

Evaluating the Usability of a Head-Up Display for Selection from Choice Lists in Cars

Garrett Weinberg

Mitsubishi Electric Research Labs
201 Broadway, Cambridge, MA, USA
+1.781.565.4814

Garrett.Weinberg@nuance.com
(now with Nuance Communications)

Bret Harsham

Mitsubishi Electric Research Labs
201 Broadway, Cambridge, MA, USA
+1.617.621.7500

harsham@merl.com

Zeljko Medenica

University of New Hampshire
Durham, NH, USA
+1.603.862.3959

zeljko.medenica@unh.edu

ABSTRACT

It has been established that head-down displays (HDDs), such as those commonly placed in the dashboard of commercial automobiles, negatively affect drivers' visual attention [1]. This problem can be exacerbated when screens are "busy" with graphics or rich information. In this paper, which is an extension of a user-preference study [23], we present the results of a driving simulator experiment where we examined two potential alternatives to HDDs for presenting textual lists. Subjects conducted a series of street name finding tasks using each of three system variants: one with a head-down display (HDD), one with a head-up display (HUD), and one with only an auditory display. We found that the auditory display had the least impact on driving performance and mental load, but at the expense of task completion efficiency. The HUD variant had a low impact on mental load and scored highest in user satisfaction, and therefore appears to be the most viable target for future study.

Categories and Subject Descriptors

H.5.2 [User Interfaces]: Graphical User Interfaces; Interaction Styles; Voice I/O

General Terms

Design, Human Factors, Experimentation

Keywords

Speech recognition, driving simulation, head-up display, HUD, head-down display, HDD, auditory display

1. INTRODUCTION

Head-up displays are partially-transparent displays that render information in a manner that allows the viewer to comprehend it while looking into the forward scene [9]. General Motors introduced the first automotive HUD in 1988 [19], and HUDs have since been deployed on a variety of vehicles from a variety of manufacturers, though mostly in the luxury segment or as costly optional upgrades. They typically display only limited, critical information such as the speed of the vehicle or the direction of an upcoming turn. However, this limitation has as

much to do with the small size and low resolution of current-generation HUDs as it does with concerns about driver distraction.

With the increased prevalence and precision of GPS and compass-based data about a vehicle's position, and the emergence of computer-vision technology that can recognize objects on and around the road and can measure the driver's head position within the cabin, there has been a dramatic upswing recently in research on richer, higher-resolution HUDs that could enable augmented-reality automotive applications. Narzt et al. [14] envision a whole-windshield HUD with navigational information overlaid onto the road surface as transparent colored paths. Charissis et al. [5] prototyped a system that superimposes virtual representations of objects onto the driving scene, providing early warnings of impending sharp turns or traffic slowdowns that might otherwise be missed due to weather conditions such as fog.

These designs, while exciting, are still in the early research phase. Their commercial deployment will depend on advancements in both object tracking/recognition algorithms and in display technology that can utilize the entire windshield as an HUD combiner, enabling full-color, high-resolution imagery to be presented at the correct focal depth without reducing the clarity and transparency of the windshield for portions of the display where there is no image.

The aim of this paper is to examine an HUD implementation that bridges the gap between the limited, low-resolution HUDs available in cars today and the full-windshield, augmented-reality HUDs envisioned for the distant future. We sought to examine what the benefits might be to a medium-sized HUD with only a few colors, occupying a region of the windshield comparable in size to current-generation dashboard LCDs, rather than the entire windshield area.

Furthermore, we wondered whether HUDs are appropriate for *non-critical textual information*, such as the destination- or media-related menus or lists that today are presented on head-down displays (HDDs). Such HDDs are typically mounted in the "center stack," as is the case with factory-installed systems.

With the proliferation of such devices and the increasing amount of rich media content available on them, there is a strong practical motivation to find safer ways of presenting this content. As we have seen with mobile phone use, just because an activity has been proven unsafe (e.g., [10]), that doesn't mean drivers will stop engaging in that activity. Thus, it is imperative that we develop displays and user interfaces that allow access to the ever-widening array of "infotainment" content with as little adverse effect on driving as possible.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

AutomotiveUI'11, Nov. 29–Dec. 2, 2011, Salzburg, Austria.
Copyright 2011 ACM 1-58113-000-0/00/0010...\$10.00.

As a foray into this research area we designed a driving simulator-based user study that compared subjects' driving performance, task performance, and reported workload when carrying out a typical in-vehicle information system (IVIS) task—street name retrieval—using three variants of a prototype IVIS: one with a HUD, one with a HDD, and one with no display at all (only audible navigation cues). We will present the design and results of this study after a brief discussion of the related literature.

2. RELATED WORK

Numerous studies have illustrated the benefits of HUDs for the presentation of information related to the operation of the vehicle itself. For example, Sojourner and Antin found that subjects operating a driving simulator that featured an HUD speedometer reacted significantly more quickly to salient cues in the driving scene versus a typical dashboard-mounted speedometer [15]. Burnett found that participants made fewer navigational errors (wrong turns) when an arrow graphic and contextual information about the surrounding roads was presented on a HUD versus on a HDD [3]. Charissis et al. used a simulator to test an HUD-based collision warning system in low-visibility conditions (typical of heavy rain or fog). They found that the HUD dramatically reduced the number of collisions and improved subjects' maintenance of following distance, when compared to a traditional HDD [4].

