Towards Automated Hesitation Detection During Support-System Enhanced Industrial Assembly

Jamil Joundi
imec-mict-UGent
Department of Industrial Systems Engineering and Product Design, Ghent University
Ghent, Belgium
jamil.joundi@ugent.be

Jonas De Bruyne
imec-mict-UGent
Department of Communication Sciences, Ghent University
Ghent, Belgium
jonas.debruyne@ugent.be

Aleksandra Zheleva
imec-mict-UGent
Department of Communication Sciences, Ghent University
Ghent, Belgium
aleksandra.zheleva@ugent.be

Wouter Durnez
imec-mict-UGent
Department of Communication Sciences, Ghent University
Ghent, Belgium
wouter.durnez@ugent.be

Jelle Saldien
imec-mict-UGent
Department of Industrial Systems Engineering and Product Design, Ghent University
Ghent, Belgium
jelle.saldien@ugent.be

Klaas Bombeke
imec-mict-UGent
Department of Communication Sciences, Ghent University
Ghent, Belgium
klaas.bombeke@ugent.be

Figure 1: Overview of virtual environment (left) and participant performing the experiment in the lab (right)

***Abstract*— Modern factories have to accommodate high flexibility, extreme customization and short product life cycles in a cost-effective way. This requires that the operators are provided with sufficient system support to aid their decision-making and hesitation. To investigate how the level of support can affect the operators' behaviour, the current study immersed 27 participants in a virtual reality factory where they were asked to complete three different assemblies with varying levels of system support (low, medium, or high). The support was provided by a collaborative robot (cobot). The participants' experience was measured via a subjective marker of difficulty and an objective eye-tracking feature (gaze switches). The results showed that when the level of cobot support was low, participants found the assembly step more difficult and were gazing at the instruction screen more often compared to the medium and high support conditions. This suggests that the number of times operators look back at the instruction screen during a step could be a promising marker to automatically detect hesitation behaviour in instruction-based assemblies. This study, therefore, presents the initial effort toward validating a behavioural marker of hesitation within this context.**

Keywords*— *Cobot support; Cognitive ergonomics; Virtual Reality; Hesitation; Industrial assembly

I. Introduction

Nowadays, factories rely on fully automated production steps. However, human creativity and adaptability are often needed to ensure smoothly running processes resulting in high quality production standards. To maintain these standards operators must be equipped with efficient support systems that aid them in making the right decisions and act optimally [17]. One way of providing support is via collaborative robots (i.e., cobots) that provide additional information to human operators and aid their decision-making [12]. However, if the cobots provide the human operators with sub-optimal support, unfavourable working conditions might be created. Research thus far has addressed this impediment by creating and optimising features of physical ergonomics [11,16]. However, this paper argues that cognitive ergonomics (e.g., hesitation) can affect the operators' behaviour in human-robot collaboration (HRC) and should be further investigated [7]. More specifically, we propose that if the operator is not provided with relevant support they will experience harder decision-making process as a result of increased hesitation. This, in turn, can lead to increased production times, and decreased quality of the final product. Hence, the main goal of the current study is to identify an objective marker of hesitation by exploring a sample HRC scenario in virtual reality (VR).

Hesitation has been described as a type of micro-movement (micro slip), which acts as "non-verbal stutter" during execution of an action [5]. It can be manifested through behavioural cues such as facial expressions, shoulder movements, head movements and prosody [22]. In the context of factory work and decision making, hesitation can be viewed as the uncertainty in following through with a task and the need for information reappraisal. This might be observed in a situation, where the additional support is inferior, in which case the operator will have to continuously refer to the provided task instructions. Therefore, they might switch their gaze back and forth between the task and the instructions in an attempt to moderate their hesitation and improve their performance. To address the issue described above, the current study aims to find an objective marker of hesitation that can be applied to a scenario in which the human operator is provided with deficient support. The objective marker that is validated in the present study is *screen gazes* (i.e., looking back at the instruction screen while performing an assembly). With this goal in mind, we access real-time eye-tracking data (i.e., gaze switches between different elements in the experimental environment), and integrate it with subjective difficulty ratings reported by participants.

These ratings were noted at each step of three assembly tasks. The difficulty level was moderated via different levels of cobot support in each step of the assemblies. Furthermore, HRC is highly specific to the factory setting and exploring it in ecologically valid conditions will greatly benefit the applicability of findings to real-life cases. For example, using AR glasses to enhance the operations on the assembly floor. Accordingly, this study immerses the user in an ecologically valid but safety-proof VR environment (see Figure 1). VR allows us to create an environment replicating in-situ training that is easily moderated to explore the effects of additional support on the operator's behaviour [1,15]. Moreover, the integration of eye-tracking technology in VR head-mounted displays (HMDs) allows us to gather eye-tracking data in an unobtrusive way that does not alter the user experience and provides uninterrupted continuous readings [13].

