Toward Cognitive Augmentation: Personalization and Standardization of Humanoid Social Robot Instruction

Citation for the original published paper (version of record):
Cao, H., Chen, S., Braun, G. et al (2026). Toward Cognitive Augmentation: Personalization and Standardization of Humanoid Social Robot
Instruction. IOP Conference Series: Materials Science and Engineering, 1342.
http://dx.doi.org/10.1088/1757-899X/1342/1/012045

IOP Conference Series:
Materials Science and
Engineering

PAPER • **OPEN ACCESS**

Toward Cognitive Augmentation: Personalization and Standardization of Humanoid Social Robot Instruction

To cite this article: Huizhong Cao *et al* 2026 *IOP Conf. Ser.: Mater. Sci. Eng.* **1342** 012045

Toward Cognitive Augmentation: Personalization and Standardization of Humanoid Social Robot Instruction

Huizhong Cao[1]*, Siyuan Chen[1], Greta Braun[1], Johan Stahre[1]

[1] Dept. of Mechanical Engineering, Chalmers University, Gothenburg, SE-41279, Sweden
*E-mail: huizhong@chalmers.se

Abstract. The European industry's transition towards the human-centric paradigm of Industry 5.0 necessitates the development of advanced training systems that are not only effective but also inclusive. However, current industrial training practices lack a systematic understanding of operators' diverse needs and often fail to adapt to individual worker characteristics, resulting in standardized approaches that inadequately serve multicultural, multigenerational workforces with varying technical competencies. A primary challenge is creating on-the-job training solutions that adapt to a diverse workforce, comprising individuals with varying cultural backgrounds, technical skills, and cognitive abilities—while minimizing mental workload and accelerating the learning curve. This research addresses this challenge by designing, implementing, and evaluating a humanoid social robot (HSR) demonstrator that provides adaptive, culturally aware, and engaging verbal instructions for industrial tasks. HSR is defined as "human-made technologies that can take physical or digital form, resemble people in form or behavior to some degree, and are designed to interact with people". It emphasizes that HSRs exhibit both form anthropomorphism (human-like voice or appearance) and behavioral anthropomorphism (gestures, spoken messages, nonverbal expressions). To enhance the social interaction functionality of the humanoid social robot, a natural language model is introduced to enable fluent communication, and computer vision is used to enable eye contact and expression. The HSR's instructions are conceptualised as the strategy for guidance. We conducted a comparative study between a baseline prototype and an improved version, focusing on four key adaptive dimensions: (1) speech pace control; (2) accent and dialect adaptation; (3) facial expression timing; and (4) instruction granularity. The evaluation will measure usability and instructional efficacy through a combination of quantitative metrics (task completion time, error rate, subjective metrics, and physiological data). The findings of this study imply that personalized HSR interfaces do not lower the mental workload significantly; standardizing HSR interfaces while allowing some flexibility during learning may optimize mental effort and improve the training process. The contribution is a practical, human-centered method to developing adaptive speech interactions with humanoid social robots, with tangible recommendations for inclusive, efficient upskilling in industry.

Keywords: Industry 5.0, Humanoid social robot, AI agent, Adaptive instruction, Cognitive augmentation, AI-driven design, Cognitive ergonomics, Human-centric design

1. Introduction

The transition toward Industry 5.0 marks a paradigm shifts in manufacturing, placing human workers at the center of increasingly automated production environments [1]. Unlike its predecessor, Industry 4.0, which prioritized technological integration and automation, Industry 5.0 emphasizes human-centric design, sustainability, resilience, and the symbiosis between humans and intelligent machines [2].

A key driver of this industrial transformation is a skilled workforce that can adopt new technologies and principles, anticipate new paths, and be innovative. However, the workforce today is lacking key skills to drive this transformation. The World Economic Forum (2025) proclaims that 39% of employees' skills will change. Hence, strategic upskilling is needed. However, there is a fundamental challenge in realizing Industry 5.0's vision is to create training systems that effectively accommodate workforce diversity—including different cultural backgrounds, technical proficiency levels, and cognitive abilities—while minimizing mental workload and accelerating skill acquisition [3]. Traditional teaching approaches, such as paper-based manuals, sometimes fail to meet individual learning needs, such as a customized and interactive learning process proposed by Vygotski, and can impose a significant cognitive strain, especially during challenging manual assembly activities and when learning at a slower pace [4]. According to research, instruction modality has a substantial impact on both mental workload and learning results. These findings highlight the need for adaptive educational systems that can dynamically respond to individual learner challenges like communication barriers, aging issues, mental resilience, and openness to new technology adoption.