Less research has been done on drivers' ability to assimilate information *not* immediately related to vehicle operation, such as the music titles or points of interest, when this information is presented via HUDs. Often such information is best conveyed by textual descriptions, or some combination of textual, visual, and audible descriptions. Tsimhoni et al. investigated the best position for presenting short text messages on a full-windshield HUD, evaluated in terms of driving performance and workload [18]. They found that message locations within 5° of a straight-ahead gaze yielded the best performance and were preferred by subjects. Note, however, that these conclusions were obtained based on very short messages only (people's names), as was also the case with [1] (where the messages were simple navigation instructions). Our work seeks to determine whether such findings hold for a multi-line, information-rich HUD.

Liu and Wen's study of commercial truck drivers found that a HUD was superior to a HDD for conveying emergency-related information (e.g., pedestrian warnings and engine temperature status), but did not find significant differences for the more text-heavy navigational and "commercial goods delivery" tasks [13]. However, this was at least in part due to issues with the design of these tasks (the navigational task had over a 95% success rate with both display types and the goods delivery task had a 100% success rate with both).

3. EXPERIMENT

In considering what kind of text-comprehension task was most interesting to study, we chose to focus on the "pick-list" screen that is often found in contemporary navigation/infotainment systems that offer voice destination entry (VDE). This is a screen that is commonly used to display the automatic speech recognition (ASR) engine's top N suggested matches for a given input utterance. A user must choose an entry in the N-best list in order to proceed through the VDE task; for example, she must confirm which city name she has spoken in order to proceed to the street name selection step.

We conducted an experiment in which subjects were asked to carry out pick-list navigation tasks (finding specific street names)



Figure 1. Driving simulator with head-up (HUD) and head-down (HDD) screen positions indicated. A jog dial controller is mounted to the steering wheel.

using both a HUD-based IVIS and two "baseline" variants: one with a HDD and one without any visual display at all—only synthesized speech output.

3.1 Hypotheses

We hypothesized (A) that our HUD implementation would provide for higher user satisfaction than either the HDD variant (which is modeled after HDDs common in vehicles today) or the audio-only variant. In terms of visual attention, we hypothesized (B) that both the HUD and audio-only conditions would be superior to the HDD, as neither the HUD nor audio-only variants required glances away from the road. In addition we hypothesized (C) that the HUD would incur the lowest task time and cognitive load among the three variants, due to this variant's having a visual modality and its salient content remaining in the driver's field of view at all times. Finally, in terms of driving performance, we hypothesized (D) that HDD would fare the worst among the three variants, due to its high visual demand.

3.2 Driving Simulator

The experiment was conducted in a mid-range driving simulator that incorporates motion feedback in the form of 2-axis tilt and vibration that varies with vehicle speed and road texture. The driving scene is rendered on a 3x1, edge-to-edge array of 50-inch (127-cm) DLP rear-projection displays with a combined resolution of 3072x768 and a refresh rate of 60Hz (Figure 1). A customized version of the rFactor racing game [11] simulates vehicle physics, renders the graphics, and communicates with both the motion chair and our data logging plug-in. For more information on the simulator's software and hardware configuration, please see [21].

The virtual roadway used in this experiment was the same one used in [22]: a gently curving highway course with a single leading vehicle traveling at a speed between 35 and 55 mph (56 and 90 km/h) depending on the road curvature.

3.3 Prototype IVIS

The prototype voice-activated IVIS with which subjects interacted during the study was similar to that used in [22]. There was a main menu representing three domains of content: Navigation, Music, and Contacts, and each of these domains contained various subdomains and functions. Users could traverse the hierarchy either by issuing voice commands or by selecting and clicking using a vertically-oriented "jog dial" device (Figure 1) mounted

on the steering wheel (note that the jog dial device was the primary way of advancing through the pick-lists and had the same behavior in all three variants). In other words, all three prototype IVIS variants had identical manual and voice interaction modalities, while two of the three of them added the visual modality, which differed only in display placement.

One important difference from the version of the prototype employed in [22] was that for this study, Point of Interest search in the Navigation mode was replaced by a simulated street name search feature. This will be further explained in the Protocol section below.

3.3.1 Audio-only Variant

The auditory-display variant we implemented was similar to one that had showed promise in a previous study [20]. The chief aspect in which the pick-list implementation in our design differs from other audio-only pick-lists is its entirely user-paced nature. Many state-of-the-art IVIS will announce the top several choices one after another and then immediately open the microphone for input¹. Our system instead maintains a logical cursor that the user controls with the wheel-mounted jog dial device. Each entry is announced—prefaced by its number—only when the cursor reaches that entry. To actually activate an entry, the user must either press the jog dial’s central momentary switch while the cursor is on that item, or press the push-to-talk button and say the number of any item.

3.3.2 HDD Variant

The head-down variant of the IVIS was positioned near the steering wheel as illustrated in Figure 1, at a slight upward tilt to allow for easy reading. This HDD design was motivated by systems in production vehicles and approximates the location of an IVIS mounted in the upper center of the dashboard.

With the simulator’s seat in its rearmost position (as it was for the tallest of subjects), the center of the display was 30° right of center, approximately 23° below the line of sight for a 5 ft. 9 in. (1.75m) person. It occupied about 11 horizontal and about 8 vertical degrees of the field of view.

