What Level of Power Should We Give an Automation?

—Adjusting the Level of Automation in HCPS —

Mehrnoush Hajnorouzi[1(✉)], Astrid Rakow[1], Akhila Bairy[2], Jan-Patrick Osterloh[1], and Martin Fränzle[2]

[1] Institute of Systems Engineering for Future Mobility, German Aerospace Center (DLR) e.V., Oldenburg, Germany
{mehrnoush.hajnorouzi,astrid.rakow,jan-patrick.osterloh}@dlr.de
[2] Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany
{akhila.bairy,martin.fraenzle}@uni-oldenburg.de

Abstract. The level of automation in human-centered systems is steadily increasing, leading to a demand for advanced design methods for automation control at the human-machine interface. This is particularly important in safety-critical applications, where the multi-faceted interaction between the automated system and humans must be carefully analyzed to identify potential risks to the overall safety. This paper presents our vision of an approach determining an appropriate level of automation taking into account the automation's impact on the human. The approach is based on a game theoretic framework where we investigate whether the automation's controller can be synthesized as a strategy considering human behavior and thus ensuring human-adaptive control.

1 Introduction

The increasing automation of human-centered cyber-physical systems (HCPS) requires advanced control technologies that not only control technical tasks, but must also interact with and support users. Adaptation to the users' explicit requirements and implicit needs is important for these systems on the one hand, but on the other hand they must not hinder users. Reconciling these goals is particularly intricate in safety-critical domains. Safety cannot be treated as an ad hoc measure but, as emphasized by Bowen, "safety must be designed into a system and dangers must be designed out of it" [6, p.4]. The importance of "the right level of automation" can be nicely illustrated e.g. in the road transportation domain. There automated vehicles (AVs) promise to reduce the number of accidents caused by driver errors and enhancement of the transport efficiency [8]. Nowadays, modern vehicles offer automated driver-assistance systems for e.g. lane-keeping and blind-spot warning. These systems have the power to intervene in a safety-critical way. One of the design challenges is to determine what control actions have to be taken and when. A lane keeping assistance system should select an appropriate level of counterforce when reminding drivers not to leave the lane. While too much counterforce might hinder drivers (e.g. to do an evasive

B. Sangchoolie et al. (Eds.): EDCC 2024 Workshops, CCIS 2078, pp. 14–21, 2024.
https://doi.org/10.1007/978-3-031-56776-6_2

maneuver, especially when they are unfamiliar with the system), applying not enough force might not be noticed (e.g. when there are strong winds).

In this paper, we outline our vision of an approach that supports the early design of human-centered automation systems. We analyze the level of permissible interference scenario-wise by varying what control actions the automation can choose and when. We then determine whether applying these control actions in the given circumstances suffice to achieve the mission goal. To this end, we define for each variant a reactive game [17,25] of the automation system interfacing with the user. The control strategies synthesized for this game accomplish the control objective. Since they also take the human reactions into account, they implement the *shared control paradigm* [24] creating a synergistic control in which both the human and the automation contribute to the task. Our approach considers the impact of the automation's control on the human by using psychological models of the human mind. Since trying to capture human's mind necessarily results in a coarse approximation, the presented approach gives guidance rather than an implementable control strategy.

We describe the approach to determine the level of automation for shared control in Sect. 2. At the core of our methodology is a comprehensive understanding of human behavior. We achieve this by modeling based on cognitive architectures, which is described in Sect. 3. Examining the landscape of existing models reveals valuable insights into their individual strengths and limitations. A single architecture is not capable of capturing a broad spectrum of human behavioral aspects. To refine our understanding, we propose to integrate a combination of different human models. We conclude in Sect. 4.

2 Determining the Level of Power

To mitigate potential hazardous situations, an AV can usually apply a spectrum of responses, from activating an audible alarm to applying the brakes. Our approach determines whether a design variant can successfully adapt its control to the human and the current situation so that the operational requirements are met without hindering the human. Thereby, the approach can help designers choose between the variants that apply different level of control.

