 DOI: 10.3217/978-3-85125-663-5-26

Evaluation of Human-Robot Collaboration Using Gaze based Situation Awareness in Real-time*

Lucas Paletta[1], Martin Pszeida[1], Harald Ganster[1], Ferdinand Fuhrmann[1], Wolfgang Weiss[1], Stefan Ladstätter[1], Amir Dini[1], Bernhard Reiterer[2], Inka Brijacak[2], Guido Breitenhuber[2], Sandra Murg[1], and Harald Mayer[1]

***Abstract*— Human attention processes play a major role in the optimization of human-robot collaboration (HRC) systems. This work describes a novel framework to assess the human factors state of the operator primarily by gaze and in real-time. The objective is to derive parameters that determine information about situation awareness which represents a central concept in the evaluation of interaction strategies in collaboration. The control of attention provides measures of human executive functions that enable to characterize key features in the collaboration domain. Comprehensive experiments on HRC were conducted with typical tasks including collaborative pick-and-place in a lab based prototypical manufacturing environment. The methodology measures executive functions and situation awareness (SART) in the HRC task in real-time for human factors based performance optimization in HRC applications.**

I. Introduction

Collaborative robotics has recently progressed to human-robot interaction in real manufacturing. Human factors are crucial as industrial robots are enabling human and robot workers to work side by side as collaborators and to assess the user's experience with a robot, while understanding how humans feel during their interaction with it [1]. Furthermore, human-related variables are essential for the evaluation of human-interaction metrics [2]. To work seamlessly and efficiently with their human counterparts, robots must similarly rely on measurements to predict the human worker's behavior, cognitive and affective state, task specific actions and intent to plan their actions. A typical application is anticipatory control with human-in-the-loop architecture [3] to enable robots to proactively perform task actions based on observed gaze patterns to anticipate actions of their human partners according to its predictions. However, measuring and modeling of the state of human factors as well as the human situation awareness based on gaze triggered information recovery is mandatory for the understanding of immediate and delayed action planning.

This work describes a novel methodology to measure the human factors state of the operator in real-time with the purpose to derive fundamental parameters that determine situation awareness as a central concept in the interaction strategies of collaborative teams. Human situation awareness is determined on the basis of concrete measures of eye movements towards production relevant processes that need to be observed and evaluated by the human. Motivated by the theoretical work of [4] on situation awareness the presented work specifically aims at dynamically estimating (i) distribution of attentional resources with respect to task relevant 'areas of interaction' over time, determined by features of 3D gaze analysis and a precise optical tracking system, and (ii) derive from this human factors in real-time, such as, (a) human concentration on a given task, (b) human mental workload, (c) situation awareness and (d) executive functions related measure, i.e., task switching rate.

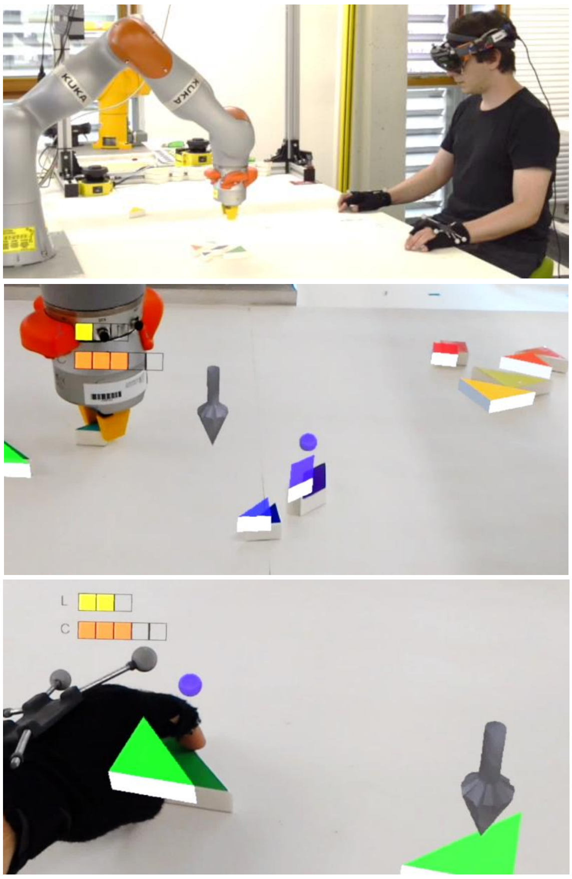
Gaze in the context of collaboration is analyzed in terms of - primarily, visual - affordances for collaboration. In this work we stress the relevance of considering eye movement features for a profound characterization of the state of human factors by means of gaze behavior, with the purpose to optimize the overall human-robot collaboration performance.

Figure 1: Human-robot collaboration and intuitive interface (HoloLens, eye tracking, markers for OptiTrack localization) for the assessment of human factors state to characterize key features in the collaboration domain.

[1]DIGITAL – Institute for Information and Communication Technologies, JOANNEUM RESEARCH, Graz, Austria.

[2]ROBOTICS – Institute for Robotics and Mechatronics, JOANNEUM RESEARCH, Klagenfurt, Austria.

