Flexible programming and orchestration of collaborative robotic manufacturing systems

Nuno Mendes
Department of Mechanical Engineering
University of Coimbra
Coimbra, Portugal
nuno.mendes@dem.uc.pt

Mohammad Safeea
Department of Mechanical Engineering
University of Coimbra
Coimbra, Portugal
ms@uc.pt

Pedro Neto
Department of Mechanical Engineering
University of Coimbra
Coimbra, Portugal
pedro.neto@dem.uc.pt

***Abstract*—A flexible programming and orchestration system for human-robot collaborative tasks is proposed. Five different interaction modes are suggested to test two Task-Managers (TMs) acting as orchestrators between a human co-worker and a robot. Both TMs rely on the task-based programming concept providing modular and scalable capabilities, allowing robot code reuse, fast robot programming and high robot programming flexibility. The TMs provide visual and audio feedback to the user about the robot task sequence being executed, guiding the user during the iterative process. The interaction modes tested were: (1) human arm gestures, (2) human hand gestures, (3) physical contact between human and robot, and (4-5) two hybrid interaction modes combining each one of the two first interaction modes with the last one. Experimental tests indicated that users prefer fast interactions with small number of interaction items to higher flexibility. Both TMs provide intuitive and modular interface for collaborative robots with a human in the loop.**

***Index Terms*—orchestration, task management, collaborative robotics, human-robot interaction, industry 4.0**

I. Introduction

When a robot is in operation, several tasks are repeated multiple times. Thus, the reuse of code that represents the repeated tasks is highly recommended. This concept becomes more important and useful when a robot is used to perform different task sequences and some of these tasks are repeated several times, occurring by different order in the sequences. A scalable and modular architecture which allows flexible and fast robot programming and on the same time introduce a human co-worker in the real-time robotic management system is a gap left on literature. The most approached topics that can be found in literature are orchestration system where the human has not/or low intervention on the real-time management loop.

In the last years, a new high-level robot programming concept called service-oriented architectures (SOAs) has been developed [1], [2]. This Plug-and-Play architecture reduces the system integration time and is better adapted to industrial robotic cell system integrators. A SOA is usually developed to provide a graphical programming interface that is intuitive and user-friendly. Some orchestrators used in SOAs have also graphical service orchestration which are easy-to-use and intuitive to understand, extending the use of industrial robots to non-robotic experts.

A task management system to perform physical services in Kukanchi is suggested by Maeda *et al.* [3]. This system, focused on the object delivery service and performs service execution without detailed instructions from the user.

Ende *et al.* suggested that the interaction between human and robot shall no longer require a separate interface such as a control panel, instead, a more intuitive and user-friendly interface can be achieved by gestures [4]. Gestures have been successfully used to interact with robots in different contexts and in different environments [5], [6]. Pedersen and Kruger also support the idea of using gestures to perform human robot interaction (HRI) as well as teaching by demonstration (TbD) [7]. This last study proposes the whole parametrization for a robotic task being carried out through intuitive human arm gestures (mainly pointing gestures). The strong points of this system are short programming time and expert knowledge not required which makes this system easy-to-use. The limitations of this system are related to the functionality of the instructor tracking (during teaching), choice of gestures (which are reported as not appealing) and system applicability to simple tasks and skills. The use of nonverbal communication between robots and humans has also been successfully explored [8].

Schou *et al.* proposed a flexible structure to use and reuse code in robot programming [9]. The proposed structure consists in three stages: Device Primitives, Skills and Tasks. Device Primitives basically are fundamental robotic instructions such as open a gripper, close a gripper, move robot to a position, etc. These Device Primitives are grouped in convenient order creating Skills, e.g. pick up a part from a given place, release a part in a specific position, etc. Tasks are in a higher programming level where Skills are put together by a desired order originating a robotic task, e.g. palletizing. The advantage of this concept is if a user has the right Skills available, he/she can reprogram a robot in an efficient way [10], [11].