II. METHOD

2.1 Participants

27 participants (15 female, M_{age}=~27.2, SD_{age}=~6.18) took part in this study. They were recruited using an online survey published on social media platforms. The only inclusion criterion was normal or corrected to normal vision. Each participant signed informed consent and received 15 euros for their participation.

2.2 Materials and equipment

The VR set-up consisted of a computer running SteamVR [20], an HTC VIVE Pro Eye [6], and manus XSens Gloves [23] that were all tracked by two Vive SteamVR Base Stations 2.0. The Vive Pro Eye has two 3.5" OLED displays with a pixel density of 1440 x 1600 pixels per eye (see Figure 1). The Manus gloves contain five 2DoF Flexible sensors and six 9DoF IMU’s with an error margin of 2.5 degrees. The HMD’s built-in eye-tracker and the Vive Eye-tracking Software Development Kit Sranipal [18] were used to collect eye-tracking data.

The virtual environment was developed in Unity [19]. It contains a workbench with the robot arm placed (left of the operator), a screen with instructions and feedback (on the right side of the operator) and the assembly (a base plate and assembly pieces) in the middle (see Figure 1).

The UI on the researcher’s screen provides a reset button in case a piece is dropped and unreachable for the participant. In-scene interactions that were performed by the Manus XSens Gloves, eye-tracking data and participants' feedback were all logged in a json file. This was accompanied by a screen recording (via Xbox game bar recorder) used for a visual check of the algorithm detecting screen gazes.

After each step of the assemblies, participants were asked to rate the experienced difficulty of the completed step by selecting one of five selection tiles that represented different levels of difficulty ranging from “very easy” to “very difficult”.

2.3 Design and procedure

This experiment used a within-subjects design. First, participants were asked to read and sign the informed consent form after which they were given instructions. After mounting the HMD and putting on the gloves, an eye tracker calibration was performed. Next, participants were allowed to familiarise with the in-scene interactions (e.g., grabbing and placing pieces, interacting with the instruction screen) and they followed a short tutorial to learn how to interact with the cobot.

Subsequently, participants were asked to complete three assembly tasks, each consisting of 13 steps during which one piece had to be added to the assembly. The instruction screen showed a picture of how the assembly should look after completing the step. It was stressed by the experimenter that each step should be completed as correct as possible. There was no time limit to the steps of the assembly. During each step of every assembly, participants were provided either with a high level, medium level or low level of cobot support. In the *high level of support* condition, the cobot handed to the participant a piece of the assembly in correct rotation and hovered over the location where the piece should be placed. In the *medium level of support* condition, the cobot handed the piece to the participant at the right location but rotated incorrectly. In the *low level of support* condition, the cobot handed the piece to the participant without suggesting the correct location and rotation. The order of this manipulation was randomised across participants.

When the assembly was completed, performance feedback was provided on the instruction screen, after which participants had to rate the experienced difficulty. Throughout the procedure, the experimenter was able to track the quality of the data signal in real-time. The two dependent variables of interest were the subjective difficulty ratings of each step and the number of times participants looked back at the instruction screen during each step.

2.4 Data analysis

The data was analysed using one within-subjects factor with three levels (i.e., *level of support* - low, medium or high level of support). Data pre-processing was performed using Python 3 and a linear mixed-effects model (LMM) was tested in R using the lme4 package [2] specifying a random intercept for each participant. Degrees of freedom for the LMM and the post-hoc pairwise comparisons are corrected using the Kenward-Rogers method [10]. Reported p-values of post-hoc pairwise comparisons are adjusted using the Bonferroni correction.

