Human Factors in Robots, Drones and Unmanned Systems, Vol. 57, 2022, 54–62
https://doi.org/10.54941/ahfe1002310

Facilitating Collaboration between Humans and Unmanned Heavy Vehicles Using Verbal Interaction and Augmented Reality

Johan Fagerlönn[1], Yanqing Zhang[2], Lina Orrell[1], and Hanna Rönntoft[1]

[1]RISE Research Institutes of Sweden AB, Kunskapsallén 14, 941 63, Piteå, Sweden
[2]Scania CV AB, Granparksvägen 10, 151 48, Södertälje, Sweden

ABSTRACT

The present explorative study examines the potential of using verbal interaction and augmented reality (AR) to facilitate collaborations between professional human operators and unmanned self-driving heavy vehicles. Concepts that support operators in loading situations were designed and evaluated with forklift operators and rock-loading operators during a video-based study. Overall, the concepts received high scores in perceived efficiency and user experience. The results from the forklift operators supported the idea that more natural and social verbal interaction between operators and unmanned vehicles could lead to increased trust and acceptance compared to using simple voice commands. However, the results from the rock-loading operators showed that extensive use of voice interaction could become disturbing. Taken together, the explorative study supports the potential of using and further exploring verbal interaction and AR to facilitate human operators' collaboration with self-driving vehicles, and the proposed concepts provide promising examples of interaction models for further investigation and implementation.

Keywords: Verbal interaction, Augmented reality, Self-driving vehicles, Autonomous vehicles, Artificial intelligence, Natural user interface, Collaboration

INTRODUCTION

There is currently a need to understand how user interfaces can be designed to facilitate collaboration between human operators and unmanned self-driving trucks in confined industrial contexts (hubs), such as logistics centers, terminals, and mines. An especially demanding situation is created when humans and machines work together in shared spaces to collaborate on common work tasks. Considering situations when truck drivers will no longer be available, the human–system interaction must provide high levels of work efficiency, safety, and user experience (UX). The In the Hub (RISE Research Institutes of Sweden, 2022) research project aims to examine and demonstrate new models of interaction between professional human operators and self-driving heavy vehicles (SAE automation levels 4–5 (Sae International, 2018)) in hubs. The project focuses especially on exploring the potential of utilizing natural user interfaces (NUI).

Previous research suggests that verbal communication may facilitate interactions between automated vehicles and humans (Kim et al. 2021; Large et al. 2019; Mahadevan et al. 2018). However, the work has mainly focused on users inside cars and pedestrians near vehicles. A possible advantage of verbal interaction is that the operator can interact with the system while keeping their hands and eyes free. Also, human-like features (anthropomorphism) may increase likeability and trust toward the system. For instance, giving the vehicle a gender, name, and voice has been shown to increase trust and likeability (Waytz et al. 2014). Furthermore, Ruijten et al. (2018) reported that conversational user interfaces resulted in higher trust compared to purely visual displays. According to Large et al. (2019), a more social and conversational interaction may result in higher pleasantness and trust compared to using verbal commands.

Visual information in work machines is often presented on conventional screens. But focusing on screens while manually operating a machine may pose a safety risk. The use of augmented reality (AR) has been examined to support human operators in industrial work contexts, such as mines (Jacobs et al. 2016). Research has indicated that technology can improve operators' productivity, confidence, and safety (Sitompul and Wallmyr, 2019). For instance, placing visualizations close to the line of sight may reduce operatorsâŁ™ workload compared to using ordinary visual displays (Wallmyr et al. 2019). By using AR, information can also be placed virtually in the work context, which means that the operator can easily perceive and interpret the information. AR also offers possibilities to "see through" obstacles (e.g., walls), which offers new possibilities for increasing situational awareness. However, an overall challenge with AR is that virtual content may obscure physical objects and humans in the physical world. Thus, using semi-transparent AR may be advantageous (Sarupuri et al. 2016). Previous research has explored combinations of verbal interaction and AR in interaction models. For example, the combination has been suggested for interaction with loader cranes (Majewski and Kacalak, 2016). However, little is known about how these emerging technologies can be used and combined to facilitate collaboration between humans and self-driving vehicles.