II. Background

Human arm/hand gestures and the physical contact between human and robot are used to interact with a collaborative robot which must perform pre-taught tasks. Two Task-Manager (TM) frameworks are proposed to carry out the HRI, namely a Parametrization Robotic Task Manager (PRTM) and a Execution Robotic Task Manager (ERTM). There are three applications running in parallel in both frameworks, i.e. a

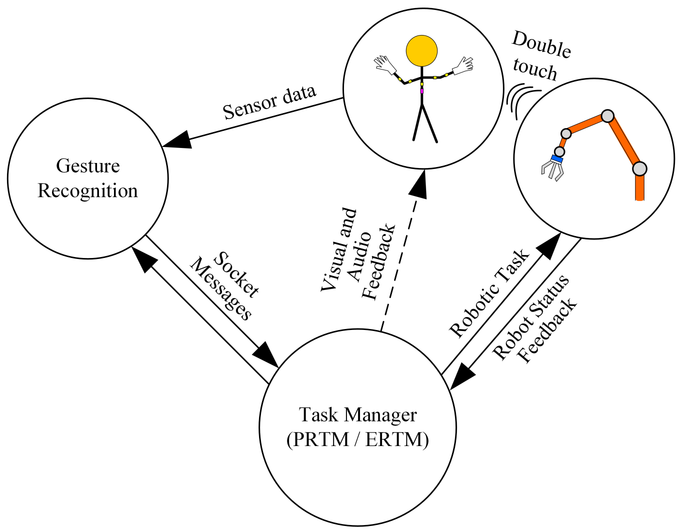

Fig. 1. Control architecture highlighting the central role of the TM. The TM receives information from the gesture recognition system and double touch detection from the robot, and sends commands to the robot. In addition, the TM manages the feedback provided to the human co-worker.

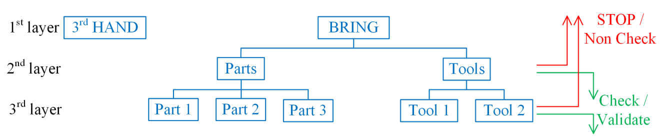

Fig. 2. The three layers of the proposed PRTM. The BRING and 3rd HAND options are in the first layer. For the BRING option we have in the second layer two options to select PARTS and TOOLS. In the third layer we have all the parts and tools available to be selected. In the second and third layers is required a validation step provided by the "Validate" gesture. On the other hand, the "Stop" gesture restarts the parametrization process.

gesture recognition application used to recognize static and dynamic human arm/hand gestures, the TM application used to define and manage the execution of task sequences, and a robot application used to execute each task and to detect double touches by the human in the robot. While the PRTM is used to parametrize online and manage online robotic tasks with the human co-worker in the loop, the ERTM is used to manage online robotic tasks with the human co-worker in the loop, Fig. 1. ERTM parametrization is performed offline previously. Both TMs provide audio feedback to the user through computer text-to-speech (TTS). The gesture recognition has implemented several methods such as data sensory acquisition, raw data processing, segmentation, and static and dynamic gesture classification. The communication among the three applications is made by sockets TCP/IP.

A. Parametrization Robotic Task Manager (PRTM)

When a gesture (static or dynamic) is recognized, a socket message is sent to the PRTM with information about the recognized gesture. It works as a phone auto attendant providing options to the human (audio feedback) which selects the intended robot service using gestures. The proposed PRTM includes in the first layer 2 options, BRING and 3rd HAND, Fig. 2. The BRING option refers to the ability of the robot to deliver parts, tools, and consumables to the human co-worker, while the 3rd HAND is related with an operation mode in which the co-worker can physically guide the robot to desired poses in space to teach a specific task or to hold a part while he/she is working on it. In the second layer, for the BRING option, the user can select Tools or Parts, with different possibilities in each one (third layer). A BRING task and operation actions related with the human co-worker, robot and user feedback are detailed in Fig. 3.