*This work has been supported by the Austrian Ministry for Transport, Innovation and Technology (BMVIT) within project CollRob (Collaborative Robotics), FLEXIFF (no. 861264) and MMASSIST (no. 858623). Lucas Paletta is corresponding author with phone: +43-316-876-1769; fax: +43-316-876-91769; e-mail: lucas.paletta@joanneum.at



(a)

(b)

(c)

Figure 2: HRC within a tangram puzzle assembly task. (a) The operator collaborates with the robot in the assembly (only robot can treat 'dangerous' pieces). (b) Egocentric operator view with augmented reality based navigation (arrow), piece localization, gaze (blue sphere), current state of mental load (L) and concentration (C) in the HoloLens display. (c) Recommended piece (arrow) and gaze on currently grabbed puzzle piece.

The estimation of situation awareness of the human worker can be crucial for the elaboration of performance analysis through measurement of executive functions, evaluation of interruption impact, as well as for the prediction of accidents.

II. Related Work

A. Human-Robot Collaboration

Human-Robot collaboration has substantially advanced, in the planning domain. [4] presented a planning executive able to handle choices by either team member while respecting causal links and temporal constraints with regard to risk bounds. In a related approach, one of the contributions of [5] defines helpfulness in terms of cost reduction resulting from the utilization of the robot. [6] split the planning process for repetitive collaborative assembly tasks into two phases, 1st is offline and establishes agents' capabilities for hierarchical plans with choices, the second one taking cost-function-based optimization decisions. [7] researched probabilistic planning for collaborative manipulation, aiming at boosting the human's trust in the robot's capabilities for better overall performance. [8] employ a probabilistic approach focused on dynamical switching and on decision-making in hierarchical assembly tasks. The aspect of making the behavior of a planning system better understandable for humans is addressed by Fox et al. [9] who pose a set of questions that the system should be able to answer. They show how to capitalize on features of planning systems that would make similar questions harder or impossible to answer based on other, currently popular AI-based decision making approaches. In this context, the cognitive state of the human gets into the focus of research studies.

B. Evaluation of Human-Robot Collaboration

Kragic et al. [10] presented results of a study featuring a human-robot interaction task using three different feedback modalities: a computer screen, projection into the workspace, and augmented reality. While projection was subjectively higher rated, the study did not yield significant performance differences between the variations. Salem et at. [11] report on an analysis of how various kinds of faulty behavior by a robot affect humans' trust. Measures for Human Robot Collaboration are of three complementary dimensions: team performance measures, measures targeting user satisfaction and experience, and safety and trust related assessments. In this line, the Huang et al. [12] evaluated different handover strategies using both objective measures (Task Completion Time (TCT), Concurrent Activity (CA), Human and Robot Idle Time (HIT and RIT) as well as subjective scales via questionnaires (Fluency, Intelligence, Awareness and Patience)

The human response to robot movements was assessed using objective measurements (TCT, CA, HIT, RIT, as well as average separation distance), and subjective criteria for perceived safety and comfort via questionnaires. Moreover, [3] evaluated the anticipatory control of a robot in user studies. Here, anticipation was derived from gaze analysis, the evaluation measures in the study included the average robot response time and the TCT. Next, [11] used a vast amount of questionnaires to evaluate the influence of robots' mistakes in the quality of the Human Robot Cooperation and Trust. Applied subjective test include the Ten Item Personality Inventory (TIPI), Godspeed Questionnaire, Human Nature Scale, and Uniquely Human Scale. Finally, [13] examines trust, interaction and safety issues of industrial workers on fenceless human robot collaboration. Here, individual questionnaires for specific interaction mechanisms (e.g. voice, gestures), and compared to objective measures such as average robot response time. Safety was regarded the most important factor for successful HRC, followed by usability and efficiency.

III. Human-Robot Collaboration Prototype

In the presented work, we apply a human factors analysis system to a complete human-robot collaboration system which is described as follows in more detail.

System Architecture

The architecture for the here-presented overall system was designed as a 3-tier application. Figure 3 shows the hierarchical system decomposition of the three layers. The hierarchy is based on task abstraction, i.e., a task at one level is achieved by invoking a set of tasks at lower levels. For connection and communication (i.e. services and messages) between the components we used the Robot Operating System (ROS).

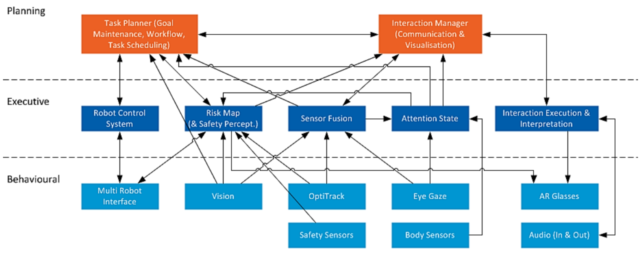


Figure 3: System architecture diagram depicting the 3-tier.

B. *Task Planner & State Description*

In order to deal with the manifold situations that may arise in a collaborative robot application, the orchestration of the robot and interaction components is performed by a task planning subsystem. It has a PDDL 2.1 [14] planner wrapped in ROSPlan [15] at its core, performant enough to be invoked after any significant change in order to always act based on current knowledge. The planning problems in our custom domain contain a mix of abstract knowledge derived from configuration, such as which agents there are and in which zones of the workspace they are allowed to act, and structures to be built at certain locations, from sensor data with respect to configuration, such as visibility and current zone of parts, and from combining the previous two with task-related aspects, such as whether a part counts as added to a structure being built. Such enriched object knowledge is also passed on to the interaction components. Our knowledge provider component keeps the knowledge about the current state and goals up to date according to the various data sources of the system. It takes care that are currently feasible and of top priority, i.e., requests by the participant to pass an object override assembly goals. The action link implementation bridges the gap between high-level plans and the robot control interface. It is able to work around robot joint limit violations and collisions on the table, also considering shape-dependent equivalent rotations of parts.

