In-Car Dictation and Driver's Distraction: A Case Study

Martin Labský[1], Tomáš Macek[1], Jan Kleindienst[1],
Holger Quast[2], and Christophe Couvreur[2]

[1] IBM Czech Republic, V Parku 2294/4, 148 00 Praha 4 Chodov
`{tomas_macek,martin.labsky,jankle}@cz.ibm.com`
[2] Nuance Communications, Guldensporenpark Building D, Merelbeke, 9820 Belgium
`{holger.quast,christophe.couvreur}@nuance.com`

Abstract. We describe a prototype dictation UI for use in cars and evaluate it by measuring (1) driver's distraction, (2) task completion time, and (3) task completion quality. We use a simulated lane change test (LCT) to assess driving quality while using the prototype, while texting using a cell phone and when just driving. The prototype was used in two modes – with and without a display (eyes-free). Several statistics were collected from the reference and distracted driving LCT trips for a group of 11 test subjects. These statistics include driver's mean deviation from ideal path, the standard deviation of driver's lateral position on the road, reaction times and the amount and quality of entered text. We confirm that driving performance was significantly better when using a speech enabled UI compared to texting using a cell phone. Interestingly, we measured a significant improvement in driving quality when the same dictation prototype was used in eyes-free mode.

1 Introduction

The popularity of using various communication, navigation, entertainment and driver assistance systems in cars is steadily increasing as more and more of these systems enter the market at competitive prices. Text entry is one domain not covered by current production systems. According to [4], about 30% of drivers surveyed in Australia sometimes entered a text message using their mobile phone while driving and 1 out of 6 drivers did so regularly. At the same time, receiving and especially sending messages is perceived by drivers as one of the most distracting tasks [11]. As a consequence, the number of distraction-related crashes is thought to be increasing. The primary constraint when developing automotive UIs is thus to keep driver's distraction minimal. The secondary aims are to minimize task completion time and maximize task completion quality. In this paper, we evaluate a prototype text dictation UI according to these constraints and aims.

2 Related Work

A significant amount of research has been done that compares driving performance degradation due to using conventional and speech-enabled UIs [1]. The general conclusion is that while speech UIs still impact driving quality, they do so significantly

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 418–425, 2011.

less than conventional UIs. Most distraction caused by conventional systems seems to be due to drivers looking away from the road, as measured by the number and duration of eye gazes. In addition, using speech was observed to be faster for most evaluated tasks.

A number of approaches were also described to perform dictation in hands-busy environments [9]. In particular, hands-free text navigation and error correction were addressed by [8]. The impact of the most prevalent correction method, re-speaking, was evaluated by [10]. Microsoft described a prototype system [5] that allowed responding to incoming text messages by matching message templates.

3 Experimental Design

The dictation prototype evaluated in this paper, code-named *ECOR* (as for error-correction), allows for open domain, unconstrained dictation augmented by a number of methods that support the tasks of error detection, location and correction. While the system has a fully-fledged GUI, it can be used completely eyes-free in order to minimize driver's distraction. The system echoes recognized text chunks as they are dictated (using clearly speaking TTS) and allows for navigating and correcting dictated text.

A standard LCT simulator [6] was used to simulate driving in an office environment. The simulator was shown on a 22" screen and the *ECOR* screen showed on a separate 8", 800x600 touch-screen, positioned on the right side of the simulator screen. A Logitech MOMO steering wheel and pedals were used to control the simulator and 4 buttons (incl. push-to-talk) on the steering wheel controlled the prototype. Our setup was very similar to that used by [3,7] and to that described as "PC condition" by [2].

One LCT trip consisted of a 3km straight 3-lane road with 18 irregularly distributed change lane signs (lane changes of both 1 and 2 lanes were included). The evaluated segment started with an extra "START" sign and ended 50m after the last change lane sign. Drivers kept a fixed maximum speed of 60km/h (16.7m/s) during the whole trip.

4 Evaluation

A group of 11 test subjects (9 male, 2 female, aged 19-43, mean 33) was used to measure the degradation in driving performance due to distraction caused by entering text.

4.1 Procedure

First, each subject was allowed to train driving until they mastered the LCT. Then, 2 **reference** LCT trips were collected. One was used to compute an adapted model of the driver's ideal path and the second was used to compute driving performance statistics using the adapted ideal path.

