Requirements for the Interaction With Highly Automated Construction Site Delivery Trucks

Mark Colley [1*†], Stefanos Mytilineos [1†], Marcel Walch [1†], Jan Gugenheimer [2†] and Enrico Rukzio [1†]

[1] Institute of Media Informatics, University of Ulm, Ulm, Germany, [2] Télécom ParisTech, Paris, France

Automated trucks for long-distance journeys seem within reach. With such automation, no human driver could be available. However, the last mile of the delivery is likely to involve humans. Therefore, either a human driver should still be present, or construction site workers must interact with the automated truck. While automated trucks capable of dealing with various construction sites could be feasible, the development could be costly and time-consuming. To define cooperative solutions for automated deliveries incorporating interaction between automated trucks and humans, a workshop with truck drivers ($N = 7$) was conducted. Based on this workshop, a model of the delivery process, including communication needs, is proposed. Requirements addressing the issues for highly automated delivery are derived from this process.

Keywords: external communication, qualitative, interview, automated vehicles (AV), delivery trucks

1. INTRODUCTION

Automated vehicles (AVs) are expected to have a significant impact on the trucking industry (Fagnant and Kockelman, 2015). Lower cost of delivery (Fagnant and Kockelman, 2015; Mersky and Samaras, 2016) *via* increased fuel economy and less need for truck drivers will probably lead to quick establishment of AVs. While (un-)loading will probably still need a human involved, long-distance journeys seem feasible (Fagnant and Kockelman, 2015). With increasing automation, human drivers will become ever more absent.

However, one major problem could be the last few meters toward an unloading spot on construction sites. Construction sites often change rapidly and extremely. Today, therefore, truck drivers, pedestrians, or construction site workers (CSWs) have to interact with each other. For example, CSWs instruct the driver *via* eye contact, gestures, voice, or moving along the way (Graham and Burch, 2006). As the human truck driver could be missing in the future, the highly automated truck (HAT) will have to substitute at least some of this communication. This includes communication of intent and receiving input from CSWs, thus, bidirectional communication. Litman believes that for unloading, humans will still be needed (Litman, 2017). While the Scania AXL (Scania, 2019) is equipped with a LED strip indicating awareness of objects around it; other forms of external communication are unexplored in this context.

To explore today's communication needs from which human-computer interaction (HCI) requirements for HATs can be derived from, we (1) gathered data on communication *via* observation on two construction sites in Ulm, Germany and Neu-Ulm, Germany and (2) conducted a workshop with truck drivers ($N = 7$). We explored their workflow and discussed potential operational modes for maneuvering HATs and the potential for enhanced safety *via* external communication concepts.

Communication, *organization*, and *size* were identified as relevant themes for construction site deliveries (CSDs). Additionally, the workshop revealed edge cases for communication with people unrelated to the construction site, such as standing halfway on the street due to a blocked entrance or having to deliver near the street. Future HATs could provide aid via external communication concepts indicating "specific instances that require additional signaling" (Moore et al., 2019, p. 304) as called for by Moore et al. Proposed interaction concepts for the induction of such HATs are discussed. Requirements include the necessity of bidirectional communication with a variety of people such as pedestrians or CSWs and as little preparation (equipment, task switching) as possible for the brief interaction between HATs and CSWs. These requirements enable practitioners and researchers to develop appropriate methods and processes for CSDs. They provide guidance to novel approaches and to better test CSDs of HATs.

Contribution Statement: This work contributes to the body of knowledge on successfully introducing automated trucks. Results from observation of two construction sites and a workshop with $N = 7$ truck drivers revealed the high need for communication of CSD today, leading to the derived requirements.

2. RELATED WORK

An overview of human-robot interaction (HRI) in the heavy machinery field is given. This research field increases safety and efficiency for CSWs. Additionally, external Human Machine Interfaces (eHMIs) could be used to enhance safety on construction sites. To define the issue, accident data and current measures against these accidents are presented.

2.1. Construction Site Accidents

One reason for construction site accidents is blind spots (Fan et al., 2019). Fosbroke (2004) stated various principles to reduce blind spot-related accidents such as limited access points, work zone layouts, buffer spaces, employment of spotters, and provision of signs. Spotters with direct communication to drivers are recommended as truck drivers need instructions on entering and exiting the work zone and how to navigate within the zone (Graham and Burch, 2006).

Several technologies were evaluated as proximity warning systems, including infrared, capacitive, ultrasonic, radar, RFID, GPS, video cameras, and magnetic sensors [The National Institute for Occupational Safety (NIOSH), 2011]. However, all have their disadvantages [The National Institute for Occupational Safety (NIOSH), 2011]. Another method to prevent such accidents is to employ proper administrative control (e.g., safety meetings, safety officers, implementation of regulations and guidelines, proper training) (Fan et al., 2019).

2.2. Heavy Machinery Interaction

HRI is concerned with the interaction between people and robots, of which HATs are a subset of. An overview of approaches to overcome system limitations is shown for voice commands and teleoperation. Teller et al. (2010) implemented a voice-commandable robotic forklift using a tablet. For interaction,

speech and stylus gestures are used. For the voice-commands, the tablet's microphone or one attached to the forklift is necessary. Commands and gestures are limited to the movement direction, to target a pallet and the desired drop-off location. The forklift uses lights, speakers, and a LED display to announce its state, imminent actions, and awareness of bystanders' presence. This multimodal interaction was also described by Correa et al. (2010). A common warning mechanism of today's forklifts is a warning and safety light (Eagle, 2019), used to indicate movement.

Teleoperation as in "operating a vehicle at a distance" (Fong and Thorpe, 2001, p. 9) has advantages such as avoiding undesirable risks (Valero et al., 2009). Vehicle teleoperation differs from remote control as it has specific characteristics such as no need for line-of-sight but requiring efficient motion command generation (Fong and Thorpe, 2001). Four categories of interfaces to control such vehicles are distinguished: *direct*, *multimodal/multisensor*, *supervisory control*, and *novel*. PdaDriver (Fong et al., 2003) uses a Personal Digital Assistant to control a vehicle in direct mode remotely. While they report high usability, especially by integrating multiple sensors, such an image-based system "sometimes fails to provide sufficient contextual cues for good situational awareness" (Fong et al., 2003, p. 5). Takayama et al. (2011) evaluated an assisted mobile remote presence system. It receives control from an operator via a web-based GUI but checks its vicinity for objects before executing the command. The system may override the command received if an object would be hit. Takayama et al. found fewer errors in assisted mode. However, the completion time was higher with assistance and that individual differences between humans have to be considered when developing such a system. A different approach called VisiCon (Hosoi et al., 2007) uses handheld projectors to guide a robot. This system was intuitively usable; however, it required twisting of the person's arms. Another related approach in the context of assistive technology is using a laser pointer to navigate a vehicle (Kemp et al., 2008). Ishii et al. (2009) enhanced this simple target designation via recognition of gestures hence enabling defining tasks such as collecting. Lasers were also shown to be useful for the definition of virtual borders for robots to operate (Sprute et al., 2019). However, in a comparison between direct physical interaction (i.e., pushing it around), person-following (i.e., following the participant), and pointing control in an indoor setting, pointing control performed and was rated significantly worse, direct physical interaction being the overall best system (Jevtić et al., 2015).

