionality.

The test environment consists of two parts, apart from the dialogue system: a driving simulator (SCANeR from Oktal) and an eye tracker (Smart Eye Pro from Smarteye).

2 Visual distraction tests

The main point of the visual distraction tests is to investigate how the "eyes-on-road" time during interaction varies between different modality conditions. The eyetracker equipment is used for capturing where the driver is looking. In addition, driving behaviour (including lane deviation) and dialogue state (including task success) is continously logged.

The following four variants were tested:

1. GUI only (haptic only in, graphics only out)
2. GUI with speech cursor (haptics only in, graphichs and speech out)
3. Multimodal with speech cursor (haptics and speech in, graphichs and speech out)
4. Speech-only with speech cursor (haptics and speech in, speech only out)

[1] The combination of Speech Cursor and spoken dialogue interaction is Patent Pending.

For each condition, there are two difficulty levels: (1) easy and (2) difficult. For both levels, the task is to drive along a softly curving road while keeping distance to one car in front of you and one car behind you. In the easy condition, the other cars have a constant speed. In the difficult condition, the other cars are speeding up and braking erratically, and the car behind you may indicate (by honking its horn) that you're going too slow.

This experimental setup, which we informally refer to as the "annoying cars" setup, differs from existing experimental setups such as the ConTRe task (Engonopoulos et al., 2008). In the latter, the driver tries to match two vertical lines representing the vehicle's position and the target (reference) position. Our setup has the advantage of being more realistic, although we acknowledge that it is still far from driving in real traffic. (On the negative side, our setup does require a full driving simulator environment, which the ConTRe task does not).

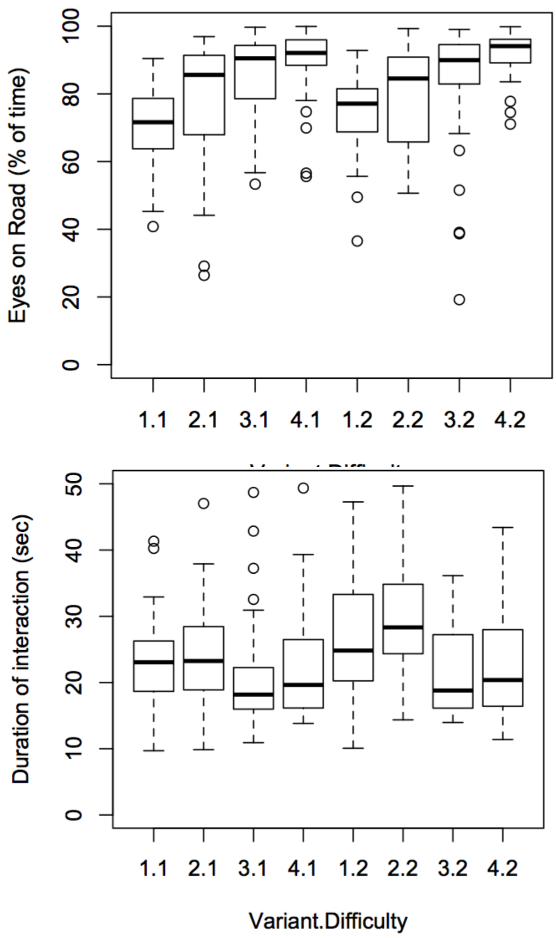
The application used in the tests has very basic phone functionality: browsing a list of contacts, and calling people up. At regular intervals, the driver receives a spoken instruction (with a voice different from the dialogue system), e.g. "You just remembered you need to call up Ashley on her mobile number.". The driver should then carry out this instruction as efficiently and completely as possible.

3 Results

This section presents results in the form of box plots[2]. The first box plot shows the % of time spent looking at the road in the different multimodality variants (the first number, as explained above), and difficulty levels (the second number, where 1=easy and 2=difficult). The second box plot shows the duration of interactions.

Even without spoken input, the Speech Cursor solution (variant 2) does better than GUI-only system (variant 1) w.r.t. visual distraction. Spoken input further (variant 3 and 4) reduces visual distraction, and reduces interaction time. The same trend was observed for both difficulty levels. The effect of modality condition on % Eyes on road has been tested with ANCOVA (with participant ID as co-variable) and was found to be significant at level $p < 0.001$.

[2] For an explanation of box plots, see e.g. http://en.wikipedia.org/wiki/Box_plot.



4 Discussion

From these preliminary observations, we can tentatively conclude that in tasks which require browsing, the Speech Cursor will significantly decrease visual distraction while browsing compared to a GUI only solution. This is true regardless of whether the system has spoken dialogue capabilities or not, at least insofar as spoken dialogue is not used for browsing[3].

The effect of this on overall visual distraction in in-vehicle interaction will depend on the amount of browsing carried out in an interaction, which in part will depend on the nature of the domain. For example, it's more common to browse for restaurants than to browse for who to call.

As the data is skewed, the normality assumption for statistical testing cannot be maintained and therefore we intend in future work to use statistical tests that are not dependent on this assumption, such as for example Generalised Linear Mixed Models (GLMMs).

[3] For example, Apple's voice-controlled CarPlay system requires the driver to look at the screen when browsing lists.

References

Nikolaos Engonopoulos, Asad Sayeed, and V Demberg. 2008. Language and cognitive load in a dual task environment. In *CogSci 2013*, pages 2249–2254.

Staffan Larsson, Alexander Berman, and Jessica Villing. 2011. Adding a speech cursor to a multimodal dialogue system. In *INTERSPEECH 2011, 12th Annual Conference of the International Speech Communication Association, Florence, Italy, 2011*, pages 3319–3320.

Staffan Larsson, Sebastian Berlin, Anders Eliasson, and Fredrik Kronlid. 2013. Integration and test environment for an in-vehicle dialogue system in the SIMSI project. In *Proceedings of the 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 2002–2004.

Staffan Larsson. 2002. *Issue-based Dialogue Management*. Ph.D. thesis, Göteborg University.