The **ideal path** was modeled using a linear poly-line. Because each driver had a slightly different driving style, we adapted several parameters of the ideal path to accommodate for individual driving styles, in order to make the driving performance statistics more comparable among drivers. The major differences among drivers during their reference drives are listed below along with the corresponding adapted parameters of the ideal path:

- Different steering angles and durations of lane changes (maneuver lengths when changing 1 or 2 lanes in both directions).
- Different reaction times to lane change signs (distance before the lane change sign where the maneuver starts).
- Different standard driving positions within each of the 3 lanes (lateral car position offset for each lane).

After the reference LCT trips, each subject was introduced to the *ECOR* prototype with display and had sufficient time to practice dictating arbitrary text; initially when car was parked, then also while driving. The time spent practicing *ECOR* was 20-30 minutes, including 1-2 training LCT trips. After mastering *ECOR*, each subject conducted a single LCT trip **with display**, during which s/he was instructed to enter a sequence of text messages with pre-defined semantic content (e.g. “instruct your partner to buy oranges, wine and chocolate” or “tell your secretary to set up a meeting, at the library, tomorrow, at 5pm.”).

ECOR display was then switched off and subjects were allowed up to 20 minutes to practice using *ECOR* eyes-free, including 1-2 training LCT trips. This was followed by a single evaluated **eyes-free** LCT trip, during which the same sequence of messages was to be entered.

Finally, the subjects ware allowed 1-2 training LCT trips to practice driving while entering text using their own cell phone. After this, a single **cell phone** LCT trip was conducted while entering messages according to the same specifications. Two out of 11 subjects said they sometimes sent text messages using their cell phone while driving.

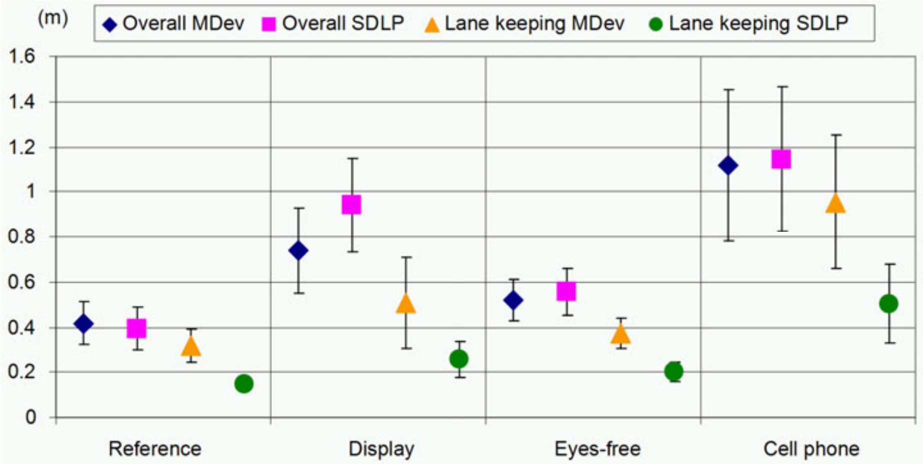

Fig. 1. Average MDev and SDLP with 95% confidence intervals

4.2 Driving Performance Results

Driving performance was measured for all 4 evaluated LCT trips using the following statistics.

- The car's mean deviation (MDev) from the ideal path in meters. This measures how much, on average, the driver drove off his/her ideal track, and is computed simply as the absolute value of subtracting the actual and ideal lateral car positions at each sampled point, and by averaging over the distance of the trip.
- The standard deviation of lateral position (SDLP) of the car in meters. SDLP measures how much the driver "weaves" within the lane and is computed as the standard deviation of the car's absolute actual deviation from its ideal path at each sampled point.
- The number of missed lane change signs per trip.
- The number of accidents during which the car's position went out of the road.
- Delay of reaction time to lane change signs.

Mean deviation and SDLP were evaluated over the whole LCT trip, and also only during its lane keeping phases. Averaged values of MDev and SDLP are shown in Figure 1. For the reference trip, both overall MDev and overall SDLP values were around **0.4m**. The next best result was achieved by the eyes-free setup, which had overall MDev and SDLP of **0.51 and 0.54m**, respectively, showing degradation of **10 and 15cm** on average. For the display trip, distraction was higher, with overall Mdev and SDLP of **0.73 and 0.93m**, with average degradation of **32 and 56cm**. The worst results were measured for the cell phone trip with overall MDev and SDLP of **1.11 and 1.15m**, with degradation of **70 and 77cm**. For the lane-keeping phases, we naturally observed lower MDev values, starting at Reference **0.32m**, followed by **0.37, 0.51 and 0.96m** for the Eyes-free, Display and Cell phone trips.

