SEMarbeta: Mobile Sketch-Gesture-Video Remote Support for Car Drivers

Sicheng Chen [1,3], Miao Chen [1,3], Andreas Kunz [2], Asim Evren Yantaç [3], Mathias Bergmark [1], Anders Sundin [1], Morten Fjeld [3]

[1] Semcon Human Factors
Theres Svenssons gata 15
SE-417 80 Gothenburg
sicheng.chen@semcon.com

[2] ICVR
ETH Zurich
CH-8092 Zurich
kunz@iwf.mavt.ethz.ch

[3] t2i Interaction Lab
Chalmers Univ. of Technology
SE-412 96 Gothenburg
morten@fjeld.ch

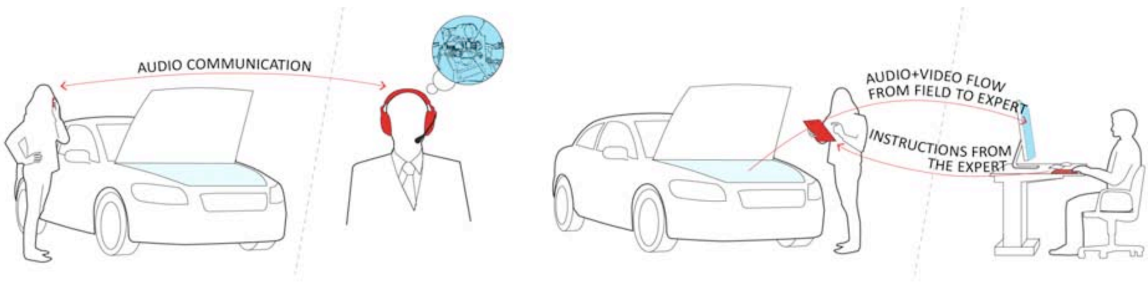

Figure 1: Remote support for car drivers is typically offered as audio instructions only (left). This paper presents a mobile solution including a sketch- and gesture-video-overlay (right).

ABSTRACT
Uneven knowledge distribution is often an issue in remote support systems, creating the occasional need for additional information layers that extend beyond plain videoconference and shared workspaces. This paper introduces SEMarbeta, a remote support system designed for car drivers in need of help from an office-bound professional expert. We introduce a design concept and its technical implementation using low-cost hardware and techniques inspired by augmented reality research. In this setup, the driver uses a portable Android tablet PC while the expert mechanic uses a stationary computer equipped with a video camera capturing his gestures and sketches. Hence, verbal instructions can be combined with supportive gestures and sketches added by the expert mechanic to the car's video display. To validate this concept, we carried out a user study involving two typical automotive repair tasks: checking engine oil and examining fuses. Based on these tasks and following a between-group (drivers and expert mechanics) design, we compared voice-only with additional sketch- and gesture-overlay on video screenshots measuring objective and perceived quality of help. Results indicate that sketch- and gesture-overlay can benefit remote car support in typical breakdown situations.

Categories and Subject Descriptors
H.5.3 [**Group and Organization Interfaces**]:
Computer-supported cooperative work, Evaluation/methodology, Synchronous interaction

General Terms
Performance, Design, User study, Human Factors, AR

Keywords
Remote support, automotive, mobile, handheld computer

1. INTRODUCTION
While troubleshooting in modern vehicles can be a challenging task for professionals, the average driver has even more problems working with automotive technologies. Even though most technical details are described in the car's handbook, many drivers are overwhelmed when having to diagnose and fix even simple problems in their car. For instance, even minor issues, such as a blown fuse, may turn out to be expensive if the driver has to call for service personnel to come and fix the problem. Physical presence of service personnel is required whenever audio communication channels such as mobile phones are insufficient. Even when audio instructions prove to be sufficient, drivers might not understand the instructions and therefore be unable to solve the problem themselves. Car drivers increasingly carry smartphones, tablets, or other devices made for sketching and video streaming. These communication channels could enable service personnel to offer help in troubleshooting without their physical presence (Fig. 1).

This paper presents remote video support technology that allows a) transferring sketch-overlaid video screenshots from the driver to the expert mechanic (the helper), and b) sending back sketch- and gesture-overlays on still images from the remote helper to the driver. While live video is streamed from the driver- to the helper-side, the driver will typically freeze the video to perform sketching. While our work is inspired by insights from Augmented Reality (AR), here we explore the use of widely used standard mobile devices. With such set-up, we design a remote instructor-operator collaborative system to support minor automotive breakdown cases.

The next section offers related work, followed by a section on system concept and design. The fourth section describes hardware implementation and the fifth software implementation. The sixth section offers system evaluation, and is followed by a section presenting a discussion and an outlook on future work.