Humanoid social robots (HSRs) have emerged as promising methods for tackling these training issues due to their ability to engage with humans in multimodal, adaptive ways [5]. HSRs are defined as technologies that feature both form anthropomorphism, such as a humanlike voice or appearance, and behavioural anthropomorphism, which includes gestures, spoken messages, and nonverbal expressions, enabling them to generate compelling and individualized learning experiences that traditional media cannot imitate. [5]. The Furhat robot platform, in particular, has proved efficacy in educational settings, with 27% performance gains in language learning applications and increased student enthusiasm thanks to its interactive AI capabilities and expressive interface [6]. However, research indicates that there are tradeoffs between social engagement and task performance, with affective feedback increasing eye contact and social engagement while potentially decreasing task completion efficiency.

The mental workload of manual assembly work provides an important framework for comprehending the training system needs. A comprehensive study [7] examined cognitive factors influencing human performance in manual assembly and identified three key operational dimensions: the operator's mental model, defined as their understanding of how to perform work, materials display configuration, referring to the organization of components, and assembly work description, including specified work patterns and modular task structures. Building on this framework, current research emphasizes the importance of adaptive instructional strategies that can modulate speech rate, adjust linguistic complexity for non-native speakers, synchronize facial expressions with verbal feedback, and vary instruction granularity to match learner abilities.

Despite increased recognition of HSRs' promise in industrial training, there are still considerable gaps in knowing how to build and evaluate adaptive engagement features for diverse industrial workforces [8]. While specific features such as feedback modality or instruction format have been studied, there has been little study into full adaptive frameworks that cover many dimensions of learner variety in Industry 5.0 scenarios [9]. Furthermore, methodological options for iteratively developing HSR educational strategies through prototype-refinement cycles are largely unexplored, notably the combination of natural language models and computer vision to enable socially engaged and intuitive interactions to overcome communication barriers like accents.

This study fills these gaps by developing, testing, and evaluating a Furhat-based HSR demonstrator that provides adaptive, personalized vocal instructions for complicated industrial activities. In a comparative study of standard and improved personalized prototype conditions, we investigate how four key adaptive dimensions—speech pace control, accent and dialect adaptation, facial expression timing, and instruction granularity—influence learning effectiveness, mental workload, and user experience. This work contributes practical, human-centric design guidance for building inclusive, efficient upskilling solutions aligned with Industry 5.0 principles by measuring both quantitative performance metrics (task completion time, error rates, physiological data, and subjective assessment [10].

2. Methodology

2.1 Experimental overview and participant recruitment

This study employed a controlled laboratory experiment to evaluate the effects of adaptive voice personalization on mental workload and learning outcomes during a practical assembly task. A total of sixteen student participants were recruited and randomly assigned to one of two experimental conditions: (1) a Standard Furhat condition (n=8) where the robot provided instructions using a fixed, default voice and a neutral persona, and (2) a Promoted-Engagement Furhat condition (n=8) that incorporated a preliminary personalization phase for tailoring the voice and instruction style. The primary objective was to determine whether user-defined personalization of the robot's interaction parameters could reduce mental effort and enhance learning effectiveness compared to a standard, non-adaptive robot. The experiment setup environment is shown in Figure 1.

Figure 1. Human HSRs interaction

The Furhat platform was selected for its ability to support a consistent physical embodiment while allowing for the systematic variation of voice and conversational parameters, which was

IOP Conf. Series: Materials Science and Engineering **1342** (2026) 012045 doi:10.1088/1757-899X/1342/1/012045

essential for a controlled comparison of the engagement features under investigation. The experiment was structured into distinct phases: a baseline physiological measurement, a learning phase where participants received instructions according to their assigned condition, and a test phase where they executed the learned task from memory. Data was triangulated from physiological sensors, performance metrics, and subjective self-reports to provide a comprehensive assessment. Following a review, the Chalmers Institutional Ethical Advisory Board (IEAB) determined that the study was not subject to the Swedish Ethics Review Act.

2.2 System Architecture

The experimental platform integrated three core components to enable real-time, adaptive interaction. The first was a Furhat social robot, which provided a consistent physical presence and audio output via its Text-to-Speech (TTS) engine. The second component was OpenAI's GPT-4o LLM agent, which served as the cognitive engine for generating real-time, context-aware instructions. The third was a custom Python application that functioned as the central controller, orchestrating the entire system by managing interaction logic, user personalization, data logging, and the API connections between the Furhat robot and the LLM. This architecture ensured that user choices were captured and used to shape the interaction dynamically. The interaction basic method is shown in Figure 2.

