Multi-Modal Visualization of Working Instructions for Assembly Operations

Josef Wolfartsberger, Michael Heiml, Georg Schwarz, Sabrina Egger

Abstract—Growing individualization and higher numbers of variants in industrial assembly products raise the complexity of manufacturing processes. Technical assistance systems considering both procedural and human factors allow for an increase in product quality and a decrease in required learning times by supporting workers with precise working instructions. Due to varying needs of workers, the presentation of working instructions leads to several challenges. This paper presents an approach for a multi-modal visualization application to support assembly work of complex parts. Our approach is integrated within an interconnected assistance system network and supports the presentation of cloud-streamed textual instructions, images, videos, 3D animations and audio files along with multi-modal user interaction, customizable UI, multi-platform support (e.g. tablet-PC, TV screen, smartphone or Augmented Reality devices), automated text translation and speech synthesis. The worker benefits from more accessible and up-to-date instructions presented in an easy-to-read way.

Keywords—Assembly, assistive technologies, augmented reality, manufacturing, visualization.

I. Introduction

THE topic of Industry 4.0 describes a trend away from mass production towards mass customization. Products are getting increasingly complex and personalized. As a result, pressure onto assembly workers has been seen to grow. Human-centered assistance systems support workers in completing their tasks by providing step-by-step working instructions [1]. In order to provide optimal support, however, these systems need to be context-sensitive regarding the current state of the product and what kind of information is currently needed by the worker. Combined with the idea of mass customization and the high variant diversity mentioned above this implies high requirements regarding visualization of assembly information. Previous publications addressing the issue of assembly assistance systems describe user studies with simplified LEGO use cases [2], [3], while disregarding the fact that this does not cover the whole complexity of real-world applications. The research gap we aim to close in this paper is how step-by-step assembly instructions of real-world products can be visualized with respect to high variant diversity in products using different devices for visualization in one application.

II. Related Work

Creating working instructions to assist operators in handling complex tasks has been topic of many research projects.

Josef Wolfartsberger, Michael Heiml, Georg Schwarz and Sabrina Egger are with University of Applied Sciences Upper Austria, Campus Steyr, Austria (corresponding author, e-mail: josef.wolfartsberger@fh-ooe.at).

For our research, we particularly looked into studies which compared the effectiveness of different technologies with focus on visualization.

The authors in [2], [4] compare the effectiveness of head-mounted displays (HMDs), tablet instructions, and baseline paper instructions to in-situ projected instructions using an abstract Lego Duplo assembly task. The results show that assembling parts is significantly faster using in-situ projection and locating positions is significantly slower using HMDs. Further, participants make less errors and have less perceived cognitive load using in-situ instructions compared to HMD. In this context, Funk et al. [5] propose a standardized experiment design named General Assembly Task Model (GTAM). The GTAM is introduced as an uniform experiment design for benchmarks to evaluate assembly instructions. Blattgerste et al. [3] compared Augmented Reality (AR)-based in-situ assistance against conventional pictorial instructions using a smartphone, Microsoft HoloLens and smart glasses as well as paper-based instructions in an experiment with a LEGO Duplo assembly task. To make the different devices comparable they used the GTAM approach based on [5]. In their study the authors evaluated the time of completion, errors, cognitive load and the qualitative results. The results show that participants solved the task fastest using the paper instructions, but made less errors with AR assistance on the Microsoft HoloLens smart glasses than with any other system.

The works above use LEGO Duplo for their experiments and use cases. Radowski et al. [6] state critically that LEGO is not the best reference for simulating assembly tasks because the low complexity of assembling LEGO cannot be compared with industrial assembly tasks. They evaluated the effectivity of AR-based working instructions based on different complexity levels. In an experiment, operators had to assemble an industrial engine, first with a paper-based instruction and second with an AR-based instruction. They identified that AR-based working instructions have only an additional benefit for complex assembly steps. Simple tasks can be done just as well with paper-based instructions. In an earlier work from Tang et al. [7] the authors present a use case where they compared paper based instructions with display-based and AR-based instructions. Their results indicate that overlaying 3D instructions on the actual work pieces reduced the error rate for an assembly task by 82 %. Measurement of mental effort indicated decreased mental effort in the AR condition, suggesting some of the mental calculation of the assembly task is offloaded to the system.