OBJECTIVES

The main aim of the explorative study was to investigate the possibility of using verbal interaction and AR to support the user experience (UX) and facilitate collaboration between professional human operators and self-driving heavy vehicles in hubs. The second aim was to gain better insights into the effects of more social and natural verbal human–machine interactions. The two main hypotheses of the study were: (1) Interaction models using verbal interaction and AR can facilitate high levels of work efficiency and UX when humans collaborate with self-driving vehicles. (2) More social verbal interaction can lead to higher levels of trust and acceptance compared to using short verbal commands.

The study focused on loading scenarios in two hub types: logistic centers and underground mines. A pre-study conducted within the In the Hub project showed that the loading of goods (e.g., rocks or pallets) is a situation that often requires complex interactions between truck drivers and loading operators in mines and logistic centers.

METHODS

To fulfill the aims of the study, concepts for interaction design were developed, implemented in 3D animated movies, and evaluated with end users.

Concept Design

The process of designing concepts involved a team of interaction designers, researchers, 3D animators, a forklift operator, and people from a mining company. The process resulted in two concept variants for each user context: natural voice interaction (NVI) and basic voice interaction (BVI). The two variants differed mainly in how the verbal interaction was implemented. The NVI variant aimed to mimic social verbal interaction between an operator and truck driver performing the loading task together. In the BVI variant, the operator uses short verbal commands to control the truck. AR was used in both variants to provide visual guidance and information to perform the loading (see Figure 1). The AR consisted of 2D and 3D elements. The 2D visualization showed information about remaining load capacity, remaining loading time (logistic center), and the vehicle and loading area. The lock icon shows if the truck can only be controlled by the operator in the forklift or loader. For the mine context, the 3D AR elements were used inform the operator about how the load was distributed in the truck. In the logistic center context, the 3D AR was used to guide the operator to where the next pallet should be placed.

The concepts were implemented in approximately three-minute-long animated movies, showing the loading being performed by the operator from a first-person perspective. More details on the design and differences between concept variants were presented by Fagerlönn et al. (2021). Videos showing the scenarios and the concepts are accessible online. For the mine, refer to (Self-driving vehicles, 2021a) for the NVI variant and (Self-driving vehicles, 2021b) for the BVI variant. For the logistic center, refer to (Self-driving vehicles, 2021c) for the NVI and for the BVI (Self-driving vehicles, 2021d).

Participants and Apparatus

Thirty-two users participated in the evaluation: 16 forklift operators and 16 rock-loading operators working in different companies in Sweden. The ages of the forklift operators ranged from 22 to 60 ($M = 36.8$, $SD = 11.0$). The ages of the rock-loading operators ranged from 21 to 60 ($M = 37.6$, $SD = 7.0$). The study was conducted using online interviews. The concept movies were experienced by watching them on the video-sharing platform Vimeo. The participants used their own computers or tablets and speakers or headphones to experience the concept videos.

Figure 1: AR display in the logistic center (left) and mine (right).

Procedure and Measurements

The subjects first provided demographic information and watched a short movie to allow calibration of the audio to a comfortable listening level. The subjects then watched the movies presenting the concept variants. The movies were presented in a counter-balanced order to reduce order effects. Ten items from the User Experience Questionnaire (UEQ) (Schrepp, 2015) and two items from the acceptance scale by Van der Laan et al. (1997) were used after each concept variant movie to obtain insights into the UX from different perspectives. The items consisted of seven-point rating scales between two opposite words. After watching both concept variants, the subjects answered statements about the perceived efficiency of the interaction technologies. Statements were answered using a scale from 1 (do not agree at all) to 7 (completely agree). To assess differences in the subjects' trust and acceptance of the different concepts, the subjects were asked to select in which concept variant they would trust the truck the most and which one they would prefer to work with. Open-ended questions were used throughout the test to gain more insights into the perceived UX of the concepts and interaction technologies.

RESULTS

Table 1 shows the results of the UX ratings. Considering that the study did not use all scales in the UEQ (Schrepp, 2015) or Van der Laan et al. (1997) methods, the results of all rating items are presented.

Efficiency of Concepts and Interaction Technologies

Among the forklift operators, both concept variants generated very high scores (above 6.0) for UEQ items related to efficiency. They also found both interaction technologies to be efficient. The statement "Communicating with the truck by talking feels efficient" generated a mean score of 5.9 ($SD = 1.7$), while the statement "Receiving information on the windscreen feels efficient" received a mean score of 6.1 ($SD = 1.4$). Analysis of the operator comments showed that the advantages in efficiency resulting from voice interaction were mainly related to the possibility of communicating/sending and receiving information without the need to use one's hands; otherwise, one might be interrupted by, for example, documentation tasks. Four operators pointed out that the interaction could be made even more efficient by using shorter

Table 1. Mean UX ratings for each context and concept variant. Scales ranged between 1 (e.g., inefficient) to 7 (e.g., efficient). Standard deviations are shown in parenthesis.