The interactive process starts with the user performing a gesture called "Attention". This gesture informs the system that the user wants to perform a given robotic task parametrization. The TTS informs the user about the selection options in the first layer. The user has few seconds (a predefined time) to perform a "Select" gesture to select the desired option. After this process, the PRTM through TTS asks the user to validate the selected option with a "Validation" gesture. If the selected option has been validated, the PRTM goes to the next layer, if the selected option has not been validated, the system continues in the current layer. If the user does not perform the "Select" gesture during the predefined time period, the PRTM continues with the other options within the layer. The procedure is repeated until the user selects one of the options or until the PRTM through TTS repeats all of the options three times. The process is similar for the second and third layer. In the third layer the PRTM sends a socket message to the robot to perform the parametrized task. At any moment the user can perform the "Stop" gesture so that the system returns to initial layer.

B. Execution Robotic Task Manager (ERTM)

The ERTM framework works in a similar way to the PRTM framework with some changes. In this case, instead of the user performs the parametrization sequence online to ask a task to the robot, all the tasks are parametrized offline and ordinated by a desired order forming a sequence of tasks. Later, the robot executes the sequence of tasks by the desired order requiring only the user permission between consecutive tasks, which is provided by a gesture. Initially, the user must choose which task sequence the robot should execute. This procedure can be done on the robot Teach Pendant. After that, the ERTM provides audio and visual feedback to the user about robot state, if it is available or which task it is executing. When the robot is available, the ERTM provides visual feedback about the next task of the sequence of tasks, which must be executed next, and on the same time, the ERTM waits for a "Next" gesture that represents the order to execute the next task. When the "Next" gesture is recognized and received by the ERTM, a socket message is sent to the robot asking the execution of the next task. When the robot ends the execution of a task, feedback is sent to the ERTM and this one provides audio and visual feedback to the user about the end of the task. The ERTM waits for a new "Next" gesture order to start the execution of a new next task. The process is repeated again and

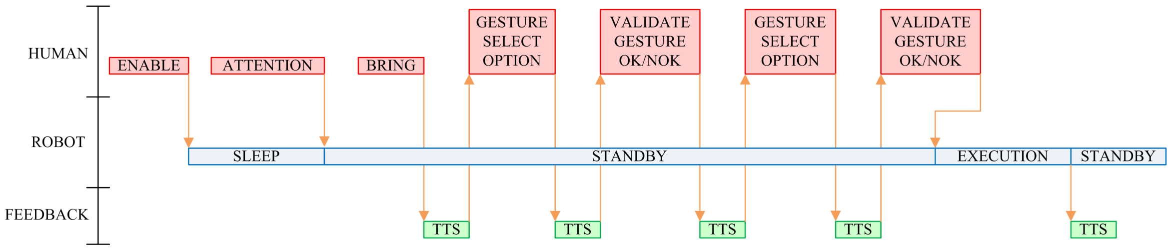

Fig. 3. PRTM framework for a BRING task with the role of the human co-worker, robot and feedback.

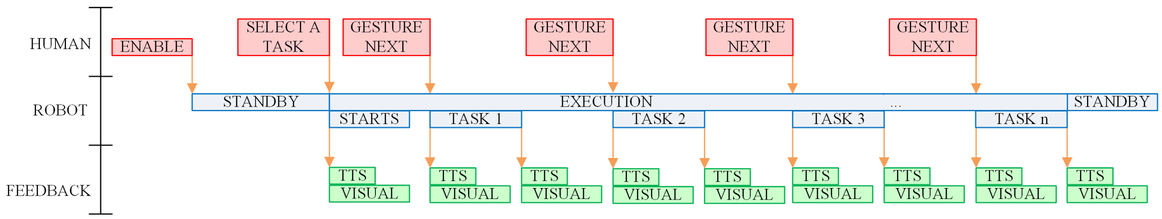

Fig. 4. ERTM framework with the role of the human co-worker, robot and feedback.

