CALLING WHILE DRIVING: AN INITIAL EXPERIMENT WITH HOLOLENS

Andrew L. Kun[1], Hidde van der Meulen[1,2], and Christian P. Janssen[2]
[1] University of New Hampshire, Durham, NH, USA
[2] Utrecht University, Utrecht, The Netherlands
Email: andrew.kun@unh.edu; hiddevandermeulen@gmail.com; C.P.Janssen@uu.nl

Summary: We investigate the visual distraction of drivers when they use an augmented reality (AR) device (HoloLens) for video calling while driving. The work is motivated by the advent of novel AR technology and by research on context sharing between callers. Both suggest that AR might soon be appropriated for 2-way video calling in cars, yet little is known on how distracting this is to the driver. Our participants drove in a simulator while engaged in a Skype conversation. We compared a condition with a video presentation (through AR), and a speech-only condition. We found that participants hardly looked at the video, perhaps because it was not visible from peripheral vision without making a head movement. In this way, HoloLens was less distracting visually than a monitor display used in earlier work. Although less distraction is desirable, using HoloLens also has a drawback: when drivers did look at the video they had to turn their head away from the road to look to the right, and down. The work makes suggestions on how to further study the safety and other issues of this new technology.

INTRODUCTION

Despite expert warnings, many drivers throughout the world use cell phones while driving. Studies also clearly indicate that this type of behavior has a negative effect on drivers' ability to drive safely (Dingus et al. 2016; Klauer et al. 2014). And while video calling using cell phones remains a potential distraction for today's drivers, in this paper we turn our attention to a novel technology: augmented reality (AR) displays. AR displays project images into the user's visual scene in such a way that those images appear to be part of the natural scene.

AR devices, such as HoloLens (Figure 1), have the potential to reduce driver distractions by presenting visual information close to the driver's visual focus, while also allowing the driver to continue to view the driving environment. However, HoloLens is a powerful computer and we

Figure 1. Participants operated a simulated vehicle and wore a HoloLens augmented reality (AR) device (left). HoloLens projected an AR Skype window for communication with a remote conversant (right).

can expect drivers to use it as such, even engaging in video calls. It is not known how distracting this is. Therefore, in this paper we assess the effects of a video call (VC) through HoloLens on the visual attention of the driver, and contrast this to the case of a speech-only (SO) call. We conducted a study in which participants controlled a simulated vehicle and at the same time engaged in a secondary task using the HoloLens device. Based on prior work with video calling while driving (Kun and Medenica 2012), our hypothesis is that on straight roads drivers' visual attention to the road ahead will be reduced when they can see the remote conversant compared to the case when they can only hear them.

METHODS

Tasks. Participants engaged in two tasks in parallel: the driving task, and a spoken task. The driving task entailed driving at 50 MPH on a two lane straight rural road and following a yellow passenger car. Apart from the lead vehicle there was no other traffic.

Participants also engaged in a spoken task: they played a series of games of Taboo with a remote conversant. Taboo is a game for two players. One player is given a target word, and attempts to make the other player utter that word. However, the player must do this without saying the target word, or five so-called taboo words. In our experiment the remote conversant was given the target word, and the driver was guessing it. For each participant the same experimenter acted as the remote conversant, to ensure that interaction with the remote conversant was relatively constant across participants. The participant and the experimenter communicated via Skype. The participant wore a HoloLens device running Skype, while the experimenter ran Skype on a laptop in another room.

Design. We conducted a one-factor within-subjects experiment in which we compare two conditions. In the speech-only (SO) condition the driver and the experimenter could hear each other, but not see each other. In the video call (VC) condition the driver could see the experimenter, and the experimenter could see the video from the front-facing camera of the driver's HoloLens. We counterbalanced the presentation order of the two conditions. The presentation order of Taboo cards was the same for each participant.

Equipment. We conducted the experiment using a high fidelity DriveSafety driving simulator offering a 180° field of view (Figure 1). The cab is surrounded by three projector displays and it is placed on a moving base to allow participants to feel bumps, acceleration and deceleration.

While operating the simulator, participants wore a HoloLens device (Figure 1). HoloLens projects visual information, such as simulated 3D objects or application windows, within a field of view that is about 40° wide by 20° high. HoloLens can pin objects and windows to specific locations in the physical world. Additionally, HoloLens supports directional sound which gives users the impression that the sound is coming from a pinned window. In the VC condition participants could see the experimenter in the Skype window (Figure 1 right). In the SO condition they only saw the Skype logo in the window. We tracked participant gazes during the experiment using a Pupil Labs eye tracker that fits underneath the HoloLens.