While these systems address unique issues for heavy machine interaction, the workflow of automated delivery of a HAT to a construction site combines these and adds complexity. Additionally, some of the assumptions made, such as continuously using devices to control a machine, lack realism as encounters between CSWs and HATs will be brief.

2.3. External Communication of Automated Vehicles

Today, there are still ongoing discussions about the necessity for external communication of AVs (Moore et al., 2019).

FIGURE 1 | (1) and (2) show the first, (3) and (4) the second investigated construction site.

However, various studies showed benefits of such concepts for people with visual impairments by providing information about the current environment (e.g., the AV's intention and other approaching vehicles) (Colley et al., 2020b) and trust toward AVs in crossing decisions (Löcken et al., 2019). Issues such as overtrust (Holländer et al., 2019) and scalability (Colley et al., 2020c) still have to be overcome.

Various external communication modalities have been evaluated. This includes displays (Florentine et al., 2016), LED strips (Florentine et al., 2016; Lundgren et al., 2017), projections (Ackermann et al., 2019; Nguyen et al., 2019) auditory (Colley et al., 2020b) or tactile cues (Mahadevan et al., 2018; Colley et al., 2020b). Concepts were grouped by used modality (Mahadevan et al., 2018; Colley et al., 2019, 2020b) or complexity (Löcken et al., 2019). Löcken et al. (2019) compared 6 concepts. The concept *Smart Road* (based on work by Umbrellium Mairs, 2017) was rated best in the relatively simple scenario. As construction sites are noisy (Kantová, 2017), according to the design space presented by Colley and Rukzio (Colley and Rukzio, 2020b), using visual clues seems to be the most promising approach for a construction site. The mentioned works address the scenario "walk over a street in front of an AV," which was shown to be the most evaluated one in current research (Colley et al., 2020a). Also, Colley and Rukzio found in their categorization (Colley and Rukzio, 2020a) that most work focuses on command- and intention-based communication. Still, they agree with Nguyen et al. (2019) that bidirectional communication should be possible (Colley and Rukzio, 2020a). However, until now, bidirectional communication was not yet addressed. In this work, we define requirements for such communication between a HAT and a CSW. These are based on an analysis of the communication needs of human truck drivers.

3. REQUIREMENT ANALYSIS

To (1) gain a better understanding of CSD challenges, (2) explore communication between truck drivers, CSWs, and other traffic participants and (3) to discuss visualization opportunities of HATs, we conducted a workshop with truck drivers ($N = 7$) in Ulm, Germany and observed two construction sites in Ulm, Germany and Neu-Ulm, Germany.

3.1. Communication on Construction Sites
3.1.1. Method
Two constructions sites depicted in **Figure 1** in Ulm, Germany and Neu-Ulm, Germany were observed by the second author *in-situ* each for $\approx$ 3 h with the criteria: (1) easily observable, (2) located close to pedestrian areas, and (3) reasonable size to ensure a high frequency of truck arrivals. Construction site 1 is a vast construction site of a multi-functional building. Construction site 2 is a demolition of a multi-story car park. As we were interested in communication between truck drivers and CSWs, communication-related observed variables were *Communication Modes*, and, with a special focus on gestures, *Gesture Count per Delivery*, and *Gesture Type* (see **Figure A1** in **Appendix** for gesture definitions and **Table A1** in **Appendix** for results).

3.1.2. Results
No tablet-based communication (e.g., showing an unloading spot) was observed. *Walking to show the way* (13) and *speech* (10) were used almost as often as gestures (15). Gesture usage was common ($\approx$ 2 per arriving vehicle) on both sites. Different gestures were used, especially indicating to come closer, show a direction, and indicate to stop. Encountered problems were a missing instructor, inconsiderate traffic, and pedestrians crossing while the truck enters the construction site.

3.1.3. Conclusion
Therefore, it is concluded that a lot of communication was needed for delivery. Nonetheless, the context has to be taken into account. While both construction sites differ in size, they are rather large. Construction site 1 requires the delivery of expensive equipment, which is why we believe that extra caution was taken, leading to increased communication. Construction site 2 was rather cluttered, therefore, especially for the entrance, communication was needed. While we believe that this is typical for larger construction sites, smaller ones might require less communication. Therefore, we conducted a workshop with truck drivers to gain additional knowledge on the communication needs for construction site deliveries.

3.2. Workshop
The research question (RQ) of the workshop was: *RQ: How is the current delivery process with special regards to communication and how could it be in the future?*

3.2.1. Procedure

The workshop consisted of four phases: *introduction*, *deliveries today*, *enactment/demonstration* and an open *discussion about AVs*. During the entire session, audio and video were recorded. The workshop started with an *introduction* of the individuals and the research field of the organizers. The goals of the workshop, defining the process of CSD, potential problems for AVs, and external communication of such AVs to overcome these, were introduced. Participants then signed informed consent. A walk-through of a CSD was described by one participant [using the critical incident technique (Flanagan, 1954)], highlighted in several aspects by other participants. Participants were encouraged to elaborate on gesture usage, communication with CSWs and pedestrians/car drivers, and major difficulties. Afterwards, an induction into a warehouse was enacted. [P2] drove a truck (moving floor) while [P1] and [P3] acted as instructors. Participants were then asked about their associations of AVs. Afterwards, the communication possibilities of AVs on construction sites were discussed. The workshop lasted about 2.5 h. Participants were compensated with 25 *currency*.

3.2.2. Participants

We recruited truck drivers ($N = 7$; see **Table A2** in **Appendix**) via a notice board at a transportation company near Ulm, Germany. Participants were required to have delivered to construction sites to take part. On average, participants were $M = 52.86$ ($SD = 4.34$; range: 48–60) years old and truck driver for $M = 27.29$ ($SD = 7.09$; minimum: 21) years. The workshop took place at the company. We recruited truck drivers as opposed to CSWs, as not all CSWs are involved in deliveries. On 5-point Likert scales (1 = *strongly disagree*, 5 = *strongly agree*), participants reported a moderate interest in AVs ($M = 3.43$, $SD = 0.79$), were neutral toward whether such a system would ease their lives ($M = 3.00$, $SD = 0.58$), and were unsure about AVs to become reality by 2029 ($M = 3.14$, $SD = 1.07$).