Manual assembly tasks are handled by workers with different preconditions and experience levels. Every operator

has various needs to an assistance system. With regard to visualization techniques, Funk et al. [8] presented a study where different kinds of visualizations for impaired workers were compared. The operators had to assemble five parts of an industrial machine. They were supported with pictorial instructions, video and contour guidance. According to their study, the contour instructions provided the best support considering error rate and time of completion. Li et al. [9] tested different kinds of assistance for assembly tasks for operators with different qualification levels. The authors measured time of completion, product quality and the workers' satisfaction. The assembly instruction were tested in a LEGO-based scenario. The results show that in comparison to text-, picture- and video-based assistance, AR support was best for product quality but not for assembly time.

Guidelines on how to present information in working instructions are proposed in [10]-[12]. Specifically, the idea of supporting assembly processes cognitively for different qualification levels is discussed in [11].

Fässberg et al. [13] specify an information demand matrix. The authors state that an unbalance in information demanded and offered will result in either dissatisfaction or assembly errors. Providing information when there is no need may frustrate the operator. Withholding important information when there is need for more detailed instruction may result in errors. The need for information depends on the respective qualification level of operators. In line with this suggestion, it is assumed that different levels of qualification (LoQ) require different detail of presentation.

It is also important how to present the information. Agrawala et al. [14] present design guidelines for designing step-by-step working instructions for assembly tasks. They defined two primary tasks for designing good working instructions: The planing and the presentation task. The planning task refers to how the object should be assembled and the presentation task describes how the information should be presented. Soederberg et al. [10] conducted a study to evaluate how assembly instructions can improve the operators' performance. The experiment showed that the presentation of information is highly important for good working instructions. They suggest five steps based on their experiment: The information layout has to be consistent. Clear and realistic pictures should be used. Third, differences between similar objects should be highlighted. In addition, complex complex steps should be visually enhanced and finally, unnecessary information has to be eliminated. Mattson et al. [11], [12] identified guidelines for information presentation in an assembly task. They found that too much information can affect the performance of the operator in a bad way. There are also differences in the operators' preferences on how information is delivered.

The results and guidelines of the papers mentioned form the basis for the implementation of a human-centered assistance system, which combines various technologies and visualization styles with focus on multi-modal support for complex assembly tasks. The following section describes the system's architecture.

III. A MULTI-MODAL ASSISTANCE SYSTEM

A. Human-Centered Workplace 4 Industry

The goal of this work is to create a human-centered, multi-modal assistance system equipped with sensors, actors and interaction/output devices to provide optimal support for complex assembly tasks. A first prototype of this system is completed, which is capable of automatically providing support for operators in manual industrial workplaces. Every operator has individual qualification profiles that can change over time. The working environment automatically adapts to the user's qualification and offers different means of assistance.
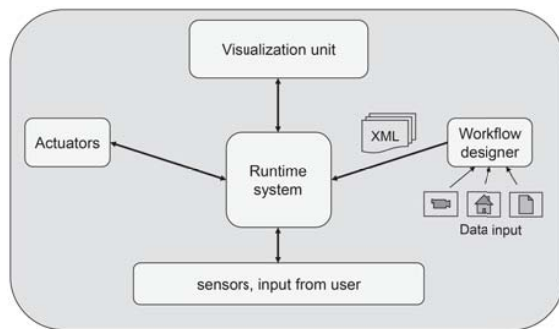

Fig. 1 System architecture

The runtime system (as depicted in Fig. 1) is the central component that acts as interface for all neighboring components. These are input signals coming either from sensors or users, actuators that communicate with the runtime system, a workflow designer (to model workflows in a graphical user interface) and a visualization component. In this paper we focus on the complexity that emerges for the visualization of working instructions from industrial product complexity combined with qualification profiles. Within this project, Lindorfer et al. [15] suggest a decision-model-based approach for the design of collaborative assembly tasks. This approach utilizes an abstract model consisting of five main elements in order to allow the creation of flexible assembly workflows.