It is important to note that adapting the ideal path dramatically reduced abs. values of both MDev and SDLP, as opposed to using a single predetermined path. In our case, using a default unadapted ideal path would cause the absolute values of Mdev to be about **1m** higher for all types of trips, and up to **1.5m** higher for SDLP. Therefore, comparing ratios of these statistics for distracted and undistracted LCT trips across studies can be misleading.

Per-subject results (for 11 subjects) for overall MDev and overall SDLP for all 4 LCT trips are shown in Figure 2. We can see that the order, by level of distraction, is for most subjects as follows: Reference < Eyes-free < Display < Cell phone. We can also see that two subjects had SDLP significantly greater for the *ECOR* Display trip than for the Cell phone trip – in both cases this was due to the subject missing a sign in the *ECOR* trip but not missing any sign in the Cell phone trip[1].

---

[1] The impact of missing a lane change sign is dramatic both to the overall MDev and overall SDLP, since the actual path diverges from the ideal path by the width of 1 or 2 lanes for the whole lane-keeping range until the next sign. When computing the lane-keeping versions of MDev and SDLP in Figure 1, these impacted segments were excluded.

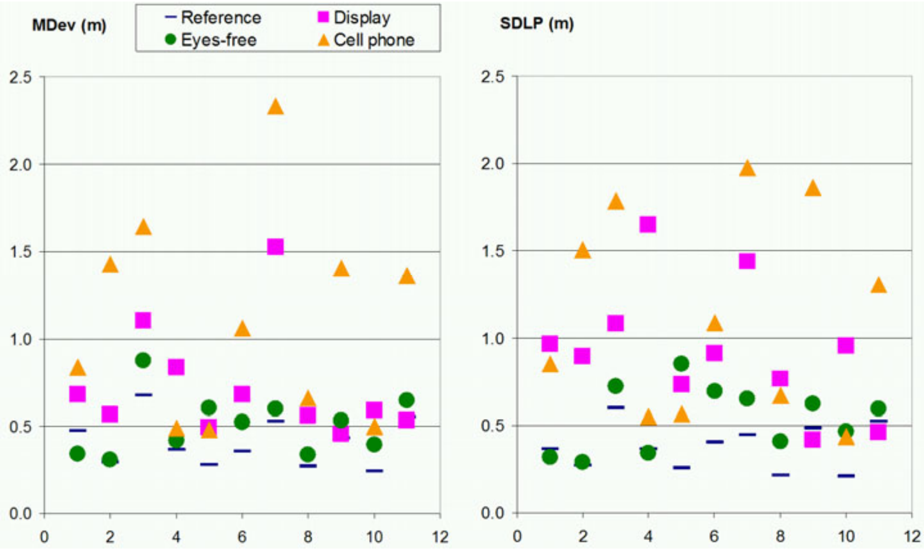

Fig. 2. Detailed values of MDev and SDLP for 11 subjects

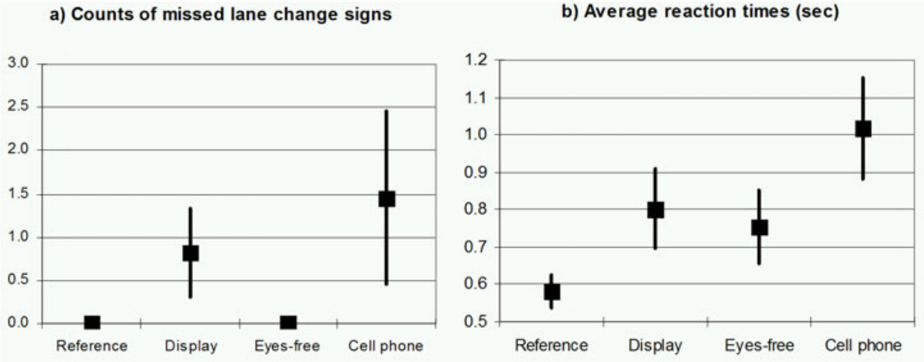

Fig. 3. Average numbers of missed lane change signs and average reaction times across all 4 LCT trips, all with 95% confidence intervals

Apart from MDev and SDLP, we also measured the number of **missed change lane signs**, computed by visual inspection of the actual and reference tracks using the LCT analysis tool; see Figure 3a. Missed signs only occurred when the secondary task included visual distraction and the number of missed signs had a high variance. The averages were **0.8 missed signs** per *ECOR* trip with display and **1.45 missed signs** per Cell phone trip. **Out-of-the road accidents** were all events when the actual car path went out of the road. These occurred only when operating cell phones (**0.7 accidents per trip**) and had a high variance – one subject had 4 accidents, another had 2, and two subjects had 1 accident each.

