Research Article

Designing a Human Machine Interface for Quality Assurance in Car Manufacturing: An Attempt to Address the "Functionality versus User Experience Contradiction" in Professional Production Environments

Nikolaj Borisov ,[1] **Benjamin Weyers** ,[2] **and Annette Kluge** [3]

[1]*University of Duisburg-Essen, 47057 Duisburg, Germany*
[2]*Computer Science Department, RWTH Aachen University, 52074 Aachen, Germany*
[3]*Department of Psychology, Ruhr-University Bochum, 44801 Bochum, Germany*

Correspondence should be addressed to Benjamin Weyers; weyers@vr.rwth-aachen.de

The complexity of nowadays car manufacturing processes increases constantly due to the increasing number of electronic and digital features in cars as well as the shorter life cycle of car designs, which raises the need for faster adaption to new car models. However, the ongoing digitalization of production and working contexts offers the chance to support the worker in production using digital information as well as innovative, interactive, and digital devices. Therefore, in this work we investigate a representative production step in a long-term project together with a German car manufacturer, which is structured into three phases. In the first phase, we investigated the working process empirically and developed a comprehensive and innovative user interface design, which addresses various types of interactive devices. Building up on this, we developed the device score model, which is designed to investigate interactive system and user interface in production context due to ergonomics, UI design, performance, technology acceptance, and user experience. This work was conducted in the second phase of the project, in which we used this model to investigate the subjective suitability of six innovative device setups that implement the user interface design developed in phase one in an experimental setup with 67 participants at two locations in south Germany. The major result showed that the new user interface design run on a smart phone is the most suitable setup for future interactive systems in car manufacturing. In the third and final phase, we investigated the suitability of the two best rated devices for long term use by two workers using the system during a full shift. These two systems were compared with the standard system used. The major conclusion is that smartphones as well as AR glasses show very high potential to increase performance in production if used in a well-designed fashion.

1. Introduction

In the last decades, the rapid development of information technology fosters growing digitization in various contexts, such as in context of home automation [1] or in context of the Internet of Things [2] and Industry 4.0 [3]. In automotive production, the increase of digitization on the one hand offers an increase in productivity but on the other increases the complexity of automotive production environments for the worker [4]. The latter implies new requirements for their working methods and tools. Additionally, workers are confronted with a continuously growing amount of information

necessary for vehicle assembling, testing, and diagnosis emerging from increasing complexity of nowadays cars [4].

For a long time, the main criteria for the design of devices in manufacturing and production including car manufacturing were *functionality* and *reliability* [5]. Meeting industry standards was the relevant criteria neglecting aesthetical aspects such as *design* and *appearance*. But since more than a decade of private smart phone use, the expectations regarding the design and appearance of manufacturing devices have been changed as discussed by Kluge et al. [6, 7]. Even if not formally allowed, many maintenance workers, for instance, in fault diagnosis and repair, use their private phone to search

for solutions and hints and use text messages to chat with colleagues to ask for support and ideas. They perceive this practice to be easier, more intuitive, and faster than using the equipment provided by the manufacturer or the suppliers [7]. In the past years, decision-makers became more open to the idea that an appealing design, user experience, hedonic qualities, intuitive use, and industry standards are not a contradiction but can be united as discussed by Vogel-Heuser [8].

Thus, the purpose of the present study was to design and comparatively evaluate six devices as "Human Machine Interfaces of the Future," which meet the criteria of *functionality*, *ergonomics*, *user acceptance*, and *user experience* in parallel. Therefore, we investigated the suitability of existing mobile devices of various sorts for testing and diagnosis of built-in electronic parts after final vehicle assembly. We identified this diagnosis as a representative digitized process in production and conducted this investigation in close collaboration with a renowned German car manufacturer between 2012 and 2016. In this three-phase project, we conducted three field studies in real production environment. In study 1, we conducted a field observation and interviews as part of Phase 1. In study 2, we conducted a field experiment with 67 workers who tested six devices conducted as part of Phase 2. In Phase 3, we conducted a third user study to investigate two devices with two workers under real working conditions for a longer duration (one full shift).

Our primary focus was to determine which different interaction techniques performed best for this part in car manufacturing and by which interaction devices workers were most satisfied. Based on these empirical results, we designed a software architecture and introduced a new user interface design that supports workers for this task.

The paper is structured as follows: First the production environment and the worker's task are introduced by presenting the status quo of the inspection process, potential errors, and the lack to flexibility in the process as well as the diagnosis system itself including the presentation of the status quo of the user interface designs and devices used. Second, in Section 3, we describe the conceptual development (Phase 1) and practical implementation of the device score model (DSM, Phase 2) which is the basis for the evaluation and selection of the devices for the final test (Phase 3).

2. Background: The Production Environment

In the outlined production environment, after final assembly, the primary goal is to verify the functionality of built-in electronic parts of a produced vehicle. One challenge is that a vehicle manufacturer may offer different vehicle models. In addition, all models may also vary in their individual configuration, which results in a large variety of component combinations. This makes tracking and testing of the present configuration challenging [4]. To manage this complexity, *diagnostic programs* are used to support workers by semi-automating the test procedure, where each component combination of a vehicle model is mapped to an individual test program. Before testing and execution of this program, a worker must connect a mobile device representing the front end of the diagnostic system to the vehicle and start the

test program. Via a unique vehicle number, the test program knows which electronic parts must be tested. Depending on the type of the test as part of the whole procedure, the test program either runs the test fully automatic with no intervention by the worker or prompts the worker with instructions describing the necessary manual intervention. Thus, there are two distinct intervention types.

Intervention Type I: workers apply certain actions to the car's electronic by, e.g., pressing a button. If the car does not react as expected, the worker records the function as not working.

Intervention Type II: workers check functions via visual inspection (e.g., the worker checks whether a lamp is enlightened). Only in case of type II interventions, the worker can make incorrect input by recording a certain function not working if it is or vice versa. Thus, in worst case the vehicle is delivered to the customer with a malfunction.

Both manual interventions include awareness of special noises, status displays, and lights which have been installed in doors, foot rooms, consoles, roofs, etc. Some electronic parts can only be tested on dedicated events. For example, the door entry lighting can only be checked when the respective door is open, which has been done by the worker on request by the test program.

2.1. Status Quo: Worker Errors during the Testing Procedure Routine. As mentioned above, during the testing procedure and the manual intervention, errors can occur. As the task requires accurate and fast responses to approximately 150 tests in 7-10 minutes (as observed in our study), and due to its repetitive character, errors tend to occur in particular in that case if many elements have to be inspected and confirmed in quick succession and if recurring in similar order. In terms of Hollnagels CREAM classification of errors [9], this represents a *decision error*. Decision errors may occur in this particular setting, as each worker inspects more than 50 vehicles per day on the production line and can therefore remember inspection steps after a short training phase. This results in workers knowledge and anticipation at which point in the test sequence the test of certain electronic parts is requested by the test program and where these parts are located in the car. This anticipation leads to a fast response in addition to the highly repetitive work at the production line, so that errors can happen without the worker realizing it. In addition to the problem of marking parts falsely functional or nonfunctional, parts may be marked as being functional although they have not been tested at all.

2.2. Status Quo: Lack of Flexibility. In addition to the problem of routine task execution and potentially resulting (decision) errors, we identified another critical problem in the system design: the test program does not provide any flexibility:

(1) If a worker has noticed an incorrect input, in most cases, the test program does not allow changing the wrong input because of technical reasons.

(2) During automatic test execution without worker intervention, the worker has to wait until the automatic test process is completed. During this time,
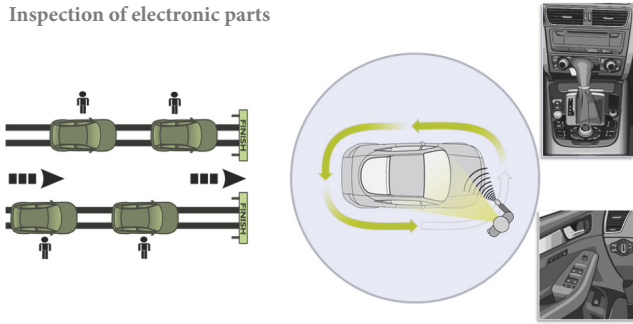
Inspection of electronic parts

FIGURE 1: Inspection of electronic parts in automotive production environments. From left to right: production line, vehicle inspection from outside, vehicle inspection inside.

FIGURE 2: Diagnostic devices present in 2012, text and line based display.

vehicles continue to move on the production line, which results in less time for type I and II interventions.

For instance, as introduced above, an inspection of a fully equipped limousine may take up to 7-10 minutes and may have over 150 type I and II interventions the worker needs to execute. To compensate for these delays, experienced workers go through untested electronic parts and remember the respective states by heart. Thus, they try to bypass the test order at some point. When these electronic components are requested for testing, workers quickly confirm results from their mind. At this point, further incorrect inputs can occur, e.g., if test steps are confirmed without waiting until the instruction is fully displayed, which may result in missing or wrong feedback if a different component is requested.

2.3. Status Quo: The Testing Device. On production line, workers use handheld devices to receive instructions from the test program and to input the outcome of type I and II interventions as feedback to the test system running the test program (see Figure 1). Currently, these systems are provided by the manufacturer that also provides the diagnosis back-end system that executes the test program and implements the communication with the car's electronic. Existing industry standards make these devices heavy and bulky where they are very versatile and have far more functions and buttons than workers need.