The five core elements of the ADAPT-model are *Action, Asset, Decision, Relationship* and *Property* (see Fig. 2).
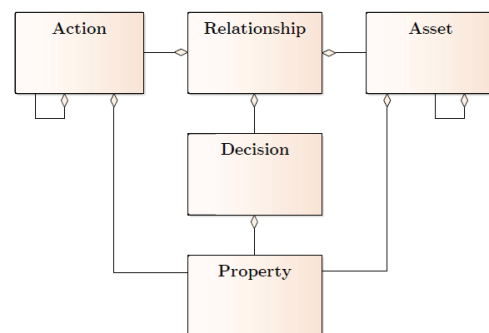

Fig. 2 Five elements of the ADAPT-modelling approach [15]

Actions are intended to model any kind of task to be performed within the process. *Assets* represent information associated with actions. *Decisions* model conditional workflows. A *Relationship* links the former three elements, enabling the representation of various connections between them (e.g. preceding and succeeding relationships of Actions). *Properties* describe Actions, Assets and Decisions in detail. Actions and Assets can be aggregations of themselves. Concretization of the model is done through instantiation of the main elements, e.g. *Reach* in the meta-model is an instantiation of *Action*, *Reach1* in the model is an instantiation of *Reach*. Regarding flexible working instructions, this approach may be used to model variant diversity in products and changing levels of qualification. Instantiating an *Asset* as *Part* and subsequently as *Part n* allows for flexible visualization and representation in working instructions. The addition of an *Level of Qualification*-asset provides a way to differentiate between beginners and experienced workers and preset instructions accordingly. The underlying assumption of this paper is that the input data in industrial applications is complex and implies challenges for the generation and visualization of working instructions.

B. Managing Variant Diversity and Operator Qualification in Working Instructions

In this paper, we choose to introduce three different levels of operator qualification: *beginner, advanced, expert*. As operators gain experience, level of detail in working instructions will continue to decrease. Currently, operators set their experience level manually at the beginning of their work task. In a future step, the operator will be able to dynamically change her experience level during executing the task or the experience level will automatically reconfigure based on image recognition, mistake indicators and other observable parameters. Hu et al. [16] conducted a comprehensive study on assembly system design with focus on product variety. They found that the larger the worker's choice complexity, the more time the operator needs to make the selection. So in order to reduce the time required for understanding instructions it is necessary to keep the amount of information low. In our step-by-step instructions complexity increases due to part variety and different levels of qualification. We address this issue by using the ADAPT-approach depicted in Fig. 2, resulting in flexible step-by-step working instructions.

To manage visualization of products with variant diversity with respect to different qualification levels, we define a set of attributes that contain all necessary information. The visualization component of the assistance system then reads the assigned values of each attribute as needed.

Fig. 3 includes the idea of part variety and different levels of qualification according to the rules of the ADAPT-approach. *Action n* represent previously defined steps in the workflow (e.g. step 1, step 2, step 3). Each Action possesses one Title-Property and a Successor-Relationship towards the following Action. In this case, an Action is required to have a Part-Asset and a LoQ-Asset. To keep the illustration simple, only two instances of the Part- and LoQ-Assets are shown.

The Part-Asset requires a Property for its name, number and task description. Having an image path attached is possible, but not mandatory. The LoQ-Asset requires a Property for its level only. The suggested set of Assets and Properties represents a minimum requirement for the model and may be extended as needed. An example of such a modification is the addition of video-based instructions[17], [9] or the inclusion of several parts or tools necessary to complete an assembly step.