Based on preliminary experiments leading up to the study, the planning subsystem was configured such that the robot keeps working in its dedicated area as much as possible and only enters the collaborative area when necessary. This principle leads to more opportunities for the human participant to safely contribute to the joint task. Including the human participant's and/or the referee's actions in the planning problem was considered in different variations and leads to promising plans in isolated tests. However, this improvement could not be integrated due to accumulating delays resulting from a temporal lack of flexibility in available planning executive configurations. We aspire to follow up on this in future work incorporating findings from [4].

C. *Computer Vision & Safe Robot Control*

For locating parts to be manipulated on the table, we use a combination of shape and color matching approaches on 2D RGB images from a single camera, the output of which is rectified and projected into the table plane in the robot's 3D workspace based on results of intrinsic camera calibration and hand-eye calibration. As a key element of our multifaceted safety infrastructure, we track the participant's head and hands with a motion capturing system and compute distances to the nearest robot parts, applying worst-case assumptions in case of bad marker visibility. Based on a combination of this evaluation and the participant's stress level, the robot's current movement speed is limited to a configurable degree, down to a full stop. Further safety measures include low overall speed, spatial limits to the robot's workspace, sufficient distribution of emergency stop buttons to the participant, referee and extra personnels.

D. *Study Goals*

Within the described user study we aim at evaluating the quality of the interaction via the here-presented interaction system within a human-robot-collaboration application. In this context, we want to assess the individual interaction components (i.e. speech, gestures, etc.) as well as the overall interaction mechanisms resulting from the interconnection of the different modalities. We particularly examine to which degree the additional interaction mechanisms provided by the interface enhance of the human-robot-interaction. Here we focus on the user perspective on the developed interaction system and emphasize the intuitiveness of the interaction. We thereby assess the targeted qualities in a multi-dimensional way applying objective and subjective measures. To compare the generated results to a baseline system we performed the same experimental procedure within two distinct evaluation settings. On the one hand, a reference system (NUL) is established without interaction assistance. The full interaction functionalities are available in a second evaluation (INT).

IV. Intuitive Multimodal Assistive Interface

Assessing the intuitiveness and performance of an interface that should be regarded as intuitive as possible appears to be an obvious objective for a human factors analysis component. In the following we present a novel intuitive interface for multimodal assistance in human-robot collaboration. This interface will serve as a testbed for the efficiency of elderly.

A. *Multimodal Interaction Design*

Conceptually, we positioned the interaction design towards the user. For intuitive interaction we opted for a human-centered approach and started from inter-human interactions and the collaborative process itself. Following these considerations we implemented an interaction system that will be described in the remainder of this section.

The here-presented interaction system is based on the following principles:

- *Natural interaction*: Mimicking human interaction mechanisms we guarantee fast and intuitive interaction processes.
- *Multi-modal interaction:* We implement speech, gaze, gestural, and Mixed-Reality interaction to offer as much interaction freedom as possible to the user.

- *Tied modalities:* We link the different interaction modalities to emphasize the intuitive interaction mechanisms.
- *Context-aware feedback:* Feedback channels deliver information regarding task, environment to the user. We pay attention at what is delivered when and where.

Figure 4 shows a schematic overview of the presented interaction system. A central component entitled 'Interaction Model' (IM) acts as interaction control and undertakes the communication with the periphery system. The IM also links the four interaction modalities and ensures information exchange between the components. It triggers any form of interaction process, both direct and indirect, and controls the context-sensitivity of the feedback. It is further responsible for dialog management and information dispatching.

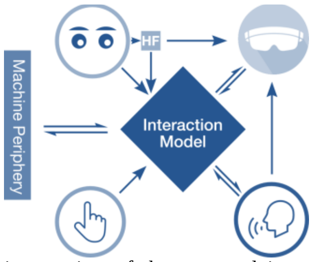


Figure 4: Schematic overview of the presented interaction system. The included interaction modalities refer to (from top left clock-wise): Gaze (including Human Factors (HF)), Mixed-Reality, Speech, and Gestures.

B. Audio Communication

Speech is the most natural communication element and acts as the main connecting element between the modalities in the developed interaction system. It is mostly used to derive the main intention of the interaction, which is strengthened by other modalities. Here, the system applies an acoustic interface to receive speech input. Context awareness is fully guaranteed since the interface is initialized and configured by the IM. It moreover delivers information about the speech interaction process (i.e. voice activity and last recognized command) to the Mixed-Reality display to increase the usability and user experience of the interaction.

Speech interaction is based on state-of-the-art Automatic Speech Recognition (ASR). We use the VoCon[1] library from Nuance, audio input is done via a Bluetooth headset. The system is set up in a constant-listening mode without a wake-up-word, i.e. the user has direct access to the voice commands without having to press a push-to-talk button. Hence, depending on the context (controlled by the IM), a various amount of speech commands is always available at a given moment. Moreover, we use a state-of-the-art speech synthesis engine to sonify the robot speech. We used the system offered by Acapella[2].