In the term of this project, we analyzed two diagnostic devices as seen in Figure 2. The main diagnostic device MFT (multifunction terminal), which runs the diagnostic program, gets attached by the worker to the steering wheel of the inspected car and is connected to the vehicle via cable before starting the testing procedure. Another mobile diagnostic device HT (handheld terminal) is used by the worker and is held in the hand for the whole testing procedure. HT is used to confirm instructions wireless outside of the vehicle. Both diagnostic devices provide a screen for displaying instructions as well as status messages and a keyboard for input. Instructions appear identically in text form on both diagnostic devices and independent of the differing screen sizes. The devices offer the following operations to the worker: OK, NOK (to notify function or nonfunction), go-back (if
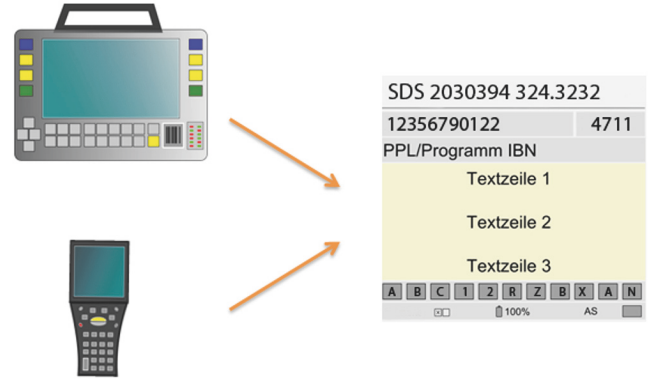
technically possible and permitted), abort test, reprint error page, and scan vehicle barcode.

In summary, based on the described current status, high potential to increase usability and user experience can be identified, which may lead to better devices and user interfaces to increase productivity (by decreasing number of errors) as well as to increase the level of ergonomics if, e.g., considering weight and size of the used devices.

3. Materials, Methods, and Results

To address the previously identified potential of redesigning the interaction concepts and devices, we conducted a research project subdivided into three phases: In Phase 1, we investigated the previously outlined status quo in more detail using empirical methods, on which basis we developed a novel device and interaction design. To consider various different potential solutions, we developed different types of interaction concepts and used various devices. In Phase 2, we developed a rating concept that enables us to rate the developed interfaces under realistic conditions, which were investigated in a user study involving 67 workers in a realistic environment. In the third and final phase, we tested the two best rated devices and design with two workers over a longer period. The next section will present each phase in more detail after presenting the ethical statement for all phases in detail. Each phase will be discussed by first introducing the used and/or developed method followed by the gathered empirical results.

3.1. Ethical Statement. Field studies in an organizational context do not need the approval of the ethical committee of the conducting department at the university. For applied research and for the present study, an approval by the car manufacturers work council was necessary for Phases 2 and 3, which is a special characteristic of the German Industrial Constitution Law that includes the right of codetermination of the companies' work council with respect to field experiments, as well as online and offline surveys or interviews.

Phase 1 was carried out with employees receiving payment over and above standard salary who do not require the approval of the works council. In addition, the participants

of the first phase were employees of the client's department. Additionally, for workers who participated in Phase 1, and for those participating in Phases 2 and 3, the approval of both the works council committee "employee surveys" and the committee "data protection" was required. Both committees approved the investigation. The participants volunteered for participation. All participants in Phases 1, 2, and 3 were informed prior to the investigation about the purpose of the investigation and their right to cancel their participation at any time without giving reasons. Participants additionally signed an informed consent. After the examination, the participants were again informed regarding the purpose of the study and we thanked them for their participation. At the end of the project, results were fed back to the workers, the work council, and the client, and recommendations were concluded for the worker-centered design of the Human Machine Interface of the future.

3.2. Phase 1—Analyzing the Status Quo, Workers' Expectations, and Selecting the User Interface Design and Evaluation Criteria, Methods. In this section we describe phase 1, which is separated into two steps (detailed below). In the first step, we observed and interviewed workers followed by the second step, which focuses on discussions with representatives of the manufacturer.

Step 1 (worker observation and interviews). In order to select the most appropriate interaction techniques for the testing procedure (and especially considering type I and II interventions), we first had to understand which activities are part of a vehicle inspection and which tools are used for each task. For this purpose, we conducted a field study, published in more detail in [10–12], in which we observed and interviewed workers in their activities. The survey method contained 23 questions addressing the following topics:

  (i) Evaluation of the equipment in terms of usefulness and performance

  (ii) Evaluation of the interface design with a focus on functionality, security, and usability

  (iii) Evaluation of multimodality, help, support, adaptability, customization, and usability of the current user interface

  (iv) The use of innovative user interfaces (data glasses, smartphone, hand gestures, etc.)

  (v) Evaluation of the diagnostic process flow

36 production workers participated in the field observations. We conducted the study in two production sites where the working experience (tenure) of the workers varied from a few days up to 40 years [10–12]. As introduced above, the study was approved by the local work council committees "employee surveys" and "data protection" and the volunteering participants signed an informed consent.

In regard of usability standards, such as ISO 9241-11, we were able observe that the used mobile devices were effective in terms of addressing requirements in production but failed in terms of perceived ease of use if subjectively rated by the worker. The worker mentioned that the diagnosis

devices provide various features to be applied in various contexts, which resulted in keyboards with many buttons structured in a standard layout. On the one hand, workers reported that this potentially results in higher error rates forced by wrong button presses. On the other hand, they also show a high degree of adaption to this circumstance that enables them to use the diagnosis system efficiently and properly. An additional outcome was that the potential acceptance of alternative devices that we presented to the workers was rather low. In the interviews, we observed that most workers had problems with imagining the use of these alternative devices in their working environment. One concern mentioned was the wearing of devices on the own body, such as data glasses or headsets, due to hygienic issues. Finally, it turned out that workers are not willing to use new interaction devices and methods despite the fact that there is a convincing increase of ergonomics and usability.

Step 2 (workshop with the representatives from the engineering department of the car manufacturer). For the selection of the relevant criteria that should drive the design and development of a new user interface, we conducted a survey using the Kano model [13] together with the responsible executives of the car manufacturer. The goal of this survey was to work out a common mental model of evaluation criteria on which basis the alternative interaction techniques can be evaluated in our field studies in Phases 2 and 3. From the knowledge gained in the preliminary study conducted in step 1 (see above, as discussed by Borisov et al. [10–12]), 55 relevant evaluation criteria (functional and nonfunctional) were selected, which address 10 topic groups. The subject groups address ISO 9241-110 (principles of dialogue design), ISO 9241-210 (process for designing usable interactive systems), user-centered interface reconfiguration [14], and aspects of emotional design. We also considered the design guidelines of Google Android (2013) and Apple iOS (2013) to be compatible with global design standards.

Results Phase 1: Specification of User Interface Design. Based on the outcome of steps 1 and 2, we created a user interface design detailed below. During the examination of the devices in use (step 1), we saw no homogeneous design for hardware and software. As already described in Section 2, we noticed that the handheld terminal had too many unnecessary keys and the keyboard layout was not designed for diagnostics and was also rather arbitrary, like the MFT device attached to the steering wheel. Important keys were not marked with distinct colour or text. Thus, wrong keys could be pressed during usage. In the worst case, the complete vehicle test could already be aborted when the handheld device was put away. The ABORT key is located very close to the OK key and could be accidentally pressed. The used software user interface was also not designed following usability requirements. If the ABORT key is pressed, the vehicle diagnosis will immediately abort the testing procedure without the possibility to cancel this action by the worker. We found also many screens with cryptic abbreviations in the instruction texts, no helpful pictures, and no feedback about the progress. As a novice, we also needed a lot of training time to understand the various

FIGURE 3: Our prototypes/interaction techniques used in the field experiment (study 2).

abbreviations and to learn the right position of each element to be checked inside and outside of the vehicle. If there are user errors, no hints are shown by the device. In stressful situations, it may happen that workers do not understand why the vehicle testing program does not continue or does not even start. The latter may happen because of connection loss to the diagnostic system or vehicle controls have to be set differently for corresponding transmissions. Beside usability, we also wanted to empirically find out what interaction techniques are most effective for this particular task.

3.3. HMI Mobile Devices Concept. For testing and implementation of the new device and user interface design, we developed a software system that enabled us to freely change devices without the need to adapt or change the existing diagnosis backend. Our client software (named HMI mobile client in the following), which we used for all the peripheral devices, offers a minimal feature set for a vehicle inspection. The HMI mobile client can simulate the diagnostic of any vehicle model and also communicate with a real diagnostic interface used by the car manufacturer. For simulation, we were able to record an entire diagnostic test of any vehicle and use the so-generated log data. Additionally, we use the log data to filter all instructions of each vehicle. The connection to the diagnostic interface was established over WLAN via TCP and the proprietary exchange protocol of the manufacturer used for the proprietary diagnosis systems (MFT and handheld device). The challenge for using our software additionally to the existing testing system was to integrate our client into the existing domain without any modifications to the actual diagnostic system. To achieve this, we had to develop our own interpreter for the proprietary protocol, which filters important data frames of the diagnostic steps from TCP communication in right order and transforms them to our HMI client. Furthermore, interactions from the HMI mobile clients needed to be communicated back into the proprietary diagnostic system.