C. Demonstration Use Case: Assembly of a PC

The theoretical model shown in Fig. 3 is used to design a flexible assembly instruction for a PC for industrial usage. To fit the requirements of a multi-modal visualization of working instructions several conditions are given. When loading the working instruction, the worker will be required to choose between different levels of qualification (i.e. beginner, advanced, expert) to determine the level of detail to be presented. It also supports visualization of text, images and animated instructions. In accordance with the aim of multi-modal visualization, instructions are suited for varying output devices. In this case we decide to use a regular TV screen, two mobile devices (tablet, smartphone) and a head-mounted display (Microsoft HoloLens). The instruction data consists of (1) material base data (part names and numbers), (2) corresponding instruction texts and media (image, video) depicting the assembly process, (3) a 3D-CAD model of the PC and its sub-components when relevant, (4) animation logic where required and (5) allocation for levels of qualification for each step. The sequence of assembly operation was derived from information received from the original manufacturer. We defined a sequence of 32 assembly steps for the assembly operation to be complete. The total number of steps depends on the qualification level of the worker, resulting in the following number of steps: 32 steps for beginners, 24 steps for advanced workers, 13 steps for experts. The working instructions were first prepared in a table and then mapped according to the model in Section III-B. The defined properties are as follows.

- instance of Action (e.g. reach, grip, move)
- title of the current action
- count of the current action
- part name
- part amount
- tools required for the assembly
- type of content like picture, video, audio and animation
- level of qualification

Table I shows a reduced example for the step-by-step instructions to be used in the system. Based on this instructions we developed a multi-modal visualization which is described in the following section.

IV. Multi-Modal Visualization

Mattson et al. [11], [12] states that the needs of operators regarding visualization of instructions strongly vary. The presentation of information has to be configurable to suit the operators' preferences. A multi-modal approach for
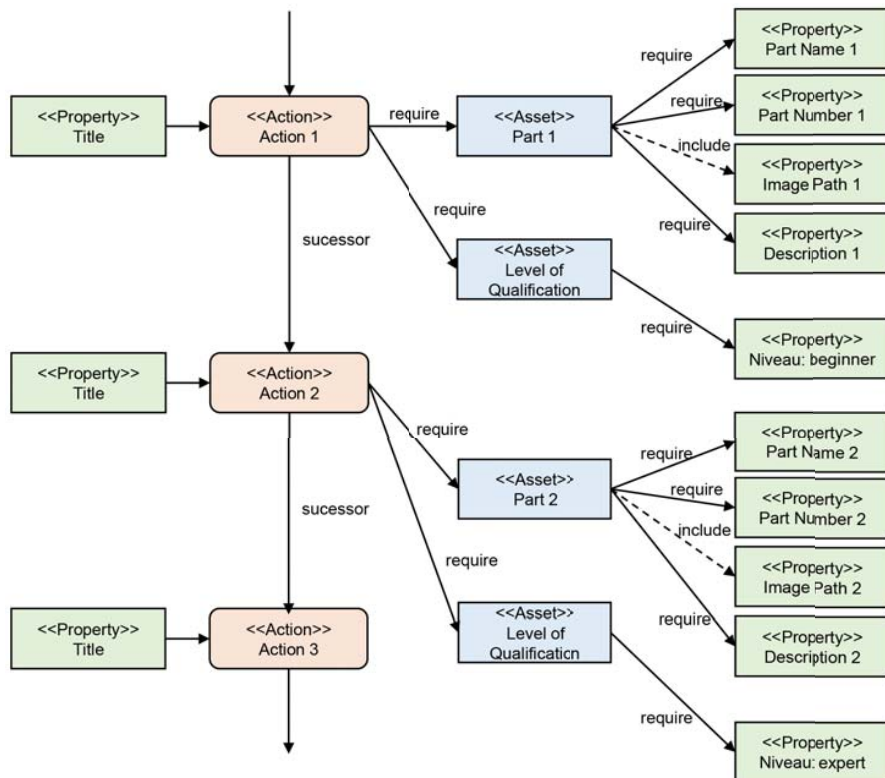
Fig. 3 Inclusion of part variety and qualification levels with the ADAPT-approach