Control Synthesis. To implement the shared control paradigm, our approach uses control synthesis in a timed game between the automation system and the user. The strategy for controlling the automation system is synthesized while the actions of the human and those of the environment are uncontrollable. In this way, the human is not hindered and the synthesized controller implements shared control. A successfully synthesized strategy guarantees that the controllable actions are applied in such a way that the control objective is achieved, even in an uncooperative environment with maximum interference. This interference is modelled by introducing uncontrollable actions that formalize all possible forms of interfering. Therefore, the control objectives of automation system are captured as winning conditions of the game. For now, we will use winning conditions that can be expressed in terms of sets of states that are either desirable or

undesirable. The reachability synthesis problem is hence important. It is about identifying a maximal subset of states and transitions in the game graph that lead into the desired states and/or avoid undesired states [17,20]. The latter corresponds to the safety control problem, where the goal is to consistently avoid a set of predefined undesired states. In a controllable game, the current state is within the set of winning states for the control player. Various algorithms to compute the set of winning states under different winning conditions are discussed in [17]. These computations yield an explicit winning strategy alongside. The problem of control synthesis can thus be reduced to computation of a winning strategy for the control player.

On the Need for a Human Model. To determine what control actions the controller should take when, we must model their impacts on the human and the technical system within their environment. While the impact on the technical system is a well-studied problem, our work focuses on the impact on the human. This aspect is currently less explored in the existing literature [15]. Our research therefore aims to fill this gap by providing specific considerations related to humans in the design of automation. We aim to explore the intricate joint dynamics of humans and automation systems in interaction through formal modelling. The objective is to determine the degree of automation, delineating the actions undertaken by the controller. The analysis is envisioned to be performed during the system design phase to facilitate the derivation of HCPS specifications. Moving along the levels of automation thoughtfully relies on a comprehensive understanding of human behavior. The inclusion of a human model becomes imperative. We use cognitive architectures as sources of human behavior models. Such models let us predict human behavior, thereby specifying what uncontrollable actions our controller has to face.

Formal Model of Human Behavior. Our approach necessitates obtaining a formalization that is amenable to game-based reactive synthesis. The construction presented by [13,14] manually translates an approximation of ACT-R cognitive model (cf. Sect. 3 for ACT-R) into a network of timed automata. To streamline the integration and translation process, we propose using model learning methods like Angluin's L* algorithm [3]. We adopt the concept of behavior from discrete-event systems [7], where the behavior is described by a temporally ordered sequence of events. Accordingly, the behavior of a dynamic system can be described as a language. Finite automata and their (ω-)regular languages are one of the convenient candidates to specify the dynamic systems. An automatic translation of arbitrary computational models into such an automata-theoretic modelling framework can be achieved by automata-learning algorithms, which provide mechanisms to derive finite automata approximating the target language with a specified degree of accuracy from finite samples. We propose using cognitive architectures as source for our human model. Moreover, we propose integrating the behavior of different psychological models of humans within a human model, HM. As illustrated in Fig. 1, various models are simulated using an identical scenario. The generated traces are then fed into a learning algorithm, in order to integrate them into one comprehensive and automata-based

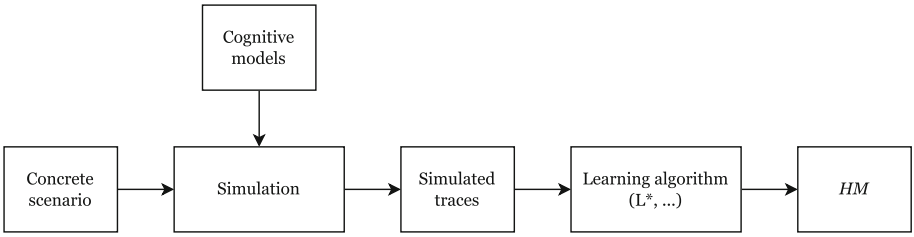

Fig. 1. Framework to learn a model of human *HM*.

human model. To enhance model learning and avoid over-fitting, we incorporate a fleet of different human models. Each model represents slight variations resulting from distinct valuation of adjustable parameters or alterations within the distributions of noisy selections. This approach enables the *HM* to capture a spectrum of possible behaviors.

Adapting Automation. The overall framework for adapting the power of automation is illustrated in Fig. 2. It describes a process that runs through a list of design variants and evaluates whether the current variant can achieve the control objective by applying its specific level of the power. On the far left, the game graph is represented that describes the HCPS. It is composed of the learned human model *HM* (including the environment) whose actions are uncontrollable, and the models of CPS components, whose actions are controllable. This game graph and the winning condition of the game, which captures the control objectives of the automation, are examined to see if it is possible to synthesize a winning strategy for the automation. If this is possible, the synthesis algorithm generates a winning strategy. However, if controlling the game is infeasible, the process ensues to update the set of control actions. The iteration process stops if either a winning strategy can be synthesized or if the last variant of control actions has been explored. If a winning strategy could be synthesized an