3.2.3. Analysis

During the focus group, the first and second author were present, could clarify uncertainties, and, therefore, were familiar with the discussions. Still, the entire focus group was recorded and analyzed using the method proposed by Mayring (2015): The first, second, and third author (as an unbiased coder) independently listened to the recordings and transcribed anchor quotes. After a discussion, the second author transcribed the parts deemed most relevant (min 5–20, min 92–122 of the workshop). Relevance was determined *via* the relatedness to the RQ. These were again read independently and coded using an open and axial coding approach (Saldaña, 2015). Afterwards, the authors discussed relevant topics with their associated anchor quotes and codings. In an iterative approach, general topics were carved out. Disagreements were discussed until an agreement was reached. Two authors filmed the demonstration. The videos were independently coded regarding the used gestures by the first and second author. We report the used gestures which, compared to the gestures used for the construction site observation (NAVY, 2019), differ slightly.

4. RESULTS

We report the practices and challenges of human CSD and put these in the context of technical and HCI solutions.

4.1. Human Delivery Process

In this section, the characteristics of construction sites and the CSD process are described.

4.1.1. Construction Site Characteristics

Participants agreed (7/7) that every construction site "is different" [P3] and change rapidly and frequently. Numerous factors define a construction site. A (non-exhaustive) list includes: size, organization, purpose, location, number of entries, underground, number of CSWs, machinery, property developer, and surroundings. Two archetypes of construction sites were distinguished by the participants: the *vast well-organized* site and the *small chaotic* site. The former is defined by clear responsibilities, a clear schedule (e.g., a defined crane date), defined unloading sites, and CSWs often already waiting for the delivery. On construction sites of the archetype *small chaotic*, the unloading spot is often unclear and the CSM/CSWs are missing.

4.1.2. Delivery Process

The delivery process was modeled according to the descriptions of the workshop participants (see **Figure 2**). Afterwards, the model was discussed with one participant that volunteered to review it.

Before delivery, contact information is essential as instructions are often unclear. Additional information required is, depending on the size, directions, crane date (if needed), and appropriate vehicle types. We want to highlight specific aspects of the delivery process. (1) Delivery without communication is possible, however, not for first-time deliveries and only for easily reachable unloading spots. Therefore, some companies "leave the keys in the vehicle" [P4] to allow truck drivers to move them. (2) Negotiation is required with CSWs/CSMs and other truck drivers. In simultaneous deliveries, one has "to discuss" [P4] about the unloading order. Truck drivers "know each other pretty well" [P4] and, therefore, avoid confrontation. Especially in larger sites, an inter-truck driver orchestration is necessary to reduce waiting times. [P4] summarizes this: "[On] big construction sites, you look who's taking longer. That's where we make arrangements. [For example:] You do it first." Orchestration is also necessary between truck drivers and CSWs. After a first check when entering the construction site, negotiation about needed arrangements (move objects or machinery; see **Figure 2** step 5) takes place. This communication is eased by a high organization as this omits the need of moving objects. (3) After successfully unloading, signing the delivery note is needed depending on the customer. [P4] states that "you take half of the delivery notes with you again". These deliveries are then invoiced to the known customers. [P2], who delivers goods such as concrete elements (see **Table A2** in **Appendix**), disagreed, as he is more involved with fitters that "are there already
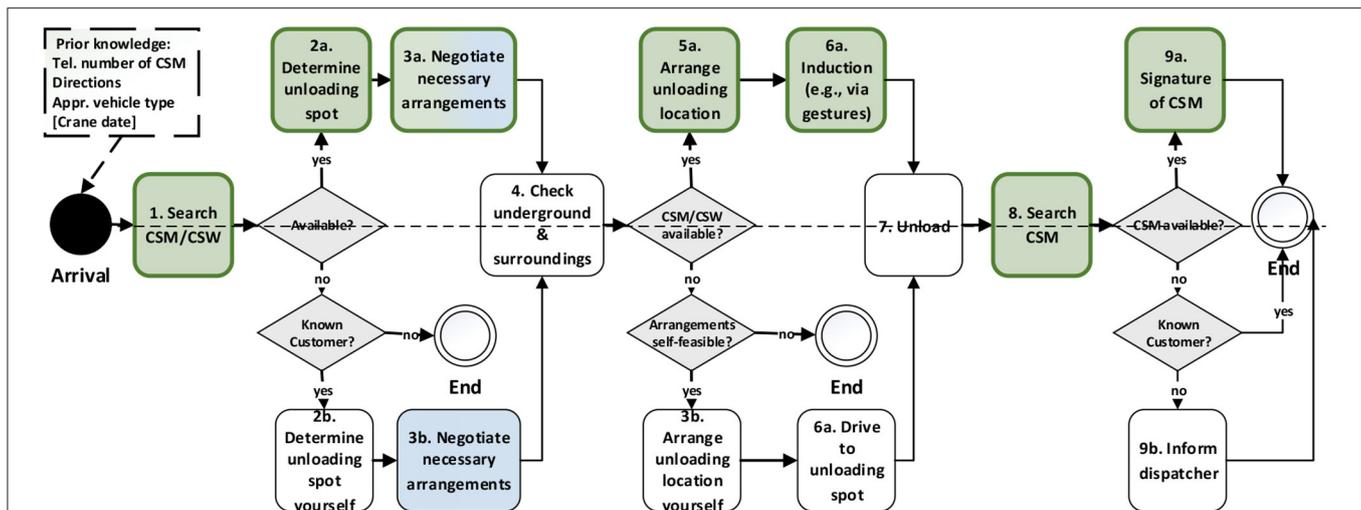
FIGURE 2 | Delivery process to a construction site. In the top half, the process is depicted with a CSM/CSW present. The lower half shows an independent delivery. During this *nine-step process*, communication between the *truck driver* and the *construction site manager* (CSM) or a CSW, if present, is needed 7 times (indicated in green). Communication with other present drivers is highlighted in blue. As construction sites tend to be chaotic, this communication is needed for making arrangements for the unloading site and for actually being able to maneuver there. If no CSW/CSM is present, the truck driver delivers independently (white background) only if (1) it is a known customer and (2) the unloading spot is reachable. The bigger and more difficult, the more communication is necessary.
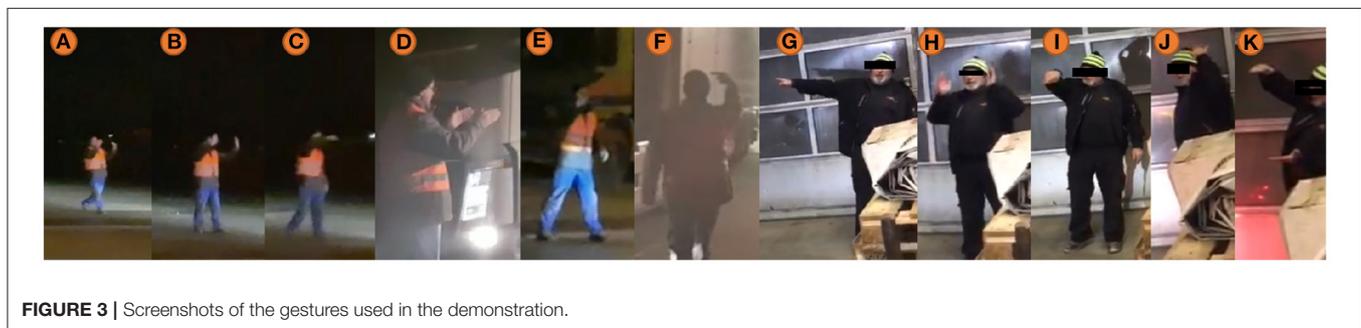