C. HoloLens Display (Augmented Reality)

Microsoft HoloLens[3] offers a state-of-the-art mixed reality (MR) development environment. By using an MR based display the system is able to augment the visual environment. We use annotations to visually mark real objects and give real-time feedback regarding the gaze and speech interaction process (e.g. gaze pointer and voice activity). We further provide indicators using visual icons to inform the user about changes in the task and the environment. Moreover, we use the interactive functionalities of the display to enable dialog interaction with virtual object selection via hand gestures. Hence, dialogs triggered by the IM can be resolved either using the Mixed-Reality display or the speech interface.

D. Human State Description and Human-in the-loop

In order to describe the current state of the human operator, a human factors measurement system (Sec. V) is integrated into the interaction system (Figure 4). It is fundamentally based on eye tracking for pervasive measurements of human cognitive and mental state.

V. Human Factors Measurement System

In human factors and ergonomics research, the analysis of eye movements enables to develop methods for investigating human operators' cognitive strategies and for reasoning about individual cognitive states [16]. Situation awareness (SA) is a measure of an individual's knowledge and understanding of the current and expected future states of a situation. Eye tracking provides an unobtrusive measure to measure SA in environments where multiple tasks need to be controlled. [17] provided first evidence that fixation duration on relevant objects and balanced allocation of attention increases SA. However, for the assessment of executive functions, the extension of situation analysis towards concrete measures of distribution of attention is necessary and described as follows.

A. Recovery of 3D Gaze in Human-Robot Interaction

Localization of human gaze is essential for the localization of situation awareness with reference to relevant processes in the working cell. [18] Firstly proposed 3D information recovery of human gaze with monocular eye tracking and triangulation of 2D gaze positions of subsequent key frames within the scene video of the eye tracking system. Santner et al. [19] proposed gaze estimation in 3D space and achieved accuracies ≈ 1 cm with RGB-D based position tracking within a predefined 3D model of the environment. In order to achieve the highest level of gaze estimation accuracy in a research study, it is crucial to track user's frustum / gaze behavior with respect to the worker's relevant environment. Solutions that realize this include vision-based motion capturing systems: OptiTrack[4] can achieve high tracking and gaze estimation accuracy (≈ 0.06 mm).

[1] https://www.nuance.com/mobile/speech-recognition-solutions/vocon-hybrid.html
[2] http://www.acapela-group.com
[3] https://www.microsoft.com/de-at/hololens
[4] http://www.naturalpoint.com/optitrack

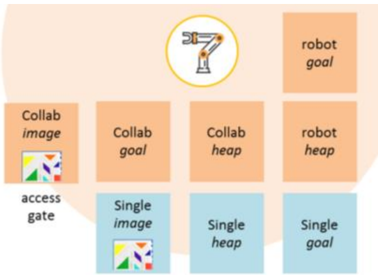


Figure 5: Schematic representation (top view) of the areas of interaction (AOIs) that are used to analyze gaze behavior and fro this derive human factors analysis. Collaborative task related AOIs are in orange, Single task related in blue color.

B. Situation Awareness

Based on the cognitive ability, flexibility and knowledge of human beings on the one hand and the power, efficiency and persistence of industrial robots on the other hand, collaboration between both elements is absolutely essential for flexible and dynamic systems like manufacturing [20]. Efficient human-robot collaboration requires a comprehensive perception of essential parts of the working environment of both sides. Human decision making is a substantial component of collaborative robotics under dynamic environment conditions, such as, within a working cell. Situation awareness and human factors are crucial, in particular, to identify decisive parts of task execution.

In human factors, situation awareness is principally evaluated through questionnaires, such as, the Situational Awareness Rating Technique (SART, [21]). Psychological studies on situation awareness are drawn in several application areas, such as, in air traffic control, driver attention analysis, or military operations. Due to the disadvantages of the questionnaire technologies of SART and SAGAT, more reliable and less invasive technologies were required, however, eye tracking as a psycho-physiologically based, quantifiable and objective measurement technology has been proven to be effective [17][22] In several studies in the frame of situation awareness, eye movement features, such as dwell and fixation time, were found to be correlated with various measures of performance. [23] have developed measurement / prediction of Situation Awareness in Human-Robot Interaction based on a Framework of Probabilistic Attention, and real-time eye tracking parameters.

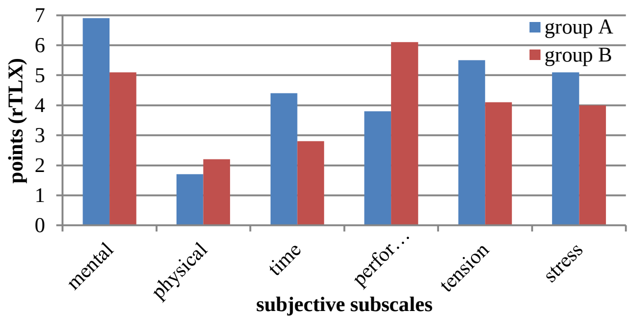
C. Stress and Concentration Estimation

For stress quantification we used cognitive arousal estimation based on biosensor data. In the context of eye movement analysis, arousal is defined by a specific parametrization of fixations and saccadic events within a time window of five seconds so that there is good correlation (r=.493) between the mean level of electrodermal activity (EDA) and the outcome of the stress level estimator [25].