2. RELATED WORK

With the "spread of wireless communication and the desire to travel 'light,' collaboration across mobile devices", such as phones, tables, and notebooks, is a likely trend for future groupware applications [3]. With the diffusion of mobile devices in the working landscape, there is an economic interest in using standard mobile devices to develop remote expert support [12]. A video image may even reinforce the affective experience in communication between geographically separated instructors and operators [11]. To model the complexity in mobile groupware systems, a graphical language describing loosely coupled work patterns has been suggested [5]. In the educational field, researchers have proposed solutions for collaborative learning outside the classroom [16]. While our work will target the use of standard devices in a mobile remote expert setting, it will not yet involve the use of AR techniques or the use of a Head-Mounted Display (HMD). Nonetheless, we find it instructive to start with a review of a few AR-related projects that laid foundation for understanding remote computer-mediated instructor-operator collaboration.

Early AR-related works on remote instructor-operator collaboration were presented in SharedView [8]. The operator wears a so-called shared camera and a HMD. The shared camera follows the operator's view and thereby jointly shows the task and the operator's field of view to the instructor. The system also transmits overlaid gestures in both directions. Extensive experiments with SharedView showed that to assure high user acceptance and effective use, it is important that the "system is an extension of an instructor's body" and that users' acceptance is high [8]. Such insights paired with observations from face-to-face instructor-operator collaboration were later used to refine design requirements. Hence, in the GestureCam project, the authors suggested either using a second camera that the instructor can control remotely, or widening the camera's field view so that many objects can be seen in the display at the same time [9]. In a more recent remote instructor-operator collaboration study, shoulder-worn active camera and laser (WACL) was compared with traditional HMD [7]. The authors showed that WACL was superior to HMD in several ergonomic aspects such as comfort and fatigue. In a follow-up study, the same team compared a WACL-HMD combination with a WACL Chest-Worn Display (CWD) combination to examine which form of visual assist is most suited for the WACL [14]. The authors found that the CWD is superior to HMD and showed that the WACL can give improved task performance when paired with a worn display. Recently, there has been an interest in identifying and tracking unknown features in an unprepared environment. Such model-free markerless tracking was examined for remote support and it was shown that the sole requirement to an unknown scene is the presence of locally planar objects [10]. Finally, Fussel et al [4] examined tools supporting "remote gesture in video systems being used to complete collaborative physical tasks-tasks" in which two or more subjects work together manipulating three-dimensional objects in the real world. They studied pointing gestures, representational gestures, and how to support erasing gestures.

Besides the AR-related issues of display, tracking, and pointing techniques, recent research into remote instructor-operator collaboration has been geared towards particular domains of application. In ReMoTe [1], for example, remote instructing for miners working underground is targeted. The helper side of ReMoTe can get a visual understanding of the working situation by viewing live-video captured by the worker's head-worn camera. Meanwhile the system also captures helpers' instructions from their display and sends them back to the workers' side. HMD- and marker-based AR has been examined for automotive assembly, inspection, and repair [19]. Also within automotive assembly, guidelines for AR-based training were suggested [2]. Widening the focus to design at large, AR can also be a meaningful technique, as it may help designers to express innovative ideas and overcome technical difficulties [13].

Group awareness is another aspect of remote collaboration. Ways to enhance awareness between remote workers has been presented in VideoArms [17] and in CollaBoard [6]. These systems employ live-video gestural overlay on shared workspaces, thereby giving remote workers a side-by-side impression. While VideoArms shows collaborators hands and arms, CollaBoard transmits the image of the collaborator's upper body.

3. SYSTEM CONCEPT AND DESIGN

The process of reaching a system concept and design consisted of four subsequent steps. First, we conducted a participatory design process. Second, the insights first gained from participatory design were intersected with insights drawn from related work. Third, hardware architecture was laid out and defined. Fourth, we investigated vision-based gesture capturing strategies.

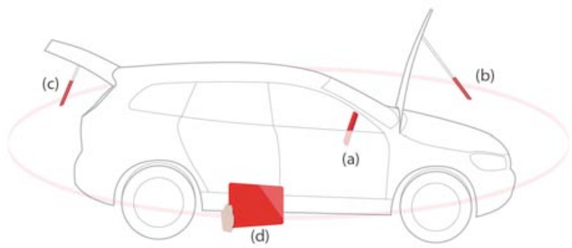

Figure 2: Our participatory design workshop indicated that mobile devices would have to be multipurpose. For example, the same device should offer a) infotainment functions, support b) engine, or c) luggage compartment instructions, and d) potentially instruct drivers how to change tires.

3.1 Participatory Design

Since the research was done in collaboration with an automotive design company, we employed a participatory design approach. Senior engineers from the company presented requirements and provided professional suggestions based on their long experience on automotive design and information presentation. Hence, we developed the idea of a system that leverages mobile devices to offer remote technical support for car drivers. We researched existing remote support services in the automotive field, as well as automotive support applications provided in iOS App stores and in the Android market. We found that the envisioned remote support system for drivers could have a major potential since the current solutions all depend on voice-only support. While video is offered as part of pre-recorded repair instructions, such products are neither interactive nor adapted to current breakdown situations. We also found that customers in the automotive industry tend to prefer multipurpose solutions where mobile devices not only offer remote support but also provide remote control for other car functionality such as an infotainment system.