FIGURE 3 | Screenshots of the gestures used in the demonstration.

waiting and they make sure that everything runs smoothly." Leaving without a signed delivery note does not happen for first-time customers.

4.1.3. Gesture Usage and Communication With Pedestrians and Car Drivers

Participants claimed gestures to be used rarely ("maybe 5% [of the time]" [P3]) by the CSW. This is in contrast to our (limited) observations (see Section 3.1). There was no agreement on the quality of these instructions (20–40% are good [P4]; "It's a larger part of them [the CSWs] that also [use gestures properly]. Can and know where they want to have their sand" [P2]). Participants referred the quality of the gestures to their dependability. For instance, CSWs might wrongly indicate that there is enough space. Regarding the demonstration, we found the following gestures to be used (see **Figure 3** and **Table 1**).

Most gestures were only used for a short duration (1 - 8s). Both one-handed and two-handed gestures were used.

Different gestures were also used for the same message (i.e., an outstretched arm and pointing finger for giving a direction; gesturing stop with one or two hands). This presents the first set of relevant gestures used today. While standardization is a possibility for future interfaces, using currently used gestures as a starting point seems reasonable as these could be used for a possible standardization and altering long-standing routines is undesirable and difficult to internalize. It also presents the first requirements for the detection capabilities if such communication were to be enabled.

As many construction sites are directly street-connected, truck drivers regularly come in contact with pedestrians and car drivers when entering the construction site (often driving backward) or because "we often stand crosswise on the road, that happens a lot with us" [P4]. Pedestrians and car drivers often react annoyed or angry. The truck drivers then try to calm the road users or let them pass, for example, via gestures. Especially with children, participants stated to try to help and prevent them from walking by behind the truck as this is dangerous.

TABLE 1 | Gestures used in the demonstration by instructor 1 (left) and instructor 2 (right).

Gesture	Figure	Usages	Total duration	Gesture	Figure	Usages	Total duration
Wave backwards both hands	**Figure 3A**	1	7s	Outstretched one arm	**Figure 3G**	3	5s
Raised both hands (stop sign)	**Figure 3B**	1	1s	Hands side by side (giving direction)	**Figure 3H**	2	6s
Steering movement one finger	**Figure 3C**	1	3s	Steering movement one finger	**Figure 3I**	2	5s
Hands side by side (giving direction)	**Figure 3D**	4	24s	Wave backwards one hand	**Figure 3J**	2	14s
Wave backwards one hand	**Figure 3E**	1	2s	Distance two hands	**Figure 3K**	1	8s
One finger pointing	**Figure 3F**	5	22s				

4.2. Automated Vehicles for Construction Site Deliveries

4.2.1. Mixed Attitudes Toward Highly Automated Trucks and Flawed Consideration of Ground Assessment

Several participants already encountered test AVs. "[You can] not recognize from the driving style that they are different" [P3]. In general, the view on AVs is positive. However, later, [P3] stated "then [when a HAT is available] your job is obsolete". [P4] disagreed and mentioned "load securing" as a human-dependent task.

Participants mentioned today's problems with sensors: "Snow and ice on the sensor, and gone it is" [P4], a concern already mentioned for AVs in the literature (Yan et al., 2016). Another concern mentioned was data acquisition above the vehicle such as "hanging branches" [P5] or "steel pipes" [P5] as well as the underground. [P4] emphasized the condition of the street banquet and whether AVs can detect these. However, this is a well-researched application of Ground Penetrating Radars (GPRs).

5. DISCUSSION

Disinterest in HATs is surprising being one of the main drivers of the development of AVs. Such vehicles are expected to reduce costs, increase safety, and improve efficiency (Dougherty et al., 2017). Being estimated "at least a decade in the future" (Nowak et al., 2016), this is still rather near. An owner could save 32 400€ per year, mostly due to reduced driver costs (Nowak et al., 2016), therefore, probably leading to quick adoption. With such negligence, younger truck drivers will fail to appropriately react to the changing job market with severe problems for the individual and the public (Hansen, 1988; Hironimus-Wendt, 2008). A potential explanation of this denial of appropriate estimation of the possibility of job loss if the *fear of being easily replaceable*, especially via automation (Au-Yong-Oliveira et al., 2019).

Regarding the capabilities of road evaluation, Saarenketo and Scullion (2000) gave a status report on the usage of GPR for road evaluation. They showed that via radar, a profile of the street (underground) can be determined from which conclusions about the ground's stability, the structure, and the road's quality can be drawn. It seems reasonable that trucks equipped with such technology should have no difficulty on construction sites. However, this problem seems to not have been addressed specifically. Therefore, the participant's assessment of the AV's capabilities seems inadequate.

5.1. Proposals for Control Over Highly Automated Trucks

Participants were introduced to the scenario: *A HAT arrives at a construction site. Due to legal reasons, there still is a truck driver present within the cabin. Please take the role of the driver. The HAT reaches a system limitation trying to enter the construction site.*

Asked to choose to either (1) drive the vehicle themselves to the final destination or to (2) somehow control the vehicle, all participants wanted to drive themselves. While these two approaches are not exclusive, in our envisioned scenarios, the CSW actually has to take over the communication with the HAT which might not be manually operable or the CSW might not have the required driver's license. Therefore, the CSW could only use option 2. Therefore, participants were asked to still imagine such control. This resembles a potential use case: a human, either a truck driver or a CSW, could induce the HAT.

