Multi-Modal Interfaces for Human–Robot Communication in Collaborative Assembly

by Gergely Horváth, Csaba Kardos, Zsolt Kemény, András Kovács, Balázs E. Pataki and József Váncza

Human–robot collaboration (HRC) in production, especially in assembly, has the potential to improve flexibility and competitiveness. However, there are still numerous challenges to address before HRC can be realised. Beyond the essential problem of safety, the efficient sharing of work and workspace between human and robot also requires new interfaces for communication. The SYMBIO-TIC H2020 project is developing a dynamic context-aware and bi-directional, multi-modal communication system that supports human operators in collaborative assembly.

Introduction

The main goal of the SYMBIO-TIC H2020 project is to provide a safe, dynamic, intuitive and cost effective working environment, hosting immersive and symbiotic collaboration between human workers and robots [1]. In such a dynamic environment, a key to boosting the efficiency of human workers is supporting them with context-dependent work instructions, delivered via communication modalities that suit the actual context. Workers, in turn, should be able to control the robot or other components of the production system by using the most convenient modality, thus lifting the limitations of traditional interfaces such as push buttons installed at fixed locations. As part of the SYMIBIO-TIC project, we are developing a system that addresses these needs.

Context-awareness in human–robot collaboration

To harness the flexibility of an HRC production environment, it is essential that the worker assistance system delivers information to the human worker that suits the actual context of production. In order to gather the information describing the context, data related to both the worker (individual properties, location, activity) and to the process under execution is required. This information is provided to the worker assistance system by three connected systems, which together form the *HRC ecosystem*, namely (1) the workcell-level task execution and control (unit controller, UC), (2) the shopfloor-level scheduling (cockpit), and (3) the mobile worker identification (MWI) systems [2].

The process execution context is defined by the state of the task execution in the UC. The identification and location of the worker by the MWI is essential in order to trigger the worker assistance system and to properly utilise the devices around the worker. Actions of the worker have to be captured either directly by devices available to the worker, or the sensors deployed in the workcell, registering the worker's activity context. The properties of the worker, such as skills or preferences of instruction delivery, define the final format of the instructions delivered.

Automatically generated work instructions

The complexity of managing work instructions in production environments characterised by shortening product life cycles and increasing product variety, as well as the requirement to fully exploit the available context data in customised instructions, calls for the automated generation of human work instructions. A method for this, relying on a feature-based representation of the assembly process [3], computer-aided process planning (CAPP) techniques, and a hierarchical template for the instructions, has been proposed in [4][3]. The method generates textual work instructions (with potential audio interpretation using text-to-speech tools) and X3D animations of the process tailored to the skill level and the language preferences of the individual worker. The presentation of the instruction can be further customised in real time by the instruction delivery system: e.g., the selection of the modality and the device, as well as the font size and the sound volume, can be adjusted according to the current context.

Multi-modal communication

Traditionally, worker assistance is provided by visual media, mostly in the form of text or images. The currently prevailing digital assistance systems hence focus on delivering visual instructions to the workers. However, in a HRC setting, it is also necessary to provide bi-directional interfaces that allow the workers to control the robots and other equipment participating in the production process.

The worker assistance system that we have developed is designed to deliver various forms of visual work instructions, such as text, images, static and animated 3D content and videos. Audio instructions are also supported: by using text-to-speech software, the textual instructions can be transformed as well. Instruction delivery is implemented as an HTML5 webpage, which supports embedding multi-media content and also allows multiple devices to be used for both visual and audio content, such as smartphones, AR-glasses, computer screens, or tablets.

Our web-based solution for input interfaces provides the classic button-like input channels, which are still required in most industrial scenarios. Potentially promising contactless technologies are also integrated into the system. Interpreting audio commands shows great potential as it is not only contactless, but also hands-free. However, in a noisy industrial environment, it could be challenging and therefore the application of two hand gesture-based technologies is also supported, one using point-cloud data registered by depth cameras and the other using a special interpreter glove that measures the relative displacement of the hand and fingers.