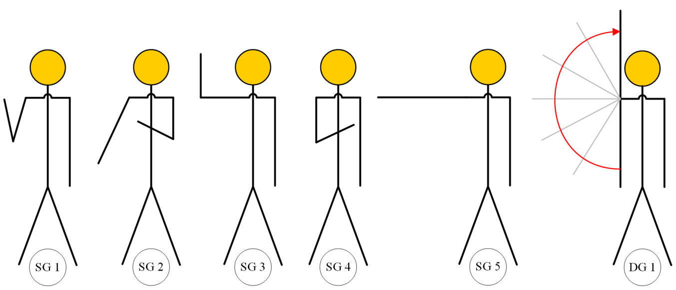

Fig. 5. Human arm gestures.

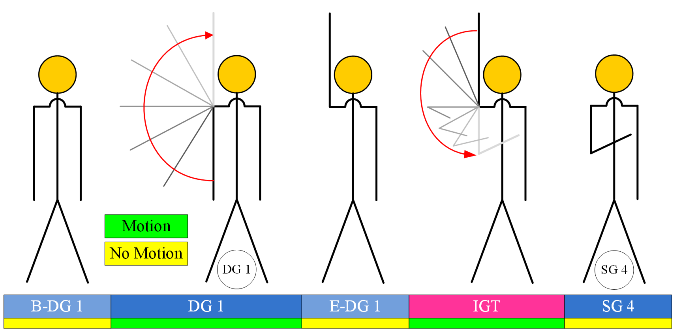

Fig. 6. Composed human arm gesture.

again till the sequence of tasks is completed, as schematically represented in Fig. 4.

C. Human arm gestures

The human arm gestures are captured by an IMU based system, Tech-MCS V3, which provides sensor data to a gesture recognition application via bluetooth at 25 Hz. The gesture information provided by this device and preprocessed by itself consists in gravity, orientation, acceleration and physical data. In order to achieve a user-friendly HRI, a small library of six gestures (five static gestures (SGs) (Attention, Select, Validate, Stop and Abort) and one dynamic gesture (DG) (Next) is created, Fig. 5. The gesture names are suggestive to its application, i.e.: "Attention" gesture is used to start a parametrization of a robotic Task, "Select" gesture is used to choose parameters, "Validate" gesture is used to confirm a parameter chosen and start executing a robotic Task, "Stop" gesture is used to cancel a parametrization of a robotic Task, while "Abort" gesture is used to cancel a robotic Task in execution (the robot is forced to stop even though it has already started executing a robotic Task). These five gestures are used for the PRTM framework while the sixth gesture, "Next" gesture, is exclusively used for the ERTM framework. The "Next" gesture is used to order the robot to execute a next task of a given sequence of tasks. The ERTM framework was also tested with a composed gesture which consists of a mix of the SGs and DGs. This kind of gesture structure avoids false positives/negatives in the gesture recognition process and allows elimination of the "validation" procedure used in the PRTM framework keeping, on the same time, high reliability. The composition of a composed gesture can be customized by each different user according to the following rules: (1) the composed gesture begins with a static pose with the beginning of a selected dynamic gesture B-DG, (2) a DG, (3) a static pose with the end of the dynamic gesture E-DG, (4) an inter-gesture transition (IGT) and (5) a SG. One example of a composed gesture is depicted in Fig. 6.

When a gesture is recognized in the gesture recognition application, a TCP/IP socket message is sent to the TM, in use (the PRTM or the ERTM), informing about which gesture was performed by the user.

D. Human hand gestures

An off-the-shelf device, Myo armband, was used to captured human hand gestures. Five gestures are recognized by the gesture recognition application (Fig. 7) taking into account 8 EMG sensor data, which are streamed at 200 Hz via bluetooth.

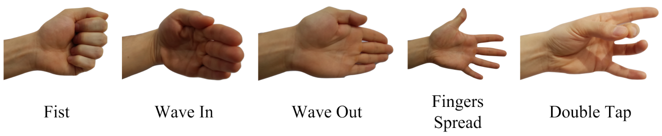

Fig. 7. Human hand gestures.