We used the following mobile devices such that we were able to investigate various interaction techniques identified during our work in steps 1 and 2 (see Figure 3):

(a) *Tablet* and *smartphone* with touch technology and different display sizes (7", 4,8", and 4" inches) to find out how size influences handling. For input, we developed a new user interface and interaction design tailored to the use of touch screens in the addressed diagnosis process. The most used actions during diagnostic are OK and NOK confirmations. Our approach was not to use fixed buttons on the display or physical buttons of the device. Instead, we used swipe gestures that can be adapted as needed and are more ergonomic (see *HMI UI Design* for detailed description). Our design also aimed at enabling the work to perform these swipe gestures regardless of where the worker touches the display.

(b) Using a *Bluetooth headset*, our HMI client sent appropriate test instructions to workers for the respective test step. For type I and II interventions, workers had to confirm instructions via the built-in microphone using special voice commands to confirm an instruction with OK, NOK, or ABORT. To prevent voice input from becoming too monotonous and boring, we have allowed several voice commands for each command. For example, to positively confirm an instruction, worker could use either the word "OK" or the word " Continue." If a worker did not understand an instruction for the first time or he was interrupted, he could also let the HMI client repeat the instruction by using a special voice command: REPEAT. Experienced workers had the possibility to confirm instructions already during the playback without hearing it to the end. At the end of an inspection, the worker received the diagnostic result via voice including further instructions. For speech recognition, we used the Microsoft Speech Recognition Framework (msdn.microsoft.com/en-us/library/office/hh361633(v=office.14).aspx).

(c) The combination of *hand gestures* for inputting the various confirmations and the previously presented Bluetooth *headset* for communicating the instructions to the work offered free hands during diagnosis. This supports safety, free movement, and ergonomics, similar to the headset only scenario described above. We placed one hand and gesture detection sensor inside and one outside of the vehicle. Inside the vehicle we used the Leap Motion Controller (leap-motion.com) and outside the Kinect controller of the Xbox (xbox.com/en-US/Kinect). A white projection panel for instructions has been installed in front of the vehicle so that it can be read from almost any position. A pocket projector, which was also mounted at the front, projected the instructions onto this white panel. To have a comparison to the display medium, we also tested headphones as output device.

(d) *Data glasses* were used as an alternative interaction device. We used data glasses from the company Optinvent (optinvent.com). These data glasses were

Figure 4: Smart watch Pebble 301RD used as input device for data glasses.

Figure 5: Example of a real photograph and the finished image of instruction.

still a prototype at the used time, but in comparison to other devices available on the market (2013/2014) it was the only one with the required projection size and a clear view for both eyes, as the projection has to be always in view. For input, we used hand gestures, a microphone, and a *smartwatch* (Pebble 301RD, see Figure 4). While the instructions were confirmed on the headset via microphone, on the smartwatch we used the built-in buttons. The display on the smartwatch was not yet relevant for our purpose but used to additionally show the type of confirmation mapped to the buttons (see Figure 4).

In summary, we investigated the following device combinations:

 (i) Smartphone

 (ii) Tablet

 (iii) Headset

 (iv) Data glasses with microphone as input device

 (v) Data glasses with hand gestures as input device

 (vi) Projector projection (visual output) with microphone as input device

 (vii) Projector projection (visual output) with hand gestures as input device

For the study conducted in Phase 3, we used

 (i) Smartphone

 (ii) Data glasses with smart watch as input device

 (iii) PDA device introduced by diagnostic system manufacturer (developed in parallel by the car manufacturer)

3.4. HMI Content. In our HMI mobile client, we designed and implemented a completely new dialog design, which is separated into visible (display) and speech/voice (headset) output. The challenge here was to find a compromise between the needs of an experienced and less experienced worker. To address this, we introduced two levels of experience, which we made selectable by the worker during the start of a vehicle diagnosis. Thus, we mapped each individual instruction emerging from the log data to a short text without

abbreviations and with an assisting image when using a visual display. For more experienced workers, we used abbreviations and left out the assisting image. For our experimental setup and for the audio-based output, we let students record all instructions by reading them out. In the future, this might be replaced by nowadays technologies such as those used in navigation systems.

We identified the following requirements for instruction content in our previous investigations conducted in steps 1 and 2:

 (i) Information needs to be presented in a very compact and consistent way (at maximum 2-3 lines on the display)

 (ii) No abbreviations should be shown for novice or user with a low level of experience

 (iii) Important words and abbreviations should be highlighted (e.g., bold typeface) for display output

 (iv) The content should be prepared to be internationalized and adaptable to the location

 (v) The content should be adapted to the HMI device

All used assisting images should fit the inspected vehicle model and therefore need to be created for each vehicle model. In our experiment, we used only one vehicle model and created 118 images. Therefore, we used photographs taken of the vehicle and the components the worker has to inspect (see Figure 5, left) such that a worker is able to recognize quickly where this component is located. Therefore, the images created need to fulfil the following requirements:

 (i) The test object must be immediately recognizable for workers in the image.

 (ii) As for the text, the images should be internationalizable and adaptable to the location.

 (iii) In the image, the current position of the worker has to be clearly visible forced by the perspective from which the photograph was taken.

 (iv) Assisting images for display output should be used that are optimized to be rendered together with text to support the understanding of the instruction and to reduce workload

To visualize the instruction, we enriched the images with pictograms (see Table 1), which represent the needed test step (see Figure 5, right). For the instruction pictograms, we

TABLE 1: Pictograms for the representation of various interventions.

touch/press	release	move/direction	push by foot	illumination	air conditioning
type I	type I	type I/type II	type I	type II	type II

used selective focus, orientation arrows, symbols, colours, and annotations to highlight the relevant components to be tested in the image and to represent the test instruction. Therefore, the pictograms are designed in a way that they can be combined.

Based on these requirements, the raw photographs were processed in three steps:

(1) *Image processing*: In this step, essential elements are extracted from the photograph using Adobe Photoshop (adobe.com/products/photoshop.html) and the images are transferred to grayscale. The result is stored in PNG/PSD format to prevent compression artefacts, which negatively influence vectorization (see next step).

(2) *Vectorization stage 1*: The PNG/PSD images are vectorized using Adobe Illustrator (adobe.com/products/illustrator.html) and the corresponding instruction pictogram is integrated.

(3) The final vector graphic gets *exported* for all HMI mobile devices considering their specific requirements such as screen size, resolution, and supported file format.

3.5. HMI UI Design. Our dialogue design for display-based systems, e.g., smartphone devices, consists of two different views as shown in Figure 6. The first view offers various operations to configure the upcoming test process where the second view is used for the vehicle diagnostic.

In the configuration view, we use clear, simple, and unambiguous pictograms as well as widgets, e.g., to select the level of expertise. Workers can perform the following actions in the configuration view:

(i) Enter his/her personal number in a text field

(ii) Scan a barcode by using the device's camera

(iii) Choose a role (Standard or Expert, see below)

(iv) Change the display language

(v) Display additional information about this screen

(vi) Go to additional settings (e.g., device settings, score-display)

(vii) Go to the next screen

During vehicle diagnostic, we use for type II intervention only simple slide inputs for OK and NOK, but also additional menu settings for advanced functions like abort vehicle diagnostic, call help, etc. (as shown in Figure 6, right). To ensure that workers perceive the readiness for input for type
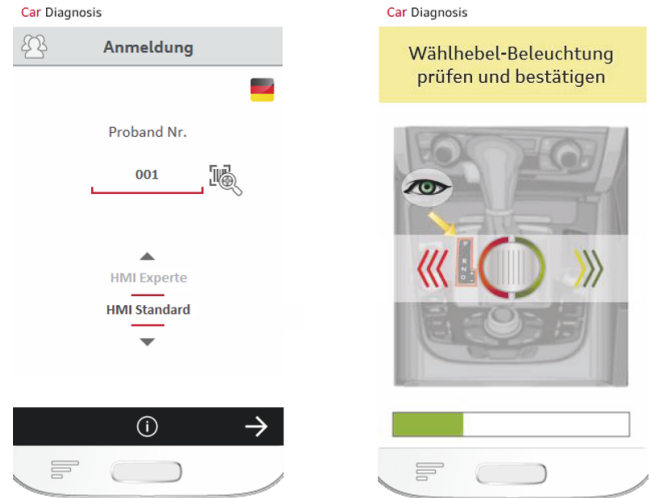
FIGURE 6: Two different interaction views: configuration and vehicle diagnostic.

II interventions, an interaction element is displayed on the screen, which uses animations to indicate the slide gesture. This interaction element disappears automatically after a few seconds to avoid overlapping the displayed elements. Once the interaction element appears, an animation is played. First the animation shows the OK direction and afterwards the animation shows the NOK direction. As soon as the animation completes, or no interaction has taken place, the interaction element disappears. On touching the display, the interaction element will be displayed again at the touched position. This allows workers to use the mobile device independently of how they hold it in their hands from any hand position. We have chosen the slide control also to avoid accidental inputs. Thus, to confirm a type II instruction, a specified distance (depending on the display size) must be reached with the finger on the display without taking it off.

For data glasses, we used only the built-in camera to scan barcodes and smart watch with built-in buttons or microphone to confirm selections.

3.6. HMI UI Assistance. Beside the new organization of the information representation of the diagnostic steps, we developed various assistive elements integrated into our user interface design, which are presented in detail in the next paragraph.