Participants. Fourteen student participants took part in the experiment; they all received course credit for participation. We discarded data from three participants who lacked the language skills to complete the Taboo task, and from one participant due to technical reasons. We analyzed data from 10 participants (8 male), between the ages of 19 and 23.

Procedure. After participants signed a consent form, we explained to them the experiment procedure, and showed them the Pupil Labs eye tracker and HoloLens. Participants read a short introduction to the game of Taboo, and practiced playing with the experimenter. Once they were confident in playing the game they were seated in the DriveSafety driving simulator and asked to drive a few minutes to become comfortable with the simulator.

Next we asked participants to wear the head mounted eye tracker and then the HoloLens. Once both devices fit comfortably, participants were asked not to touch the devices. We then calibrated the eye tracker, and then powered on the HoloLens. Participants were asked to start the Skype application in HoloLens and we then initiated a Skype call from the laptop. We instructed participants to pin the Skype window to the top of the center console, just underneath the windscreen. After this setup, the experimenter moved to another room, such that the experimenter and participant could only communicate through Skype. Next, participants completed the two experimental conditions (SO and VC). In each condition we started with a practice session, and then proceeded with data collection. In the practice sessions participants operated the simulated vehicle and played 8 Taboo cards with the experimenter. During the subsequent data collection, the participants played 20 Taboo cards with the experimenter. Thus, participants played $2 \times (8 + 20) = 56$ cards. For each practice and data collection session participants started a new simulated drive. We started the Taboo game 20 seconds after the start of a drive, to allow participants to settle into the driving task. If participants were unable to guess the word within 60 seconds the card was skipped. We kept track of the number of skipped cards.

After participants completed both experimental conditions, we asked them to complete a digital questionnaire using Limesurvey. In addition to demographic information we asked for their views on driving with the HoloLens and the two experimental conditions (SO and VC). The total experiment lasted approximately 50 minutes.

Measures. All measures were obtained for each participant and interaction type (SO and VC), and then averaged over all participants. We collected the following measures:

- Percent dwell time (PDT) on the road ahead (i.e., percent of time drivers spent looking at the forward road). Decreased PDT on the road indicates reduced visual attention.
- Standard deviation of lane position (SDLP), as defined in SAE J2944. Increased SDLP can indicate worse driving performance.
- Number of missed cards in Taboo. Missed cards indicate poor performance in Taboo.
- Levels of agreement with preferential statements on a 5-point Likert scale.

We calculated PDT and SDLP over 3 minute-long segments that started 20 seconds after the beginning of an experiment. We did this regardless of how long it took to complete the 20 Taboo cards for an experiment. Eye tracker data was collected at 30 Hz, while driving simulator data was collected at 10 Hz.

RESULTS

Visual attention, driving, and game performance

For technical reasons we had to exclude eye tracking data for three participants. For one participant the HoloLens covered the world camera, thus we have no way to establish where this participant directed his gaze during the experiments. For two other participants gaze tracking was poor for most of the experiment. We compared the PDT values for the remaining 7 participants using a paired t-test. In contrast to findings by Kun and Medenica (2012) we did not observe a significant difference between the speech-only (SO) (*M*=95.7%, *SD*=3.2%) and video call (VC) (*M*=96.8%, *SD*=2.8%) conditions (t(6) = -1.525, *p*=.178). These high PDT values are in line with those observed by Kun and Medenica (2012). Also in agreement with Kun and Medenica (2012), a paired t-test for all 10 participants did not reveal any differences in SDLP for the SO (*M*=0.21 m, *SD*=0.09 m) and VC (*M*=0.22 m, *SD*=0.12 m) conditions (t(9) = -.879, *p*=.402).

Participants successfully guessed most of the 20 taboo cards. The number of words they could not guess was low for both the SO (*M*=1.5, *SD*=1.4), and the VC (*M*=1.8, *SD*=1.8) condition.

Preferential statements

To assess participants' attitudes toward using the SO and VC modes of interaction in real driving, we asked participants to provide their agreement with two statements: "I would engage in a [speech-only/video call] phone conversation in my own car." While 70% of participants indicated they would engage in a SO conversation, the same percentage (70%) indicated they would not engage in a VC conversation in their own vehicles. We performed a Wilcoxon Signed Rank test with respect to the type of interface, and found that participants' attitudes were different toward engaging in SO and VC conversations in their own vehicles (*p*=0.010). We also found that 80 % of participants chose VC in response to the following question: "Which phone conversation distracted you more from driving?"