	Forklift operators		Rock-loading operators	
	NVI	BVI	NVI	BVI
UEQ - Efficiency				
Inefficient - Efficient	6.4 (1.1)	6.4 (0.9)	4.8 (1.4)	5.1 (1.5)
Impractical - Practical	6.1 (1.2)	6.1 (1.1)	5.7 (0.7)	5.5 (1.0)
Cluttered - Organized	6.3 (1.2)	6.3 (1.3)	5.3 (1.7)	5.3 (1.3)
UEQ - Perspicuity				
Complicated - Easy	5.6 (1.6)	5.6 (1.7)	5.9 (1.2)	5.5 (1.4)
Confusing - Clear	5.8 (1.6)	5.9 (1.2)	5.6 (1.5)	6.0 (1.0)
UEQ - Dependability				
Unpredictable - Predictable	5.8 (1.0)	6.0 (0.8)	5.2 (1.4)	5.4 (1.5)
Not secure - Secure	5.6 (1.3)	5.1 (1.4)	5.8 (1.3)	5.5 (1.5)
UEQ - Stimulation				
Boring - Exciting	5.9 (1.0)	5.3 (1.7)	5.4 (1.5)	5.5 (1.6)
Not interesting - Interesting	5.8 (1.3)	5.6 (1.4)	6.0 (1.0)	6.0 (1.0)
UEQ - Attractiveness				
Unpleasant - Pleasant	5.4 (1.4)	5.1 (1.5)	5.3 (1.6)	5.9 (1.1)
Van der Laan et al. items				
Useless - Useful	5.9 (1.2)	6.2 (1.2)	6.0 (0.7)	6.3 (0.7)
Irritating - Likeable	5.6 (1.3)	5.3 (1.4)	4.2 (1.6)	4.7 (1.5)

voice commands. For the AR, the identified advantages were mainly related to the possibility of easily accessing information when working, and that one does not need to focus visually on something else (e.g., a separate screen) when operating the machine.

Compared to the forklift operators, the rock-loading operators generated somewhat lower scores in items related to the efficiency of the concepts. However, the mean scores for the respective interaction technologies were quite high: 5.6 ($SD = 1.7$) for voice interaction and 5.8 ($SD = 1.7$) for AR. Analysis of operator comments related to efficiency indicated that the loading task took too long when using the concepts compared to how long it takes in their daily work. This indicates that even though the concepts and interaction technologies scored high in terms of efficiency, the concept should be better adapted to the mine context and workflow.

User Expereince

Both concept variants generated mean scores above 5.0 in the UEQ items related to perspicuity, dependability, stimulation, and attractiveness, indicating overall high levels of UX for the interaction. Also, the users provided high ratings for "usefulness." In terms of "irritation," the concepts resulted in somewhat lower scores for the rock-loading operators, especially in the NVI variant, with a mean score of 4.2. Comments regarding what was perceived as frustrating or stressful in the concepts revealed some insights that may explain this neutral mean score. A common theme in the comments was concern that the extensive use of verbal interaction would be "too much speech" and feel like nagging.

Effects of Verbal Interaction on Trust and Preference

Participants were asked, "In which concept would you feel that you can trust the vehicle the most?" Twelve forklift operators (75%) chose the NVI, while only two (13%) answered the BVI. Two other forklift operators did not report any difference. The reasons provided for selecting the NVI were mainly related to the fact that it felt more human-like, comforting, social, and safe. Among the rock-loading operators, five subjects (31%) chose the NVI, while six (38%) chose the BVI. Five rock-loading operators did not report any differences.

Participants were asked, "If you were given a chance to work with one of the concepts in the future, which would you prefer?" Ten (63%) of the forklift operators chose the NVI, while five (31%) chose the BVI. One forklift operator reported no difference. An analysis of comments related to their choice of preference, as well as how they experienced the voice interaction in the two concept variants, was conducted. The results showed that most operators preferred the voice interaction in the NVI, and they provided the following reasons: it felt more like a real dialogue and was human-like, social, natural, and a more pleasant experience. For the rock-loading operators, five subjects (31%) chose the NVI, while nine subjects (56%) selected the BVI. Two rock-loading operators reported no differences in preference. Analysis of operator comments revealed that many rock-loading operators seemed concerned about the extensive use of verbal interactions in the NVI, providing reasons such as that the BVI would be less disturbing, less irritating, involve less "unnecessary talking," be more "straight to the point," and offer a calmer and quieter work environment.