In the process of gathering requirements, we used a so-called scenario method to retrieve customers' needs [19]. We produced a short video showcasing a typical use of our envisioned system; it shows a driver facing an engine breakdown. She picks up her mobile device and starts an app provided by the car manufacturer. After having connected with the helper side, the expert mechanic helper provides instructions using sketching and gesturing. Based on this demonstration it became easier for the senior engineers to understand the overall concept of our system, and thus be able to suggest relevant design goals, such as reducing the cost of use of the helper side, and defining the handheld device as a future standard automobile feature (Fig. 2).

3.2 Related Work—Design Specs Intersect

The concept of the ReMoTe system is similar to our research objective, in which the helper side can provide additional information to the worker side that is not limited to simple linguistic instruction. However, the work of conceptualizing, designing, and implementing equipment for professional users is not the subject here. That is, while HMD, head-worn camera, and laser pointer may work for professionals in ReMoTe, SharedView, or GestureCam [1, 8, 9], such devices are most likely not apt for an average end-user of mobile devices. Moreover, for cases where the HMD is not a see-through device, combining hand images with live video becomes even more intricate. Google's glass project [22] might be a future device from which our envisioned system could conditionally benefit [18].

Other AR solutions in the automotive field, like ARVIKA [21], depend on pre-defined fiducial markers for position track and 3D animation. This kind of solution is not suitable for remote support, since everyday real-time diagnosis and instruction takes place in an environment without visual markers [13].

Furthermore, VideoArms [17] and CollaBoard [6] inspired our system concept. In CollaBoard, the full upper body of a worker is displayed on the remote side. Thus, all information like postures or deictic gestures is in context with the underlying content of the whiteboard. Although we do not need postures, transferring of deictic gestures is crucial for our system, since it is the most natural explanatory gesture.

Finally, since we want to present a mobile solution for the driver side, an important design aspect is device size. While the device should have a screen large enough for the driver to unequivocally recognize the helper's gestures in relation to the underlying image, it should still be a handheld portable device.

3.3 Hardware Architecture

Our system consists of a handheld device (tablet) and a stationary computer (PC) (Fig. 3). Both the tablet and the stationary computer offer sketching and audio communication. The driver side can transmit live video streams to the helper side in order to describe the problem (e.g. check oil, locate fuse). The helper receives live video streams from the remote situation (e.g. the car engine or fuse board) and can give back instructions on how to check oil or fuses, either by outlining directly on his screen or by using gestures that are captured by a camera. His gestured instructions and sketched outlines are offered as separate layers directly overlaid on the still image. Using another color for sketching, the driver can also outline to clarify problems.

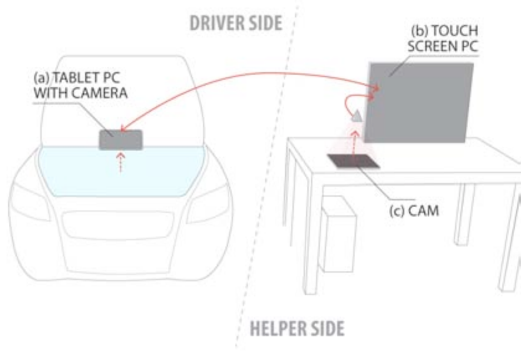

Figure 3: SEMarbeta hardware architecture

3.4 Gesture Capturing Strategy

Gesture capturing in front of a highly dynamic background such as a live video is a delicate task. While VideoArms used color segmentation algorithms to capture the deictic gestures in front of a screen, CollaBoard used a linearly polarizing filter in front of the camera and thus benefitted from the fact that an LC-screen already emits linearly polarized light. However, each method has specific shortcomings. While color segmentation only works well if no skin-like colors exist on the screen, the solution with polarized light cannot detect dark objects unless they differ significantly from the dark-gray of the captured image of the screen.

Although our system could be adapted to both of these segmentation strategies, we adopted another solution to capture the helper's gestures. We did not use any polarizing filter nor color segmentations. Instead, we mounted the camera next to the stationary computer facing downwards towards a black mat (16"x12") on the table. Thus, the helper can put his hand between the camera and the mat for his gestures to be captured. Here, we make use of the fact that we are used to this spatially distributed interaction. With a mouse, we regularly control the cursor, click a button, or draw a line while keeping our eyes on the screen (Fig. 4, bottom). However, there are some limitations to such gesture capturing that we will discuss in later sections.

4. HARDWARE IMPLEMENTATION

We set out to employ inexpensive hardware only. We also took into account that a driver's mobile unit should be low-weight for ease of handling and should be usable for other tasks. Unlike VideoArms or CollaBoard, we propose an asymmetric setup while maintaining several CollaBoard features. Two sides are involved in our system, helper side and driver, but each side works in a different context (Fig. 3). This partly asymmetric system set-up is presented next.

4.1 Helper Side

On the helper side, streamed or captured images from the driver side are combined with detected sketching (Fig. 4). Thus, standard touch screens are sufficient. In the prototype, a 22" touch screen is used. Since we use a touch screen, no further input devices, such as mouse or keyboard are needed, and the helper can easily use a pen or finger to interact with the software. However, since gestures should also be captured and transferred, an additional input capability is required.