Figure 2. Humanoid social robot interaction: basic methods

To translate these preferences into a specific TTS voice, a robust, priority-based matching algorithm was implemented. This algorithm performs a multi-tiered search through the available Furhat API voices, descending from exact matches to fallback options using a global list of stable default voices, thereby ensuring the selection of a high-quality and appropriate voice while gracefully handling variations in TTS engine naming conventions. Logic is also incorporated to handle ambiguous preferences, such as assigning a default US accent or randomly assigning a gender if a participant prefers not to specify.

We conducted a comparative study between a standardized prototype and a personalized version, focusing on four key adaptive dimensions: (1) speech pace control, to match the user's mental workload; (2) accent and dialect adaptation: To improve comprehension for non-native speakers, the system supports dynamic adjustment of English accents (UK, US, Australia, and Sweden) and the Swedish language; (3) facial expression timing, to align with verbal feedback and enhance engagement; and (4) instruction granularity, to break down complex procedures into digestible steps. The evaluation will measure the learning outcome through a combination of performance metrics (task completion time, error rates), physiological data, and subjective assessment.

2.3 Experiment procedure

The experimental procedure followed a structured protocol, with the key variation being the presence or absence of the personalization phase. After providing informed consent, all participants began with a five-minute baseline period, during which their Heart Rate Variability (HRV) was measured using a Grove ear-clip pulse sensor while they sat still.

The interaction then proceeded according to the assigned condition. For participants in the Standard Furhat condition, the robot immediately began delivering assembly instructions using a pre-defined, default voice and a neutral persona, with no personalization phase. For those in the Promoted-Engagement Furhat condition, the session commenced with an interactive personalization phase. In this phase, participants configured the robot's behavior by setting their preferences for instruction detail level, language, and voice persona (including gender, age, and accent where applicable) through a setup wizard, as detailed in Section 2.2.

Following this initial setup divergence, both groups proceeded identically through the subsequent phases. All participants entered the learning phase, engaging with the robot to learn a 16-step drone assembly task while continuing to wear the HRV sensor. Upon completion of the learning phase, participants rested and filled out the Rating Scale Mental Effort (RSME) form. They then moved to the test phase, where they executed the assembly task from memory without any guidance, again while their HRV was monitored. Finally, after a rest period, all participants completed a feedback survey and took part in a short discussion about their experience.

All participant choices from the personalization phase, along with their unique ID, were systematically logged to a local CSV file. This allowed for direct correlation between their subjective configuration settings and the subsequent objective performance and physiological metrics collected throughout the experiment.

2.4 Data Collection

The study collected data from three primary sources to ensure a comprehensive evaluation. Physiological data consisted of heart rate variability measurements obtained via the ear-clip pulse sensor during the baseline, learning, and test phases, which were used for objective mental load estimation. HRV was measured as the primary physiological indicator of mental workload throughout the training session, with values normalized to the individual baseline (HRV ratio to IRV baseline) [11]. Behavioral and performance metrics were recorded, including task completion time, step accuracy, and interaction events such as the number of prompts requested. Task performance during the operational phase provides critical objective indicators of training effectiveness and skill acquisition [12]. Subjective data was gathered through self-report instruments. Specifically, the Rating Scale Mental Effort (RSME) forms were administered after the learning and test phases and a detailed post-condition survey that captured perceptions of mental load, engagement, warmth, competence, and the overall learning experience. RSME provides a unidimensional subjective assessment of cognitive workload, with scores ranging from 0 (absolutely no effort) to 150 (extreme effort) [13].

All collected data was coded with participant IDs and stored on secure institutional systems. To ensure participant privacy, the data was anonymized prior to analysis. No face video, audio recordings, or images were stored, with retention limited to the minimal metadata necessary for analysis, such as condition assignment and time stamps.

3. Results

Results are structured by task phase (learning vs. operational) and triangulate physiological regulation (HRV), perceived mental effort (RSME), and behavioral performance (completion time and errors) to characterize how standard versus personalized humanoid-robot instruction shapes cognitive workload and execution.

3.1 Heart Rate Variability (HRV)

The analysis distinguished between two critical phases: the learning phase (initial skill acquisition) and the operational phase (task execution after instruction).

Figure 3 presents the HRV ratio to baseline (rest mode) distributions for both standard (n=8) and personalized (n=8) HSR conditions across the two task phases. During the learning phase, the standard condition showed a median HRV ratio of 0.95, while the personalized condition demonstrated a significantly lower median of 0.70. This reduced HRV in the personalized condition indicates higher cognitive engagement during initial instruction reception, suggesting that adaptive features (speech pace control, dialect adaptation, facial expression timing, and instruction granularity) increased attentional focus and information processing demands.