DISCUSSION

Overall, the evaluated concepts and interaction technologies generated high efficiency and UX scores in both user contexts. The forklift operators especially highlighted the advantages of these interaction technologies—that is, that voice interaction does not require them to use their hands and that AR makes it easy to access information while working. The results support the use of interaction models utilizing verbal interaction and AR to facilitate collaboration between human operators and vehicles. Moreover, the proposed concepts combining interaction technologies provide promising examples of interaction models for further investigation and implementation. Previous research has explored the use of new interaction models to facilitate interaction with self-driving vehicles, for instance, by using human–machine interfaces mounted on vehicles (Mahadevan et al. 2018; Habibovic et al. 2018; Faas and Baumann, 2019). The present work provides new insights and suggestions on how voice interaction and AR can be used and combined to facilitate interaction in complex collaborative tasks, e.g., in loading tasks.

Regarding the impact of social voice interaction on preference and trust, the two contexts revealed interesting results and differences. The UX ratings revealed no significant differences between concept variants in the two contexts. Still, more forklift operators preferred the concept designed to offer a more extensive social verbal dialogue between the operator and the machine, providing arguments such as that it felt more human-like, comforting, social, and pleasant. Also, the results showed that more extensive social interaction made most operators trust the vehicle more. These results indicate the potential for allowing more natural and social dialogues when developing future verbal interfaces for human–vehicle collaborations. These results are in line with previous research suggesting that more social and conversational interfaces may result in higher pleasantness and trust when compared to using verbal commands (Large et al. 2019).

The rock-loading operators reacted somewhat more negatively to the more extensive use of voice interaction, and more subjects preferred the concept using simple voice commands. The results indicate that extensive use of verbal interaction can be perceived as unnecessary, irritating, and a potential disturbance among these users. Overall, these results highlight the importance of being careful when designing and implementing verbal interactions. Considering that sound is omnidirectional and hard to ignore, it can become a distraction or annoyance. It also supports the importance of carefully adapting the design of verbal interactions to the needs and requirements of specific users and contexts to gain high acceptance and productivity. A possible reason for the rock loaders' preference for simple voice commands could also be that the working environment in mines can include much radio communication with colleagues. Therefore, they might prefer a less verbal alternative when communicating with the system.

The analysis of operator comments also revealed differences in individual opinions and concerns within hub types. For instance, while most forklift operators liked more social verbal interaction, some also pointed out that it could become a disturbance in various situations. A possible way to handle individual preferences and contextual requirements in the design of the system could be to allow users to select the type of verbal interaction to use in different situations. However, a more enticing approach would be to enable the AI system to automatically adapt its interaction style to the human operator's preferences, needs, work experience, and work situation (e.g., current workload). If the loading operator is performing repetitive tasks (which is common in, for example, mines), the system may need to adapt verbal interaction to avoid becoming irritating in the long run.

The results of this explorative study should be interpreted with caution. Experiencing solutions on video is not the same as experiencing and being able to interact with the solutions in a realistic context. In the next study, a refined concept will be implemented and evaluated in an interactive virtual reality (VR) environment. In future work, we hope to be able to implement and test solutions with users in real contexts over longer periods. Also, the solutions were only evaluated with Swedish users, and cultural differences may have impacted the results.

CONCLUSION

This explorative study investigated the potential for combining verbal interaction and AR to support UX and facilitate collaborations between professional human operators and self-driving heavy vehicles. The results support the idea that emerging interaction technologies have promising potential to facilitate high levels of work efficiency and UX in collaborative tasks. Results from forklift operators also support that successful implementation of social interaction could increase the operator's trust in the automated vehicle and acceptance of the interaction. However, rock-loading operators tended to prefer simple voice commands, which demonstrates the importance of adapting future implementations to the needs and desires of specific users in specific contexts. The results provide insights for use by practitioners and researchers with the intention of researching and implementing interaction design models for collaborations with self-driving vehicles.

ACKNOWLEDGEMENTS

The work was funded by the Swedish partnership program Strategic Vehicle Research and Innovation (FFI) (Grant no: 2019-05898). It was conducted in a collaboration between RISE Research Institutes of Sweden, Scania, Boliden Mineral, Icemakers and AFRY.