First, we investigated the subjective feedback participants provided after each step's completion. The responses were scaled to a continuous measure between 1 and 5 (1 = “very difficult” and 5 = “very easy”). The other dependent variable we investigated was *number of screen gazes*, that is, the number of times a participant looked at the instruction screen. This was calculated based on the eye-tracking data that categorized the virtual object the participant's gaze collided at almost any moment. The gaze collision data was collected at a rate of 50 Hz and was later filtered on the event “watching the instruction screen” and

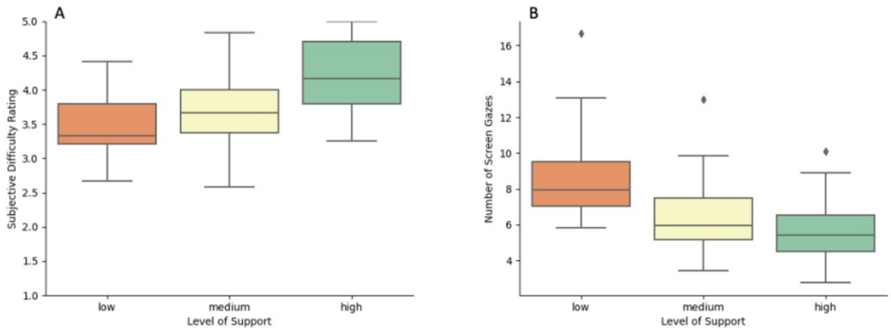

Figure 2: Box plots showing the main effects of level of support on both subjective difficulty rating (A) and number of screen gazes (B).

converted to data representing the mean number of screen gazes during each condition. For this measure, one participant was excluded due to critical eye-tracking data loss.

III. RESULTS

An effect of level of support on subjective feedback was found, $F(2, 52) = 47.04$, $p < 0.001$. According to post-hoc pairwise comparisons, all three levels of support differed significantly as for the reported experienced difficulty. Specifically, the low level of support condition was found more difficult than the medium level of support condition, $t = -3.65$, $p = .002$, the medium level of support condition more difficult than the high level of support condition, $t = -5.96$, $p < .001$ and the low level of support condition more difficult than the condition in which there was a high level of support provided by the cobot, $t = -9.61$, $p < .001$ (Figure 2 A).

We found an effect for level of support on number of screen gazes in the constructed LMM, $F(2, 50) = 57.85$, $p < 0.001$. Post-hoc pairwise comparisons revealed that there were significant differences between all three of the comparisons. In detail, not only the difference between the low and high level of support conditions was significant, $t = 10.39$, $p < .001$, but also were the difference between the low and medium level of support conditions and the difference between the medium and high level of support conditions, respectively $t = 7.61$, $p < .001$, $t = 2.78$, $p = 0.023$ (Figure 2 B).

A weak, yet significant correlation of -0.22 was found between the subjective feedback on experienced difficulty and number of screen gazes, $t = -2.00$, $p < 0.05$. In other words, there exists a negative relationship between these two variables in that the more difficult a step was judged, the more times a participant looked at the instruction screen.

IV. DISCUSSION & CONCLUSION

The current study immersed participants in a VR factory floor where they were asked to complete three assemblies during which they received a different amount of support from a cobot. The study found a possible objective marker that signifies hesitation in a system-supported factory setting. Specifically, it was tested whether the number of times participants looked back at the instructions provided for a current step was influenced by the amount of additional support they received from the cobot. To do so, three different levels of cobot support were simulated during a VR assembly task. During the assembly, participants' gaze position was continuously logged using the HMD's built-in eye-tracker. Additionally, they were asked to rate each step of the assemblies on their experienced difficulty right after completing that step. In doing so, it was expected to gain insight into whether the variance in the provided support was related to a different difficulty perception and number of times participants looked at the instruction, both indicative of hesitation behaviour.

There are two main outcomes of this study. First, the level of support of the cobot influenced the perceived difficulty of the task. More specifically, when the cobot gave more exhaustive support, the human operator found the task easier. This is in line with the assumption that the perception of the complexity of the task at hand is related to the amount of system support [9,12]).

Second, the objective marker *number of screen gazes* showed promising results as it significantly differed between all three conditions (i.e., low, medium or high level of support). Consequently, the number of times people looked back at the instruction screen was negatively related to the amount of provided support by the cobot. Thus, the increase in gaze switches could be reflective of the hesitation of participants induced by insufficient cobot support. This finding shows that it might be interesting to use gaze metrics in studies in which operator support systems are evaluated. It has been shown, for example, that gaze entropy (i.e., the extent to which visual exploration patterns become less stereotyped/more random) is indicative for task load in a surgical setting [4]. It is likely that this can be translated to assembly contexts as well. As such, efficacy of operator support systems can be evaluated on the basis to which they decrease or increase task load. Therefore, in the continuation of the current project, we plan to explore and validate the eye-tracking metric gaze entropy as a measure of cognitive load in order to evaluate various operator support systems (e.g., light-guiding systems, cobots, augmented reality instructions etc.) at a later stage.