[P3] and [P4] mentioned markers to navigate the HAT ("Then we make a point [with a color]." [P4]; "with a laser pointer" [P3]). This approach, however, misses relevant information such as "[...] unload sideways, unload backwards " [P4]. A more complex interaction seems necessary in a more complex environment in which the HAT has to maneuver around curves and even has to switch from driving forwards to backwards or vice versa. Such an approach is using a device to guide the HAT (see Kemp et al., 2008). [P3] proposed voice commands or a laser pointer to indicate waypoints. [P4] proposed using a device followed by the HAT and defines the needed orientation via its orientation. Gestures were thought to miss the needed range and expressiveness of commands ("There aren't that many movements for everything we have to do" [P2]).

5.1.1. Proposals for External Communication of Highly Automated Trucks

Participants agreed that a visualization of the *driving space* would be beneficial to increase safety ("You're only dead once" [P4]). They agreed (7/7) that some visualization of *intent* and *awareness* would be beneficial ("I'd like to know what he [the truck] is doing" [P4]). Comparisons to forklifts with blue lights indicating their trajectory were drawn. [P4] mentioned the visualization of the unloading spot ("You'll see it before it is put down, before it tips over."). This has the advantage of maneuvering to the correct spot more precisely and for CSWs to be alerted of danger.

Another proposal was to visualize or give auditory feedback of recognized obstacles (e.g., a metal beam dangling aloft) to remove them ("[the trucks] says obstacle" [P3]; "hurdle right rear, 2m" [P4]). Afterwards, some ideas and concepts from related work were introduced. *Awareness* of people and objects was greeted enthusiastically. No preference regarding the kind of visualization was given, however, a combination was requested ("Then maybe both [LED strip and projection of awareness], to be 100% sure." [P4]) Regarding visualizations relevant to pedestrians and car drivers, there were no clear preferences ("maybe like a warning light" [P2]; "are gimmicks" [P4]).

6. COMMUNICATION REQUIREMENTS FOR HIGHLY AUTOMATED TRUCKS

As construction sites are likely to stay complex and disorganized, at least at times, it seems unlikely to assume that HATs can operate under all circumstances imaginable. Therefore, being able to communicate is a prerequisite. Thus, we discuss each communication relevant step of the delivery process (see **Figure 2**), which was derived by the workshop, regarding automation or interaction possibilities. Additionally, we present communication possibilities to enhance safety and efficiency. Several assumptions are made: (1) Relevant prior knowledge is available. (2) Due to quickly changing constructions sites, no high resolution a priori maps are available. (3) It is a first time delivery, therefore, the site and customer are unknown. Three approaches seem valid: (1) Automate every task. (2) Use teleoperation (continuously or solely for the communication and induction part) as a subset of HCI. (3) Use an HCI approach without including a distant human operator, which we focus on. Based on the workshop, the two main foci of communication requirements concern (1) the delivery process and (2) simultaneously increasing safety for CSW and other people.

6.1. Automating Process Steps or Employing Human-Computer Interaction Principles

1. Search CSM/CSW. This is a necessary step to obtain information about the unloading spot and to confirm the delivery. For this, a CSW/CSM is needed. As the telephone number is available, a text message or a call could be made in advance (e.g., with technology such as Google Duplex O'Leary, 2019). Upon arrival, the HAT should communicate its load, its origin, and other relevant information to the CSM/CSW (e.g., via speech or a display).

Req. (1): Notify the CSM of the arrival of the HAT and be able to communicate relevant data of the delivery.

2a. Determine Unloading Location. In this task, usually, the truck driver is shown the unloading spot. For this, the driver walks with the CSW/CSM. This is not possible for a HAT. In this step, the human CSM/CSW could either (1) gather contextual

data and determine directions for the HAT to execute or (2) determine continuous directions to support precise navigation.

3a. Negotiate Necessary Arrangements. In this task, the truck driver communicates with other drivers and CSWs to define who delivers first. In mixed traffic, i.e., deliveries from HATs and manually driven ones, this should be taken over by the responsible CSM. For totally automated deliveries, the HATs could communicate with each other to be able to switch positions if, for example, a traffic jam occurs.

5a. Arrange Unloading Location. As task *2a* is not possible for HATs, it is likely that the vehicle will be somehow lead to the unloading spot (e.g., gestures, person-following, laser pointer, transmitter, etc.). While there could be technical solutions to this problem, these are likely time-consuming, expensive, or difficult to achieve. For example, for smaller deliveries, a different approach is to use HATs equipped with drones as proposed by Daimler (Korosec, 2016). Such a design was shown to be cost-reducing under certain assumptions such as using multiple drones and certain costs per mile for trucks and drones (Campbell et al., 2017; Ham, 2018). This seems unfeasible for heavy deliveries. However, such drones could be used for an assessment of the situation and its conditions. OpenDroneMap (2020) was shown to provide good results compared to commercially available software (Burdziakowski, 2017). In this case, clear markings for unloading spots are needed. Still, the HAT would have to communicate necessary arrangements (e.g., "remove the container"). For an HCI solution, this is also a requirement. Additionally, receiving information about directions has to be possible, leading to requirements 2 and 3.

Req. (2): Be able to communicate to the CSM/CSW about issues preventing reaching the unloading spot (e.g., steel pipes, unconsolidated ground, etc.).

Req. (3): Be able to receive relevant information on directions.

6a. Induction. Req. (3) opens up possibilities to engage with the HAT. Numerous ways to communicate are possible, some of them were mentioned in the Workshop or in literature. Methods that implicitly assume that an operator uses a robot for a longer time could not be applicable as a CSW will probably have to abandon current work to address the delivery. Therefore, such approaches would require the CSW to take off gloves (also for stylus usage) and get a tablet. Such prerequisites could diminish the usage of these methods (see Fong et al., 2003; Hosoi et al., 2007). A laser pointer (Ishii et al., 2009) could be used as CSWs could constantly carry one. The increased cost could be a negative factor. Additionally, orchestration for multiple deliveries with multiple CSWs using laser pointers at the same time could reduce effectiveness. However, in other domains, it was shown that physical interaction or person following were preferred (Jevtić et al., 2015).

Therefore, speech and gestures/person following seem to be more promising approaches. Gestures used today were shown in the demonstration (see **Figure 3** and **Table 1**). The duration of the gestures in the demonstration also gives some indication about needed recognition times of algorithms ($\approx$ 1s).