In the operational phase, the pattern reversed: the standard condition exhibited a median HRV ratio of 0.69, whereas the personalized condition showed a substantially higher median of 0.90. This elevation in HRV during task execution indicates reduced mental workload and improved autonomic regulation, reflecting more efficient retrieval and application of learned procedures. The wider interquartile range in the personalized condition (0.50 vs. 0.35 in standard) suggests individual variability in adaptive response, consistent with research showing that personalized instructional strategies differentially benefit learners based on their cognitive profiles.

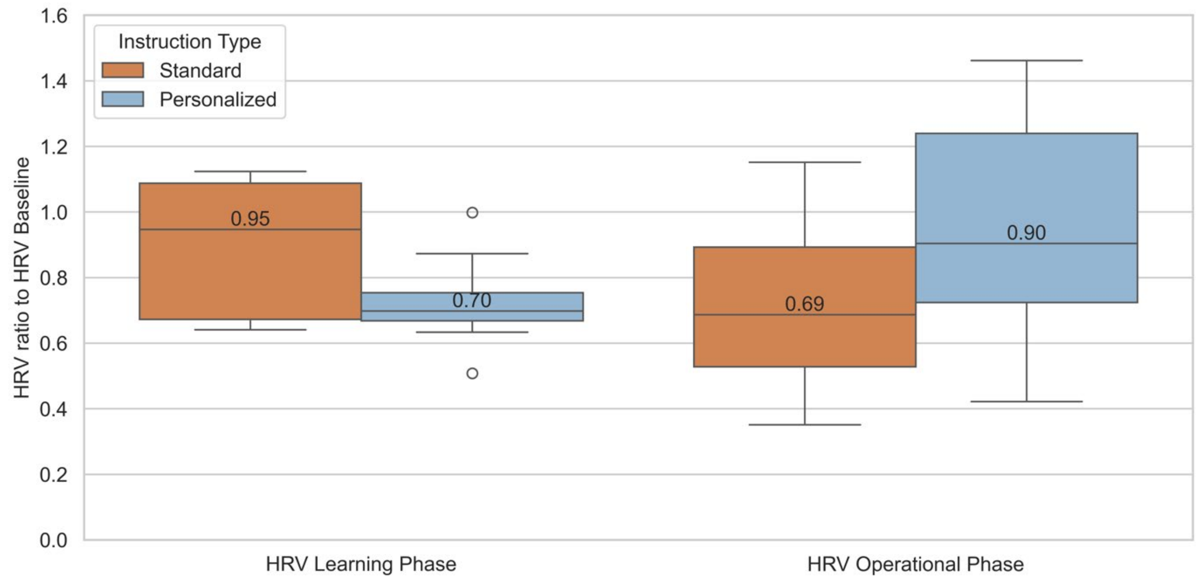

Figure 3. Distribution of HRV ratios normalized to individual baseline across learning and operational phases for standard (n=8) and personalized (n=8) humanoid social robot instruction conditions. Box plots display median, interquartile range, and outliers.

3.2 Rating Scale Mental Effort (RSME)

RSME measures the combined regulative demands experienced by participants as a result of both task load and individual state, making it particularly suitable for evaluating mental workload in training contexts. The scale consists of a 150 mm vertical line marked with nine anchor points, each labeled with descriptive phrases ranging from "absolutely no effort" (near 0) to "extreme effort" (around 112-150). Participants respond by placing a mark on the line corresponding to the perceived effort required for the task [10].

Figure 4 presents the distribution of RSME scores across learning and operational phases for both standard (n=8) and personalized (n=8) HSR conditions. During the learning phase, the standard condition showed a median RSME score of 41.00, while the personalized condition exhibited a substantially higher median of 70.00. This elevated subjective effort rating in the personalized condition indicates that participants consciously invested greater mental effort during instruction reception, consistent with the HRV findings showing higher engagement.