3.3.3 HUD Variant

The head-up variant was superimposed on the driving scene above the hood of the virtual car using a video overlay mixer. The original black background of the IVIS interface was removed by the mixer, resulting in a partially transparent interface. Only the pick-list text and the selection bar remained fully opaque. The HUD occupied a very small area of the driving scene. According to the reports of subjects in the post-experiment discussion, the opacity did not interfere significantly with the driving scene. As shown in Figure 1, the HUD was positioned to the right of the virtual roadway and the lead vehicle (14° to the right of center and approximately even with the vertical line of sight). It occupied about 8 horizontal and about 7 vertical degrees of the field of view.

Our HUD is displayed in the same plane as the driving scene, and thus represents an idealized heads-up display. We believe this is a fair representation of state-of-the-art HUDs, because they are designed such that their image appears to the viewer to be within the driving scene (at bumper depth or beyond). This positioning allowed subjects to glance away from the driving scene as often as necessary in short bursts, and then attend to it once again, all without changing their focal depth.

The three IVIS variants were identical in terms of their menu structure and their tactile and audio interaction affordances. The only variable factor was the presence and location of the visual feedback. This means that any artifacts of our particular IVIS implementation—such as the length of the pick-lists or the behavior of the jog dial device—are equally distributed among the conditions.

3.4 Experimental Design

Our study used a within-subject, repeated measures design with *interaction-technique* (visual interface style) and *repetition* as independent variables, and a number of dependent variables that measured the subjects’ interactions with the interfaces as well as the interfaces’ effects on driving and eye glance behavior.

In order to measure the performance of the user interfaces themselves, we recorded task time and subjective preference. With respect to the interfaces’ effects on driving performance, we measured headway maintenance (distance behind the lead vehicle), lateral deviation, driving speed, throttle depression, steering angle, and the number of glances and percent time looking away from the forward roadway. In addition we measured total subjective workload using the NASA-TLX survey [8].

In short, our design was: 24 participants x 3 conditions (head-up, head-down, audio-only) x 7 interactions = 504 trials in total. The data from all 24 subjects was used for analyzing the driving performance. However, due to the difficulties that the eye-tracking system had with two subjects, we had to reject their data and use only 22 subjects for the analysis of visual attention. Since driving performance and eye-tracking data were collected by separate systems, using different numbers of subjects in those two analyses is acceptable.

3.5 Protocol

Subjects were recruited from the local area. All were licensed drivers and native English speakers. The latter requirement was introduced to reduce variations in the number of speech recognition errors between subjects. In total, twenty four subjects, 10 female and 14 male (age M=22.45, SD=3.64), participated in this study. Each experimental session was about one hour and thirty minutes long, and subjects were compensated \$40 for their participation.

The experiment consisted of four drives: control, head-up, head-down, and audio-only. In the control drive the subjects operated the simulator without any IVIS interactions, whereas in the other drives, they interacted multimodally (voice + manual controls) with the given IVIS variant while simultaneously driving. Each drive lasted for seven minutes, and all subjects completed all four drives. All drives took place on the same simulated roadway, a gently curving coastal highway. To account for learning effect, we counterbalanced the order of the drives among subjects. The main task in all four drives was to follow a pace vehicle and to maintain a constant distance behind it, even if the pace vehicle slowed down unpredictably (which it in fact did at certain places on the

¹The current Ford Sync IVIS, for example, presents an audible pick-list in this fashion when recognition confidence is low. This consists of synthesized speech such as the following: “Say one after the tone for *play artist Vanilla Ice*. Say two after the tone for *play artist Vanilla Fudge*. Say three after the tone for *play artist Milli Vanilli*. <Beep>”

Table 1. Example interaction: Searching for the street Sunset Boulevard

User	System
<presses the push-to-talk button>	<beep>
<i>“Enter destination”</i>	<i>“Please press the listen button and say the name of the street”</i>
<presses the push-to-talk button>	<beep>
<i>“Sunset Boulevard”</i>	<shows the results in the list and reads the first entry>
<selects the desired street name from the list using the jog dial or voice commands>	<confirms the selection by reading the selected item>

course). Subjects were instructed that following the vehicle and driving safely had the highest priority, while all other tasks (i.e., listening to the experimenter’s prompts and operating the IVIS) had secondary importance. They were also encouraged to suspend any interaction with the IVIS if they found that it was distracting them from the primary task of driving.

Before starting the experiment itself, subjects had a five minute training period to get accustomed to the driving simulator. Before the first street-retrieval (i.e., non-control) condition, the subjects were trained on how to operate the IVIS until they were comfortable searching for street names on their own (the baseline HDD variant was always used for this training).

3.5.1 Street Finding Tasks

During the three IVIS drives, the experimenter prompted subjects to find various randomly-chosen streets from a database of all non-numerically named streets in California. Subjects were instructed to use whichever combination of voice commands and manual controls they found easiest; however, there was no purely-manual affordance for inputting the street names themselves, only voice. Table 1 gives an overview of the steps subjects followed to complete a typical street finding task. The same steps were used in all three conditions—only the means of presenting the result pick-list varied. The text-to-speech (TTS) messages issued by the system (such as instructions and the names of items under the cursor) were interruptible; if desired, subjects could “barge in” by pressing any button rather than waiting until the end of the message read-out.