Req. (4): Interaction should need as little preparation (technological overhead, cost, equipment, and task switching) by the CSW as possible.

8a. Search CSM & 9. Signature of CSM. Here, the CSM has to be found for a signature, if still needed in the future. This could either (1) be done as in step 1 (see **Figure 2**), including a stylus capable touch screen on the HAT.

6.2. Increase Safety for Construction Site Workers and Other People

To increase safety, a HAT should convey its intention and awareness to the surrounding CSWs. Multimodal communication is crucial as a construction site holds many distractors.

Req. (5): A HAT should be able to communicate its intention and awareness to surrounding people multimodally.

(1) Auditory cues could be used for gaining the attention of relevant CSWs (e.g., via beam loud speakers Olszewski et al., 2005 to reduce noise) while (2) visual cues can indicate recognition of objects or people and where the vehicle is heading to see for a forklift (Correa et al., 2010; Teller et al., 2010; Walter et al., 2015). Additionally, unloading spots could be highlighted (e.g., where the vehicle will unload and how much space this will occupy) to avoid people walking into them.

The Workshop showed today's need for truck drivers to communicate with a variety of people: CSWs, other truck drivers, pedestrians of varying age, and car drivers. It seems reasonable to define appropriate strategies to communicate with each group. or example, for pedestrians wanting to walk past a HAT standing halfway on the street, the HAT could aid in displaying warnings about oncoming traffic or using a see-through display (Zhang et al., 2018).

Req. (6): The HAT should be capable of communicating appropriately with a variety of construction site related people: CSW, pedestrians, car drivers.

7. LIMITATIONS

The relatively small size of the workshop ($N = 7$) and especially the absence of younger truck drivers limit the generalizability of the findings, e.g., the claim that there is the denial of the job loss possibility. However, the validity of the findings is not necessarily decreased by this (Toner, 2009). Also, the cultural background of the participants was European and deliveries were mostly within Europe (see **Table A2** in **Appendix**). As communication is culture-dependent (Nishimura et al., 2008), this Additionally, actually involving CSWs would increase external validity as the attending truck drivers could have been unwilling to present relevant information out of fear of job loss. Still, we believe our approach to be valid as truck drivers are today **always** involved in the communication process. Also, CSWs could sympathize with truck drivers and, therefore, try to sabotage HATs. However, such findings would be difficult to obtain in a focus group. Thus, we did not include a requirement for being not sabotageable.

8. CONCLUSION AND FUTURE WORK

Overall, this work provides data about current delivery practices of truck drivers and proposes clear requirements for the future interaction between CSW/CSM and HATs. We show the challenges truck drivers face when delivering to construction sites by conducting a workshop ($N = 7$). A detailed process for deliveries was presented (see **Figure 2**) showing communication needs between truck drivers and CSWs, other truck drivers, as well as other road users such as pedestrians. This detailed process highlights the challenges HATs will encounter for the last mile of CSD. Based on this process, requirements for a cooperative approach were deducted. These show the need to include interaction aspects for the delivery of HATs to construction sites. This work shows that even for an uncomplicated delivery, a lot of parameters and various people have to be accounted for. In a next step, we intend to implement scenarios of such a HAT with a particular focus on communication with car drivers. We also plan to evaluate external communication concepts targeted toward a construction site and especially toward CSWs.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

REFERENCES

Ackermann, C., Beggiato, M., Schubert, S., and Krems, J. F. (2019). An experimental study to investigate design and assessment criteria: what is important for communication between pedestrians and automated vehicles? *Appl. Ergon.* 75, 272–282. doi: 10.1016/j.apergo.2018.11.002

Au-Yong-Oliveira, M., Almeida, A. C., Arromba, A. R., Fernandes, C., and Cardoso, I. (2019). "What will the future bring? the impact of automation on skills and (un)employment," in *New Knowledge in Information Systems and Technologies*, eds Á. Rocha, H. Adeli, L. P. Reis, and S. Costanzo (Cham: Springer International Publishing), 206–217.

Burdziakowski, P. (2017). "Evaluation of open drone map toolkit for geodetic grade aerial drone mapping-case study," in *17 International Multidisciplinary Scientific GeoConference SGEM 2017: Informatics, Geoinformatics and Remote Sensing Strony*, 101–109.

Campbell, J. F., Sweeney, D., and Zhang, J. (2017). *Strategic Design for Delivery With Trucks and Drones.* Supply Chain Analytics Report SCMA (04 2017), University of Missouri, St. Louis, United States. Available online at: https://www.villagereach.org/wp-content/uploads/2018/07/StrategicDesignforDeliverywithDronesandTrucks_4-17-17_SCMA-2017-0201.pdf (accessed April 17, 2017).

Colley, M., Mytilineos, S. C., Walch, M., Gugenheimer, J., and Rukzio, E. (2020a). "Evaluating highly automated trucks as signaling lights," in *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '20* (New York, NY: Association for Computing Machinery), 111–121.

Colley, M., and Rukzio, R. (2020a). "A design space for external communication of autonomous vehicles," in *Proceedings of the 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '20* (New York, NY: ACM, Association for Computing Machinery).

Colley, M., and Rukzio, R. (2020b). "Towards a design space for external communication of autonomous vehicles," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20* (New York, NY: ACM, Association for Computing Machinery).

Colley, M., Walch, M., Gugenheimer, J., Askari, A., and Rukzio, R. (2020b). "Towards inclusive external communication of autonomous vehicles for pedestrians with vision impairments," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20* (New York, NY: ACM, Association for Computing Machinery).

Colley, M., Walch, M., Gugenheimer, J., and Rukzio, E. (2019). "Including people with impairments from the start: external communication of autonomous vehicles," in *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings, AutomotiveUI '19* (New York, NY: Association for Computing Machinery), 307–314.

Colley, M., Walch, M., and Rukzio, R. (2020c). "Unveiling the lack of scalability in research on external communication of autonomous vehicles," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20* (New York: ACM, Association for Computing Machinery).

Correa, A., Walter, M. R., Fletcher, L., Glass, J., Teller, S., and Davis, R. (2010). "Multimodal interaction with an autonomous forklift," in *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY: IEEE Press), 243–250.

Dougherty, S., Ellen, P., Stowell, J., and Richards, A. (2017). Perceptions of fully autonomous freight trucks. doi: 10.2139/ssrn.3027143

Eagle, C. (2019). *Forklift Warning Safety Lights.* Available online at: https://www.cisco-eagle.com/catalog/category/7099/forklift-safety-light.