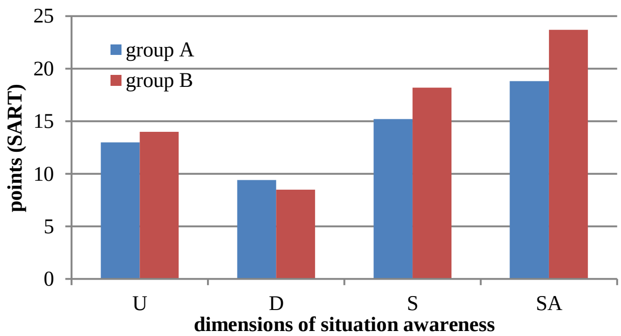
For the estimation of concentration or sustained attention, we refer to the areas of interaction (AOI) in the environment as representing the spatial reference for the task under investigation. Maintaining the attention on task related AOI is interpreted as the concentration on a specific task [26], or on session related tasks in general. Various densities of the fixation rate enable the definition of a classification of levels of actual concentration within a specific period of time, i.e., within a time window of five seconds.

D. Estimation of Task Switching Rate

Task switching, or set-shifting, is an executive function that involves the ability to unconsciously shift attention between one task and another. In contrast, cognitive shifting is a very similar executive function, but it involves conscious (not unconscious) change in attention. Together, these two functions are subcategories of the broader cognitive flexibility concept. Task switching allows a person to rapidly and efficiently adapt to different situations [27].



(a) Task Load indeX (rTLX)



(b) SART

Figure 6: Results of the questionnaires from after the session without (group A) & with assistance (group B), respectively. (a) The rTLX subjective subscales show significant reduction of mental workload in group B. (b) The SART results in significant increase in U (understanding) and S (support of attention) and decrease in D (attentional demand) for group B.

In a multi-tasking environment, cognitive resources must be shared or shifted between the multiple tasks. Task switching, or set-shifting, is an executive function that involves the ability to unconsciously shift attention between one task and another. Task switching allows a person to rapidly and efficiently adapt to different situations. The task-switching rate is defined by the frequency by which different tasks are actually operated. The difference between tasks is defined by the differences in the mental model which is necessary to represent an object or a process in the mind of the human operator. Mental models are subjective functional models; a task switch requires the change of the current mental model in consciousness and from this requires specific cognitive resources and a load.

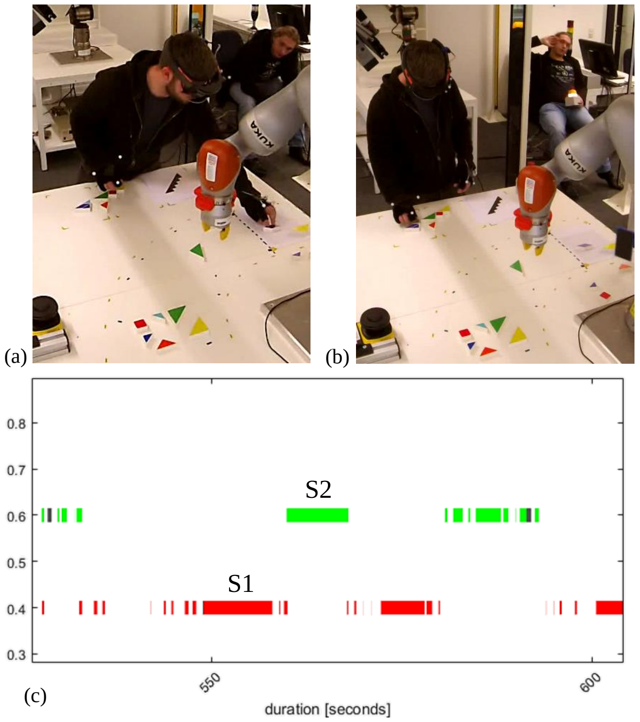


Figure 7: Task switching between collaborative and the single task. (a) Placing a puzzle piece to goal area in the collaborative task. (b) The operator places a puzzle piece to the goal area of single task.(c) Switch between collaborative task (S1) and single task (S2). Task duration is determined by the human gaze being focused within an AOI related to the specific task.

In the presented work, processing of a task is determined by the concentration of the operator on a task related area of interaction (AOI). Interaction is defined by areas in the operating environment where the operator is manipulating the location of puzzle objects, i.e., grabbing puzzle pieces from a heap of pieces, or putting pieces onto a final position in order to form a tangram shape. Whenever the gaze of the operator intersects with an AOI that belongs to a specific task, then it is associated with an on-going task. The task switch rate is then the number of switches between tasks per period of time, typically the time of a whole session (see Figure 7 for a visualization). Task switching has been proposed as a candidate executive function along with inhibition, the maintenance and updating of information in working memory, and the ability to perform two tasks at the same time. There is some evidence not only that the efficiency of executive functions improves with practice and guidance, but also that this improvement can transfer to novel contexts. There are demonstrable practice-related improvements in switching performance [29][30].