Figure 1: Schematic architecture of the HMIC implementation and its immediate environment in the production system.

Implementation and use case

A complete server–client-based solution for the human–machine interface system was implemented in accordance with the aforementioned requirements and technologies. The system is named Human Machine Interface Controller (HMIC). Figure 1 shows its major structure (backend/frontend design) and its connections to other elements of the ecosystem. The implemented HMIC system was successfully demonstrated in the laboratory simulation of an automotive assembly use case, where 29 parts were assembled in 19 tasks (see Figure 2). The research project is now in its closing phase, where the focus is on the development of demonstrator case studies and the evaluation of the perceived work experience with the use of the generated content and the multi-modal content delivery system.
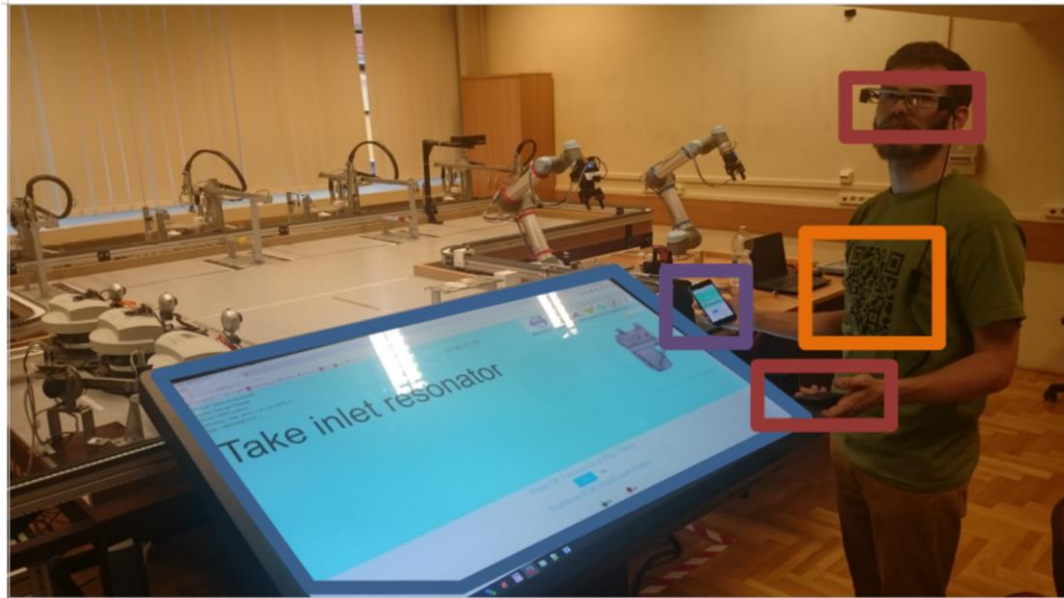
Figure 2: Demonstration of an automotive assembly case study using the HMIC system. The devices available for the user are a large touchscreen, a smartphone and an AR-glass.

Acknowledgement:

References:

[1] SYMBIO-TIC EU H2020 Project, http://www.symbio-tic.eu

[2] Kardos, Cs.; Kemény, Zs.; Kovács, A.; Pataki, B.E.; Váncza, J.: Context-dependent multimodal communication in human–robot collaboration. 51st CIRP International Conference on Manufacturing Systems, 2018.

[3] Kardos, Cs.; Kovács, A.; Váncza, J.: Decomposition approach to optimal feature-based assembly planning. CIRP Annals – Manufacturing Technology, 66(1):417-420, 2017.

[4] Kardos, Cs.; Kovács, A.; Pataki, B.E.; Váncza, J.: Generating human work instructions from assembly plans. 2nd ICAPS Workshop on User Interfaces and Scheduling and Planning (UISP2018), Delft, The Netherlands, 2018.

Please contact:

Csaba Kardos

MTA SZTAKI: Institute for Computer Science and Control, Hungarian Academy of Sciences

+ 36 1 279 6189

csaba.kardos@sztaki.mta.hu