Novice vs. Expert Assistant: The developed HMI design provides two different experience roles for affecting the displayed information in the used mobile devices (see Figure 7). The major difference between the *HMI-Standard* or

HMI - Standard

HMI - Expert

Visual inspection with image
and instruction text

Help text for
current instruction

Visual feedback
after interaction:
▪ OK green
▪ NOK red

Visual inspection without picture
and with original instruction text

Progress bar is displayed independently of the user role

Awareness signals:
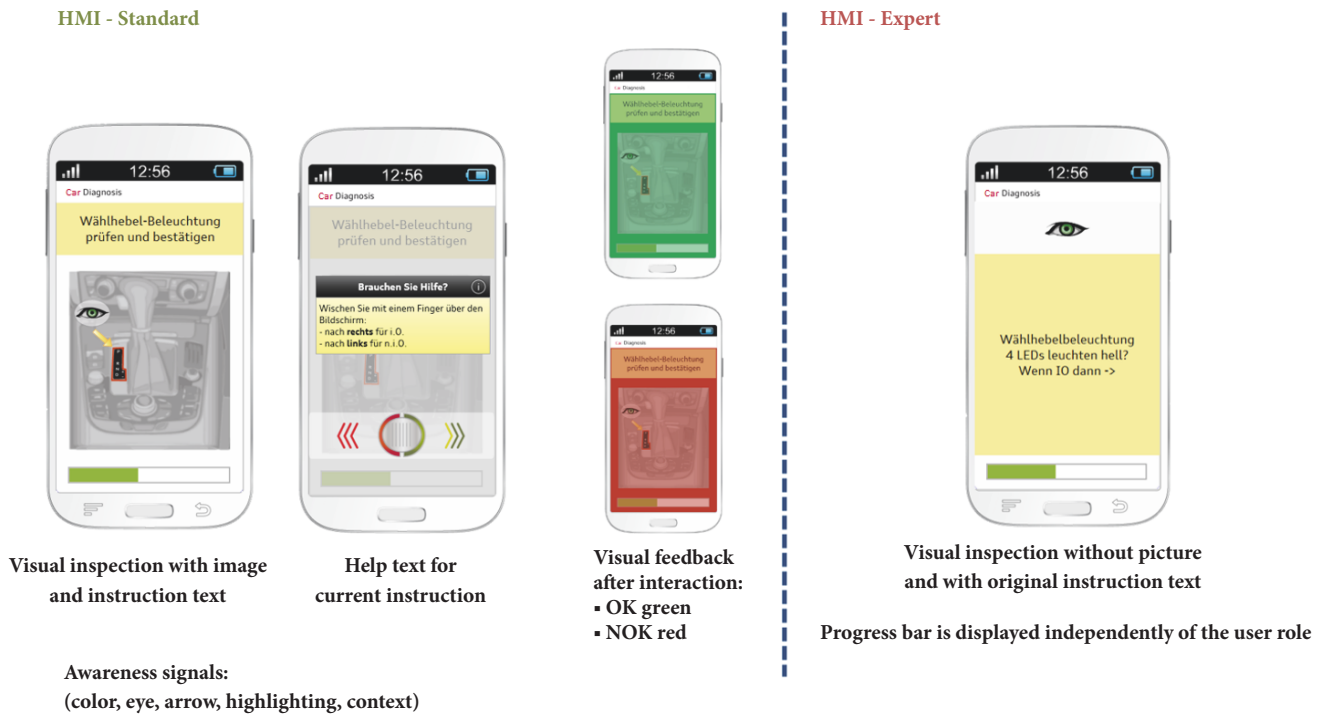(color, eye, arrow, highlighting, context)

FIGURE 7: Two different roles of experience.

novice role and the *HMI-Expert* role is the type of content that is presented to the worker. For experienced workers (HMI-Expert), the instruction text is most important because they know it for a very long time. For unexperienced or less experienced workers (HMI-Standard), image and text are presented where the image takes up the most space of the display. For experts, the only image-based information (shown in Table 2) shown is to distinguish the intervention type of the instruction.

Memory Assistant: Many built-in electronics are already automatically tested in the background by the testing system. During the automatic test time, no inspections of vehicle components are allowed and requested. During this time, the worker has to wait for the automated process to be completed. Nevertheless, in step "Worker Observation and Interviews," we observed that experienced workers would like to perform the visual inspections while the automated tests are running. This makes it necessary to remember all inspection steps and their result in mind to confirm them later on request. With the memory assistant integrated into the UI design (see Figure 8) it is now possible to access a view that lists all known next visual inspection steps (type II interventions). In this list, workers have the possibility to confirm future visual inspections before requested by the diagnostic system. If a saved instruction is requested by the diagnostic system, the HMI client confirms this instruction with the stored result automatically without the need for further action by the worker.

Gamification: Gamification can be defined as "…applying game design elements to non-game contexts. The integration of gamification into the workplace adds a stimulating

TABLE 2: Visual signals of interaction types for expert role.

touch/press	release
type I	type II

and captivating game-like layer to the working experience of employees" [15]. Therefore, we integrated gamification artefacts into the HMI design, which should support the motivation of the worker and thereby increase attention and improve performance during the inspection. We implemented a statistic that shows the worker's own performance by the current number of vehicles inspected (during the current day), the average time of all inspections, and the inspection time of the current inspection. This gives workers an overview about their own performance. After finishing a vehicle inspection, a new dialog appears with the test result as well as a presentation of the statistic (see Figure 9). Furthermore, if the car was recorded NOK, the screen is highlighted red and green otherwise. The red colour is supposed to raise worker's attention.

Awareness Assistant: By pressing buttons inside a vehicle or open and close doors (type I intervention), the worker always perceives this haptic and aural feedback. However, if an instruction is to be confirmed via the mobile device, we also wanted to transmit a recognizable signal to the worker after interaction. Therefore, we added the following feedback implementations to the devices:
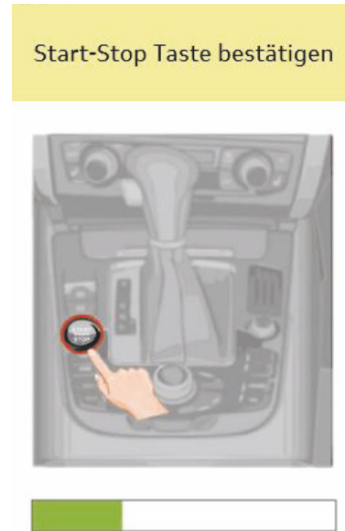
FIGURE 8: Memory assistant for visual inspections.
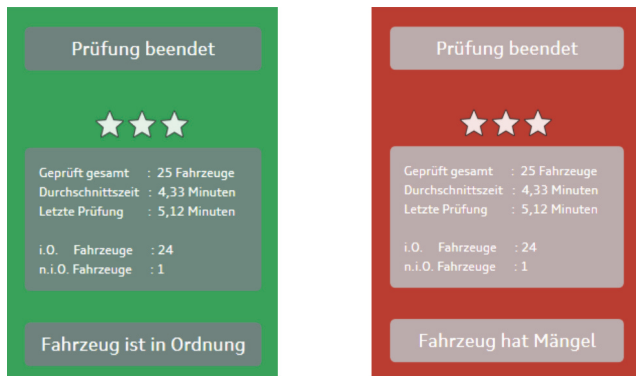

FIGURE 10: Progress indicator during vehicle inspection.

FIGURE 9: Statistic shown after vehicle inspection.

(i) For the audio headset, we used two different short noises depending on the instruction result (OK/NOK).

(ii) For smartphones or mobile devices with display support like data glasses, we overlapped the screen for a short time with a transparent colour: green for OK and red for NOK. In addition, depending on the device, we also used different vibration types for a haptic feedback: one short vibration for OK and two short vibrations for NOK.

Help Assistant: Especially for workers with little experience, but also for experienced workers in stressful situations, sometimes the diagnosis process is system-side interrupted. A hint for the current instruction can be requested immediately or after a time out it will be displayed on a display or transmitted via voice message over headset. This is done either automatically by the HMI client by a timeout or manually either via a voice command with headset or via menu command on smartphone.

Battery Assistant: Before starting a vehicle diagnosis, the battery status of the current HMI mobile device is checked.

If the battery is too low for an inspection, a warning dialog appears and the worker has the option to start the diagnosis and risk that the device will switch off during inspection or take another available device.

Progress Indicator: During diagnosis, the worker always sees the current inspection progress (as shown in Figure 10). For this purpose, the HMI system must first learn dialog sequences of all instructions for the respective vehicle model as has been described above.

3.7. UI Design for Data Glasses. The UI design for the used data glasses needed to be adapted from the mobile device version. While making experiments by using data glasses, we discovered that bright colours limit the worker's view. Furthermore, dark colours mix up the projection with the field of view and the displayed elements became very difficult to separate from their background. Since the projection in the data glasses is horizontally aligned, we have experienced that the use of the entire projection area makes spatial orientation in the close production environment very difficult. In addition, the risk of accidents increases because workers can collide or injure themselves with equipment or other vehicles standing around while moving. To minimize the risk of injury, we decided to use the horizontal projection only partially. The result is shown in Figure 11. The small boxes mark the area in the projection that is transparent for the user. Additionally, the visible projection is provided with a white frame to separate it clearly from the environment. The only difference when using the data glasses is that we found it much more pleasant for the eyes if the text instruction is displayed at the bottom line of the projection and the instruction image is displayed above. The other HMI design elements were identical to other display devices and vary only according to screen resolution.

3.8. Phase 2—Development of the Device Score Model. Based on the workshop results in Phase 1 step 2, we developed the

TABLE 3: Description of sample.