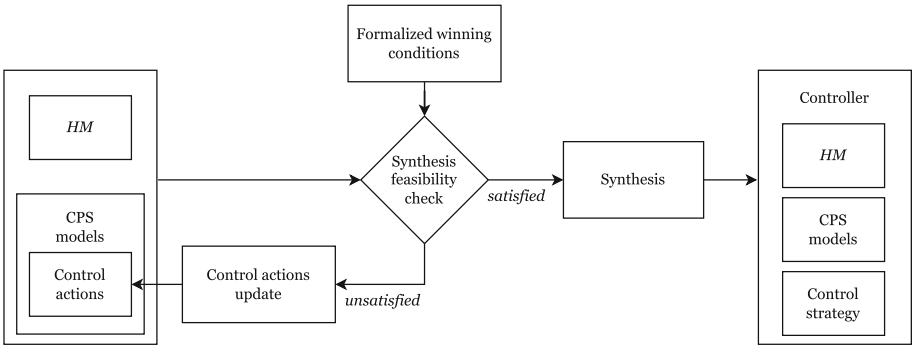

Fig. 2. Framework to design interactive controller.

adaptive design [12] has been achieved, which takes the human into account. The adaptive nature of the generated strategies makes it possible to adjust the degree of automation to the human and environmental conditions. For example, let us consider automation variants in the automotive sector. According to the SAE classification [22], the level of automation can range from 0 (no automation) to 5 (full automation). Each level represents different human involvement. Within this spectrum depending on the automation systems in place, the degree of human involvement varies from *hands-on* to *mind-off*. Our approach can be employed to determine when what level of automation is appropriate.

One shortcoming of the approach as described so far is that the feasibility check of the control synthesis may yield unsound results due to incompleteness of the learned HM. Since the HM is trained using a finite number of observed traces, HM might not reflect the original cognitive models sufficiently well. We hence co-simulate the CPS with the synthesized control strategy CS_i and the cognitive model in order to refine the learned HM_i (see Fig. 3) and subsequently to eliminate insufficient control. Therefore, the traces of the simulation that violate the control objectives are fed into the automaton learning algorithm to refine the HM_i. The result is a new learned human model, HM_{i+1}. After learning HM_{i+1}, we proceed the control synthesis following Fig. 2 yielding a new control strategy CS_{i+1}. As HM_{i+1} has been derived from the observed traces of a CPS controlled by CS_i, CS_{i+1} main purpose is to deal with the newly discovered behaviors. Thereby our approach realizes a divide a conquer approach of the control synthesis. Within the current framework, there is no guarantee that the refinement process will be terminated. Therefore, a termination criterion needs to be defined, e.g. when the synthesized controller has sufficient performance. The development of termination conditions is part of our future work. Additionally, by collecting real-world data, the generated behavior of the learned HM can be

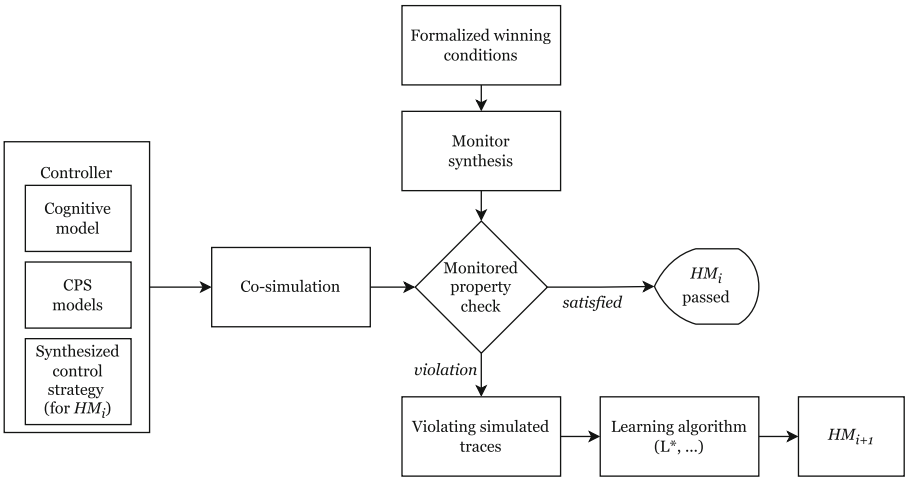

Fig. 3. Framework to refine the learned human model HM.

tuned by adjusting its parameters to improve its accuracy and adaptability to human behavior.