In this interaction mode, just the fifth gesture, double tap, was used to order the robot to perform a next task of a sequence of tasks. The procedure followed to send information between the gesture recognition application and the ERTM is the same as described in the above section.

E. Physical contact

A third interaction approach followed in this study consists in a double touch performed by the user in the collaborative robot. This touch is captured by load cells integrated in the robot itself and processed into the robot controller. A double touch detection occurs at 10 Hz, when it is recognized/detected, a TCP/IP socket message is sent to the ERTM which proceed by the same way as described above (for the other two interaction modes). As in the human hand gesture approach, the double touch is used to order the robot to execute a next task of a sequence of tasks in the ERTM.

III. Experiments

In order to assess the performance of the proposed approaches, five users have tested the system. A test was carried out by two users (user A and B) that contributed to the development of the system and created the gesture training data set, and three users (user C, user D and user E) that are not robotics experts and are using the system for the first time. This test consists in performing an assembly task which is composed of subtasks: manipulation of parts, tools, consumables, holding actions and screw. Some tasks are more suited to be executed by humans, others by robots, and others by the collaborative work between human and robot. When requested by the human co-worker (using gestures), the robot has to deliver in the human workplace the parts, consumables (screws and washers) and tools for the assembly process. The parts and tools are placed in known fixed positions. The KUKA Sunrise Toolbox for MATLAB is used to interface with a KUKA iiwa robot [12]. Moreover, the human can setup the robot in precision hand-guiding movement mode [13] to manually guide it to hold workpieces while tightening the elements, Fig. 8.

All the interactive process is performed by human arm gestures, by human hand gestures, or by physical contact between human and robot (double touch). Furthermore, two hybrid solutions consisting in the combination of the third interaction mode with one of the two others interaction modes were also tested. After a gesture is recognized it serves as input for one of the TMs that interface with the robot and provide audio feedback to the human co-worker (section II), Fig. 1.

The two TMs were tested by the five users mentioned above. Subjects C, D and E received a 15 minutes introduction to the system by subjects A and B that contributed to the system development and created the gesture dataset. Finally, users C, D and E were briefed on the assembly sequence and components involved. After the users tried the three different interaction modes and are familiarized with them, which took about 10 minutes for each interaction mode and for each user, each user performed a sequence of tasks three times repeating it for each interaction mode. Then, each user filled a quiz to identify weaknesses and added value of the solution.

IV. Results and Discussion

The use of gestures to interact with robots was considered by all the users as a good procedure, intuitive and easy to use, having the users A and B referred that it must be used as a complement to other interaction mode. All the users pointed in favor of the PRTM framework the great flexibility in choosing any robotic task at any time. On the other hand, the great disadvantage of this framework is the long time spent to parametrize a robotic task. This is felt by the users as discouraging from an industrial point of view. The static gestures are also stated as not attractive, the users reported some gestures are difficult to be performed (requiring challenge user limb configurations and causing fatigue) and are not discreet. In spite of small size of the gesture library, the users would rather a system that requires fewer gestures to interact with the robot.

Although the gesture recognition rate is relatively high, the occurrence of false positives and negatives was analysed. Our experiments demonstrated that if a given gesture is wrongly classified the “validation” procedure allows the user to know from the audio feedback that it happened, so that he/she can adjust the interactive process.

On average, the time that passes between the recognition of a gesture and the completion of the associate PRTM/robot command is about 1 second. If the setup of the PRTM is taken into account, with the selection of the desired options, it takes more than 5 seconds.