Fagnant, D. J., and Kockelman, K. (2015). Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Trans. Res. A Policy Pract.* 77, 167–181. doi: 10.1016/j.tra.2015.04.003

Fan, W., Carroll, C., Radley, L., Hostetler, E., Choe, S., Leite, F., et al. (2019). *Prevention of Backing Fatalities in Construction Work Zones.* TxDOT Report 0-6703-1, Texas, United States. Available online at: https://library.ctr.utexas.edu/hostedpdfs/uttyler/0-6703-1.pdf (accessed February 01, 2019).

Flanagan, J. C. (1954). The critical incident technique. *Psychol. Bull.* 51, 327. doi: 10.1037/h0061470

Florentine, E., Ang, M. A., Pendleton, S. D., Andersen, H., and Ang, M. H. (2016). "Pedestrian notification methods in autonomous vehicles for multi-class mobility-on-demand service," in *Proceedings of the Fourth International Conference on Human Agent Interaction, HAI '16* (New York, NY: Association for Computing Machinery), 387–392.

Fong, T., and Thorpe, C. (2001). Vehicle teleoperation interfaces. *Auton Robots* 11, 9–18. doi: 10.1023/A:1011295826834

Fong, T., Thorpe, C., and Glass, B. (2003). "Pdadriver: a handheld system for remote driving," in *IEEE International Conference on Advanced Robotics, Number CONF* (New York, NY: IEEE).

Fosbroke, D. (2004). *Niosh Reports! Studies on Heavy Equipment Blind Spots and Internal Traffic Control.* Available online at: https://www.workzonesafety.org/files/documents/news_events/wz_conference_2004/heavy_equipment.pdf.

Graham, J. L., and Burch, R. (2006). Internal traffic control plans and worker safety planning tool. *Transp Res. Rec.* 1948, 58–66. doi: 10.1177/0361198106194800107

Ham, A. M. (2018). Integrated scheduling of m-truck, m-drone, and m-depot constrained by time-window, drop-pickup, and m-visit using constraint programming. *Transp. Res. CEmerg. Technol.* 91, 1–14. doi: 10.1016/j.trc.2018.03.025

Hansen, G. B. (1988). Layoffs, plant closings, and worker displacement in america: Serious problems that need a national solution. *J. Soc. Issues* 44, 153–171. doi: 10.1111/j.1540-4560.1988.tb02097.x

Hironimus-Wendt, R. J. (2008). The human costs of worker displacement. *Human. Soc.* 32, 71–93. doi: 10.1177/016059760803200105

Holländer, K., Wintersberger, P., and Butz, A. (2019). "Overtrust in external cues of automated vehicles: an experimental investigation," in *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '19* (New York, NY: Association for Computing Machinery), 211–221.

Hosoi, K., Dao, V. N., Mori, A., and Sugimoto, M. (2007). "Visicon: a robot control interface for visualizing manipulation using a handheld projector," in *Proceedings of the International Conference on Advances in Computer Entertainment Technology, ACE '07* (New York, NY: Association for Computing Machinery), 99–106.

Ishii, K., Zhao, S., Inami, M., Igarashi, T., and Imai, M. (2009). "Designing laser gesture interface for robot control," in *Human-Computer Interaction-INTERACT 2009*, eds T. Gross, J. Gulliksen, P. Kotzé, L. Oestreicher, P. Palanque, R. O. Prates, M. Winckler (Berlin; Heidelberg: Springer Berlin Heidelberg), 479–492.

Jevtić, A., Doisy, G., Parmet, Y., and Edan, Y. (2015). Comparison of interaction modalities for mobile indoor robot guidance: direct physical interaction, person following, and pointing control. *IEEE Trans. Hum. Mach. Syst.* 45, 653–663. doi: 10.1109/THMS.2015.2461683

Kantová, R. (2017). Construction machines as a source of construction noise. *Procedia Eng.* 190, 92–99. doi: 10.1016/j.proeng.2017.05.312

Kemp, C. C., Anderson, C. D., Nguyen, H., Trevor, A. J., and Xu, Z. (2008). "A point-and-click interface for the real world: laser designation of objects for mobile manipulation," in *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (New York, NY: IEEE), 241–248.

Korosec, K. (2016). *Here's Why Mercedes Is Betting on Drones and Self-Driving Robots.* Available online at: https://fortune.com/2016/09/07/mercedes-vans-drones-matternet/.

Litman, T. (2017). *Autonomous Vehicle Implementation Predictions.* Victoria, BC: Victoria Transport Policy Institute.

Löcken, A., Golling, C., and Riener, A. (2019). "How should automated vehicles interact with pedestrians? a comparative analysis of interaction concepts in virtual reality," in *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '19* (New York, NY: Association for Computing Machinery), 262–274.

Lundgren, V. M., Habibovic, A., Andersson, J., Lagström, T., Nilsson, M., Sirkka, A., et al. (2017). *Will There Be New Communication Needs When Introducing Automated Vehicles to the Urban Context?* Cham: Springer International Publishing.

Mahadevan, K., Somanath, S., and Sharlin, E. (2018). "Communicating awareness and intent in autonomous vehicle-pedestrian interaction," in *Proceedings of the*

2018 CHI Conference on Human Factors in Computing Systems, CHI '18 (New York, NY: Association for Computing Machinery), 1–12.

Mairs, J. (2017). *Umbrellium Develops Interactive Road Crossing That Only Appears When Needed.* Available online at: lhttps://www.dezeen.com/2017/10/12/umbrellium-develops-interactive-road-crossing-that-only-appears-when-needed-technology/.

Mayring, P. (2015). "Qualitative content analysis: theoretical background and procedures," in *Approaches to Qualitative Research in Mathematics Education: Examples of Methodology and Methods*, eds A. Bikner-Ahsbahs, C. Knipping, and N. Presmeg (Dordrecht: Springer Netherlands), 365–380.

Mersky, A. C., and Samaras, C. (2016). Fuel economy testing of autonomous vehicles. *Transp. Res. C Emerg. Technol.* 65, 31–48. doi: 10.1016/j.trc.2016.01.001

Moore, D., Currano, R., Strack, G. E., and Sirkin, D. (2019). "The case for implicit external human-machine interfaces for autonomous vehicles," in *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '19* (New York: Association for Computing Machinery), 295–307.

NAVY (2019). *Navy Gestures, Appendix A.* Available online at: https://www.globalsecurity.org/military/library/policy/army/fm/1-564/AA.HTM.

Nguyen, T. T., Holländer, K., Hoggenmueller, M., Parker, C., and Tomitsch, M. (2019). "Designing for projection-based communication between autonomous vehicles and pedestrians," in *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '19* (New York, NY: Association for Computing Machinery), 284–294.