Each attempt at a street retrieval was considered to be a separate trial for the purposes of our analysis (including multiple subsequent attempts to find the same street, in cases where it did not appear in the list at all or the subject did not notice it and repeated the search by issuing a new *“Enter destination”* command). The time between attempts to find different streets was not less than ten seconds. This allowed the subject some down-time to resume following the lead vehicle at the preferred distance and/or to correct any deviations from the travel lane.

A task was considered to be successfully completed when the sought item was found and activated, either by pressing the “listen” button and saying the number of the sought item within the match list, or by using the manual jog dial controller on the steering wheel to highlight and open the item. In case of HDD and

HUD conditions, a maximum of 7 items were displayed per page, with 2 pages in total. This gives a total of 14 possible positions where the sought item can be located. The choice of 7 results per page is based on multiple contemporary IVIS that show this number of results per page during voice dialogs (viz. current Mercedes C-Class and E-Class).

The position of the target item within the list was artificially controlled; in other words, the actual output of the speech recognizer was ignored in the street name phase of the voice dialog. 50% of the time, the target street name appeared at the top of the list, 25% of the time in a random position elsewhere on the first page (entries 2 – 7), and 20% of the time in a random position on the second page of results (entries 8 – 14). 5% of the time, the target item did not appear on the list at all, forcing the subjects to issue the *“Enter destination”* command again (at which point the target item was re-shuffled into the list according to the given distribution). The non-target entries on the list were filled with similar sounding street names chosen automatically using the Soundex algorithm, in order to maintain the appearance of real ASR while controlling for error rate.

The relatively low proportion of exact matches (50%) was chosen to emphasize any differences between the display variants in terms of the relative ease or difficulty of assimilating textual information and choosing the target item. Real ASR systems have top-1 rates better than 90% for tasks such as this (at least under favorable background noise conditions). However, accidents usually happen when a driver is distracted, e.g., after an ASR error where a driver has to compensate by finding the sought item from somewhere in pick-list (if it’s there at all). These error-remediation cases are exactly where real systems’ varying visual modalities will matter, so we artificially increased the ASR error rate for our tested systems in order to obtain more data from these interesting cases.

3.5.2 Task Timing

The subjects had a maximum of 90 seconds to complete a retrieval task. If time elapsed before successful completion, the task was marked as unsuccessful. In addition, as a means of reducing user frustration, we instructed subjects that they could at any point declare the task to be a failure if they could not find the target item. In either of these cases, the user was given ten seconds of recovery down-time, and then the experimenter began the next task.

Following the completion of each drive, subjects were instructed to complete the NASA-TLX survey considering the driving task alone or the combination of driving and retrieval tasks, as appropriate. At the end of the entire experiment, a user interface preference questionnaire was administered.

3.6 Data Analysis

Our driving performance metrics (Following Distance, Lateral Position, Speed, Steering Angle, and Throttle Percentage) were logged by the simulator software in real-time. The above variables are often reported together in literature because they can be differently influenced by the various experimental conditions. Thus, it is best to consider them in concert when drawing conclusions about driving performance.

Eye-tracking measurements were modeled after those discussed in [6]. We divided eyes-off-the-road glances into short (< 0.5 seconds), medium (0.5 – 2 seconds), and long (> 2 seconds) bins, as in previous work ([2], [22]). Eyes-off-the-road glances were

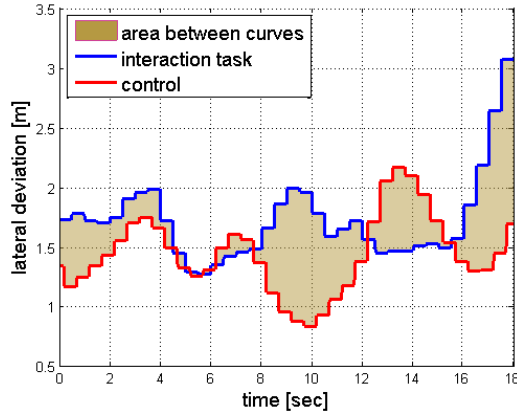


Figure 2. An example of calculating the area between the curves for lane deviation.

defined as those that targeted any area besides the three front simulator screens, such as those falling on the steering wheel or the HDD. Since in the HUD condition the visual feedback was presented within the central simulator screen, we delineated the area it covered in our eye-tracking software and counted glances onto that area as off-road glances.

In terms of usability analysis, our primary metric was retrieval task duration, as there were very few unsuccessful retrieval tasks.

3.6.1 Calculation of Driving Metrics

We extracted the driving data relevant to the street retrieval tasks using the “task windowing” methodology described in [22]. The driving data collected during a given retrieval task is compared with the driving data collected from the relevant portion of the same subject’s control drive. This approach normalizes across tasks for the difficulty of the roadway portion on which they occurred, as well as across subjects for varying driving ability. This way each subject acted as her own control (both with respect to road curvature and interaction type), enabling us to compare the IVIS interfaces by way of comparing the degree to which they affected the subject’s driving.

This was accomplished using a metric that we call “area between the curves,” which is similar to the “mean deviation per average task portion” metric discussed in [7]. As its name implies, our metric returns the area between two curves: one coming from the control session and the other coming from the particular retrieval task in the particular IVIS interaction session that is being analyzed (see Figure 2). The beginning and endpoint of each curve is obtained through the “task windowing” technique described above. The area between the curves is calculated, and then normalized by dividing by the duration of the task. In the ideal case, the secondary task has no impact on driving performance, so there is no difference between the conditions (control and IVIS), and the area between the curves is zero. Thus, for each driving metric, a smaller area indicates better driving performance. This enables us to compare the IVIS interfaces by quantifying the degree to which the interfaces cause drivers to diverge from their unencumbered driving.