Like in CollaBoard and VideoArms, the SEMarbeta system captures the helper's deictic gestures using a camera. As for the cost and quality of such a camera, we selected a high-resolution webcam with an auto-focus function for our prototype (Logitech QuickCam Vision Pro 9000). For our prototype, there was no need to apply polarizing filters to eliminate the background image of an LC-screen, since our setup was different to CollaBoard and VideoArms. This is further discussed in the software design section below.

No specific setup for the audio channel was required. The default microphone in the camera is used for audio input. For clearer audio quality, headphones are connected to the audio output. The helper side application is running under Windows 7 OS. Since the application runs an image processing function as well as a video transmission, a powerful CPU (Intel i5 CPU) is required. The computer is connected to the LAN.

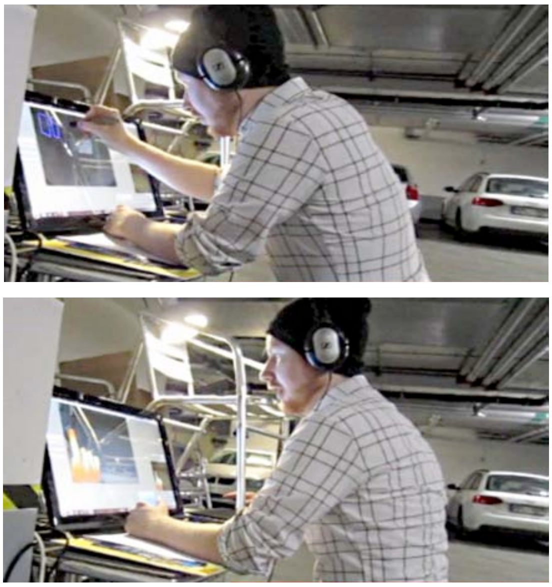

Figure 4: Implementation of the helper side: Sketching (top) and input of deictic gestures where the helper's right hand is captured and overlaid on screen (bottom).

4.2 Driver Side

In a mobile setting, a user must be able to pick up and hold a device easily and conveniently. For our application, the device has to run an image processing in order to guarantee a smooth video transfer. Here, we chose a Samsung Galaxy Tab 10.1 [23]. This product provides multiple network connections (Wi-Fi and 3G), so that the driver can connect at any time as long as a network is available. The device also has two cameras: one in front and on the back; the back camera is used to capture breakdown situations, while the front one is left unused.

5. SOFTWARE DESIGN AND IMPLEMENTATION

To our knowledge, Microsoft does not support the ConferenceXP [24] remote presentation software anymore (as it was used in the CollaBoard system). Therefore, new software was developed for our prototype, which can be used for the remote support system (Fig. 5). This software provides three different information layers at each side (audio, sketching, and image capturing). In our software architecture, the driver will use Layers 1-3, while the helper will use Layers 4-6. Next, we describe implementation and functionality of our software running on both helper and driver sides. We also present and analyze our implementation of user interaction.

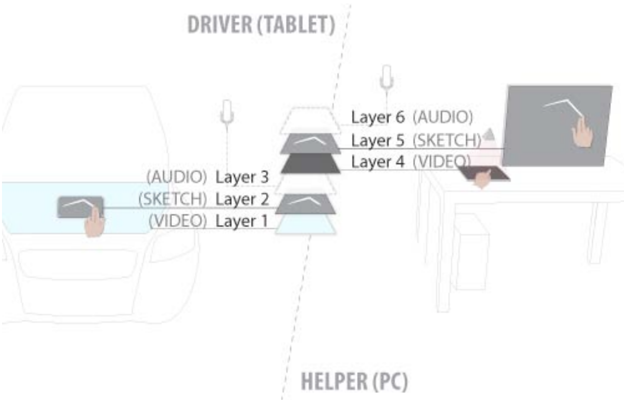

Figure 5: SEMarbeta software architecture

5.1 Audio and Video Connection

In order to realize a smooth video and audio transfer, the UDP [25] protocol is adopted for the transmission. A driver with a technical problem in the car can directly start a VoIP [24[] call to the helper side, while the helper can decide whether to accept the call or not. When accepted, audio is first transmitted to the other side. After the helper and the driver have established the initial communication, they can activate a live-video stream in case the helper thinks the problem is too difficult to be explained by voice-only, or if the driver considers the problem too difficult to describe. In this case, the Samsung Galaxy Tab transmits a video or a still image of the working scene to the helper side.

5.2 Screenshot Functionality

The reason for having the screenshot functionality is obvious. In our use case, the driver has to hold the device on one hand while performing the repairs with the other. Live video would result in a very unsteady image, which is not suitable for sketch or gesture overlay. In our design process, we observed that it is difficult for both the driver and the helper to point at a certain object or to outline an object on live video. Even the slightest movement of the device would disturb the analysis of the problem by the helper and thus hinder the discussion. Consequently, we designed and implemented screenshot functionality in the driver side application, where the Android tablet can temporally freeze the screen, so that the helper and driver can discuss the issues based on still image.