Driver's reaction times to lane change signs were measured as follows: the actual car path was examined in the range between 35m before the sign (sign visibility) and 20m after the sign. The first noticeable steering wheel movement of angle greater than 3° was identified and its start was considered as the time of driver's reaction to the sign. Reaction times shown in Figures 3a and 4 were computed by subtracting the time the sign became visible from the time of driver's reaction. When undistracted, drivers on average reacted in **0.58s**. The next best reaction times were **0.75s** for the Eyes-free setup and **0.8s** for the system with display. Cell phone users reacted in **1s**. Figure 4 shows 198 lane change reactions (11 subjects × 18 signs) for each of the 4 LCT trips. Note that Eyes-free users had less extreme delays in their individual reactions.

Finally, we analyzed subject's eye gazes during the 3 distracted LCT trips by manually annotating their videos. Table 1 shows highest road attention for the Eyes-free drivers. Subjects spent **9%** of their driving time looking at the display when it was available. For cell phone, the average time spent looking at its display was extreme **44%**. As cell phone typing consisted of many elementary operations, the number of gazes at its UI was also 3 times higher compared to speech interface with display.

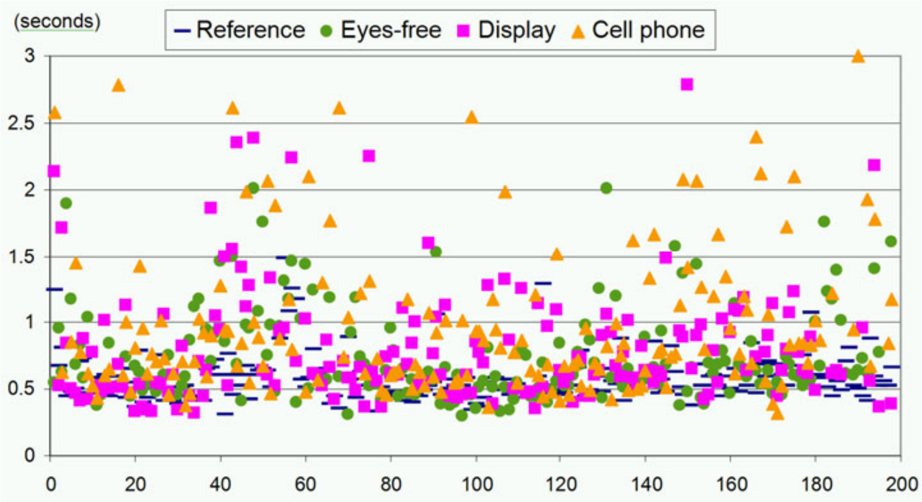

Fig. 4. Per-sign reaction times for all 4 LCT trips

4.3 Dictation Performance Results

In addition to driving performance, we also evaluated the subjects' performance of composing messages. For all 3 distracted trips, we collected texts of all composed messages. These texts were analyzed manually by a single test conductor and each message was scored on a 0-1 scale. High scores were assigned to messages that contained all prescribed semantic content and did not contain undesired text. The major scoring factor was the semantic understandability of the message. Typos that could be easily decoded had minimal impact on the scores.

In Figure 5 we can see that using voice to dictate messages was on average **significantly faster** than typing using mobile phone. On the other hand, **message quality was lower** for messages entered by voice. This is due to ASR errors being more destructive to message semantics as they typically mistake one or more whole words for other valid words. If not noticed and corrected by the user, ASR errors cause semantic mismatches that are much harder to decode by the addressee than isolated character typos often produced when typing messages manually.

When comparing the *ECOR* tests with and without display, message quality was slightly worse for the messages entered without display. Surprisingly, subjects were able to send slightly more messages without display, which we attribute to the subjects not noticing some recognition errors and therefore getting less “stuck” correcting dictated text.

4.4 Discussion

According to MDev, SDLP, missed sign counts and reaction times, the eyes-free dictation system resulted in significantly better results than the same system with display switched on. Based on the gaze data, we conclude that the visual distraction caused by reading text appearing on the display has significant adverse effects on driving. When available, subjects looked at the display even though all information was available via acoustic feedback.

Unlike [5], we consider the eyes-free use of an automotive dictation system possible by solving the problems of error detection and correction. Several features aiding both problems were already part of the evaluated prototype and several more are yet to be evaluated.

Table 1. Portions of time spent looking at the road, at the text entry UI, and elsewhere. Average counts and durations in seconds are shown for out-of-the-road gazes.