VI. Experimental Results

A. Study Setup

The experimental procedure was performed in a robotics laboratory. 20 persons (f=8, m=25) aged 25.4±4.7 years were engaged, mostly with education at university level (BSc). 8 of them were with technical or even scientific background, but all naïve to the task. All participants were introduced into the tasks which consisted of assembling 8-piece tangram puzzles according to target images. The overall task was to finalize three puzzles in collaboration with a robotic arm that performed pick and place actions according to a plan. Tt was possible to assemble in parallel a difficult ‘extra’ puzzle without any external support. The session was finalized upon the finalization of the 3rd collaborative puzzle, and performance was measured by completion time minus a bonus time for a single finalized ‘extra’ puzzle. Participants were told that the best 3 performances would earn special gifts (for competitive pace)

Robotic system. The robot arm used was a Kuka iiwa LBR 7 R800, equipped with a Zimmer R800 2-finger gripper prototype and customized 3D-printed fingers with metal reinforcement and soft cover. The RGB camera used for vision was a Basler acA2440-20gc with an f8mm F1.4 lens. It was connected via Gigabit PoE to the computer for vision and safety perception applications (Fujitsu Celsius, Intel Xeon 72-core 2.3GHz, 128GB RAM, Nvidia Quadro P4000 GPU). The motion capturing PC worked with Intel dual core i7 3.6GHz CPU, 16GB RAM and Nvidia Geforce GT 630 2GB, receiving data from 9 OptiTrack Prime 17W cameras.

Wearable system. For the intuitive assistance device we used a Microsoft Hololens AR headset with Pupil Labs binocular eye tracking with 200 Hz eye tracking cameras and USB connector clip that connected to a Microsoft Surface Pro 6 with 8GB/128GB RAM/SSD for the gaze analysis.

B. Outcome Measures

The major outcome parameters of the study were on the one hand related to standardized questionnaires that are commonly used to specify important human factors in the human-machine interaction domain. At the same time, several eye movement features were measured in real-time in order to derive human factors on-site directly from the human-robot collaboration task. The objective of the study was to investigate if on-site measured data would correlate with the results of the questionnaires.

The subjects were asked to complete several questionnaires during the experimental procedure. Specifically, the standardized questionnaires raw Task Load indeX (rTLX) and the Situation Awareness Rating Technique (SART) were used. Furthermore, we designed individual questionnaires to cover general interaction, mixed reality interaction and visualization, speech interaction, and human-robot-collaboration. These include user input via scale ratings as well as open questions relating to good and bad experiences within the respective area. The real-time assessment of the interaction was built upon a mixture of recorded data. The HMD video was captured as well as data from an external video camera for post study analysis. Eye tracking data were collected in real-time and analyzed with respect to the measures presented in Sec. V: the human gaze was positioned in 3D space and intersections with various areas of interaction (AOI) analyzed with respect to level of concentration; all data were recorded in a further study analysis.

C. Descriptive Statistics

The overall performance comparison resulted in a total completion time of M=709.7, SD=193.4 sec. without and M=714.6, SD=159.5 sec. with assistance, respectively, and ANOVA analysis resulted in $F_{crit} > F$ which identifies identical distributions. The bonus addendum was negligible. From the data of twenty participants, only fifteen were used for analysis, for reasons of data loss and loss of eye tracking calibration during the session when persons accidentally changed the orientation of eye tracking cameras when tipping on the helmet. However, the analysis of the questionnaires indicated substantial and interesting differences in the outcomes. Figure 6a depicts subjective subscales of rTLX indicating that users with the intuitive assistance system (group B) subjectively experienced *significantly reduced* levels of *mental workload*, time pressure, tension and stress, respectively. The results of the SART questionnaire (Figure 6b) depict in U (*understanding*) and S (*support of attention*) dimensions with *significant increase* and decrease in D (attentional demand) for group B.

Furthermore, the gaze based outcome measures of the experimental study showed further evidence for interesting differences with respect to without and with intuitive assistance, as documented in Table I: Firstly, the mean *concentration level* (‘M(C)’) with the use of an intuitive assistance technology was *significantly above the one without assistance*. Figure 8 depicts sample tracks of attentional concentration on tasks during sessions.

In addition, results showed a *significant increase* in the *situation awareness* in terms of attentional attributed to the collaborative AOIs (‘Collab’), as well as to the single task (‘Single’) related AOIs. This clearly shows that the assistance mode successfully *supports the channeling of attentional resources on the task related interaction* areas.

The execution function related feature based on eye movements, i.e., the expected time for a next task switch (‘ExpTS’), significantly decreased in case of intuitive assistance. Figure 7a,b shows task switching between the collaborative and the dual, single task, where the operator places a puzzle piece to the goal tangram area as contribution to the collaborative task or to the goal tangram area of the single task. Figure 7c depicts the switches between the collaborative task (S1) and the single task (S2) over time. Substantial switch times are crucial in order to enable meaningful task completion.

The results clearly show an *increased task-switching ability* of the *human operator* while the stress level (‘M(S)’) factor was not increased (ANOVA shows $F_{crit} > F$ comparing the two distributions of S with and without assistance). Performance on task switching, a paradigm commonly used to measure executive function, has been shown to improve with practice and as a consequence of knowledge [30] .