5.3 Sketch Overlay

The sketch overlay we implemented is one of the essential functions in our system. When the helper and the driver collaborate, it is difficult for the helper to explain technical issues by audio only, even when the helper is fully aware of the problem. This can mainly be ascribed to an uneven helper-driver knowledge distribution. Since sketches can help towards shared understanding, we developed the sketch overlay. It has basic painting tools for both the helper and the driver, and allows outlining the issues directly, which will be subsequently transmitted to the other side (Fig. 6).

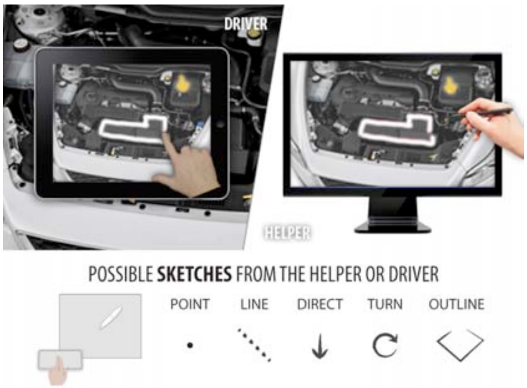

Figure 6: Two-way sketching function (top) and potential sketches for use by the driver or helper (bottom).

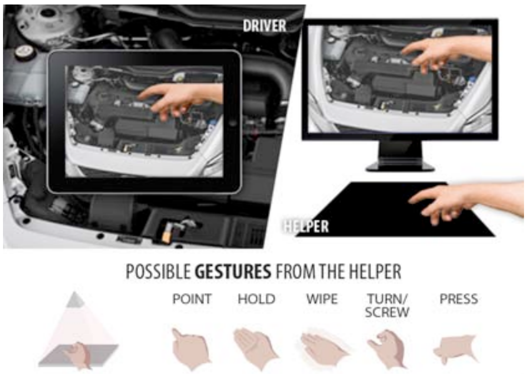

Figure 7: One-way gestures capture and overlay function (top) and potential gestures for use by the helper (bottom).

5.4 Gesture Overlay

While gesture capturing implemented has a distinct layer only on the helper side (Fig. 7), gestures are captured and shown on both sides. When a troubleshooting strategy is hard to explain, using hand gestures may help in clarifying the situation. However, deictic (e.g. "this handle" or "that fuse") gestures are only relevant when shown in relation to the problem, that is, the underlying image. To capture a hand gesture, but not the local background, it must be segmented from the background. We chose an image processing function where the software captures an image of the hand in front of a unique black background. A gray-scale function is then used to transform the whole image into gray-scale image. After this, a mask function converts the gray-scale image into a mask image, depending on a specific threshold value. Pixel gray-scale values below the threshold are set to 0 (black); values above are set to 255 (white). Since the background of the gesture is a black mat, the gesturing hand is white against a black background within the mask image. In a next step, another function named 'processing' compares the original image with the mask image. If the color of pixel is black in the mask image, then the color of that pixel in the original image will become transparent. The final result of this segmentation pipeline is an image of the hand in full color against a transparent background, which will then be overlaid atop the still image from the driver side.

6. SYSTEM EVALUATION

We carried out a user study in order to assess the usability, functionality, and performance of the system. Two tasks and a test environment were designed to come close to a minor car breakdown situation in an authentic environment. The goal of the system evaluation was to find out whether the new functionalities, such as sketch overlay and gesture overlay, could improve communication, collaboration, and thereby problem solving. To assess the SEMarbeta system's capacity, we compared the system with an industry-standard voice-only assistance implementation. Hence, the first condition was the SEMarbeta system with video streaming, sketch overlay, and gesture overlay. The second condition was voice-only working on the same hardware as the SEMarbeta setup, but without sketching and gesture overlays. We hypothesized that task completion time for the two repair tasks would be shorter with SEMarbeta than when using voice-call only. We also hypothesized that the sketch and gesture functions would support helper-driver communication and thereby collaboration.

6.1 SEMarbeta vs. voice-only condition

The SEMarbeta condition was carried out in a wireless local area network. For the study, the driver subject worked on a Volvo V70 car while holding a Samsung Galaxy Tab 10.1 tablet PC. The SEMarbeta application for Android ran on the tablet and allowed the driver to start a video-call, transmit duplex painting information, and receive gestural information. On the helper side, subjects were presented with the SEMarbeta application running on a stationary computer. The hardware consisted of a PC, an Acer 23" touchscreen monitor, and a fixed camera sitting on top to capture helper-side deictic gestures (Fig. 4). In the voice-only condition, we use the same experimental setup as for the SEMarbeta condition, so that this condition had the same quality of speech transmission and portability of devices. Since the VoIP functionality and video streaming functionality were separated in our design, the voice-only condition could be achieved by turning off the video streaming functionality on the tablet at the driver side. Thus, the helper side cannot get any video images.