Age group (in years)				
< 20	20-30	31-40	41-50	> 50
-	27	15	11	14
General Work experience at car manufacturer (in years)				
< 1 year	1-5	6-15	> 15	
7	8	19	33	
Education (nominal)				
Vocational training	Technician	supervisor	Engineer	Other
46	1	1	6	13
Experience with devices for test procedure in general (in years)				
< 1	1-5	6-15	> 15	
14	27	19	7	

FIGURE 11: Dialogue design for data glasses.

so-called *device score model* (DSM), which aims at quantifying the subjective quality and suitability of an interactive device in the outlined production context. Therefore, the DSM includes evaluation criteria measuring ergonomics, software design, performance, technology acceptance, and user experience of the newly designed HMI and devices.

Ergonomics and *software design* were rated in terms of the German school grades system as the workers are familiar with that scheme (from 1 = *very good* to 5 = *very bad*); thus, the smaller the grade, the better the result. Criteria for measuring the *ergonomic* aspects were as follows: How many hands are free when working with this device? How heavy is the device? How large is the device (compared to the old one or other future HMIs)? Does the device need to be put away during the testing procedure? Does the device require eye movements? According *to the UI design,* questions were as follows: Is the interaction flexible? How fast is the interaction speed? How accurate is the interaction understood and received?

Performance was measured by analyzing gathered eye tracking data and log files. We calculated the mean of actions and visual tests, the number of errors made (either a wrong OK confirmation or no visual test before OK confirmation), and motivation that we measured as the subjectively rated motivation a worker experienced while working with the device.

A questionnaire was developed to measure *usability and attractiveness* based on the questionnaire measuring user experience (based on [16]). Items addressed hedonic quality (identification, attractive features), hedonic stimulation, pragmatic quality, and attractiveness (altogether 28 item poles). Sample poles of the polarity profile are (7-point profile) good-bad, convenient-inconvenient, easy-difficult, ugly-beautiful, stylish-bad style, predictable- unpredictable, uninspired-creative; confusing-giving overview, repellent-attractive, fractious-manageable, human-technical, and motivating- discouraging.

Additionally, a questionnaire was developed to measure *technology acceptance* based on the Technology Acceptance Model 3 (TAM 3, [17]). Items addressed perceived usefulness (e.g., "The device supports the execution of my task"; "The device enhances my productivity"), user friendliness (e.g., "The device is easy to use"; "The devices does exactly what I want it to do"), level of being self-explaining (e.g., "An introduction how to use the device is not needed"), appeal of the test instructions (e.g., "I had no difficulties understanding the instruction given by the device"), health and hygiene aspects (e.g., "I do not mind wearing the device close to my body"; "I think the device is not problematic concerning hygiene aspects"), input quality (e.g., "Inputs are easy to learn"; "The device can be used intuitively"), first impression (e.g., " I think I will talk to other workers and share my experiences about the device"; "I would like to use this device"), and general use of technology (e.g., "electronic devices make my life easier"; " I know most of the functions of the technical device I own").

We measured *sociodemographic data* (age, working experience; see Table 3. Sample description) and individual motivation while participating in the study (e.g., "While I was executing the testing procedure I forgot that I participate in a study"; "I was concentrated as if I am in a real car testing situation").

With this model, any changes to the UI design or the introduction of new prototypes and devices can be reevaluated and compared with other prototypes or existing devices in use or developed before. Additionally, the model allows each individual requirement to be weighted and evaluated even during the planning or introduction of new diagnostic devices. The DSM evaluation system is dynamic and changes with new adaptations and requirements. Further evaluation

factors can be defined in the model. This allowed us to continuously reevaluate the entire HMI concept. As a result, a test device that has been rated very well in the past may no longer be suitable for practical use or may no longer be state of the art due to new requirements. Thus, the DSM not only enables the evaluation of individual systems but also documents changing requirements over time.

The weightings for individual categories and criteria according to the model are determined by the responsible authority. Since each production line can have different specifications and requirements, it is recommended to duplicate the model for different work areas and define an individual scaling.

For the application of the DSM, we entered all technical, determined, and collected values for each device into a MS

Excel data sheet, which implements the calculation schema as outlined below. In this sheet, each column represents a sub-evaluation-criteria of the DSM. Furthermore, each column is also divided into two subcolumns: value and school grade. The school grade is calculated based on corresponding values of compared devices by a specific formula given as formula (1)-(4), below. Considering the possible best and the possible worst case of each criteria ΔC_{φ} and each evaluated value of a device D depicted as ΔV_D, we consider 2 different cases:

Case 1 (local value comparison between evaluated devices). (a) Higher value (max) is better (e.g., for scale rating) (see (1))

(b) Lower value (min) is better (e.g., for time performance rating) (see (2))

$$C_{\varphi_1}(D_\delta) = 1 + \frac{5 * \left(\max\left(\{V_{D_1}(\varphi_1), \ldots, V_{D_n}(\varphi_1)\}\right) - V_{D_\delta}(\varphi_1)\right)}{\max\left(\{V_{D_1}(\varphi_1), \ldots, V_{D_n}(\varphi_1)\}\right) - \min\left(\{V_{D_1}(\varphi_1), \ldots, V_{D_n}(\varphi_1)\}\right)} \tag{1}$$

$$C_{\varphi_2}(D_\delta) = 1 + \frac{5 * \left(V_{D_\delta}(\varphi_2) - \min\left(\{V_{D_1}(\varphi_2), \ldots, V_{D_n}(\varphi_2)\}\right)\right)}{\max\left(\{V_{D_1}(\varphi_2), \ldots, V_{D_n}(\varphi_2)\}\right) - \min\left(\{V_{D_1}(\varphi_2), \ldots, V_{D_n}(\varphi_2)\}\right)} \tag{2}$$

Case 2 (global vs. local value comparison between evaluated devices based on custom defined possible best value V_B and possible worst value V_W ror each devive D_δ). (a) Higher value (max) is better (e.g., for scale rating) (see (3))

(b) Lower value (min) is better (e.g., for time performance rating) (see (4))

$$C_{\varphi_1}(D_\delta) = 1 + \frac{5 * \left(\max\left(\{V_{D_1}(\varphi_1), \ldots, V_{D_n}(\varphi_1), V_B(\varphi_1), V_W(\varphi_1)\}\right) - V_{D_\delta}(\varphi_1)\right)}{\max\left(\{V_{D_1}(\varphi_1), \ldots, V_{D_n}(\varphi_1), V_B(\varphi_1), V_W(\varphi_1)\}\right) - \min\left(\{V_{D_1}(\varphi_1), \ldots, V_{D_n}(\varphi_1), V_B(\varphi_1), V_W(\varphi_1)\}\right)} \tag{3}$$

$$C_{\varphi_2}(D_\delta) = 1 + \frac{5 * \left(V_{D_\delta}(\varphi_2) - \min\left(\{V_{D_1}(\varphi_2), \ldots, V_{D_n}(\varphi_2), V_B(\varphi_1), V_W(\varphi_1)\}\right)\right)}{\max\left(\{V_{D_1}(\varphi_2), \ldots, V_{D_n}(\varphi_2), V_B(\varphi_1), V_W(\varphi_1)\}\right) - \min\left(\{V_{D_1}(\varphi_1), \ldots, V_{D_n}(\varphi_2), V_B(\varphi_2), V_W(\varphi_2)\}\right)} \tag{4}$$

For all these formulas, the result in DSM will always be a decimal scale of German school grade (1–6): 1, very good (best case), up to 6, insufficient (worst case). Custom defined values for the best and worst case need to be adjusted manually in case of changing requirements. For example, a customer can define by himself a best value of x-milliseconds and a worst value of y-milliseconds to be used as the basis for the grade calculation for the performance criteria for all instructions.

In certain cases, grades must be entered manually in the current model, as these depend on other underlying factors. For example, if the weight of a device (e.g., Smartphone ca. 135g) is very light to hold in the hand, it can be too heavy for data glasses worn on head or nose. The model cannot currently take such classifications into account. However, this can be corrected by further improvements to the model in future work. For each superset criteria (ergonomics, performance, technical acceptance, and user experience), the

developed Excel sheet provides a separate table containing the calculated grades.

3.9. Field Experiment. In order to collect data to be used for the DSM, a field experiment was conducted with 67 workers that worked with the newly designed devices.

3.10. Description of the Sample. To compare and evaluate the six devices as outlined above, 67 workers (47 males) from the car manufacturing partner located in the south of Germany participated in a field experiment. Per day, 5-6 workers were tested individually. The complete field experiment lasted 2.5 weeks and 12 working days.

As presented in Table 3, the sample has several years of work experience and is experienced with several devices used for the testing procedure. We can therefore assume that the sample is able to execute the testing procedure proficiently and to compare the various devices and device combination using the new UI design with the current devices.

TABLE 4: Device score model: results overview (the lower the better).

DEVICE SCORE MODEL		Smartphone	Tablet	Headset	AR-Mic.	AR-Gest.	Proj.-Mic.	Proj.-Gest.
Overall grade		2,1	2,2	3,0	2,6	3,5	3,6	4,2
Criterion	Weight.	Grade	Grade	Grade	Grade	Grade	Grade	Grade
Ergonomics	25%	2,4	3,4	3,2	3,2	3,1	2,9	3,0
Performance	40%	1,7	1,9	1,7	2,9	4,4	5,8	2,8
Technology acceptance	15%	1,7	1,1	3,3	2,0	5,1	4,4	5,0
User Experience	20%	2,7	2,3	3,2	6,0	1,7	2,6	1,8

AR-Mic = augmented reality: data glasses with microphone as input device.
AR-Gest. = augmented reality: data glasses with hand gesture as input device.
Proj.-Mic. = display instructions over projection with microphone as input device.
Proj.-Gest. = display instructions over projection with hand gesture as input.