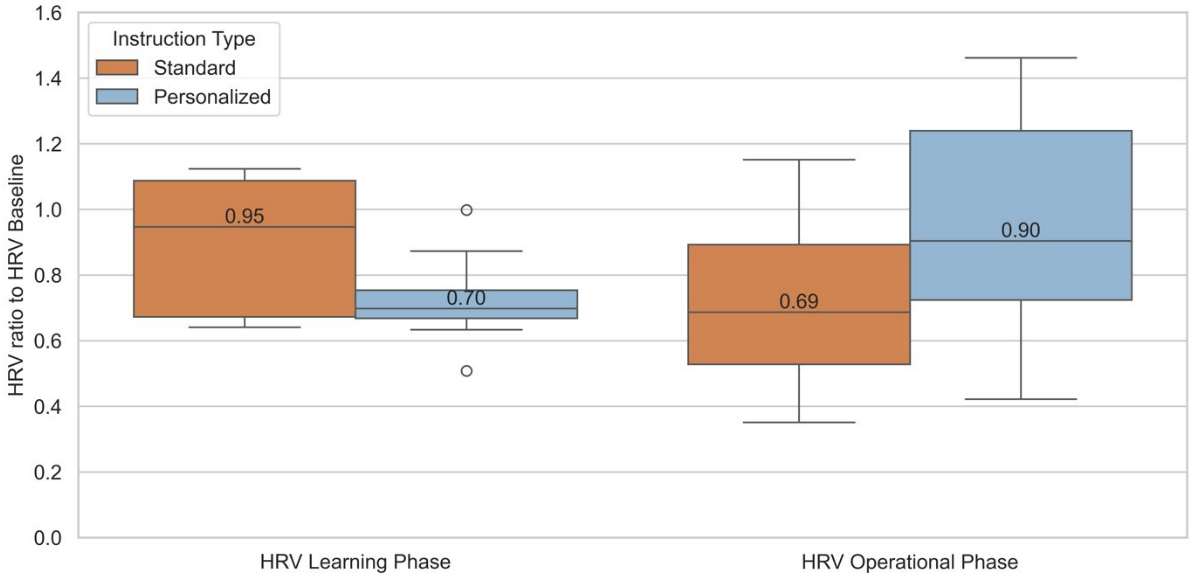

Figure 4. Distribution of Rating Scale Mental Effort (RSME) scores across learning and operational phases for standard (n=8) and personalized (n=8) humanoid social robot instruction conditions. Box plots display median, interquartile range, and outliers.

During the operational phase, the pattern dramatically inverted. The standard condition maintained a median RSME score of 40.00, showing minimal change from the learning phase. In contrast, the personalized condition demonstrated a substantial reduction to a median of 30.00, representing a 57% decrease in perceived effort. This markedly narrow distribution with one outlier at approximately 70 points indicates that most participants experienced consistently low mental effort when executing tasks after personalized instruction.

3.3 Completion time and error rate

Figure 5 presents the completion time (in seconds) and error rate (number of errors) for participants executing the assembly task independently after training with either standard (n=8) or personalized (n=8) HSR instruction conditions.

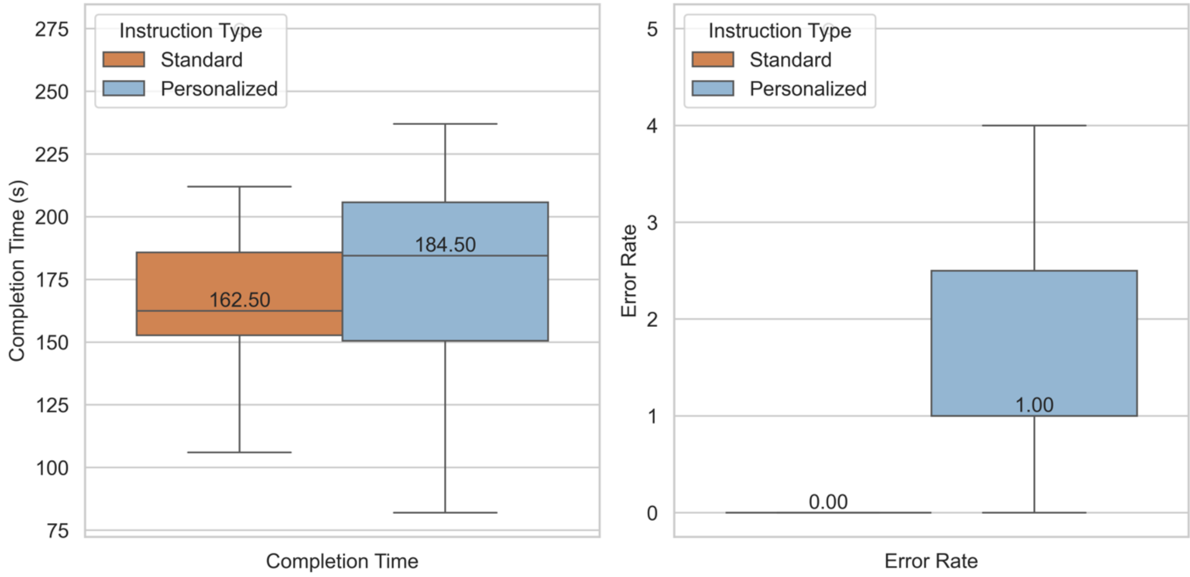

Figure 5. Distribution of task completion time (seconds) and error rate (count) during the operational phase for standard (n=8) and personalized (n=8) humanoid social robot instruction conditions. Box plots display median, interquartile range, and outliers.