3.7 IVIS Usability Results

This section presents the results of the usability analysis and the participants’ subjective opinion about the desirability of the three interaction techniques.

3.7.1 ASR Errors

Since the recognition accuracy of the search tasks was simulated according to the distribution specified in section “Street Finding Tasks,” there were no ASR related errors except the ones intentionally introduced. The only time when the ASR was actually used in this experiment was for the “Enter destination” command, for which the recognition was always perfect.

3.7.2 Task Time

Task time was defined as the amount of time which elapsed between the beginning and the end of a task. The beginning of each task was marked at the end of the experimenter’s task prompt (e.g., “Please find Sunset Boulevard”). As for the end of the task there were a few possibilities: the participant selected the desired street name, the participant declared the current task to be a failure, or the participant issued a new “Enter destination” command to start a new search.

Our results show a significant main effect of the *interaction-technique* on *task time* ($F_{2,46} = 17.250$, $p < 0.001$). The mean task times for audio-only, HDD, and HUD conditions were 22.9s, 17.1s, and 16.6s, respectively. A post-hoc comparison indicated significant differences between audio-only and HDD ($p < 0.001$) and audio-only and HUD ($p < 0.001$). No difference was observed between HUD and HDD ($p = 0.547$). Even though the HUD required the shortest task time among the three variants, the lack of statistical significance in the HUD/HDD pairing indicates that hypothesis C is only partially supported for this metric.

3.7.3 Subjective Preference Rankings and Opinions

After concluding the study, each participant had a chance to express her subjective opinion about each of the *interaction-techniques* through a Likert scale questionnaire (1: fully disagree - FD, 2: disagree - D, 3: neutral, 4: agree - A, 5: fully agree - FA). Table 2 shows the number of participants and their agreement with statements about ease-of-use and driving distraction for each interface. The statistics for the main effect of the *interaction-technique* are obtained using the Friedman non-parametric test. We can see that a significant main effect of the *interaction-technique* exists for both statements.

Post-hoc comparisons were conducted for both questions using the Wilcoxon Signed Rank test and Bonferroni correction for multiple comparisons. The results revealed the following significant differences: for ease-of use – HDD vs. HUD ($p = 0.003$) and for distraction from driving – audio-only vs. HUD ($p = 0.006$), audio-only vs. HDD ($p < 0.001$), and HDD vs. HUD ($p < 0.001$).

Table 2. Subjective opinions about *interaction-techniques*. White cells show number of subjects who fully agreed (FA) or agreed (A) with each statement. Shaded cells show number of subjects who fully disagreed (FD) or disagreed (D) with each statement

Statement	Agree -ment	Audio -only	HDD	HUD	p (χ^2)
The ___ interface was easy to use.	FA/A	19	15	23	0.003 (11.4)
	FD/D	4	4	1	
The ___ interface distracted me from driving.	FA/A	3	19	8	0.001 (27.9)
	FD/D	17	2	7	

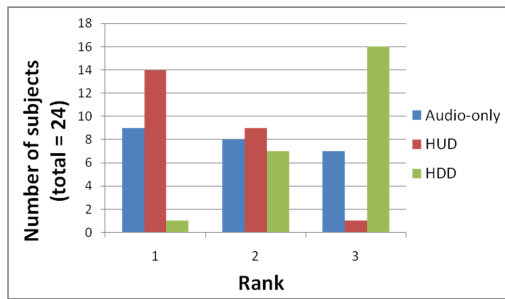


Figure 3. Overall preference ranking of the interaction-techniques.

0.001). This supports hypothesis A above, since for ease of use HUD obtained the highest score (although not significantly so compared to audio-only). As for the users' opinions about their distraction from driving, audio-only fared the best (contrary to hypothesis A), followed by HUD and HDD.

Figure 3 shows how subjects ranked the interfaces with respect to the desirability of using them in their own vehicles (1: most desirable, 3: least desirable). The results show a significant main effect of the *interaction-technique* on the subjective ranking ($\chi^2=16.583$, $p < 0.001$). A post-hoc comparison revealed significant differences for audio-only vs. HDD ($p = 0.033$) and HDD vs. HUD ($p < 0.001$). As we can see, HUD was ranked highest by the majority of subjects, although not significantly so compared to audio-only. This indicates that hypothesis A is partially supported for this subjective ranking measure as well.

3.7.4 Subjective Mental Load Evaluation

Immediately after completing each of the sessions, subjects filled out a NASA-TLX questionnaire in order to assess their workload while interacting with each of the interfaces. The mean reported workload was 40.6, 29.6, 49.3, and 42 for audio-only, control, HDD and HUD sessions, respectively (Figure 4). We found a significant main effect of the *interaction-technique* on mental load ($F_{3,69} = 16.484$, $p < 0.001$). A post-hoc comparison revealed highly significant differences ($p < 0.008$) for all pairings except audio vs. HUD ($p = 0.699$), so hypothesis C is partially supported.

3.8 Driving Performance Results

In this section we present the results obtained through the objective driving performance metrics. To put these results in perspective, we will later discuss how well they correspond to the participants' subjective opinions presented in the previous section.