TABLE I
A (REDUCED) EXAMPLE FOR STEP-BY-STEP INSTRUCTIONS TO BE USED FOR THE VISUALIZATION

Type	Level	Title	Description	Name	Qt.	Step	Animation	Image
place	Beginner	Place PC	Place the PC in front of you on the desk.	PC	1	0	anim_placePC	
pick	Advanced	Take heatpipe	Take the heatpipe from the bin.	heatpipe	1	1		img_pick-bin
place	Expert	Insert heatpipe	Insert heatpipe. CAUTION: heatpipe has to touch the cooling block.	heatpipe	1	2	anim_ insertpipe	
pick	Advanced	Take cover 1	Take cover 1 from the bin.	cover_1	1	3		img_pickbin
place	Expert	Insert cover 1	Insert cover 1 as shown on the screen.	cover_1	1	4		img_cover1
pick	Advanced	Take screws	Take the screws from the bin.	screw_tx10	4	5		img_pickbin

Fig. 4 The assembly table equipped with a tablet-PC, a projector, a screen below the working surface, light-indicators, sensors and a collaborative robot

visualizations has been subject to several studies (see Section II on Related Work).

Fig. 4 shows the working table. Information can be displayed on tablet, smartphone, projector, touch-panel and AR-glasses like *Microsoft HoloLens*.

Some examples for assistance devices we integrated so far include:

- a pick-by-light system for bins to support picking tasks using LED light strips and *Arduino*
- a *LEAP Motion* gesture controller for hands-free interaction
- a standing mat to recognize the workers presence
- a NFC controller to read serial numbers of parts and identify workers
- a bone-conduction headset to support voice control and audio output while still being able to sense the acoustic surroundings
- a traffic light system to visualize safety status using *Philips Hue Light*
- an ultrasound distance metering tool to recognize grasps in the bins
- an USB video camera to recognize parts ans the worker's

position

- and a button glove for controlling the visualization.

The user interface is responsive, thus the layout adapts automatically to the screen resolution and aspect ratio of the current device. Text is automatically scaled to a readable size depending on the devices' physical screen size. As shown in Fig. 5 the user can customize the level of information detail presented (e.g. hiding the needed tool since the worker already knows that from past experiences). Additionally, the worker is able to interact with the 3D object (zoom, rotate, animate, change between single isometric and 4-side-view) and can instantly provide feedback via email to the current working step, including image annotations and text comments.
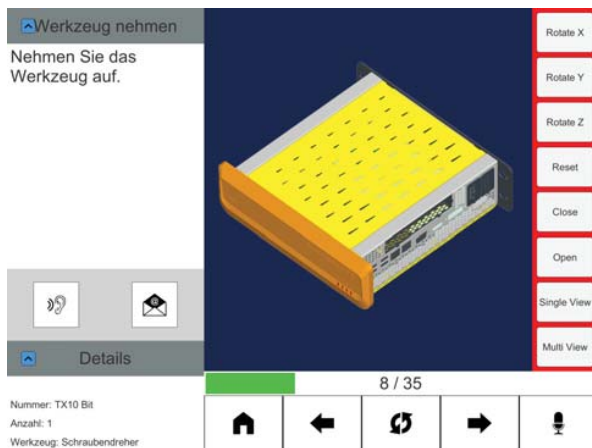

Fig. 5 The layout showing text instructions, a 3D model and a set of options

All the content is streamed from a cloud service and can therefore easily be hot-swapped and updated. Video streaming is possible in standard definition, HD, UHD and 4k. Videos can be zoomed in and rotated for a better view. Pictures up to 4096x4096 in default formats are supported and can also be rotated, zoomed and panned. In addition, pictures can contain annotations like exclamation marks or highlights. The visualization also supports streaming of 3D models (including animations) with single isometric or 4-side-view. The objects can be transformed by touch input and animations can be replayed.