Even though the current study found evidence suggesting that gaze switching can be an objective marker of hesitation, it should be noted that this is a first step in the formulation of such a measure. As previously outlined, hesitation is a cognitive state that is expressed via a range of behavioural cues [5,22]. For instance, studies on the effect of mental workload in an assembly task have illustrated that hand and/or head freezing, head re-positioning and attempted trials are just a few of the behavioural components observed when a person is hesitating [21]. Hence, it is strongly recommended that future research integrates various methodologies (e.g., video coding, full body tracking) to explore an array of behavioural cues experienced during hesitation.

Detecting moments of hesitation is vital for creating support systems that can reduce the operators' cognitive load, which in turn has shown to increase their mental well-being and efficiency [14]. Indeed, if support systems can adapt to the operators' behaviour and provide them with the appropriate support in targeted moments (i.e., when they hesitate), then the operators are less prone to mistakes and tend to rate their work as more enjoyable [8]. To this point, better work conditions - in the current context, facilitated by an adaptive support system - have been shown to have a positive influence on the operators' well-being and health, which has advantages both for the operator and the organisation (e.g., less physiological, psychological or emotional costs) and the organisation (e.g., higher productivity, less absence from work, etc.) [3]. A disputable limitation of the present study is that after completing a step during the task, participants were first given feed- back on their performance (i.e., "you connected the piece correctly/incorrectly") and only afterward asked to rate their perceived difficulty. Because participants were first informed on their performance, their subjective difficulty rating might have been biased. A solution might be to reverse the order of the feedback and the experienced difficulty assessment. However, in that case when performing a step incorrectly, that step might be judged as very easy, although the step had not been completed as it should, leaving much unknown about the implemented support manipulation.

In conclusion, the present study has shown that the behaviour of looking back at the provided instructions can be used to detect hesitation in assembly operators. This, in turn, can be used to evaluate different (settings of) support systems that are provided to the operator with the final aim to create better working conditions and increase operator well-being. Even though we need to further fine-tune and investigate the methods with which hesitation during an assembly can be detected, this study demonstrates how objective measures that mark specific cognitive states can be explored and validated in an industrial cognitive ergonomics context.

ACKNOWLEDGMENT

We thank Ward Dehairs, a freelance Unity developer, for his support in building the virtual environment.

REFERENCES

[1] ST Engineering Antycip. 2019. How Virtual Reality can Revolutionise Project Research. https://steantycip.com/blog/vr-in-research/

[2] Douglas Bates, Martin Mächler, Benjamin M. Bolker, and Steven C. Walker. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67, 1 (oct 2015), 1–48. https://doi.org/10.18637/jss.v067.i01 arXiv:1406.5823

[3] Karen Danna and Ricky W Griffin. 1999. Health and well-being in the workplace: A review and synthesis of the literature. Journal of management 25, 3 (1999),357–384.

[4] Leandro L Di Stasi, Carolina Diaz-Piedra, Héctor Rieiro, Jose M Sanchez Carrion, Mercedes Martin Berrido, Gonzalo Olivares, and Andrés Catena. 2016. Gaze entropy reflects surgical task load. Surgical endoscopy 30, 11 (2016), 5034–5043.

[5] Roberta Etzi, Siyuan Huang, Giulia Wally Scurati, Shilei Lyu, Francesco Ferrise, Alberto Gallace, Andrea Gaggioli, Alice Chirico, Marina Carulli, and Monica Bordegoni. 2019. Using virtual reality to test human-robot interaction during a collaborative task. In Proceedings of the ASME Design Engineering Technical Conference, Vol. 1. American Society of Mechanical Engineers Digital Collection. https://doi.org/10.1115/DETC2019-97415

[6] HTC. 2020. VIVE Pro Eye Specs & User Guide. (2020). https://developer.vive.com/resources/hardware-guides/vive-pro-eye-specs-user-guide/https://developer.vive.com/resources/knowledgebase/vive-pro-eye-specs-user-guide/

[7] International Ergonomics Association. 2021. What Is Ergonomics? | The International Ergonomics Association is a global federation of human factors/ergonomics societies, registered as a nonprofit organization in Geneva, Switzerland. https://iea.cc/https://iea.cc/what-is-ergonomics/

[8] P. E.C. Johansson, G. Eriksson, P. Johansson, L. Malmsköld, Fast-Berglund, and L. Moestam. 2017. Assessment based information needs in manual assembly. In 24th International Conference on Production Research, ICPR 2017. DEStech Publications, 366–371. https://doi.org/10.12783/dtetr/icpr2017/17637