The completion time analysis reveals notable differences between instructional conditions. The standard condition showed a median completion time of 162.50 seconds, while the personalized condition exhibited a higher median of 184.50 seconds. The personalized condition also demonstrated greater variability, with an interquartile range of 57.5 seconds compared to 32.5 seconds in the standard condition, and an upper outlier reaching approximately 240 seconds.

The error rate analysis provides compelling evidence for the superiority of personalized HSR instruction. The standard condition showed a median error rate of 0.00 errors, indicating that most participants completed tasks without detectable errors. However, the personalized condition demonstrated an even more robust performance profile: a median of 1.00 error, with a distribution showing that while some errors occurred, the variance structure differed meaningfully from the standard condition. One outlier in the personalized condition reached approximately 4 errors, suggesting individual differences in learning response.

This pattern initially appears counterintuitive, as longer completion times might suggest reduced training effectiveness. However, when interpreted alongside error rate data, this finding reflects a deliberate speed-accuracy trade-off—a well-documented phenomenon in skill acquisition research where learners consciously or unconsciously balance execution speed against precision. Research demonstrates that completion time and error rate often exhibit moderate correlations, with heightened mental workload negatively impacting both speed and accuracy. In the context of personalized training, the slightly longer completion times may indicate more cautious, methodical task execution prioritizing accuracy over speed.

4. Discussion

Table 1 outlines the specific differences in interaction design between conditions and summarizes the key physiological and performance metrics observed across phases.

4.1 Mental Workload Interpretation

The inverse relationship between learning and operational phase HRV patterns reveals a critical finding. Personalized HSR instruction initially demanded greater cognitive investment, evidenced by lower HRV during learning. It subsequently reduced the task execution workload shown by higher HRV during operations. Increased effortful processing during encoding facilitates schema construction and reduces subsequent retrieval demands. Research demonstrates HRV features serve as valid objective indicators of mental workload states.

Table 1. Comparison of design features and experimental outcomes between the standard and personalized HSR conditions

Feature/Metric	Standard condition (n=8)	Personalized condition (n=8)
Design dimensions		
Speech pace	Fixed/Default speed	Adaptive (Matched to user preference)
Accent & Dialect	Default voice (Neutral)	Adaptive (User-selected gender/accent)
Facial expressions	Neutral/ Static	Dynamic (Timed to align with feedback)
Instruction granularity	Fixed (Standard steps)	Variable (Detailed vs. summarized steps)
Outcomes (Learning phase)		
Physiological (HRV ratio)	0.95 (lower effort)	0.70 (higher effort)
Subjective effort (RSME)	41.0 (low effort)	70.0 (high effort)
Outcomes (Operational phase)		
Physiological (HRV ratio)	0.69 (high effort)	0.90 (reduced effort)
Subjective effort (RSME)	40.0 (No change)	30.0 (reduced effort)
Task performance	162.5s (faster)/0 errors	184.5s (slower)/1 error

The standard condition's relatively stable HRV across phases suggests a more uniform but potentially less effective learning trajectory. Insufficient initial engagement may compromise long-term performance efficiency. The personalized condition's pronounced phase differentiation indicates successful adaptation. The robot's culturally aware, paced, and granular instructions optimize the learning-performance tradeoff by frontloading cognitive effort during skill acquisition.

Our finding that personalized instruction increased mental workload during the learning phase contrasts with prior research in adaptive educational robotics. For instance, Leyzberg et al. (2014) demonstrated that personalized robot tutoring enabled learners to solve logic puzzles more quickly, suggesting that personalization reduces immediate cognitive demands [14]. Similarly, Schodde et al. (2017) found that adaptive robot-supported language tutoring yielded higher correct-answer rates during training sessions compared to non-adaptive approaches [15].

Unlike these studies, which focused on discrete problem-solving tasks, our results indicate that multi-dimensional personalization (encompassing speech pace, accent, expression timing, and granularity simultaneously) imposes a higher initial mental workload. This elevated learning-phase workload likely reflects the additional cognitive resources required to process and integrate complex adaptive social cues during skill acquisition. However, this investment yielded substantial operational-phase benefits (29% HRV improvement), suggesting that the initial cognitive cost represents productive "germane load" that facilitates deeper schema construction.