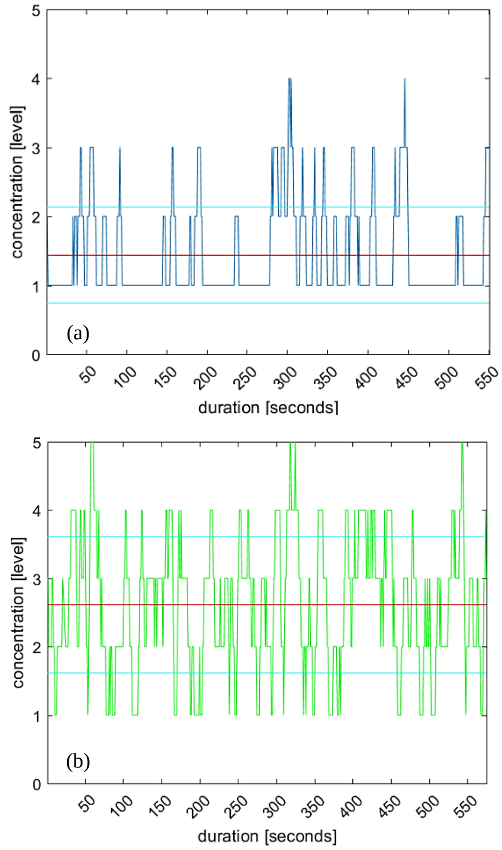


Figure 8: Measure of attentional concentration on tasks. (a) Concentration level during a session without assistance (red line is the mean; red is standard deviation), and (b) concentration level during the session with assistance. On average, the concentration increased when using the intuitive assistance.

D. Inferential Statistics and Discussion

An extensive correlation analysis was performed between real-time gaze based human factors analysis and the questionnaire results. The first important observation is the substantial correlation (Pearson r=-0.404) between the expected task switching time (‘ExpTS’) and the concentration level (‘M(C)’) which expresses the fact that an *increased concentration level* at the same time means an *increase in the task switching ability* (shorter expected durations between task switches with the same performance). The second observation is related to SART based situation awareness: M=63.0, SD=10.7% without and M=71.2, SD=16.1% with intuitive assistance, showing significant benefit in applying the assisting technology. *SART correlates positively with attentional resources on collaborative* AOIs (r=0.311 with ‘Collab’) and negatively on single tasks (r=-0.605 with ‘Single’).

Discussion. The conclusion with respect to the inferential statistics is that the assessment derived from real-time gaze based human factors analysis is capable to fully quantify the distribution of attentional resources on task relevant space, in real-time, and in this manner not only correlates but represents the performance of executive functions, i.e., the task switching ability in a human-robot collaborative scenario. This will enable in the future on-site measured attentional resources to represent standard questionnaires on situation awareness which provides a basis to evaluate basic human factors, such as, concentration of attention, complexity of the situation, familiarity of the situation, focusing of attention, information quantity, information quality, instability of the situation, variability of the situation, arousal, and spare mental capacity.

TABLE I. GAZE BASED OUTCOME MEASURES OF THE EXPERIMENTAL STUDY WITH RESPECT TO WITHOUT AND WITH INTUITIVE ASSISTANCE.

user groups	**Outcome measures**					
	M (C)	*M (S)*	*Collab**	*Single*	*ExpTS*	*AcGa*
group A	1.40±0.3	1.35±0.3	23.5	16.1	2.03±1.2	51.15
group B	1.94±0.5	1.52±0.4	125.0	29.8	1.60±0.6	23.7

*$\times 10^{-3}$. *C=concentration level; S=stress level; Collab=mean attentional time on mixed areas in seconds filtered by concentration, Single= mean time on single (human operator specific) areas in seconds filtered by concentration; ExpTS=expected time for next task switch; AcGa=mean time access gate observation per session.*

VII. CONCLUSION

We presented a novel methodology for the assessment of gaze based human factors which provides a potential to measure executive functions performance, such as, task switching ability, in real-time. Within a typical human-robot collaboration scenario and the study setup including absence/application of state-of-the-art intuitive assistance technology for the performance of collaborative tasks, we illustrated the potential of interpretation from gaze based human factors data in order to evaluate the MRC system.

Future work will include a study with a larger number of participants and evaluate the executive function measures under more competitive pressure for performance. Furthermore we will study the potential of more complex eye movement features to enable a more detailed analysis of the dynamic distribution of attentional resources during the tasks.

REFERENCES

[1] L. Tiberio, A. Cesta, and M. O. Belardinelli, "Psychophysiological methods to evaluate user's response in human robot interaction: a review and feasibility study", *Robotics*, 2(2):92–121, 2013.

[2] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich, „Common metrics for human-robot interaction", *Proc. ACM SIGCHI/SIGART Human-robot interaction*, 2006.

[3] C.-M. Huang, and B. Mutlu. "Anticipatory robot control for efficient human-robot collaboration", *Proc. ACM/IEEE HRI 2016,* 2016.

[4] S.J. Levine and B.C. Williams, "Watching and Acting Together: Concurrent Plan Recognition and Adaptation for Human-Robot Teams", *JAIR*, vol. 63, pp. 281–359, 2018.

[5] R. G. Freedman and S. Zilberstein, "Integration of Planning with Recognition for Responsive Interaction Using Classical Planners," in *Proceedings of AAAI* 2017, pp. 4581–4588.

[6] L. Johannsmeier and S. Haddadin, "A Hierarchical Human-Robot Interaction-Planning Framework for Task Allocation in Collaborative Industrial Assembly Processes," *IEEE Robotics and Automation Letters*, vol. 2, no. 1, pp. 41–48, Jan. 2017.