3.11. Procedure. Each worker used two devices (out of 6) one after the other. For each worker, the field experiment took approx. 90 min. After workers were welcome in the test center of the car manufacturer, each worker was introduced to how to operate device 1 (approx. 10 min) and then the eye tracker was calibrated (5 min). The worker used device 1 for three different testing scenarios consecutively (approx. 10 min) and rated the device according to the criteria in a questionnaire described below and was also interviewed to learn about his/her impressions which were eventually not captured by the questionnaire (10 min). Subsequently, he/she was introduced to device 2 (10 min), the eye tracker was calibrated again (5 min), and after three additional test scenarios (10 min) the worker filled in the questionnaire, this time referring to device 2 (10 min).

3.12. Results of Field Experiment. All results are shown in Table 4 (overview) and Table 5 (all criteria). They are based on weightings as defined by the automobile manufacturer's managers. The greatest attention is paid to performance at 40%, followed by ergonomics at 25%, user experience at 20%, and technical acceptance at 15%.

In terms of ergonomics, smartphones were rated higher. The weightings were given to us as follows: safety of operation with 35%, hand-free work with 30%, weight and size each of the device with 15%, and interaction distance, which can be defined as number of single user actions needed to generate a specific input to the system (e.g., confirm OK or NOK), with only 5%. Workers had only one hand free when wearing a mobile device, but the ergonomic size and weight, interaction safety, and short latency times for instructions made them preferable in comparison to the other devices and device combination. Interaction distance, on the other hand, was negatively affected. For example, the device always had to be taken off when the seatbelt was put on and always requires eye contact for every instruction. Both potentially cost more time compared to the other solutions.

The tablet received the worst marks in ergonomics for size and weight compared to the other devices. For devices with gestures and voice input, the biggest problem was the reliability. Both interaction techniques work based on recognition algorithms that do not work fully accurately. For example, gestures and speech input had to be repeated several times in case of background noise to confirm an instruction. Another

disadvantage of both interaction techniques is if overlapping communication between worker-device and worker-operator occurs, gesture or talking can be interpreted as input and thus causes unwanted actions.

Additionally, we measured user errors and the worker's motivation as performance indicators for both intervention types I and II. User errors were counted if the worker did not look correctly or not at all in the right direction during an instruction. Another error was counted if the worker confirmed an instruction incorrectly. We measured both with using an eye tracking system. In this device category (including data glasses with voice input), the smartphone also received the best marking. Additionally, we observed shortest interaction distance if the data glasses were used. This resulted in efficient tests of intervention type I. However, in the case of type II instructions, time was lost due to lower reliability of speech input.

In terms of technical acceptance, the tablet scored best in the technical acceptance survey. The smartphone came in second place. The result may vary, since the workers who worked with the smartphone were less technically sophisticated (technology skills: grade 2.5) than the workers who worked with the tablet (technology skills: grade 1.2). This would at least explain the differences in the subcriteria *self-explaining* and *representation*. We found very little acceptance with the headset as well as with the projection devices using the two interaction techniques of speech input and hand gestures. All criteria were equally weighted by those responsible for technical acceptance. Only the single criterion *technology skills* was weighted at 0%, as it is not important to the responsible persons for this activity.

Like technical acceptance, user experience was evaluated as a subjective opinion of each device via a questionnaire. All criteria were equally weighted by those responsible persons. Except for the data glasses with hand gestures, almost all workers had a positive experience with the alternative interaction techniques.

In Table 6, we present examples of workers qualitative responses and reaction to the six devices, which reflects the personal experience and confirms the quantitative ratings.

3.13. Phase 3—Comparison between the Two Prototypes and the Diagnostic Device in Use. In Phase 2, we applied the innovative HMI standard mode for the new mobile devices

TABLE 5: Device score model: for all criteria (lower is better).

Ergonomics		Smartphone	Tablet	Headset	AR-Mic.	AR-Gest.	Proj.-Mic.	Proj.-Gest.
Overall grade		2,4	3,4	3,0	3,2	3,2	3,1	2,9
Criterion	Weight.	Grade	Grade	Grade	Grade	Grade	Grade	Grade
hands-free	30%	3,5	3,5	1,0	1,0	1,0	1,0	1,0
weight	15%	2,2	6,0	1,2	1,8	1,6	1,2	1,0
size	15%	3,0	6,0	2,0	3,0	3,0	2,0	1,0
interaction distance	5%	3,5	3,5	2,3	2,3	2,3	3,5	3,5
interaction safety	35%	1,0	1,0	6,0	6,0	6,0	6,0	6,0
Performance		Smartphone	Tablet	Headset	AR-Mic.	AR-Gest.	Proj.-Mic.	Proj.-Gest.
Overall grade		1,7	1,9	2,8	1,7	3,5	4,4	5,8
Criterion	Weight.	Grade	Grade	Grade	Grade	Grade	Grade	Grade
Type I	20%	2,1	2,6	3,2	1,0	3,8	3,5	6,0
Type II	30%	1,3	1,4	1,0	2,6	5,2	4,4	6,0
User error rate	40%	1,5	1,9	3,7	1,6	1,0	4,5	6,0
Motivation	10%	3,4	1,9	3,5	1,0	2,2	6,0	4,0
Technical acceptance		Smartphone	Tablet	Headset	AR-Mic.	AR-Gest.	Proj.-Mic.	Proj.-Gest.
Overall grade		1,7	1,1	5,0	3,3	2,0	5,1	4,4
Criterion	Weight.	Grade	Grade	Grade	Grade	Grade	Grade	Grade
Useful	14%	1,0	1,1	3,1	2,4	2,2	4,2	6,0
User friendly	14%	1,1	1,0	4,4	1,8	2,0	6,0	5,0
Self-explaining	14%	2,6	1,0	4,7	3,1	3,1	4,0	6,0
Representation	14%	2,9	1,0	6,0	3,2	1,8	5,0	3,8
Health care	14%	1,0	1,1	5,6	5,0	2,1	6,0	2,2
Input quality	14%	2,0	1,0	5,3	2,2	2,0	6,0	5,7
First impression	14%	1,4	1,6	6,0	5,3	1,0	4,4	2,4
Technology skills	0%	2,5	1,2	4,8	3,3	4,8	6,0	1,0
User Experience		Smartphone	Tablet	Headset	AR-Mic.	AR-Gest.	Proj.-Mic.	Proj.-Gest.
Overall grade		2,7	2,3	1,8	3,2	6,0	1,7	2,6
Criterion	Weight.	Grade	Grade	Grade	Grade	Grade	Grade	Grade
HQ identity	25%	1,6	1,5	1,0	5,0	6,0	1,2	3,4
HQ stimulation	25%	3,7	3,2	2,7	1,0	6,0	3,4	2,5
Pragmatic quality	25%	3,5	3,5	2,7	4,2	6,0	1,0	2,6
Attractivity	25%	2,0	1,1	1,0	2,4	6,0	1,4	1,9

and evaluated their performance in the study considering the DSM. Until now, workers have not been able to choose between HMI standard and HMI expert mode. Everyone had to use the HMI standard mode with image preview. Our primary objective in the study was to identify the two best vehicle diagnostic devices in cooperation with workers. After evaluation, the two mobile devices were selected via DSM: smartphone and data glasses with microphone as input source.

In Phase 3, the goal of a final study was to evaluate a direct comparison with the new diagnostic device (PDA) already in use on the production line. The new diagnostic device was developed in parallel to our project directly by the diagnostic system manufacturer starting in 2012 and introduced to us in 2016. It was much smaller than the devices used before (see discussion above) and instead of a keyboard it also had a touch screen and only 5 fixed keys for vehicle diagnostics.

However, the instructions were also displayed as pure texts and with abbreviations on 3-4 lines.

Beside the new *reference device*, the HMI clients were upgraded for the final study with the following functionality: choice between HMI standard and HMI expert mode, visual feedback (OK: green, NOK: red), user statistics, instruction notice, battery notification, and memory assistant from the HMI dialog design as described above. The DSM has also been upgraded by these new criteria. In Phase 2 study, we observed that microphone input had very low reliability, which had a negative effect on efficiency. Cross-talk and noise on the production line led to many misinterpretations of speech input and commands were either misinterpreted by the speech system or not recognized at all. Thus, we decided to use a smartwatch as input device with the data glasses instead of a microphone. After the review of available devices on the market in 2014, we selected a smartwatch manufactured by

TABLE 6: The worker's statements about the devices.

Device	Pro	Contra
Smartphone	The best Super, easy-to-handle Much more comfortable than the existing device	The put away has disturbed Strap is missing for fixation
Tablet	Tablet was cool Handling very good Touch is super	Too bulky Handicapped at work
Headset	Favourable, because both hands are free Perfect, great, pleasantly light	Annoying instructions Can't talk to people
AR – Mic.	Very good Easy to understand Hands-free	Too heavy Partially restricted view
AR – Gest.	Good idea Must be improved Hands-free	Safety of interaction must be provided
Proj. – Gest.	Gesture control is ok Satisfied	Not practical Difficulty in handling
Proj. – Mic.	OK Better because of the visualization	Better be confirmed by yourself No personal dialogues

Pebble because it had a very long battery life (more than 3 days) and four physical keys for input. The use of these physical keys solved the problem of a too small touch area offered by other smartwatches.