3 Cognitive Modeling

Cognitive architectures encode dynamic models of cognition based on established theories about the structure of mind [2,18]. Empirically validated, these architectures constitute plausible cognitive theories that enable predictions about imminent human behavior. Above that, emotions have a significant impact on decisions and influence individuals' daily choices [5]. Various configuration parameters affect the overall behavior of the cognitive model, such as the learning rate, the decay rate of memory, the retrieval threshold, and the noise distribution of knowledge selection, to name but a few. In the following, we briefly survey the research landscape of human models.

ACT-R [1] is a neurally plausible architecture with interconnected modules through buffers and a central production-pattern matching module. The input/output system consists of visual, auditory, and motor modules. The declarative and procedural modules constitute the central cognitive component by storing knowledge. The goal and imaginal modules track the agent's intentions and the internal representation of the world.

CASCaS [9,16] is designed for real-time simulation of human behavior in traffic scenarios. It has been validated extensively in aviation and automotive application, e.g. by [19,28,29]. It includes perception and motor modules as well as memory module for storing declarative (e.g., current speed limit) and procedural (e.g. driving instructions) knowledge. The memory incorporates processes for retrieval and forgetting.

SEEV [26] designed for predicting pilot attention, has been applied to predict driver attention in road traffic, e.g. by [4,10,27]. Modifying the SEEV parameters (*S*alience, *E*ffort, *E*xpectancy and *V*alue) leads to human modes, e.g. anxious, calm or bored.

There are a few cognitive architectures that capture the interplay of emotions and behavior as reviewed by [21], such as e.g. MAMID [11] and SAMPLE [30]. Cognitive appraisal theory [23] focuses on the processing of input stimuli to infer emotional states by taking the individual history, personality and current affective state into account.

4 Conclusion

This paper presents our vision on how to determine an appropriate level of automation in a human-centered system in the early phase of system design. The proposed approach aims to support designers to evaluate the automation variants in a structured way to choose the variants that guarantee safety objectives while realizing the shared control paradigm. Therefore, we propose to derive a human model *HM* from cognitive architectures using automata learning methods. The *HM* encodes human limitations and changing behavior patterns and is used

in a game-theoretic analysis. The appropriateness of the level of automation is determined by testing whether a control strategy can be synthesized that implements the shared control paradigm and establishes the control objectives.

Acknowledgments. This research is partially funded by the German Federal Ministry of Education and Research (BMBF) within the projects "ASIMOV" and "TRANSACT", and by Universität Oldenburg within RTG SEAS.

References

1. Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y.: An integrated theory of the mind. Psychol. Rev. **111**(4), 1036–60 (2004)
2. Anderson, J.R., Lebiere, C.: The atomic components of thought. Psychol. Press (1998). https://doi.org/10.4324/9781315805696
3. Angluin, D.: Learning regular sets from queries and counterexamples. Inf. Comput. **75**(2), 87–106 (1987). https://doi.org/10.1016/0890-5401(87)90052-6
4. Bairy, A., Fränzle, M.: Optimal explanation generation using attention distribution model. Hum. Interact. Emerg. Technol. (IHIET-AI 2023): Artif. Intell. Future Appl. **70**(70) (2023). https://doi.org/10.54941/ahfe1002928
5. Bechara, A., Damasio, H., Damasio, A.: Emotion, decision making and the orbitofrontal cortex. Cerebral cortex (New York, N.Y. : 1991) **10**, 295–307 (2000). https://doi.org/10.1093/cercor/10.3.295
6. Bowen, J., Stavridou, V.: Safety-critical systems, formal methods and standards. Softw. Eng. J. **8**, 189–209 (1993). https://doi.org/10.1049/sej.1993.0025
7. Cassandras, C.G., Lafortune, S.: Introduction to Discrete Event Systems. Springer, Cham (2008). https://doi.org/10.1007/978-0-387-68612-7
8. ERTRAC Working Group: Connected automated driving roadmap (2019)
9. Frische, F., Osterloh, J.P., Lüdtke, A.: Simulating visual attention allocation of pilots in an advanced cockpit environment. In: Selected Papers and Presentations Presented at MODSIM World 2010 Conference Expo, pp. 713–721. MODSIM World Conference, Hampton, VA, USA (2011)
10. Horrey, W.J., Wickens, C.D., Consalus, K.P.: Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. J. Exp. Psychol. Appl. **12**(2), 67–78 (2006). https://doi.org/10.1037/1076-898X.12.2.67
11. Hudlicka, E.: This time with feeling: integrated model of trait and state effects on cognition and behavior. Appl. Artif. Intell. **16**, 1–31 (2002). https://doi.org/10.1080/08839510290030417
12. Landau, I.D., Lozano, R., M'Saad, M., Karimi, A.: Adaptive Control: Algorithms, Analysis and Applications. Springer Science & Business Media, Cham (2011). https://doi.org/10.1007/978-0-85729-664-1
13. Langenfeld, V., Westphal, B., Albrecht, R., Podelski, A.: But does it really do that? Using formal analysis to ensure desirable ACT-R model behaviour. In: Cognitive Science (2018)
14. Langenfeld, V., Westphal, B., Podelski, A.: On formal verification of ACT-R architectures and models. In: CogSci, pp. 618–624 (2019)
15. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. Hum. Factors **46**(1), 50–80 (2004). https://doi.org/10.1518/hfes.46.1.50_30392
16. Lüdtke, A., Osterloh, J.P., Frische, F.: Multi-criteria evaluation of aircraft cockpit systems by model-based simulation of pilot performance. In: Embedded Real Time Software and Systems (ERTS2012). ERTS, Toulouse, France (2012)