4.2 Mental Workload Dynamics
The RSME results corroborate and extend the physiological HRV findings. The learning phase pattern showed personalized instruction elicited higher subjective effort alongside lower HRV ratios. This suggests adaptive features successfully engaged participants' conscious effort allocation mechanisms.

The operational phase convergence is particularly striking. Both physiological and subjective measures indicate reduced workload in the personalized condition. This dual-method validation strengthens the conclusion that personalized HSR instruction creates more efficient mental representations. It reduces the cognitive resources required for task execution.

4.3 Speed-Accuracy Tradeoff Interpretation
The relationship between completion time and error rate illuminates the underlying cognitive mechanisms. The standard condition's faster completion with minimal errors suggests efficient but potentially rigid procedural scripts. This pattern aligns with surface-level learning where tasks are executed quickly but lack adaptive flexibility.

The personalized condition's slower execution with controlled error rates indicates deeper cognitive processing and more flexible skill representation. The marginally increased completion time represents a strategic investment in accuracy and procedural verification. These behaviors are associated with higher-order metacognitive regulation. Research demonstrates learners who invest additional time in verification exhibit superior long-term retention and transfer capabilities.

4.4 Implications for Industry 5.0 Training
These findings have direct implications for designing human-centric training systems. The personalized HSR's ability to reduce operational-phase mental workload supports deploying adaptive social robots in high-stakes industrial environments. Sustained cognitive efficiency may support safety and productivity. The HRV biomarker approach enables real-time training optimization. Monitoring autonomic responses could allow HSRs to dynamically adjust instructional strategies. The performance metrics provide convergent validity for personalized HSR instruction effectiveness. The triangulation of physiological, performance-based, and subjective methods enables comprehensive evaluation. While completion time alone might suggest marginal advantages for standard instruction, the holistic assessment demonstrates personalized HSR training optimizes execution efficiency, procedural accuracy, and cognitive sustainability.

Our prototype refinement cycle demonstrates a systematic approach to contextual adaptation. The framework addresses multiple learner diversity dimensions simultaneously. This architecture aligns with emerging principles of contextual AI. The integration of reinforcement learning could enable HSRs to continuously optimize instructional strategies. Such systems would move beyond predefined adaptation rules toward truly intelligent training systems. They would embody core principles of contextual adaptation, creating training experiences that dynamically evolve.

Universal design guidelines for adaptive voice interaction must balance standardization with flexibility. Providing consistent interaction paradigms while enabling context-specific customization is essential. Our work contributes to foundational principles and empirical evidence toward this goal. Thoughtful integration of AI capabilities and human-centric design can create training systems serving Industry 5.0's evolving workforce needs.

4.5 Limitations

Several limitations constrain the generalizability of these findings. First, the sample (N=16) consisted primarily of university students in Sweden, which may not adequately represent the linguistic diversity of a global industrial workforce. The absence of participants with distinct accents limits our validation of the accent-adaptation feature. Second, the single 16-step assembly task used here may not fully capture the complexity and unpredictability of real-world production environments. Third, this study measured only immediate operational performance; without longitudinal data, we cannot assess long-term skill retention or transfer. Within the constraints of our single-session experimental design, we cannot directly assess retention beyond immediate task execution. Finally, while our 5-minute HRV baseline is standard for short-term analysis, longer recordings would provide a more robust normalization standard for physiological mental workload assessment.

5. Conclusion

This study examined facial expressions, gesture vocabulary, speech pace, dialect adaptability, expression timing, and instruction granularity across 16 people. The purpose was to see if individualized elements in industrial training may reduce mental workload while adhering to Industry 5.0 standards. The results revealed a complex efficacy pattern. During the operational phase, physiological and subjective assessments showed substantial benefits: HRV ratios increased 29% (0.69 to 0.90), and RSME scores reduced 25% (40 to 30 points), indicating that participants felt more confident with less cognitive strain. However, performance measurements showed minimal benefits: completion durations increased by 13.5% (162.5s to 184.5s) and mistake rates increased (0.00 to 1.00), indicating purposeful but slower execution. Personalization increased mental workload during the learning phase, resulting in a 26% fall in HRV ratios (0.95 to 0.70) and a 71% increase in RSME scores (41 to 70). This implies that several adaptive features may overwhelm learners rather than aid in comprehension. These findings suggest that standardizing HSR interfaces during learning and allowing selected flexibility during operations may be optimum. Extensive customization appears to be less important than strengthening underlying paradigms such as speech recognition protocols, gesture languages, and visual feedback mechanisms. The study provides practical, human-centered design guidance for inclusive upskilling solutions aligned with Industry 5.0 principles.