[7] M. Chen, S. Nikolaidis, H. Soh, D. Hsu, and S. Srinivasa, "Planning with Trust for Human-Robot Collaboration," in *Proceedings of the 2018 ACM/IEEE HRI'18*, Chicago, IL, USA, 2018, pp. 307–315.

[8] A. Roncone, O. Mangin, and B. Scassellati, "Transparent role assignment and task allocation in human robot collaboration" in *Proc. IEEE ICRA*, Singapore, Singapore, 2017, pp. 1014–1021.

[9] M. Fox, D. Long, and D. Magazzeni, "Explainable Planning", in *Proc. IJCAI-17 Workshop on Explainable AI (XAI)*, 2017, pp. 24–30.

[10] D. Kragic, J. Gustafson, H. Karaoguz, P. Jensfelt, and R. Krug, "Interactive, Collaborative Robots: Challenges and Opportunities," in *Proceedings of IJCAI*, Stockholm, Sweden, 2018, pp. 18–25.

[11] M. Salem, G. Lakatos, F. Amirabdollahian, and K. Dautenhahn, "Would You Trust a (Faulty) Robot?: Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust," in *Proceedings of the Tenth HRI '15*, 2015, pp. 141–148.

[12] C. Huang, C. Maya, and B. Mutlu, "Adaptive Coordination Strategies for Human-Robot Handovers", *Robotics: science and systems*, 2015.

[13] I. Maurtua, A. Ibarguren, J. Kildal, L. Susperregi and B. Sierra, "Human–robot collaboration in industrial applications: Safety, interaction and trust", *Intl. Journal of Advanced Robotic Systems*, vol. 14, no. 4, 2017.

[14] M. Fox and D. Long, "PDDL2.1: An Extension to PDDL for Expressing Temporal Planning Domains," *J. Artif. Intell. Res. (JAIR)*, 2003.

[15] M. Cashmore et al., "ROSPlan: Planning in the Robot Operating System," in *Proceedings of the Twenty-Fifth International Conference on Automated Planning and Scheduling*, 2015., 2015, pp. 333–341.

[16] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhusrt, H. Jarodzka, and J. van de Weijler. *Eye Tracking – A Comprehensive Guide to Methods and Measures*, Oxford University Press, 2011, pp. 187

[17] K. Moore, and L. Gugerty, "Development of a novel measure of situation awareness: The case for eye movement analysis," *Human Factors & Ergon. Soc. Ann. Meeting*, vol. 54:19, 2010

[18] S. M. Munn, and J.B. Pelz, "3D point-of-regard, position and head orientation from a portable monocular video-based eye tracker", *Proc. ETRA 2008*, pp. 181-188.

[19] K. Santner, G. Fritz, L. Paletta, and H. Mayer, "Visual recovery of saliency maps from human attention in 3D environments", *Proc. ICRA 2013*, pp. 4297-4303.

[20] C. Heyer, "Human-robot interaction and future industrial robotics applications," *Proc. IEEE/RSJ IROS*, pp. 4749–4754, 2010.

[21] R. M. Taylor, "Situational awareness rating technique (SART): The development of a tool for aircrew systems design", in *Situational Awareness in Aerospace Operations*, 1990, pp. 3/1–3/17.

[22] Stanton, N.A., P.M. Salmon, G.H. Walker, and D.P. Jenkins. Genotype and phenotype schemata and their role in distributed situation awareness in collaborative systems. *Theoret. Issues in Ergon.Sci*, 2009.

[23] Dini, A., Murko, C., Paletta, L., Yahyanejad, S., Augsdörfer, U., and Hofbaur, M. (2017) Measurement and Prediction of Situation Awareness in Human-Robot Interaction based on a Framework of Probabilistic Attention, *Proc. IEEE/RSJ IROS 2017*, 2017.

[24] Paletta, L., Pittino, N., Schwarz, M., Wagner, V., and Kallus, W., Human factors analysis using wearable sensors in the context of cognitive and emotional arousal, *Proc. 4th International Conference on Applied Digital Human Modeling*, AHFE 2015, July 2015.

[25] L. Paletta, M. Pszeida, R. Marton, T. Haspel, and Nauschnegg, B. (2019). Stress Measurement in Multi-Tasking Decision Processes Using Executive Functions Analysis, *Proc. AHFE 2019 International Conference on Human Factors and Ergonomics*, 2019

[26] Bailey, Brian P.; Konstan, Joseph A. (July 2006). On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior*. 22 (4): 685–708.

[27] Stephen Monsell, "Task Switching", *Trends in Cognitive Sciences*, Vol.7 No.3 March 2003.

[28] A. Miyake, N.P. Friedman, M.J. Emerson, A. Witzki, A. Howerter, T.D. Wager. "The unity and diversity of executive functions and their contributions to complex 'frontal lobe' tasks: A latent variable analysis". *Cognitive Psychology*, 41, 49-100, 2000.

[29] A.T. Jersild. "Mental set and shift". *Archives Psychol.*, 89.1927.

[30] Kramer, A. F., Hahn, S., & Gopher, D. (1999). Task coordination and aging: Explorations of executive control processes in the task switching paradigm. *Acta Psychologica*, 101, 339-378.