DISCUSSION

In this experiment we found no evidence that participants' visual attention to the road was negatively affected by engagement in a spoken task with a remote conversant, even when the video of the remote caller was shared. This result is in contrast to findings in the work of Kun and Medenica (2012), where (on straight roads) in the video call condition participants spent more time looking away from the road than in the speech-only condition.

The reason for the difference in the findings is likely to be in the difference in the visibility of the display when the participant is looking straight ahead. When the display is a physical display, participants can observe it with their peripheral vision, and they can bring the display into their focal vision with a slight turn of their head to the right combined with an additional rotation of the eyes to the right and down. But this is not the case with HoloLens. HoloLens has a small field of view: approximately 40° horizontally, and 20° vertically. We placed the Skype window such that it was outside of this field of view when the participants were focused on the road ahead. Such a placement was necessary in order to avoid blocking any part of the road by the

Skype window. In this sense, a HoloLens display that is placed at this location is perhaps less distracting than an in-car display as it does not compete for attention from the periphery.

Nonetheless, the HoloLens display can be distracting in its own right, as with HoloLens users can only turn toward a displayed item by rotating their head, and they cannot combine a head rotation with an eye rotation to bring an item into focal vision. This means that to see the video call (VC) display, our participants had to turn their heads to the right and down. Visual attention data indicates that our participants rarely engaged in such head rotation. We further confirmed this finding by transcribing the videos from the two excluded participants for whom the tracking was poor. In the videos we marked instances where there was head motion to the right and down, assuming that such head motion would occur whenever the participants looked at the Skype window in HoloLens. We found that one of the participants made no head motions indicating gazes at the Skype window. The other participant made only three such head motions, all in the VC condition, with each head motion taking up about 0.5 seconds. This indicates that the visual behavior of these two participants did not differ from that of the seven participants for whom we were able to calculate the PDT at the road ahead.

AR devices might improve the safety of talking to a remote conversant while driving, because an AR device could make talking to a remote conversant more like talking to a passenger, and less like talking on the phone. Research indicates that in many instances talking to a passenger is much less distracting than talking on the phone (Charlton 2009). For example, an AR device could render a remote conversant as a life-like hologram sitting in the passenger seat (similarly to (Pejsa et al. 2016)). Alternatively, the AR device could project an avatar in the passenger seat. And just like HoloLens, the AR device could make the sound directional, so that the speech of the remote conversant would appear to emanate from the hologram in the passenger seat.

Furthermore, future work could also look at two-way sharing of information. That is, just like HoloLens, the AR device could provide the remote conversant a video feed of the driver's view of the world, which might help the conversants negotiate dialogue turns, taking into account the workload associated with driving. Thus, when the remote conversant sees that the driver is engaged in a complicated driving maneuver, they might stop talking, or switch the topic of conversation to traffic, in order to reduce the driver's overall workload.

LIMITATIONS

While our experiment produced encouraging results, there are several limitations that must be taken into account. One important limitation is that our participants might not have felt the need to look at the display, because they were able to complete the task without doing so. Furthermore, it is possible that not all of our participants found Taboo to be an engaging task. It is possible that a different task might lead to more gazes towards the remote conversant. However, results from the experiment of Kun and Medenica (2012) indicate that on straight roads participants engaged in a Taboo game will look at a physical display showing the video of the remote conversant. We should also point out that each participant engaged in the Taboo game with the same experimenter, and this experimenter was a stranger to them. When talking to someone they know participants might be more inclined to look at the HoloLens display (and away from the road) because they do not feel that they are being observed or evaluated. Also, the

experimenter was aware of the experimental conditions (SO or VC). Thus, it is possible that the way the experimenter talked to the participants influenced their visual behavior.

Another limitation is the narrow field of view of HoloLens. If our device had a wider field of view perhaps our participants would have looked at the remote conversant more often, more in line with the results of the experiment of Kun and Medenica (2012). In fact, the narrow field of view might also reduce the potential positive impact of our proposed idea of projecting a hologram onto the passenger seat: since drivers cannot see the hologram without turning their head, they might not perceive the hologram as being present in the vehicle. However, we might be able to alleviate this problem by modifying HoloLens to present visual information in the user's visual periphery using an array of LEDs, as in the approach of Xiao and Benko (2016).