3.8.1 Lateral Position

Our results show a significant main effect of the *interaction-technique* on the area between the curves for lateral position ($F_{2,46} = 3.705$, $p = 0.032$). The mean values for each *interaction-technique* are shown in Figure 5 (left). A post-hoc comparison

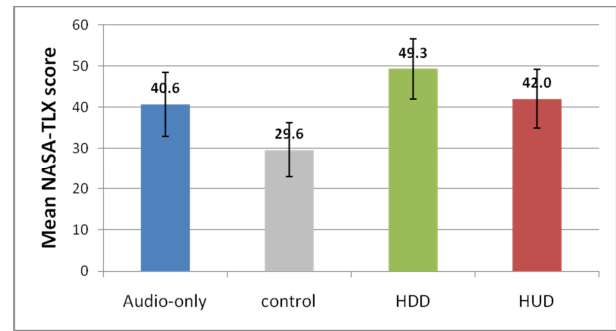


Figure 4. Mean workload values for each interaction-technique obtained through NASA-TLX. Higher values indicate higher mental load. Note: Error bars represent 95% confidence intervals in all figures.

revealed significant differences for audio-only vs. HDD ($p = 0.023$), and audio-only vs. HUD ($p = 0.019$). No difference was observed for the HUD vs. HDD pairing ($p = 0.884$).

3.8.2 Speed

We found a significant main effect of the *interaction-technique* on the area between the curves for speed ($F_{2,46} = 6.580$, $p = 0.003$). The means for all three *interaction-techniques* are depicted in Figure 5 (center). After performing post-hoc comparisons we discovered significant differences for audio-only vs. HDD ($p = 0.026$) and audio-only vs. HUD ($p = 0.001$). No difference was observed for HUD vs. HDD ($p = 0.237$).

3.8.3 Steering

A significant main effect of *interaction-technique* on area between the curves for steering was found ($F_{2,46} = 3.686$, $p = 0.033$). The mean values for all three *interaction-techniques* are shown in Figure 5 (right). A post-hoc comparison revealed a significant difference for audio-only vs. HDD ($p = 0.006$). No significant differences were observed for either audio-only vs. HUD ($p = 0.322$) or HUD vs. HDD ($p = 0.136$).

3.8.4 Throttle and Following Distance

Our analysis did not show any significant differences between our three *interaction-techniques* in terms of throttle depression and following distance.

In general, the driving performance results are contrary to our hypothesis (D). Lateral position, steering, and speed results are all consistent in that drivers performed better with the audio-only variant than with either of the other two variants (with the exception of steering, where there was not a significant difference between audio-only and HUD).

3.9 Visual Attention Results

Using our eye-tracking system we were able to calculate the

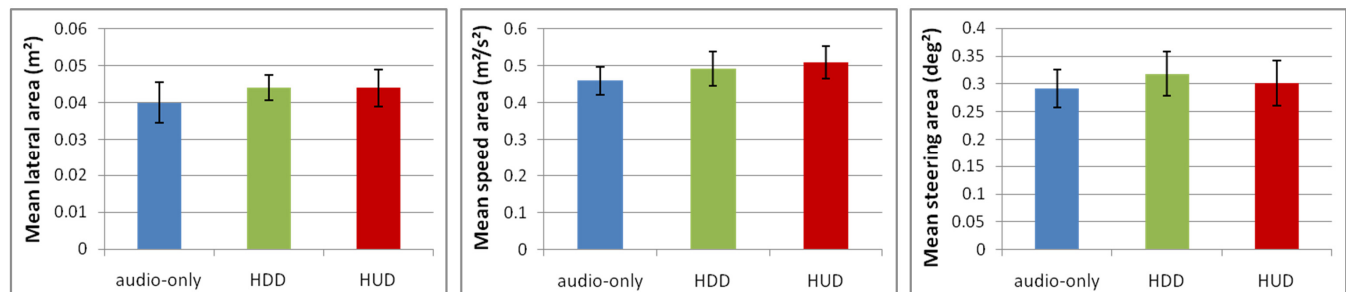


Figure 5. Mean area between the lateral position, speed and steering curves for all interaction-techniques.

percentage of time that the participants looked away from the forward roadway for each of the *interaction-techniques*. With respect to this metric we found a significant difference between the three *interaction-techniques* ($F_{2,46} = 26.033$, $p < 0.001$). On average, participants looked away from the road 3.01% of time for audio-only, 11.24% of time for HDD, and 2.81% of time for HUD. Note that these means correspond to looking away from the road while actually interacting with the three interfaces on test; downtimes between the tasks are not included. In agreement with hypothesis B, a post-hoc comparison showed significant differences for audio-only vs. HDD ($p < 0.001$) and HUD vs. HDD ($p < 0.001$).

Figure 6 shows the average number of glances away from the forward roadway that subjects made when using each *interaction-technique*. The glances are grouped by their duration into short (less than 0.5s), medium (between 0.5 and 2s), and long (longer than 2s). The following significant differences were observed using the post-hoc comparisons: for medium glances – audio vs. HDD ($p < 0.001$) and HUD vs. HDD ($p < 0.001$), and for short glances – audio vs. HDD ($p = 0.011$) and HUD vs. HDD ($p = 0.006$). This again supports hypothesis B. In general there were very few long glances off road, so no significant differences were observed there.