6.2 Subjects

A total of sixteen subjects (9 males and 7 females) took part in the evaluation. All subjects hold a valid driving license. Thus, we assumed that all subjects have basic knowledge of automotive issues. For each round of the user study, one subject acted as a driver who has to fix some problem at the car, while the other subject acted as a helper who provided support. In total, 8 groups took part in the evaluation. Subjects on the helper side were instructed to act as experts who can provide support at a professional level. In the subject-recruiting process, we found it difficult to find actual experts in automotive repair. Thus, all the helper subjects got to use an operation manual that holds enough information for troubleshooting and fixing the relevant problem, thus enabling them to act like real experts.

6.3 Experimental Design

The tasks in the user study were designed to be similar to a real-life scenario. Since all subjects in the evaluation were not professional in repairing cars, we had to provide basic tasks that could be finished by normal drivers with instructions or manuals. In the official manual of the test car, the manufacturer provided some instructions on basic works, such as changing tires or examining fuses, changing the battery, checking the engine oil, and installing a child seat. However, we eliminated the task of changing wheels because it is dirty and may take more than half an hour to complete. We also did not include changing the battery because of safety issues.

Figure 8: Tasks evaluated: SEMarbeta used for checking oil in the engine compartment (top) and voice-only used for examining rear fuses (bottom).

To counter-balance possible learning effects, the kind of task was changed in each round. This required that the two tasks examined in the user study had to be approximately of the same difficulty level. Hence, the two task chosen were i) checking oil engine compartment and ii) localizing the fuse for the rear audio system (Fig. 8). The two tasks are comparable in terms of complexity and difficulty; they also have three steps in common:

- Step 1: Opening a closed container
- Step 2: Locating the right component
- Step 3: Reading and confirming a value (or: state)

6.4 Procedure

Before starting a task of the user study, a short task description was given to both subjects, and the driver got an additional list with task requirements. Both subjects were told that they have to accomplish one task through voice-only communication, and the other task with our SEMarbeta system. Then, the two subjects received short instructions on how to use the SEMarbeta system. The helper side was introduced to the features of the sketch overlay and the gesture overlay, while the driver side learned the process of starting a video support call. In addition, the user manual of the test car was handed out to the helper in order to help him solve the driver's requests easily.

The driver went to the car in the garage where the helper cannot see or hear him directly without communication devices. At the beginning of each test round, the experimental leader started the voice-only call or the SEMarbeta video streaming, depending on the schedule, and then gave the devices to the subjects. In addition, the supervisors started a video camera to record the processes at each side, and so the video recording could provide the completion time of the test as well. After finishing the first task, the supervisors switched the system to the other mode. After the subjects took a short break, they started the next task. Once both tasks were completed, the supervisors interviewed the subjects and asked questions related to the comparison between the voice-only call and the SEMarbeta system, the user experience of SEMarbeta, and suggestions on the future demands as well.

6.5 Results

In this section, we present both objective subjective results, based on 16 subjects, corresponding to 8 groups, where each group included one driver and one helper.

6.5.1 Objective Measures

Mean task completion time and standard deviation per condition and task are presented in Table 1; all data are given in minutes rounded to two decimals. Since the number of subjects was limited, we could not gather sufficient data to prove a significant difference on the time performance between the two conditions tested. According to mean completion time, SEMarbeta performed better on both tasks. However, as mentioned before, the variance of the means was quite high. It is partly caused by the variability in repair skill across the subjects and the unsteady performance of our system affected by the test environment. While no significant differences can be presented, we still observed interesting trends in the data.

Table 1. Task completion time: Means and standard deviations by condition (column) and task (row). All values given in minutes rounded to two decimals.

	VoIP		SEMarbeta	
	mean	*sd*	*mean*	*sd*
Engine task	10.75	3.53	10.42	3.67
Fuse task	8.22	4.13	6.95	4.21

Two important environment factors that influenced our user study are: network quality and lighting conditions. As the user study took place in a basement garage, the wireless network signal could easily be affected or blocked by other facilities. The garage lighting was not bright enough for the camera of the tablet PC to capture clear images. In one round of the user study, a subject could not see the switch of the engine hood clearly, which prolonged the completion time of that round severely and pulled up the average completion time of SEMarbeta system to some degree. However, this also showed the limitations of our system, which cannot be used at night or under poor lighting conditions.

6.5.2 Subjective Measures

After each group of two subjects, the subjects were interviewed on system usability. Some questions were tailored to driver subjects or helper subjects, and some were common for both kinds of subject. Table 2 presents all the questions and some representative answers.

Table 2. Subjective results: All questions (left) and some representative answers (right).