At the beginning of a task, the user chooses her experience level manually. As shown in Fig. 6 the operator can choose the language, qualification level and the device she wants to use for the assistance.

The visualization app is connected to a cloud-based machine translation service that also incorporates text-to-speech synthesis. Therefore, instructions only need to be written once in their default language (currently in German) and are translated and spoken out automatically without a noticeable delay. Instructions can be translated and speech-synthesized into 70 languages. In a future publication, we will evaluate the usefulness of automatic translations for working instructions.

The worker also has the possibility to use voice control rather than touch or gesture input for hands-free interaction. A voice recognition service translates natural voice into input request, like "next", "back" or "repeat". This is especially

Fig. 6 The start screen: Selecting language, level of experience and the type of device (e.g. screen, tablet, smart-glass

useful in situations where both hands are occupied or device interaction is uncomfortable.

All instruction information is transmitted via MQTT protocol, where the visualization app acts as client and the runtime as server. With this architecture, it is possible to use several visualization instances together - e.g. using a tablet or big TV screen only for showing images and videos while controlling the system with a separate smartphone or bluetooth-buttons placed on the work surface. Additional sensors (e.g. a distance sensor in the bins for picking tasks) also communicate via MQTT with each other and the runtime environment. Therefore, information can not only be parsed and used by the central runtime, but also by other clients that can respond immediately by bypassing the central runtime (e.g. pausing the assembly process and showing a notification on the visualization devices when the worker leaves the workplace).

V. FUTURE WORK

The visualization is in a state where all the described features can be demonstrated and evaluated. Our future work focuses on completing and upgrading the following features:

1) Workflow Modeler: The workflow modeler is a tool which allows to easily design, layout and define the work instructions and stream the output of this tool via the described MQTT server to the visualization clients. It works like a visual editor (WYSIWYG-principle) to create each step of the work instruction. Future work includes to upgrade the WYSIWYG functionality to enable a quick definition of workflows without any knowledge of programming.

2) Device Manager: The device manager enables communication between local assistance tools. Currently, the device manager sends and receives MQTT commands via network and controls assistance devices via Bluetooth, NFC, WIFI, USB and Serial Port. The device manager keeps track of all user interactions and forwards this information via the MQTT content server to visualization devices. Additional devices will be added in future, such as collaborative robots and tracking sensors.

3) Analytics: The visualization shows progress insights and other status reports to the workers. In a next step, workers could e.g. assess their own performance in anonymous comparison to others. Workers could also have better insights on their assignments and overall progress, e.g. on how much objects have to be assembled on the particular day by the whole group of people assigned.

4) Ergonomics and Accessibility: Workers could be reminded of planned breaks and posture issues by sensing parameters like movement, heart-rate or time of day. Also, day/night light cycles for eye-friendly operation could be considered automatically via colored ambient light systems. Environment parameters like heat, moisture, sound level or even more dangerous parameters like radiation or ultraviolet light could be analyzed and forwarded to the worker. In terms of accessibility the environment could adapt to the needs of visually or hearing impaired workers.

5) Remote Assistance: Workers could share the camera image of their Augmented Reality HMD with a geographically distant expert, who can expand the field of view with digital annotations. The field of remote support (also known as "Maintenance 4.0") is another promising use cas in industrial context.

Altogether, the presented assistive system provides a promising testbed for future evaluations. We plan to investigate the effectiveness of different assistive technologies (such as novel Mixed Reality supported HMDs) according to their strengths and weakness for completing manual assembly tasks. In addition, we plan to evaluate the quality and usefulness of automatic translations for working instructions. We think that technology-driven assistive systems will significantly influence the future of manual assembly tasks for complex and diversified products. What it still needs are high quality evalutations of technologies and methods in realistic working environments on real-world objects.

References

[1] S. Büttner, H. Mucha, M. Funk, T. Kosch, M. Aehnelt, S. Robert, and C. Röcker, "The design space of augmented and virtual reality applications for assistive environments in manufacturing: A visual approach," in *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*, ser. PETRA '17. New York, NY, USA: ACM, 2017, pp. 433–440.