3.10 Discussion and Conclusions

Given the design of the tested interfaces, it is not surprising that there would be a twofold increase in the number of off-road glances when using the HDD versus the HUD. What might be more surprising is that the HUD and audio-only variants required as many off-road glances as they did. Bear in mind, though, that all three interfaces required interaction with the steering wheel-mounted controller in order to activate the speech recognizer, move the selection bar among the pick-list items, and to select an item. Given the relatively short duration of the testing session, subjects did not have time to gain the motor memory that would allow them to operate this controller confidently without occasionally looking down at the steering wheel.

The results show that HUD was the most preferred by participants, although not significantly so in comparison to audio-only. Even though the audio-only condition resulted in less driver distraction in comparison to HUD, the fact that audio-only tasks took so much longer might be the reason that it did not achieve the highest subjective rank. It is very interesting to see in Figure 3 that similar numbers of subjects ranked the audio-only interface 1st, 2nd, and 3rd, which means that the opinions about its effectiveness vary widely. On the other hand, trend lines for the HUD and HDD rankings have very steep slopes, which show that in general participants liked the HUD and disliked the HDD interface.

If we compare the user preference results to the driving performance measures, we notice interesting differences between the audio-only and HUD system variants. Participants considered HUD to be the most desirable interface (Table 2). This preference seems to be founded upon this interface’s reported ease-of-use and its relatively low required task time. However HUD does not win out in terms of the objective driving performance metrics or in subjects’ opinions about the distraction from driving. These show the audio-only condition to be the most favorable.

One might speculate that subjects preferred the HUD in spite of its greater impact on driving because of its novelty, or because its visual modality afforded a working memory “crutch” that was missing in the purely auditory variant.

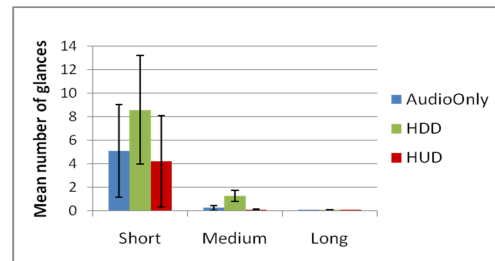


Figure 6. Mean number of glances away from the forward roadway for all *interaction-techniques*.

The head-down variant was the least favorable interface in the subjective categories and in all objective categories except for lateral deviation, speed, and steering, where it did not differ significantly from head-up. Even though tasks took far shorter to execute than with the audio-only interface, the need to look away from the road was seen as a great disadvantage—and not only on the user preference survey carried out at the end of the entire study. This trend is also clear in the NASA-TLX workload data gathered after each interaction session.

Based on the presented results we can conclude that displaying pick-lists using either an auditory display or a head-up display is a much better approach than using a head-down display, even though HDDs similar to the one we tested are ubiquitous in today’s factory-installed navigation systems. If we consider driving performance alone, audio-only is the preferred approach. However, if we broaden our focus to consider user satisfaction and task efficiency as well, HUD appears to be the best alternative.

One interesting finding is that even though each interface with a visual output modality (HUD and HDD) offered audible (TTS) output as well, subjects still made extensive use of the visual part. In other words, they preferred looking at the GUI even though the equivalent information was delivered audibly in addition. A very similar phenomenon was observed in [12], where the authors compared a standard map-based personal navigation device with the one which provided only voice navigation. Their results showed that even though the audio-only navigation device made for better driving performance and visual attention, participants still preferred the audiovisual interface. We believe that the reason for such findings is that people can assimilate a much larger amount of data through the visual channel than through the auditory channel given the same amount of time. Subjects took the opportunity to combine the two channels for better perceptual throughput, even at the expense of attention to the driving task.

4. FUTURE WORK

Is there a “middle ground” that offers the attention and safety benefits of an audio-only approach but allows tasks to be completed as efficiently as they can with a multimodal HUD?

A potential way forward might be to augment the audio-only interface with a “sometimes-on” HUD. The HUD would only be displayed when a content-retrieval task is active, and would slide out of view after a short timeout. Alternatively, one could display the HUD pick-list only in situations when the recognition quality is poor (as denoted for example by low confidence scores from the ASR engine).

Since our initial HUD variant offered an efficient interface (in terms of task completion time) and introduced a similar cognitive load to the audio-only condition, it would be valuable to investigate how we might improve upon the HUD design in order

to further reduce the system's impact on driving and mental load. In this study the HUD's screen design was identical to that of the HDD—merely relocating the display to within the driver's line of sight offered very encouraging results. We wonder whether we could do even better by adapting the HUD's layout, fonts, or graphical/iconic elements to better address the specific needs and constraints of the head-up use case (e.g., partial transparency, non-interfering colors).

There are many other factors involved in the design of HUDs that can affect drivers' mental workload. These include the size of the area into which the HUD is projected and the degree to which the HUD occludes the driving scene. In our experiment the HUD was simulated such that the pick-list text was equally visible across road and lighting conditions. However, in reality, differing ambient light levels might affect the legibility of the projected information, causing an impact on mental load and content comprehension. Perhaps this effect can be simulated in a laboratory setting by adjusting the transparency level of the HUD background or by partially blurring the text.