Driver subject questions:	**Driver subject answers:**
1. Do you think SEMarbeta can provide better quality of help compared with voice call?	• Users may have different backgrounds, different languages and different abilities to fix the car. The system can help people in such situations. • It performed badly in the dark environment, but others are good, which can easily show something to the other side.
2. What do you suggest on the sketch function?	• It is easy to use, but it is difficult to indicate the position if the worker is moving. • It is a little complex to select the eraser function, and only one color is available.
3. What do you suggest on the gesture function?	• I could see the helper's instruction with his finger pointing things out. • I saw the gesture from other side, but it was too big. It could be more helpful if the image was smaller.
4. Would you repair your car by yourself if you have SEMarbeta system?	• I would like to use it if the problem is more complicated, such as electronic repairing. I can solve them according to instructions, rather than read manuals.
5. Which device would you prefer to run this system? Phone or Tablet PC?	• Technically, having a tablet is better for the bigger screen to see the instructions, but I will not buy a tablet just because of having this system in my car.
Helper subject questions:	**Helper subject answers:**
6. Have you helped your friend on fixing things via phone before?	• I used to help them on computer problems by phone call. It was hard to understand what the other side was doing.
7. Do you think SEMarbeta can provide better quality of help compared with voice call?	• Yes, it enables users to see more. Sometimes it is difficult to describe a thing with only oral explanations, since people always don't know how to call it. With this kind of system, I can point it out and verify the other's actions.
8. What do you suggest on the sketch function?	• I could paint something on the screen to show what I mean to the other side. • It should have some transparent properties. Because sometimes the sketches from two sides may overlap.
9. What do you suggest on the gesture function?	• It is easier to point out than paint with fingers or the cursor. However, sometimes I forgot this function. It may take time to learn and accept it. The Interaction way is good and intuitive. • Since it is the first time to use it, people may be confused by the hand showing on screen and consider it belongs to the other side.
10. Would you try to help your friends on fixing things if you have SEMarbeta system?	• If I know how to fix the things, I will use the system to fix the problem. For example, I can fix computer programs, point out the button that my father need to click and so on. Video streaming is a good feedback for instructors while helping others who do not have enough knowledge to fix the problem.

Besides the positive feedback, other interviews raised some problems related to the system performance, the GUI of the SEMarbeta application, and the interaction using gesture input. As expected, subjects expressed their positive attitude towards having SEMarbeta as a remote support solution to take the place of the former telephone solution (100% positive). However, from the observation and the final interviews, it also turned out that subjects using the system the first time had some recognition problems on the additional overlays.

7. DISCUSSION AND FUTURE WORK

Only 25% of driver subjects used the sketch overlay to outline objects on their tablet PC. There are three main reasons for ignoring this function. The most frequently given reason was that it is more comfortable to point to objects directly with the finger instead of outlining things on the screen. Another reason was that the outlining would not work when moving the tablet PC, and thus users wanted to have predefined drag-and-drop shapes to highlight things more easily.

On the helper side, there was a big difference in the usage of the sketch overlay and gesture overlay. 87.5% of the helper subjects (or: helpers) used the sketch overlay to outline things for the driver. In contrast, only 25% of helpers provided gesture instructions to the drivers in a sufficient quality. Subjects expressed two main reasons for this. Three helpers did not remember this function although a short introduction of this functionality was given prior to the test. Two helpers thought the video communication and sketch overlay were sufficient to solve the problem.

Since only a few subjects used the gesture function, we discussed possible reasons for this. Most helpers stated that the sketching function is a more common and natural way of interaction than using gestures. Some of the helpers focused on looking at the touch-screen and sketching on it, rather than moving their right hand to the black pad and providing gestures. This kind of inconsistency of interactive actions may have reduced the cognitive support provided by the system.

It is possible that the gesture capturing techniques used in CollaBoard may perform better than the solution we currently employed. This is mainly because CollaBoard uses pointing at the position where the user also sees the object, while SEMarbeta requires indirect pointing, so that input and output are not collocated, but coupled through the captured image of the finger. This indirect pointing could be avoided if a camera would capture the helper's interaction directly on the screen. However, as we mentioned in previous sections, the required segmentation for this has some limitations, and this kind of interaction may cause fatigue for professional helpers who have to work long hours.

In future work, we plan to improve information presentation, which has the end of reducing cognitive load. Current gesture presentation is still quite limited by the segmentation algorithm and image processing ability of mobile devices. We believe the interaction concept of Kinect can be a way to realize gesture recognition in front of the screen. The concept of capturing body movement to control the movement of a 3D skeleton and then rendering the skeleton to a 3D avatar could bring two benefits to our system. From the user study, we learned that helpers always make gestures during their communication, even though they knew their gestures could not be captured by our system. It is a natural response when helpers meet some difficulties in speaking out the exact word they want to describe the operation. If a device similar to Kinect would be used in our system, the helpers could provide their gestures more naturally during the support process, with no limitations from the system on providing gestures in some fixed area. Moreover, when using a Kinect instead of our image capturing process, gestural input may be less affected by ambient light conditions or background images. Haptic cues, for instance using actuated faders [15], could enable eyes-free control [20].

8. ACKNOWLEDGEMENTS

We are highly grateful for funding, work environment, and advise on user testing offered by the Human Factors group of Semcon Caran AB. This help enabled us to implement our idea in the automotive domain and provided valuable supervision. We also thank members of the CollaBoard team for their support. Morten Fjeld thanks the Swedish Foundation for Strategic Research for enabling him to work at the Innovation Center Virtual Reality (ICVR) at INSPIRE AG, Zurich (SSF grant: SM11-0009).