Nishimura, S., Nevgi, A., and Tella, S. (2008). "Communication style and cultural features in high/low context communication cultures: a case study of finland, japan and india," in *Teoksessa A. Kallioniemi (toim.)*, *Uudistuva ja kehittyvä ainedidaktiikka. Ainedidaktinen Symposiumi, Vol. 8.* (Helsinki), 783–796.

Nowak, G., Maluck, J., Stürmer, C., and Pasemann, J. (2016). *The Era of Digitized Trucking Transforming the Logistics Value Chain.* Available online at: https://www.strategyand.pwc.com/gx/en/insights/2016/the-era-of-digitized-trucking/the-era-of-digitized-trucking-transforming.pdf.

O'Leary, D. E. (2019). Google's duplex: pretending to be human. *Intell. Syst. Account. Finance Manag.* 26, 46–53. doi: 10.1002/isaf.1443

Olszewski, D., Prasetyo, F., and Linhard, K. (2005). "Steerable highly directional audio beam loudspeaker," in *Ninth European Conference on Speech Communication and Technology* (Lisbon: Interspeech), 137–140.

OpenDroneMap (2020). *Opendronemap.* Available online at: https://www.opendronemap.org/odm/.

Saarenketo, T., and Scullion, T. (2000). Road evaluation with ground penetrating radar. *J. Appl. Geophys.* 43, 119–138. doi: 10.1016/S0926-9851(99)00052-X

Salda na, J. (2015). *The Coding Manual for Qualitative Researchers.* London, UK: SAGE Publications Ltd.

Scania (2019). *A New Cabless Concept – Revealing Scania AXL.* Available online at: https://www.scania.com/group/en/a-new-cabless-concept-revealing-scania-axl/.

Sprute, D., Tönnies, K., and König, M. (2019). "This far, no further: Introducing virtual borders to mobile robots using a laser pointer," in *2019 Third IEEE International Conference on Robotic Computing (IRC)* (New York, NY: IEEE), 403–408.

Takayama, L., Marder-Eppstein, E., Harris, H., and Beer, J. M. (2011). "Assisted driving of a mobile remote presence system: system design and controlled user evaluation," in *2011 IEEE international conference on robotics and automation* (New York, NY: IEEE), 1883–1889.

Teller, S., Walter, M. R., Antone, M., Correa, A., Davis, R., Fletcher, L., et al. (2010). "A voice-commandable robotic forklift working alongside humans in minimally-prepared outdoor environments," in *2010 IEEE International Conference on Robotics and Automation* (New York, NY: IEEE), 526–533.

The National Institute for Occupational Safety (NIOSH). (2011). *Engineering Considerations and Selection Criteria for Proximity Warning Systems for Mining Operations.* https://www.cdc.gov/niosh/mining/content/pwsselection.html.

Toner, J. (2009). Small is not too small: Reflections concerning the validity of very small focus groups (vsfgs). *Qualitat. Soc. Work* 8, 179–192. doi: 10.1177/1473325009103374

Valero, A., Randelli, G., Saracini, C., Botta, F., and Mecella, M. (2009). "The advantage of mobility, mobile tele-operation for mobile robots," in *Proc. AISB-HRI Symposium New Frontiers in Human-Robot Interaction*, 8.

Walter, M. R., Antone, M., Chuangsuwanich, E., Correa, A., Davis, R., Fletcher, L., et al. (2015). A situationally aware voice-commandable robotic forklift working alongside people in unstructured outdoor environments. *J. Field Robot.* 32, 590–628. doi: 10.1002/rob.21539

Yan, C., Xu, W., and Liu, J. (2016). Can you trust autonomous vehicles: contactless attacks against sensors of self-driving vehicle. *Def. Con.* 24, 109. Available online at: https://files.laggy.org/infoconatdc27/contents/cons/DEF%20CON/DEF%20CON%2024/DEF%20CON%2024%20presentations/DEF%20CON%2024%20-%20Liu-Yan-Xu-Can-You-Trust-Autonomous-Vehicles-WP.pdf

Zhang, B., Wilschut, E. S., Willemsen, D. M. C., Alkim, T., and Martens, M. H. (2018). "The effect of see-through truck on driver monitoring patterns and responses to critical events in truck platooning," in *Advances in Human Aspects of Transportation*, ed N. A. Stanton (Cham: Springer International Publishing), 842–852.

APPENDIX

A. REQUIREMENT ANALYSIS

Table A1 | Collected data at construction sites 1 (15 deliveries) and 2 (7 deliveries) in [omitted for review].

Site	Time spent on site per truck (min)	Gesture count per delivery	Gesture type (see Figure A1 in Appendix)	Communication modes	Problems encountered	Cargo type
1	M = 9.93, SD = 4.62	M = 3.33, SD = 2.81	a (10), b (8), c (10), d (3), e (2)	Gesture (11), Walking (13), Speech (5), Standing (1)	Waiting, loose cable, inconsiderate traffic, pedestrians crossing, no instructor (2)	Liquid concrete (12), gravel (3)
2	M = 7.42, SD = 8.68	M = 2.00, SD = 2.08	a (4), b (3), c (2), d (1), e (3)	Gesture (4), Walking (0), Speech (5), Standing (0)	Multiple approaches needed, cooperation with excavator	Dirt (5), rubble (1), gravel (1)

Table A2 | Demographic information of workshop participants.

ID	Age	Driver since	Truck model	Current truck since	Delivered countries	Delivered components
P1	57	1996	Tipping	2017	GER	Sand, gravel
P2	54	1992	Moving floor	2015	AUT, BEL, CHE, ITA, GER, LUX, NLD, GB, USA	Concrete elements, shuttering formwork, roof tiles, lumber, climbing wall, prefabricated house
P3	60	1980	Tipping	2018	ITA, FRA, GER, NLD	Steel, sand, gravel, concrete elements, formwork
P4	51	1998	3-axle stacker	2017	GER	Garbage, construction debris, excavation, scrap
P5	50	1991	Tipping, canvas	2019	AUT, BEL, CHE, GER, LUX, NLD	Bagged goods, roof tile, bricks, steel, roof plate
P6	50	1997	Tipping, canvas	2017	BEL, CHE, GER	Pallet goods with lifting platform
P7	48	1996	Multiple	n.a.	GER	All mentioned

All participants are male and employed by the same company. Five participants are German, one is from the USA, and one from Greece.

FIGURE A1 | (a–e) Show the NAVY-defined gestures (NAVY, 2019) used for coding of the gestures observed at the construction sites.