

Safety of Autonomous Cognitive-oriented Robots

Philipp Ertle¹

Published online: 15 October 2015
© Springer-Verlag Berlin Heidelberg 2015

Keywords Autonomous system · Robot · Safety · Dynamic risk assessment

1 Future Intelligent Systems

Companion-systems (CS) are a vision of future technical systems, which take into account the user's abilities, preferences, requirements and current needs, reflecting the user's situation and emotional state. They are competent, always available, cooperative, trustworthy, and cooperative service partners [7]. For the realization of CS, cognitive processes like reasoning and planning are utilized in order to infer executable actions from a knowledge-base, required to assist solving the user's intended task. Depending on the kind of a CS, these actions may be executed directly (by the underlying technical system) or by (instructing) the user. In both cases, the CS may be involved in initiating hazards; either directly, by execution of actions itself, or indirectly, by guiding a user who probably lacks knowledge to foresee the impact of his/her actions. This relates to a system safety process. From the user's perspective, system safety is related to trustworthiness of CS. From the manufacturer's perspective, the declaration of conformity to relevant standards and at least product liability aspects oblige to perform safety considerations.

✉ Philipp Ertle
philipp.ertle@gmx.net

¹ Mühlbachstraße 17, 72189 Vöhringen, Germany

2 System Safety Problem of Intelligent Systems

Safety is typically defined as relative freedom from danger or the risk of harm. System safety is the engineering discipline which aims at developing safe systems and products. Traditionally, system safety is considered as a discipline with the attempt to identify potential hazards during the system design phase. Intelligent systems are used in complex situations for which it is difficult to design and deploy conventional systems. Hence, the hazard identification can be complex, as well as the balancing between risk mitigation and task performance. For instance, for an intelligent system in a complex environment that perceives, reasons, and plans its interactions on basis of a (probably adjustable) knowledge base it is challenging to ensure system safety: For hazard identification, establishing countermeasures, and identifying effects on system performance it has to be taken into account mutual interaction of system state, knowledge base state, and resulting reaction of environment state and vice versa. This problem is addressed in outlined dissertation thesis exemplary in the scope of service robots [1].

3 Safety of Service Robots

Future service robots shall autonomously provide services in different spheres of life by executing demanding and complex tasks in a dynamic environment. Typical visions of useful tasks are serving beverages and foods, cleaning, or guiding through museums, shops, buildings, and the like. These tasks are commonly a composition of basic but differently parameterized actions, which are arranged by planning and often related to a manipulation of environment objects. The manipulation aspect has a significant

meaning for the complexity of the system safety: The so-called object interaction hazards occur when environment objects interact with objects that are manipulated by a robot, e.g., a burning kandle and a paper napkin, hot cooking plate and a wooden cutting board or humans and boiling water. This problem area is so far not addressed in current research work or in the relevant standards.

Hardware design or motion control are assumed to be less relevant for object interaction hazards. Rather, these hazards have to be considered at a deliberative level. In order to sufficiently consider the environment and operation context, the system has to be aware of such interactions. In the context of autonomous vehicles, a dynamic risk assessment (DRA) concept is outlined [6], which postulates that risks have to be continuously assessed by the system itself during runtime. This concept is adopted and integrated into a cognitive architecture in order to combine it with cognitive functions, such as anticipation, planning, and learning [2]. The strategy is pursued that anticipated situations can be assessed with regard to risks and the information about risks can be taken into account within the planning process of a robot [5].

The safety knowledge base of a DRA component is derived using a suggested procedural model. In this connection, the concept of the so-called ‘Safety Principles’ is introduced. Safety Principles are formalizations of risks abstracted to root causes of hazard actuation. These root causes are referenced to object properties and shall remain as well valid for future unknown situations [4]. In addition, Safety Principles can be considered as meta-structure, which may integrate already available safety-related approaches (adaptive collision avoidance, adaptive compliant actuation, etc.).

However, it can be assumed that the safety knowledge potentially lacks completeness. For this reason, light is shed as well on selected learning methods in the scope of safety-critical systems: For instance, a ‘learning from demonstration’ approach is investigated, constituting interesting potential for improving or simplifying the generation process of the safety knowledge [3].

Finally, the thesis reveals a general perspective on potential hazards of intelligent autonomous systems. It describes the generation, integration, utilization, and maintenance of an explicit system-internal safety knowledge base for DRA. It denotes an overall concept toward solving the advanced safety problem of intelligent autonomous systems.

References

1. Ertle P (2013) Safety of autonomous cognitive-oriented robots. Ph.D. thesis, University of Duisburg–Essen, Faculty of Engineering
2. Ertle P, Gamrad D, Voos H, Söffker D (2010) Action planning for autonomous systems with respect to safety aspects. In: Proceedings of the IEEE International Conference on Systems Man and Cybernetics (SMC) 2010, Istanbul, Turkey, pp 2465–2472
3. Ertle P, Tokic M, Voos H, Söffker D (2012) Towards learning of safety knowledge from human demonstrations. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2012), Vilamoura, Algarve, Portugal, pp 5394–5399
4. Ertle P, Voos H, Söffker D (2010) On risk formalization of on-line risk assessment for safe decision making in robotics. In: Proceedings of the 7th IARP/IEEE-RAS joint workshop on technical challenge for dependable robots in human environments, Toulouse, France, pp 15–22
5. Ertle P, Voos H, Söffker D (2012) Utilizing dynamic hazard knowledge for risk sensitive action planning of autonomous robots. In: Proceedings of the IEEE International Symposium on Robotic and Sensors Environments ROSE, Magdeburg, Germany, pp 162–167
6. Wardziński A (2008) Safety assurance strategies for autonomous vehicles. In: Proceedings of the 27th International Conference on Computer Safety, Reliability, and Security SAFECOMP, Springer, Berlin, pp 277–290
7. Wendemuth A, Biundo S (2012) A companion technology for cognitive technical systems. In: Esposito A, Esposito A, Vinciarelli A, Hoffmann R, Müller V (eds) Cognitive behavioural systems, vol 7403. Lecture Notes in Computer Science. Springer, Berlin Heidelberg, pp 89–103



Philipp Ertle was born in 1976 in Germany. He studied Physical Engineering, Technology Management, and Mechatronics and received his diploma (2005) and master degree (2007) from the University of Applied Sciences Ravensburg-Weingarten. He started his dissertation 2009 at the University of Duisburg–Essen (Chair of Dynamics and Control) and within the Collaborative Center for Applied Research on Service Robotics at the University of Applied Sciences Ravensburg–Weingarten. He received his doctorate in 2013.