9. REFERENCES

[1] Alem, L., Tecchia, F. Huang, W.: Remote Tele-assistance System for Maintenance Operators in Mines, In: 11th Underground Coal Operators' Conference, 2011, 171-177.

[2] Anastassova, M. Burkhardt, J.M.: Automotive technicians' training as a community-of-practice: Implications for the design of an augmented reality teaching aid. In Journal of Applied Ergonomics. 2009 Jul; 40(4), 713-21.

[3] Cohen, M., Fernando, O.N.N.: Awareware: Narrowcasting Attributes for Selective Attention, Privacy, and Multipresence. In P. Markopoulos, B. de Ruyter, and W. Mackay, editors, Awareness Systems: Advances in Theory, Methodology and Design, Human- Computer Interaction Series, ch. 11, 259-289, Springer, 2009.

[4] Fussell, S.R., Setlock, L.D., Yang, J., Ou, J., Mauer, E., Kramer. A.D.I. Gestures over video streams to support remote collaboration on physical tasks. Hum.-Comput. Interact. 19-3, 2004, 273-309.

[5] Herskovic, V., Ochoa, S. F., Pino, J. A.: Modeling Groupware for Mobile Collaborative Work. In Proc. CSCW 2009, 384-389.

[6] Kunz, A., Nescher, T., Küchler, M.: CollaBoard: A Novel Interactive Electronic Whiteboard for Remote Collaboration with People on Content. In Proc. CW 2010, 430-437.

[7] Kurata, T., Sakata, N., Kourogi, M., Kuzuoka, H., Billinghurst, M.: Remote collaboration using a shoulder-worn active camera/laser. In Proc. ISWC, 2004, 62–69.

[8] Kuzuoka, H.: Spatial workpace collaboration: A SharedView video support system for remote collaboration capability. In Proc. CHI 92, 1992, 533–540.

[9] Kuzuoka, H., Kosuge, T., Tanaka, M.: GestureCam: a video communication system for sympathetic remote collaboration. In Proc. ACM CSCW '94, 1994, 35-43.

[10] Ladikos, A., Benhimane, S., Appel, M., Navab, N.: Model-free markerless tracking for remote support in unknown environments. In Proc..VISAPP, 2008, 627-630.

[11] O'Hara, K., Black, A. Lipson, M.: Everyday Practices with Mobile Video Telephony. In Proc. CHI 2006, 871-880.

[12] Papadopoulos, C.: Improving Awareness in Mobile CSCW. In IEEE Trans. on Mobile Computing, Vol. 5.10, 1331-1346.

[13] Ran, Y., Wang, Z., Zhu, F.: Trends of mixed reality aided industrial design applications. In Proc. ESEP 11, 3144-3151.

[14] Sakata, N., Kurata, T., Kuzuoka. H.: Visual assist with a laser pointer and wearable display for remote collaboration. In CollabTech, 2006, 66–71.

[15] Shahrokni, A., Jenaro, J., Gustafsson, T., Vinnberg, A., Sandsjö, J., Fjeld, M. (2006): One-dimensional force feedback slider: going from an analogue to a digital platform. In Proc. NordiCHI '06. 453-456.

[16] Spikol, D., Milrad, M., Maldonado, H., Pea, R.: Integrating Co-Design Practices into the Development of Mobile Science Collaboratories. In Proc. 9th IEEE ICALT, 2009, 393-397.

[17] Tang, A., Neustaedter, C., Greenberg, S.: VideoArms: embodiments for mixed presence groupware. In People and Computers XX—Engage, 2007, 85–102.

[18] U.S.News & World Report: Google Glass Unlikely to Be Game Changer in 2013; accessed 14.01.2013 http://www.usnews.com/news/articles/2013/01/02/google-glass-unlikely-to-be-game-changer-in-2013

[19] Villacorta, P. J.: Sensitivity analysis in the scenario method: A multi-objective approach; 11th Int. Conf. on Intelligent Systems Design and Applications (ISDA), 2011, 867-872.

[20] Yi, B., Cao, X., Fjeld, M., Zhao, S. (2012): Exploring user motivations for eyes-free interaction on mobile devices. In Proc. CHI '12, 2789-2792.

[21] http://www.arvika.de/www/e/home/home.htm; accessed 14.01.2013

[22] https://plus.google.com/111626127367496192147#111626127367496192147/posts; accessed 14.01.2013

[23] http://www.samsung.com/ch/consumer/mobile-phone/tablets/tablets/GT-P7500UWDITV; accessed 14.01.2013

[24] http://research.microsoft.com/en-us/projects/conferencexp/; accessed 14.01.2013

[25] http://en.wikipedia.org/wiki/User_Datagram_Protocol; accessed 14.01.2013

[26] http://en.wikipedia.org/wiki/Voice_over_IP; accessed 14.01.2013