System \ Gazes at	Road	Text UI (%, #, sec)			Other (%, #, sec)		
Display	85%	9%	19	0.74	6%	23	0.46
Eyes-free	94%	-	-	-	6%	20	0.47
Cell phone	54%	44%	59	1.31	2%	4	0.70

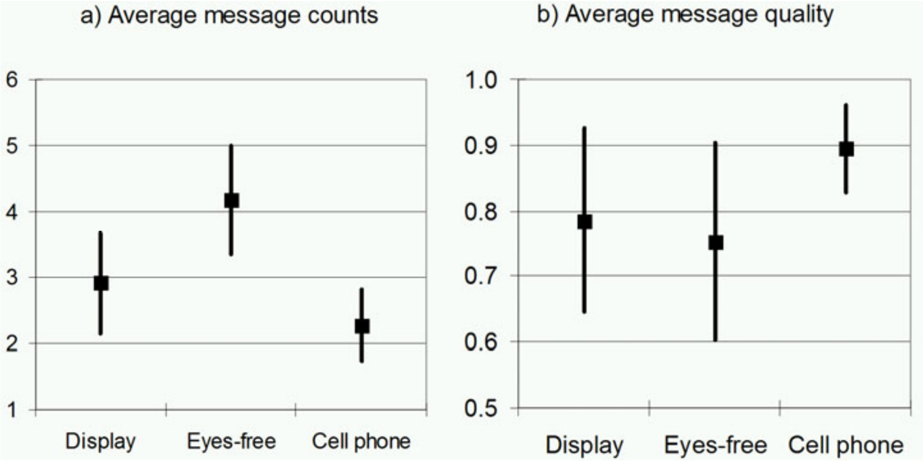

Fig. 5. Numbers of sent messages and message quality

5 Conclusion and Future Work

We presented results of evaluating an automotive text dictation system using a LCT simulator. We showed that when operating without a display, driving performance was significantly better compared to the same system with display. Both versions of the dictation system greatly outperformed cell phone typing in all aspects except for message quality, which we attribute to the character of ASR errors.

Acknowledgements. We would like to thank to our test subjects who participated in the evaluation (honored by hand-made wooden sticks allowing them to use the degree "Subject").

References

1. Barón, A., Green, P.: Safety and Usability of Speech Interfaces for In-Vehicle Tasks while Driving: A Brief Literature Review. Technical Report UMTRI-2006-5. University of Michigan Transportation Research Institute (2006)
2. Bruyas, M.P., Brusque, C., Auriault, A., Tattegrain, H., Aillerie, I., Duraz, M.: Impairment of Lane Change Performance due to Distraction: Effect of Experimental Contexts. In: Proc. European Conf. on Human Centered Design for Intelligent Transport Systems (2008)
3. Harbluk, J.L., Mitroi, J.S., Burns, P.C.: Three Navigation Systems with Three Tasks: Using the Lane-change test to Assess Distraction Demand. In: Proc. 5th Intl. Driving Symp. on Human Factors in Driver Assessment, Training and Vehicle Design (2009)
4. Hosking, S., Young, K., Regan, M.: The effects of text messaging on young novice driver performance. In: Faulks, I.J., et al. (eds.) Distracted Driving, pp. 155–187. Australasian College of Road Safety, NSW (2007)
5. Ju, Y.C., Paek, T.: A Voice Search Approach to Replying to SMS Messages in Automobiles. International Speech Communication Association (2009)
6. Mattes, S.: The Lane-Change-Task as a Tool for Driver Distraction Evaluation. In: Proc. Annual Spring Conference of the GFA/ISOES (2003)
7. Petzoldt, T., Bär, N., Krems, J.F.: Gender Effects on Lane Change Test Performance. In: Proc. 5th Intl. Driving Symp. on Human Factors in Driver Assessment, Training and Vehicle Design (2009)
8. Suhm, B., Myers, B., Waibel, A.: Multimodal error correction for speech user interfaces. In: Proc. ACM Transactions on Computer-Human Interaction, TOCHI (2001)
9. Oviatt, S., et al.: Designing the user interface for multimodal speech and pen-based gesture applications: State-of-the-art systems and future research directions. HCI 15(4), 263–322 (2000)
10. Vertanen, K.: Speech and Speech Recognition during Dictation Corrections. In: Proc. Interspeech (2006)
11. Young, K.L., Regan, M., Hammer, M.: Driver Distraction: A Review of the Literature. Monash University Accident Research Centre. Report no. 206 (2003)