REFERENCES

Faas, S.M. and Baumann, M., 2019, November. Light-based external human machine interface: Color evaluation for self-driving vehicle and pedestrian interaction. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 63, No. 1, pp. 1232–1236). Sage CA: Los Angeles, CA: Sage Publications.

Fagerlönn, J., Sirkka, A., Orrell, L., Zhang, Y., Larsson, S., Tybring, E. and Rönntoft, H., 2021, September. Designing Collaboration between Human Beings and Self-driving Heavy Vehicles with Emerging Interaction Technologies. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 123–127).

Habibovic, A., Lundgren, V.M., Andersson, J., Klingegård, M., Lagström, T., Sirkka, A., Fagerlönn, J., Edgren, C., Fredriksson, R., Krupenia, S. and Saluäär, D., 2018. Communicating intent of automated vehicles to pedestrians. *Frontiers in psychology*, p.1336.

Jacobs, J., Webber-Youngman, R.C.W. and van Wyk, E., 2016. Potential Augmented Reality Applications in the Mining Industry. [unpublished manuscript] Researchgate.

Kim, K., Park, M. and Lim, Y.K., 2021, June. Guiding Preferred Driving Style Using Voice in Autonomous Vehicles: An On-Road Wizard-of-Oz Study. In *Designing Interactive Systems Conference 2021* (pp. 352–364).

Large, D.R., Harrington, K., Burnett, G., Luton, J., Thomas, P. and Bennett, P., 2019, September. To please in a pod: employing an anthropomorphic agent-interlocutor to enhance trust and user experience in an autonomous, self-driving vehicle. In *Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications* (pp. 49–59).

Mahadevan, K., Somanath, S. and Sharlin, E., 2018, April. Communicating awareness and intent in autonomous vehicle-pedestrian interaction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–12).

Majewski, M. and Kacalak, W., 2016. Conceptual design of innovative speech interfaces with augmented reality and interactive systems for controlling loader cranes. In *Artificial Intelligence Perspectives in Intelligent Systems* (pp. 237–247). Springer, Cham.

RISE Research Institutes of Sweden. 2022, RISE, viewed 14 February 2022, < https://www.ri.se/en/what-we-do/projects/in-the-hub >

Ruijten, P.A., Terken, J. and Chandramouli, S.N., 2018. Enhancing trust in autonomous vehicles through intelligent user interfaces that mimic human behavior. *Multimodal Technologies and Interaction*, *2*(4), p. 62.

Sae International, 2018. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. *SAE*.

Sarupuri, B., Lee, G.A. and Billinghurst, M., 2016, November. Using augmented reality to assist forklift operation. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction* (pp. 16–24).

Schrepp, M., User experience questionnaire handbook: All you need to know to apply the ueq successfully in your projects (2015). *URL http://www.ueq.online.org*.

Self-driving vehicles (2021a). *Natural Voice Interaction (NVI) mine*. [online video] Available at: https://vimeo.com/571617671 [accessed 1/2/2022]

Self-driving vehicles (2021b). *Basic Voice Interaction (BVI) mine*. [online video] Available at: https://vimeo.com/571616251 [accessed 1/2/2022]

Self-driving vehicles (2021c). *Natural Voice Interaction (NVI) logistic center*. [online video] Available at: https://vimeo.com/559904314/e3699a5e7d [accessed 1/2/2022]

Self-driving vehicles (2021d). *Basic Voice Interaction (BVI) logistic center*. [online video] Available at: https://vimeo.com/559904431/56646ca46e [accessed 1/2/2022]

Sitompul, T.A. and Wallmyr, M., 2019, November. Using augmented reality to improve productivity and safety for heavy machinery operators: State of the art. In *The 17th International Conference on Virtual-Reality Continuum and Its Applications in Industry* (pp. 1–9).

Van Der Laan, J.D., Heino, A. and De Waard, D., 1997. A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies*, *5*(1), pp. 1–10.

Wallmyr, M., Sitompul, T.A., Holstein, T. and Lindell, R., 2019, September. Evaluating mixed reality notifications to support excavator operator awareness. In *IFIP Conference on Human-Computer Interaction* (pp. 743–762). Springer, Cham.

Waytz, A., Heafner, J. and Epley, N., 2014. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, pp. 113–117.