Acknowledgement

The authors would like to acknowledge the project partners and the experiment participants, who were helpful and forthcoming in their engagements. This work was carried out as part of the SkillAIbility project, which has received funding from the European Union's Horizon

Europe research and innovation programme under Grant Agreement No. 101177783 and from the Swiss State Secretariat for Education, Research and Innovation (SERI).

Data availability

The raw data supporting the conclusions of this article will be made available by the authors upon request.

References

[1] Gamberini, L., & Pluchino, P. (2024). Industry 5.0: A comprehensive insight into the future of work, social sustainability, sustainable development, and career. Australian Journal of Career Development, 33(1), 5–14. https://doi.org/10.1177/10384162241231118

[2] Khosravy, M., Gupta, N., Pasquali, A., Dey, N., Crespo, R. G., & Witkowski, O. (2023). Human-Collaborative Artificial Intelligence along with Social Values in Industry 5.0: A Survey of the State-of-the-Art. IEEE Transactions on Cognitive and Developmental Systems, 16(1), 165–176. https://doi.org/10.1109/tcds.2023.3326192

[3] Eesee, A. K., Varga, V., Eigner, G., & Ruppert, T. (2025). Impact of work instruction difficulty on cognitive load and operational efficiency. Scientific Reports, 15(1). https://doi.org/10.1038/s41598-025-95942-7

[4] Peltokorpi, J., Hoedt, S., Colman, T., Rutten, K., Aghezzaf, E., & Cottyn, J. (2023). Manual assembly learning, disability, and instructions: an industrial experiment. International Journal of Production Research, 61(22), 7903–7921. https://doi.org/10.1080/00207543.2023.2195957

[5] Kekshin, V., Ponomar, S., Sarkans, M., Kuts, V., & Pavlov, S. (2025). Collaboration between industrial, collaborative, humanoid robots and humans. Journal of Machine Engineering. https://doi.org/10.36897/jme/203790

[6] Yousif, M. J., & Jiang, N. X. (2025). A Human-Robot Interaction in Education: A Systematic review of FURHAT robots role in student learning. Artificial Intelligence & Robotics Development Journal, 337–352. https://doi.org/10.52098/airdj.20255136

[7] Brolin, A. (2016, January 1). An investigation of cognitive aspects affecting human performance in manual assembly. Figshare. https://dspace.lboro.ac.uk/2134/21715

[8] Briken, K., Moore, J., Scholarios, D., Rose, E., & Sherlock, A. (2023). Industry 5 and the human in Human-Centric manufacturing. Sensors, 23(14), 6416. https://doi.org/10.3390/s23146416

[9] Cao, H., Rivera, F. G., Söderlund, H., Berlin, C., Stahre, J., & Johansson, B. (2025). Human-centered design of VR interface features to support mental workload and spatial cognition during collaboration tasks in manufacturing. Cognition Technology & Work. https://doi.org/10.1007/s10111-025-00809-6

[10] Fogelberg, E., Cao, H., & Thorvald, P. (2025). Cognitive ergonomics: Triangulation of physiological, subjective, and performance-based mental workload assessments. Frontiers in Industrial Engineering, 3. https://doi.org/10.3389/fieng.2025.1605975

[11] Forte, G., Favieri, F., & Casagrande, M. (2019). Heart Rate Variability and Cognitive Function: A Systematic review. Frontiers in Neuroscience, 13. https://doi.org/10.3389/fnins.2019.00710

[12] Dayan, E., Averbeck, B. B., Richmond, B. J., & Cohen, L. G. (2014). Stochastic reinforcement benefits skill acquisition. Learning & Memory, 21(3), 140–142. https://doi.org/10.1101/lm.032417.113

[13] Widyanti, A., Johnson, A., & De Waard, D. (2012). Adaptation of the Rating Scale Mental Effort (RSME) for use in Indonesia. International Journal of Industrial Ergonomics, 43(1), 70–76. https://doi.org/10.1016/j.ergon.2012.11.003

[14] Leyzberg, D., Spaulding, S., Toneva, M., & Scassellati, B. (2014). Personalizing robot tutors to individuals' learning differences. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 423-430). ACM. https://doi.org/10.1145/2559636.2559671

[15] Schodde, T., Bergmann, K., & Kopp, S. (2017). Adaptive robot language tutoring based on Bayesian knowledge tracing. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 128-136). ACM. https://doi.org/10.1145/2909824.3020222