17. Maler, O., Pnueli, A., Sifakis, J.: On the synthesis of discrete controllers for timed systems. In: Mayr, E.W., Puech, C. (eds.) STACS 95. Lecture Notes in Computer Science, vol. 900, pp. 229–242. Springer, Berlin, Heidelberg (1995). https://doi.org/10.1007/3-540-59042-0_76
18. Newell, A.: Unified Theories of Cognition. Harvard University Press, USA (1990)
19. Osterloh, J.P., Rieger, J.W., Lüdtke, A.: Modelling workload of a virtual driver. In: Proceedings of the 15th International Conference on Cognitive Modeling. ICCM, Warwick, UK (2017)
20. Pnueli, A., Rosner, R.: On the synthesis of a reactive module. Automata Lang. Program. **372**, 179–190 (1989). https://doi.org/10.1145/75277.75293
21. Rakow, A., Hajnorouzi, M., Bairy, A.: What to tell when? - Information provision as a game. Electron. Proc. Theor. Comput. Sci. **395**, 1–9 (2023). https://doi.org/10.4204/eptcs.395.1
22. SAE International: J3016: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (2021)
23. Scherer, K.R., Schorr, A., Johnstone, T.: Appraisal Processes in Emotion: Theory, Methods, Research. Oxford University Press, Oxford (2001)
24. Sheridan, T.B., Verplank, W.L., Brooks, T.: Human and computer control of undersea teleoperators. In: NASA. Ames Res. Center The 14th Ann. Conf. on Manual Control (1978)
25. Thomas, W.: On the synthesis of strategies in infinite games. In: Mayr, E.W., Puech, C. (eds.) STACS 95, 12th Annual Symposium on Theoretical Aspects of Computer Science, Munich, Germany, March 2–4, 1995, Proceedings. Lecture Notes in Computer Science, vol. 900, pp. 1–13. Springer, Cham (1995). https://doi.org/10.1007/3-540-59042-0_57
26. Wickens, C., Helleberg, J., Goh, J., Xu, X., Horrey, W.: Pilot task management: testing an attentional expected value model of visual scanning. Savoy, IL, UIUC Institute of Aviation Technical Report (2001)
27. Wortelen, B.: Das Adaptive-Information-Expectancy-Modell zur Aufmerksamkeitssimulation eines kognitiven Fahrermodells. Ph.D. thesis, Carl von Ossietzky Universität, Oldenburg, Germany (2014)
28. Wortelen, B., Baumann, M., Lüdtke, A.: Dynamic simulation and prediction of drivers' attention distribution. Transport. Res. F: Traffic Psychol. Behav. **21**, 278–294 (2013). https://doi.org/10.1016/j.trf.2013.09.019
29. Wortelen, B., Unni, A., Rieger, J.W., Lüdtke, A., Osterloh, J.P.: Monte Carlo methods for real-time driver workload estimation using a cognitive architecture. In: Klempous, R., Nikodem, J., Baranyi, P.Z. (eds.) Cognitive Infocommunications, Theory and Applications, pp. 25–48. Springer International Publishing, Cham, Switzerland (2019). https://doi.org/10.1007/978-3-319-95996-2_2
30. Zacharias, G.L., Miao, A.X., Illgen, C., Yara, J.M., Siouris, G.: SAMPLE: situation awareness model for pilot in-the-loop evaluation. In: Proceedings of the 1st Annual Conference on Situation Awareness in the Tactical Air Environment. Citeseer (1996)