Also, we asked participants to position the Skype window, and through Skype we visually confirmed that they were successful in this. However, we did not make adjustments to carefully match the Skype window placement between participants. Thus, it is possible that different participants had somewhat different visual experiences. Furthermore, this experiment was the first time for all of our participants to engage with HoloLens. It is an open question whether they would be more willing to make head motions while driving if they grew accustomed to making such head motions while using HoloLens in other settings, such as playing games, or talking on Skype in the workplace. Such longitudinal effects could be explored in a separate study.

Finally, we conducted our experiment with a relatively small number of participants (N=10), and due to technical issues our PDT averages are based on only seven of these. However, our high PDT results are very consistent over the participants: the standard deviation of the PDT calculated for the seven participants is quite low (3.2% and 2.8% for SO and VC, respectively). Furthermore, the transcription of video data from the eye tracker for two more participants showed almost no glances at the Skype window. These results taken together indicate that our participants indeed cast very few glances at the AR display.

CONCLUSION

We started this research by proposing the hypothesis that drivers, who use an AR device such as HoloLens to call a remote conversant, will look away from the road more when they can see the remote conversant compared to the case when they can only hear them. Our results do not support this hypothesis. A key reason for the lack of glances at the HoloLens display is likely the narrow field of view of the device. With such a narrow field of view drivers would have to move their head to the right and down in order to see the display. It is possible that our participants were simply not comfortable making such a head motion. Further work could assess if this is the case, and if participants would look away from the road more if they had prior exposure to an AR device, or if the remote conversant was someone familiar instead of an experimenter.

Looking ahead, AR devices could be explored for navigation applications, given that prior work has indicated that AR can improve visual attention to the road compared to other navigation devices (Medenica et al. 2011). AR could also be explored for applications by drivers in professional settings. For example, first responders need to interact with a number of in-vehicle

devices while driving, and we can expect that this trend will continue as connected vehicles become part of driving, and as we move towards automated vehicles (Kun et al. 2015).

Importantly, we feel that AR devices can play a significant role in automated vehicles, such that we can transform vehicles into places for productivity and play, and exploit new mobility options while preserving user privacy and data security (Kun et al. 2016; Riener et al. 2016). For example, AR devices might be useful in presenting motion cues that can reduce motion sickness in vehicles as passengers look away from the road and onto a source of information.

ACKNOWLEDGEMENT

REFERENCES

Charlton, S.G. (2009). Driving while conversing: Cell phones that distract and passengers who react. *Accident Analysis & Prevention*, 41: 160-73.

Dingus, T.A, Guo, F., Lee, S., Antin, J.F., Perez, M., Buchanan-King, M. & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data, *Proceedings of the National Academy of Sciences*: 201513271.

Klauer, S.G., Guo, F., Simons-Morton, B.G., Ouimet, M.C., Lee, S.E. & Dingus, T.A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers, *New England journal of medicine*, 370: 54-59.

Kun, A.L., Boll, S. & Schmidt, A. (2016). Shifting Gears: User Interfaces in the Age of Autonomous Driving, *IEEE Pervasive Computing*, 15: 32-38.

Kun, A.L., & Medenica, Z. (2012). Video call, or not, that is the question, *Proceedings of the 2012 ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts*, 1631-36. ACM.

Kun, A.L., Wachtel, J., Miller, W.T., Son, P. & Lavallière, M. (2015). User interfaces for first responder vehicles: views from practitioners, industry, and academia, *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 163-70. ACM.

Medenica, Z., Kun, A.L., Paek, T. & Palinko, O. (2011). Augmented reality vs. street views: a driving simulator study comparing two emerging navigation aids, *Proc. of the 13th Int. Conf. on Human Computer Interaction with Mobile Devices and Services*, 265-74. ACM.

Pejsa, T., Kantor, J., Benko, H., Ofek, E. & Wilson, A. (2016) Room2Room: Enabling Life-Size Telepresence in a Projected Augmented Reality Environment, *CSCW '16*. San Francisco, CA: ACM.

Riener, A., Boll, S. & Kun, A.L. (2016) Automotive User Interfaces in the Age of Automation (Dagstuhl Seminar 16262). *Dagstuhl Reports*. Vol. 6. No. 6.

Xiao, R. & Benko, H. (2016). Augmenting the Field-of-View of Head-Mounted Displays with Sparse Peripheral Displays, *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 1221-32. ACM.