Field Experiments. In order to collect data to be used for the DSM, a small field experiment was conducted with only two workers, who did not know the HMI mobile devices.

Description of the Sample. To compare and evaluate the three devices as outlined above, two workers (male) from the vehicle manufacturing partner located in the south of Germany participated in this field experiment. Per day, the two workers were tested individually. Both workers already have very good knowledge about vehicle diagnostics. The complete field experiment lasted 3 days and about 100 minutes per person and per device.

Procedure. Each worker used all three devices. For each worker, the field experiment took approx. 300 min. On the first day, workers were welcome in the test center of the vehicle manufacturer and got an introduction to the test procedure (approx. 10 min). We presented the smartphone to the worker, which gave him a first impression of this device. Subsequently, we gave the worker time to get used to the reference device PDA (approx. 15 min) and asked him to rate it according to the criteria in the questionnaire described above. Additionally, we interviewed him to learn about his impressions of the device (10 min). Subsequently, he was introduced to the smartphone (10 min) and conducted three test runs with it (each 15 min), the first in HMI Expert-Mode, the second in HMI Standard mode, and the third in HMI

Standard mode in combination with the memory assistant. After the last run, the worker filled out the questionnaire again (10 min).

On the second day, we conducted a performance test using the smartphone and the reference device PDA. First, the worker tested a vehicle with the PDA (15 min) and rated this device with a performance questionnaire (5 min). After a break of one hour, the worker did the same test with the smartphone, but this time in three runs (30 min). For the performance test, the worker was able to choose the HMI mode (Standard/Expert) and the use of memory assistant as he wished. After the last run, the worker filled out the performance questionnaire (5 min). Finally, the worker was introduced to the data glasses such that he was able to gain a first impression of the device (10 min). Then he ran the same vehicle diagnostic (35 min). This was followed by an interview (10 min).

On the third day, we conducted a performance test for the data glasses (15 min), which again was followed by filling out a questionnaire (10 min).

Results of Field Experiment. All results are shown in Table 7 (overview) and Table 8 (all criteria). For the evaluation of the results, only the first performance measurement of the vehicle test was evaluated for each person, because only in this case the persons tested all test steps correctly (repetition error). Due to health concerns, data glasses could only be used once a day.

The results show that the smartphone is the better diagnostic device. The smartphone met all criteria regarding ergonomics, acceptance, and performance and stands out significantly from the two other devices. The used data glasses

TABLE 7: Device score model: results overview (the lower the better).

DEVICE SCORE MODEL		Smartphone	Data glasses	Current PDA
Overall grade		1,7	3,9	2,8
Criterion	Weight.	Grade	Grade	Grade
Ergonomics	13%	1,7	3,6	2,3
Software-Design	13%	1,3	1,3	4,4
Performance	40%	1,7	5,5	2,2
Technology acceptance	15%	1,7	3,2	2,8
User Experience	20%	1,8	3,1	3,5

had the worst performance in this field study and were not well accepted by the workers. Regarding ergonomics, smartphone performs better and with a special wristband the rating can be increased even further. Data glasses fail due to health grade, weight, and size. Due to the introduced software design, the new HMI design concept offers a lot of potential for improvement in usability. The wiping on the smartphone still needs some adaptation. Using the data glasses, a complete test took about 1 minute longer compared to the other devices. The workers accepted the smartphone immediately; also they were able to use the reference device after the first briefing. Overall, the smartphone was very positively rated due to user experience.

The workers had filled out the questionnaire twice for each device. We wanted to know if there is any change of opinion about the device after the performance test. After the second run, the workers found the smartphone and PDA even more interesting and stimulating in terms of increased HQ stimulation. However, there were no visible changes for the data glasses.

In Table 9, we present examples of workers qualitative responses and reaction to the three devices which reflects the personal experience and confirms the quantitative ratings.

4. Discussion

The purpose of the present study was to design and comparatively evaluate 7+2 devices as "Human Machine Interfaces of the Future," which meet the criteria of *functionality*, *ergonomics*, *user acceptance*, and *user experience* in parallel, which represents an attempt to address the "Functionality versus User Experience Contradiction" visible in various professional production environments. Taking all results together, we can draw the following conclusions.

All workers saw the *smartphone* as generally suitable and as a nice new approach for vehicle diagnostics. The new graphical representation of the instructions was experienced positively and the touch gestures including vibration are at least as good as tactile buttons. The workers perceived the smartphone as intuitive and all were able to work with it without extensive training. The introduced memory assistant increased productivity and led to higher satisfaction compared to the new reference device investigated in the final study.

In the Phase 2 study with 67 workers, the *tablet* performed as good as the smartphone. The workers perceived it as very intuitive to use and as a useful tool in contrast to

currently used devices (MFT and handheld). Only the groups of older and very experienced workers did not see any positive improvement for their work using the tablet.

The use of the *Bluetooth audio headset* was mainly supported by the workers. We noticed that workers' concentration was even higher by using the headset compared to other mobile devices. Due to the higher motivation and concentration of the headset, the number of errors was significantly lower. However, listening to an instruction via headset may reduce the working speed but, simultaneously, the use of natural language was perceived as self-explaining and the learning effort was reduced. The first impression of the headset differs between men and women and is more positive among men. One potential reason for this effect might be that women were more concerned about health issues raised by using these devices than men.

Data glasses (AR) combined with a microphone require a longer learning period for the workers. This device has been perceived as attractive and useful, and workers were more motivated by using this device during the field study. Workers who had experience with it had fewer problems with the visualization of the instructions through the data glasses and saw no health or hygiene issues. We noticed that the younger group of workers reported a more positive experience. The performance and error rate were independent of gender, age, and work experience. However, well experienced workers were rather less convinced by the combination of data glasses with audio input.

For data glasses (AR) combined with hand gesture, we made similar experiences compared to data glasses with audio input. Workers who could handle gestures better also saw more potential in this device for their work. The results showed us that portable equipment must be designed very ergonomically in terms of hardware, software, and hygiene to satisfy workers in a production context. Overall, data glasses and gesture control are very new concepts to use in this environment and require an additional adaptation phase. This device was rated equally positive and/or negative by both genders and regardless of the device experience. Particularly, younger workers were more comfortable with gesture handling than older workers.

The Phase 3 study confirmed the results from the Phase 2 study regarding doubts on health issues if it comes to the use of data glasses (AR). The continuous use of data glasses was experienced as stressful and very dependent on the ambient light situation. Due to change of eye focus between

TABLE 8: Device score model: for all criteria (lower is better).

Software design		Smartphone	Data glasses	Current PDA
Overall grade		1,3	1,3	4,4
Criterion	Weight.	Grade	Grade	Grade
Corporate design	20%	1,0	1,0	2,7
User-adaption	20%	1,0	1,0	4,8
HMI assistant	20%	1,4	1,4	4,5
Dialog design	20%	2,3	2,3	4,8
Multimedia content	20%	1,0	1,0	5,4
Ergonomics		Smartphone	Data glasses	Current PDA
Overall grade		1,7	3,6	2,3
Criterion	Weight.	Grade	Grade	Grade
hands-free	20%	3,5	1,0	3,5
weight	10%	1,3	6,0	3,1
size	15%	1,0	6,0	3,5
interaction distance	5%	3,5	3,5	4,8
interaction safety	25%	1,0	1,0	1,0
Health	25%	1,0	6,0	1,0
Performance		Smartphone	Data glasses	Current PDA
Overall grade		1,7	5,5	2,2
Criterion	Weight.	Grade	Grade	Grade
Type I	33%	1,0	6,0	3,9
Type II	50%	1,9	6,0	1,0
User error rate	0%	-	-	-
Motivation	17%	2,3	3,0	2,5
Technical acceptance		Smartphone	Data glasses	Current PDA
Overall grade		1,7	3,2	2,8
Criterion	Weight.	Grade	Grade	Grade
Useful	14%	1,0	3,3	3,1
User friendly	14%	1,6	2,5	3,1
Self-explaining	14%	2,9	3,3	3,5
Representation	14%	1,6	3,1	2,5
Health care	14%	2,0	4,3	2,7
Input quality	14%	1,6	2,7	2,0
First impression	14%	1,2	3,5	2,5
Technology skills	0%	1,6	1,0	1,2
User Experience		Smartphone	Data glasses	Current PDA
Overall grade		1,8	3,1	3,5
Criterion	Weight.	Grade	Grade	Grade
HQ identity	17%	1,6	2,9	2,9
HQ stimulation	17%	2,6	2,6	4,4
Pragmatic quality	17%	1,3	3,0	3,0
Attractivity	17%	1,3	3,5	3,8
Ergonomics	17%	2,0	3,0	3,6
Worker grade	17%	1,8	3,8	3,3

the projection surface and the test element, eye pain and headache occurred. However, workers who had experienced the data glasses positively tend to see no health or hygiene issues.

The projector device with the projection panel for visual output and microphone as input device required a longer training period similar to the use of data glasses. Workers felt unreliable with this interaction device as communication with colleagues becomes more difficult. The more motivated the workers were, the more efficiently they completed visual inspections. Workers who rated projection and microphone as user-friendly and useful also saw the health benefits. There

TABLE 9: The worker's statements about the devices.