[9] Virpi Kalakoski, Sanna Selinheimo, Teppo Valtonen, Jarno Turunen, Sari Käpykangas, Hilkka Ylisassi, Pauliina Toivio, Heli Järnefelt, Heli Hannonen, and Teemu Paajanen. 2020. Effects of a cognitive ergonomics workplace intervention (CogErg) on cognitive strain and well-being: A cluster-randomized controlled trial. A study protocol. BMC Psychology 8, 1 (jan 2020), 1–16. https://doi.org/10.1186/s40359-019-0349-1

[10] Michael G. Kenward and James H. Roger. 1997. Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. Biometrics 53, 3 (sep 1997), 983.https://doi.org/10.2307/2533558

[11] Ilias El Makrini, Kelly Merckaert, Dirk Lefeber, and Bram Vanderborght. 2017. Design of a collaborative architecture for human-robot assembly tasks. IEEE International Conference on Intelligent Robots and Systems 2017-September (2017), 1624–1629. https://doi.org/10.1109/IROS.2017.8205971

[12] Olessia Ogorodnikova. 2008. Human Weaknesses and strengths in collaboration with robots. Periodica Polytechnica Mechanical Engineering 52, 1 (2008), 25–33. https://doi.org/10.3311/pp.me.2008-1.05

[13] Suja Sreeith Panicker and Prakasam Gayathri. 2019. A survey of machine learning techniques in physiology based mental stress detection systems. , 444–469 pages. https://doi.org/10.1016/j.bbe.2019.01.004

[14] Davy D. Parmentier, Bram B. Van Acker, Jan Detand, and Jelle Saldien. 2020. Design for assembly meaning: a framework for designers to design products that support operator cognition during the assembly process. Cognition, Technology and Work 22, 3 (aug 2020), 615–632. https://doi.org/10.1007/s10111-019-00588-x

[15] Ruth Prijls. 2020. Virtual reality: The future of experimental research? https://ocean.sagepub.com/blog/tools-and-tech/virtual-reality-the-future-of-experimental-research

[16] Daniel Rücker, Rüdiger Hornfeck, and Kristin Paetzold. 2019. Investigating ergonomics in the context of human-robot collaboration as a sociotechnical system. In Advances in Intelligent Systems and Computing, Vol. 784. Springer Verlag, 127–135. https://doi.org/10.1007/978-3-319-94346-6_12

[17] Anna Syberfeldt, Magnus Holm, Oscar Danielsson, Lihui Wang, and Rodney Lindgren Brewster. 2016. Support Systems on the Industrial Shop-floors of the Future - Operators' Perspective on Augmented Reality. In Procedia CIRP, Vol. 44. Elsevier B.V., 108–113. https://doi.org/10.1016/j.procir.2016.02.017

[18] Tobii SDK. 2021. Get started developing with the HTC VIVE Pro Eye using Tobii XR SDK. https://vr.tobii.com/sdk/develop/unity/getting-started/vive-pro-eye/

[19] Unity. 2022. Get your Unity Pro subscription today | Unity. https://unity.com/pages/unity-pro-buy-now?ds_rl=1295837&gclid=CjwKCAjwrqqSBhBbEiwAlQeqGr7yWw3lbXwiXrhNAbjx8KgTS2aNELzix8rlMJAfO_MLsubQV5qhoRoC-vsQAvD_BwE&gclsrc=aw.ds

[20] Valve Corporation. 2020. SteamVR on Steam. https://store.steampowered.com/app/250820/SteamVR/

[21] Bram B. Van Acker, Davy D. Parmentier, Peter D. Conradie, Stephanie Van Hove, Alessandro Biondi, Klaas Bombeke, Peter Vlerick, and Jelle Saldien. 2021. Development and validation of a behavioural video coding scheme for detecting mental workload in manual assembly. Ergonomics 64, 1 (2021), 78–102. https://doi.org/10.1080/00140139.2020.1811400

[22] Tomaž Vodlan, Marko Tkalčič, and Andrej Košir. 2015. The impact of hesitation, a social signal, on a user's quality of experience in

multimedia content retrieval. Multimedia Tools and Applications 74, 17 (2015), 6871–6896. https://doi.org/10.1007/s11042-014-1933-2

[23] Xsens. 2021. Xsens Gloves by Manus. https://www.xsens.com/products/xsens-gloves-by-manus https://motionshop.xsens.com/index.php/products/motioncapture/accessories/xsens-gloves-by-manus?hsCtaTracking=29eb25c5-0377-4e43-9111-0029ec8696e7%7C96c41353-2324-4699-a0d8-791838c4e1