[2] M. Funk, T. Kosch, and A. Schmidt, "Interactive worker assistance: comparing the effects of in-situ projection, head-mounted displays, tablet, and paper instructions," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp16*. ACM Press, 2016.

[3] J. Blattgerste, B. Strenge, P. Renner, T. Pfeiffer, and K. Essig, "Comparing conventional and augmented reality instructions for manual assembly tasks," in *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments - PETRA17*. ACM Press, 2017.

[4] S. Büttner, M. Funk, O. Sand, and C. Röcker, "Using head-mounted displays and in-situ projection for assistive systems: A comparison," in *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, ser. PETRA '16. New York, NY, USA: ACM, 2016, pp. 44:1–44:8.

[5] M. Funk, T. Kosch, S. W. Greenwald, and A. Schmidt, "A benchmark for interactive augmented reality instructions for assembly tasks," in *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia - MUM15*. ACM Press, 2015.

[6] R. Radkowski, J. Herrema, and J. Oliver, "Augmented reality-based manual assembly support with visual features for different degrees of difficulty," *International Journal of Human-Computer Interaction*, vol. 31, no. 5, pp. 337–349, jan 2015.

[7] A. Tang, C. Owen, F. Biocca, and W. Mou, "Comparative effectiveness of augmented reality in object assembly," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '03. New York, NY, USA: ACM, 2003, pp. 73–80.

[8] M. Funk, A. Bchler, L. Bchler, O. Korn, C. Krieger, T. Heidenreich, and A. Schmidt, "Comparing projected in-situ feedback at the manual assembly workplace with impaired workers," in *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments - PETRA15*. ACM Press, 2015.

[9] D. Li, S. Mattsson, Å. Fast-Berglund, and M. Åkerman, "Testing operator support tools for a global production strategy," *Procedia CIRP*, vol. 44, pp. 120–125, 2016.

[10] M. Sderberg, Johansson, "Development of simple guidelines to improve assembly instructions and operator performance," in *Proceedings of the sixth Swedish Production Symposium*. Chalmers Research, 2014.

[11] S. Mattsson and Å. Fast-Berglund, "How to support intuition in complex assembly?" *Procedia CIRP*, vol. 50, pp. 624–628, 2016.

[12] S. Mattsson, Å. Fast-Berglund, and D. Li, "Evaluation of guidelines for assembly instructions," *IFAC-PapersOnLine*, vol. 49, no. 12, pp. 209–214, 2016.

[13] T. Fässberg, Å. Fasth, S. Mattsson, and J. Stahre, "Cognitive automation in assembly systems for mass customization," in *Proceedings of the 4th Swedish Production Symposium (SPS), Lund, Sweden*, 2011.

[14] M. Agrawala, D. Phan, J. Heiser, J. Haymaker, J. Klingner, P. Hanrahan, and B. Tversky, "Designing effective step-by-step assembly instructions," *ACM Transactions on Graphics*, vol. 22, no. 3, p. 828, jul 2003.

[15] R. Lindorfer, R. Froschauer, and G. Schwarz, "Adapt - a decision-model-based approach for modeling collaborative assembly and manufacturing tasks," in *2018 IEEE 16th International Conference on Industrial Informatics (INDIN)*, 2018, pp. 559–564.

[16] S. Hu, J. Ko, L. Weyand, H. ElMaraghy, T. Lien, Y. Koren, H. Bley, G. Chryssolouris, N. Nasr, and M. Shpitalni, "Assembly system design and operations for product variety," *CIRP Annals*, vol. 60, no. 2, pp. 715–733, 2011.

[17] L. Gong, D. Li, S. Mattsson, M. Åkerman, and Å. F. Berglund, "The comparison study of different operator support tools for assembly task in the era of global production," *Procedia Manufacturing*, vol. 11, pp. 1271–1278, 2017.