Another area that we plan on investigating in relation to HUDs is the potential for cognitive capture (inattention blindness) and perceptual tunneling [17]. As described in [17], cognitive capture can be thought of as being "lost in thought;" a good example is a highly emotional conversation on the phone where one loses a sense of one's immediate physical surroundings. We believe that there is a low risk of cognitive capture in the display of pick-lists on HUDs. However, we can envision that perceptual tunneling may indeed occur, for example if graphics or effects that draw too much visual attention are employed. Given the negative impact that perceptual tunneling has on driving, it is important to keep such considerations in mind when designing future in-car HUDs. As suggested in [16], a detailed analysis of drivers' eye behavior and reaction times to external events may highlight such issues.

5. REFERENCES

1. Ablassmeier, M., Poitschke, T., Wallhoff, F., Bengler, K., and Rigoll, G. Eye Gaze Studies Comparing Head-Up and Head-Down Displays in Vehicles. In *Proc. of ICME 2007* (China, 2007). IEEE, 2250-2252.
2. Bach, K., Jæger, M., Skov, M. B., and Thomassen, N. You can touch, but you can't look: interacting with in-vehicle systems. In *Proc. of CHI 2008* (Italy, 2008). ACM, 1139-1148.
3. Burnett, G. E. A Road-Based Evaluation of a Head-Up Display for Presenting Navigation Information. In *Proc. of the Tenth International Conference on Human-Computer Interaction* (2003). Lawrence Erlbaum Associates, 180-184.
4. Charissis, V., Papanastasiou, S., and Vlachos, G. Comparative Study of Prototype Automotive HUD vs. HDD: Collision Avoidance Simulation and Results. *SAE World Congress & Exhibition* (USA, 2008).
5. Charissis, V., Papanastasiou, S., and Vlachos, G. Interface Development for Early Notification Warning System: Full Windshield Head-Up Display Case Study. In *Proc. of the 13th International Conference on Human-Computer Interaction. Part IV: Interacting in Various Application Domains*, Julie A. Jacko (Ed.) (USA, 2009). Springer-Verlag, 683-692.
6. Chittaro, L. and De Marco, L. Driver Distraction Caused by Mobile Devices: Studying and Reducing Safety Risks. In *Proc. of the 1st Int'l Workshop Mobile Technologies and Health: Benefits and Risks* (2004). 1-19.
7. Harbluk, J.L., Mitroiu, J.S., and Burns, P.C. Three Navigation Systems With Three Tasks: Using the Lane-Change Test (LCT) to Assess Distraction Demand. In *Proc. of Driving Assessment 2009* (USA, 2009). 24-30.
8. Hart, S. G. and Staveland, L. E. 1988. Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. *Human Mental Workload*, 139-183.
9. Head-up display. http://en.wikipedia.org/wiki/Head_up_display.
10. Horrey, W.J. and Wickens, C.D. Examining the Impact of Cell Phone Conversations on Driving Using Meta-Analytic Techniques. *Human Factors* 48, 1, 196-205.
11. Image Space Incorporated. rFactor (2007). www.rfactor.net.
12. Kun, A. L., Paek, T., Medenica, Ž., Memarović, N., and Palinko, O. Glancing at personal navigation devices can affect driving: experimental results and design implications. In *Proc. of AutomotiveUI '09* (Germany, 2009). ACM, 129-136.
13. Liu, Y.C. and Wen, M.H. 2004. Comparison of head-up display (HUD) vs. head-down display (HDD): driving performance of commercial vehicle operators in Taiwan. *Int. J. Human-Computer Studies* 61, 679-697.
14. Narzt, W., Pomberger, G., Ferscha, A., Kolb, D., Müller, R., Wiegardt, J., Hörtner, H., and Lindinger, C. A New Visualization Concept for Navigation Systems. In: *Stary, C., Stephanidis, C. (eds.) UI4ALL LNCS 3196*, Springer (2004), 440-451.
15. Sojourner, R. J. and Antin, J. F. 1990. The effects of a simulated head-up display speedometer on perceptual task performance. *Human Factors* 32, 3, 329-339.
16. Tönnis, M. Time-Critical Supportive Augmented Reality – Issues on Cognitive Capture and Perceptual Tunneling. In *Proc. of ISMAR 2006* (USA, 2006). IEEE.
17. Tönnis, M., Broy, V., and Klinker, G. A Survey of Challenges Related to the Design of 3D User Interfaces for Car Drivers. In *Proc. of 3DUI 2006* (USA, 2006). IEEE.
18. Tsimhoni, O., Green, P., and Watanabe, H. Detecting and Reading Text on HUDs: Effects of Driving Workload and Message Location. In *Proc. of ITSA 2001* (2001).
19. Weihrauch, M., Meloeny, G.G., and Goesch, T.C. The first head-up display introduced by General Motors. *SAE Paper* (1989).
20. Weinberg, G. and Kondili, D. Display Style Considerations for In-Car Multimodal Music Search. In *Proc. of IHCI 2008* (The Netherlands, 2008).
21. Weinberg, G. and Harsham, B. Developing a low-cost driving simulator for the evaluation of in-vehicle technologies. In *Proc. of AutomotiveUI '09* (Germany, 2009). ACM, 51-54.
22. Weinberg, G., Harsham, B., Forlines, C., and Medenica, Z. Contextual push-to-talk: shortening voice dialogs to improve driving performance. In *Proc. of MobileHCI '10* (Portugal, 2010). ACM, 113-122.
23. Weinberg, G., Harsham, B., and Medenica, Z. Investigating HUDs for the Presentation of Choice Lists in Car Navigation Systems. In *Proc. of Driving Assessment 2011* (USA, 2011). 195-202.