Device	Pro	Contra
Smartphone	Wiping is great The idea with the list assistant is great (++) Wiping is state of the art Statistics are very good Device feels faster than current device	Hand strap missing. Gives an additional feeling of safety at work Statistics do not increase motivation You have to adapt to the wiping first
Data glasses		Very stressful due to poor projection quality Burning eyes Uncomfortable and heavy to wear it The noseband is very pressing Problematic projection with white-coloured vehicles Hardware is not ready-made to use Not suitable for people with glasses
Current PDA		Sideways keys are difficult to operate The device could be a little lighter Slow screen refresh, hangs and is bulky

was no impact rating negative and/or positive by both sexes or age group.

With projection device and with hand gesture as input investigated in the Phase 2 study, we noticed that workers who had experienced the device positively were more skilled in technology and much more motivated to use this device. Workers who rated this device as easy-to-use also saw the health benefits of this immaterial interaction. Similar to projection with microphone, age and gender had no negative or positive influence. However, the novel immaterial interaction concept initially causes confusion among participants. In general, the performance was negatively affected by hand gesture interaction.

In summary, the project has shown that, for future use, smartphones and data glasses seem to be the better diagnostic devices for the presented production step in the automotive industry. The devices using projection with interaction by hand gestures performed most poorly. We assume that the result is strongly influenced by the use and habit of the old technology. Above all, the workers were highly enthusiastic concerning the combination of hand gestures and projection from a distance. This has something to do with the closeness and distance of interaction with the diagnostic system, operating accuracy, and the feedback when an instruction is confirmed. We noticed that some workers tried to confirm the instructions to the direction of the projection panel by hand gesture and not in needed field area of the installed hand gesture device inside and outside of the vehicle. This caused the problem that instructions were not confirmed at all and we had to remind the workers to perform the hand gestures inside the sensor area. The new technique must first be understood and learned. In our experiment, the workers only got 15 minutes to become familiar with the new interaction technology before the start of experiment. The results are all based on a very short usage time to confirm that the established diagnostic devices in this project are absolutely applicable for the production line. For a recommendation,

we would first have to make a long-term study in which the workers at least complete one complete using the new equipment, best several days in a row.

5. Conclusions

This paper summarizes a project of 3.5 years during which we performed various field studies to develop and evaluate the application of an innovative Human Machine Interface (HMI) in automotive manufacturing together with a German car manufacturer. In this work, we focused on a representative production step in which the electronic system of a produced vehicle gets inspected by the worker. The project was conducted in three phases. In the first phase, we investigated the working process empirically and developed a comprehensive and innovative user interface design, which addresses various types of interactive devices. Building up on this, we developed the device score model (DSM), which is designed to investigate interactive system and user interfaces in production context due to ergonomics, UI design, performance, technology acceptance, and user experience. This work was conducted in the second phase of the project. We used this model to investigate the subjective suitability of six innovative device setups that implement the user interface design developed in Phase 1. The experimental setup was executed with 67 participants at two locations in south Germany. The major result showed that the new user interface design run on a smart phone is the most suitable setup for future interactive systems in car manufacturing for the selected production step. In the third and final phase, we investigated the suitability of the two best rated devices resulting from the Phase 2 study over a longer term. These two systems were compared with the standard system used at this time. The outcome showed that light and ergonomic smartphones have a very high potential to be used in the future of production. Nevertheless, if considering addressing health and hygienic issues of data glasses, also AR technology

showed high potential to be used in future production environments. The study has clearly shown that functionality already provided by current devices is sufficient for quality testing, but does not take the user into account at all. Instead, the Human Machine Interfaces for quality assurance in car production environments presented here have noticeably increased the satisfaction among workers.
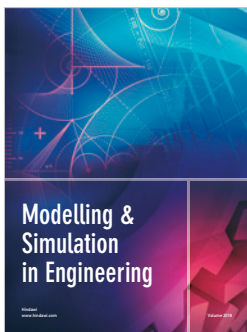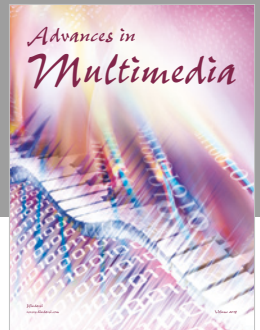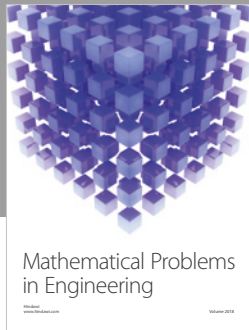
Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] M. Friedewald, O. Da Costa, Y. Punie, P. Alahuhta, and S. Heinonen, "Perspectives of ambient intelligence in the home environment," *Telematics and Informatics*, vol. 22, no. 3, pp. 221–238, 2005.

[2] l. Kong, Z. Gu, and T. Li, "Biomechanical optimization of implant diameter and length for imeddiate loading: a nonlinear finite element analysis," *The International Journal of Prosthodontics*, vol. 22, no. 6, pp. 607–615, 2009.

[3] H. Lasi, P. Fettke, H.-G. Kemper, T. Feld, and M. Hoffmann, "Industry 4.0," *Business & Information Systems Engineering*, vol. 6, no. 4, pp. 239–242, 2014.

[4] H. Schleich, J. Schaffer, and L. F. Scavarda, "Managing complexity in automotive production," in *Proceedings of the International Conference on Production Research*, vol. 100, Valparaiso Chile, 2007.

[5] J. Backman and H. Helaakoski, "Mobile technology to support maintenance efficiency Mobile maintenance in heavy industry," in *Proceedings of the 2011 9th IEEE International Conference on Industrial Informatics*, pp. 328–333, 2011.

[6] A. Kluge and V. Hagemann, "Neue und Soziale Medien in der Fertigung und der Personalentwicklung – am Beispiel von Industrie 4.0 und E-Coaching aus Sicht der AOW Psychologie," *Zeitschrift Wirtschaftspsychologie*, pp. 5–21, 2016, https://www.wiso-net.de/document/BLIS_7922D1C7D809CEAE5E73179D2927D839.

[7] A. Kluge and A. Termer, "Human-centered design (HCD) of a fault-finding application for mobile devices and its impact on the reduction of time in fault diagnosis in the manufacturing industry," *Applied Ergonomics*, vol. 59, pp. 170–181, 2017.

[8] B. Vogel-Heuser, "Herausforderungen aus Sicht der IT und der Automatisierungstechnik," in *Industrie 4.0 in der Produktion, Automatisierung und Logistik. Anwendung. Technologie. Migration*, T. Bauernhansl, M. ten Hompel, and B. Vogel-Heuser, Eds., pp. 37–48, Springer, Wiesbaden, Germany, 2014.

[9] E. Hollnagel, *Cognitive Reliability and Error Analysis Method (CREAM)*, Elsevier, Oxford, UK, 1998.

[10] N. Borisov, A. Kluge, W. Luther, and B. Weyers, *Integrating Production Workers into User Interface Design for Diagnosis Devices in Automotive Production Environments: Field Experiences and Lessons Learned*, C. Stephanidis, Ed., Springer, Berlin, Heidelberg, Germany, 2013.

[11] N. Borisov, B. Weyers, A. Kluge, and W. Luther, *Kontextabhängige Modellierung der Mensch-Maschine-Interaktion in Mehrbenutzersystemen - eine explorative Untersuchung*, Kognitive Systeme 1, 2013.

[12] N. Borisov, A. Kluge, W. Luther, and B. Weyers, "User interface design for test and diagnosis software in automotive production environments," in *Ubiquitous Computing and Ambient Intelligence. Personalisation and User Adapted Services, LNCS XX*, pp. 372–375, 2014.

[13] N. Kano, "Attractive quality and must-be quality," *The Journal of the Japanese Society for Quality Control*, pp. 39–48, 1984.

[14] B. Weyers, D. Burkolter, A. Kluge, and W. Luther, "User-centered Interface Reconfiguration for Error Reduction in Human-Computer Interaction," in *Proceedings of The Third International Conference on Advances in Human-Oriented and Personalized Mechanisms, Technologies, and Services (CENTRIC)*, R. Hariprakash, E. Egyed-Zsigmond, and M. Hoffmann, Eds., pp. 50–55, IEEE Verlag, 2010.

[15] M. Liu, Y. Huang, and D. Zhang, "Gamification's impact on manufacturing: Enhancing job motivation, satisfaction and operational performance with smartphone-based gamified job design," *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 28, no. 1, pp. 38–51, 2018.

[16] M. Hassenzahl, M. Burmester, and F. Koller, "AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität," in *Mensch & Computer*, pp. 187–196, Vieweg+ Teubner Verlag, 2003.

[17] V. Venkatesh and H. Bala, "Technology acceptance model 3 and a research agenda on interventions," *Decision Sciences*, vol. 39, no. 2, pp. 273–315, 2008.

Applied Computational Intelligence and Soft Computing

The Scientific World Journal

Mathematical Problems in Engineering

Journal of Engineering

Advances in Multimedia

Modelling & Simulation in Engineering

Advances in Artificial Intelligence

International Journal of Reconfigurable Computing

Advances in Fuzzy Systems

Hindawi

Submit your manuscripts at

www.hindawi.com

Scientific Programming

Advances in Human-Computer Interaction

International Journal of Engineering Mathematics

Advances in Civil Engineering

Journal of Computer Networks and Communications

Journal of Robotics

International Journal of Computer Games Technology

International Journal of Biomedical Imaging

Journal of Electrical and Computer Engineering

Computational Intelligence and Neuroscience