The five users filled a questionnaire about the proposed interfaces, resulting in the following main conclusions concerning to the PRTM framework:

1) The gesture-based interface is intuitive but delays the interactive process. It can be complemented with a tablet to select some robot options faster;
2) It was considered by all the subjects that the “validation” procedure slowly the interactive process. Operating a version of the PRTM without all the validations proved to be faster (ERTM);
3) The composed gestures are more complex to perform compared to SGs and DGs. Nevertheless, they are more reliable than SGs and DGs;
4) The automatic audio feedback is considered essential for a correct understanding of the interactive process. Also, it was indicated that the co-worker feedback can be complemented with visual feedback (for example from a screen installed in the robotic cell);

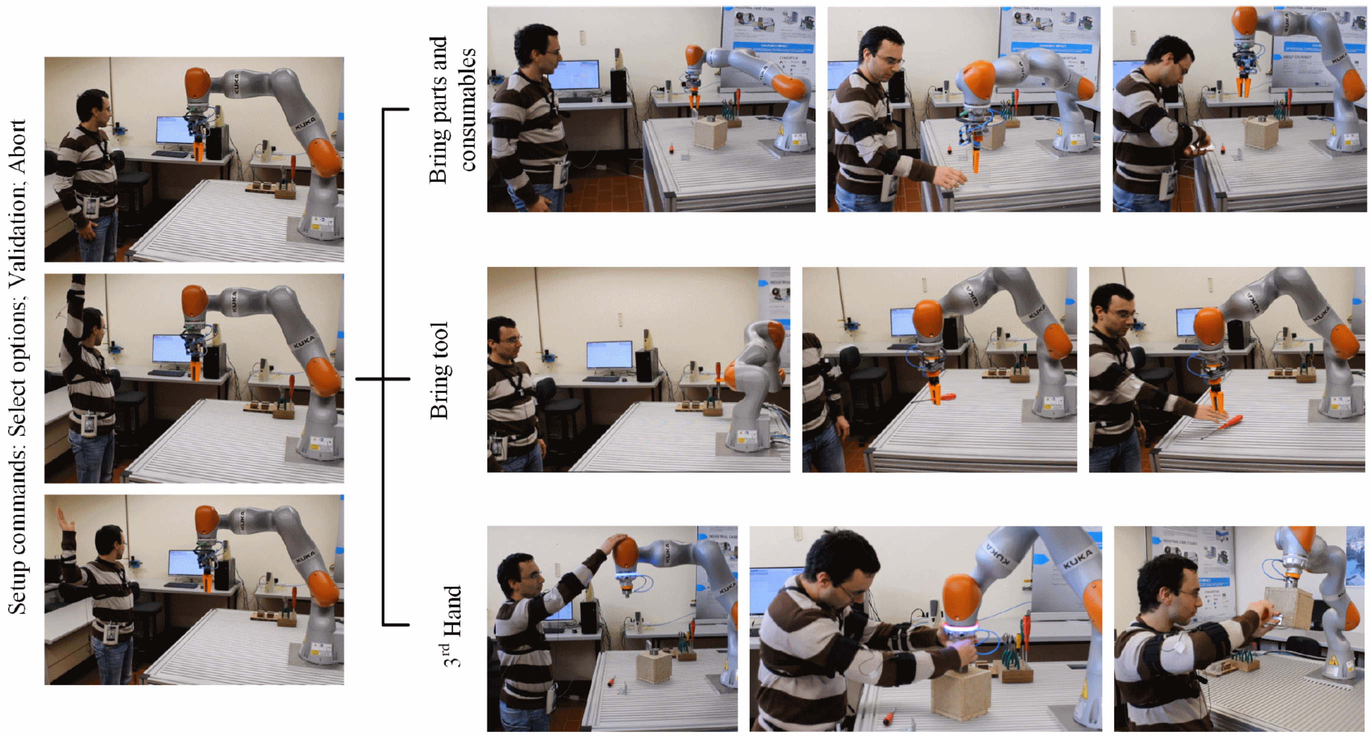

Fig. 8. Human-robot collaborative process. In this use case the robot delivers tools and parts to the human co-worker (top and middle) and the robot holds the workpiece while the co-worker is working on it (bottom). For better ergonomics the co-worker adjusts the workpiece position and orientation through robot hand-guiding.

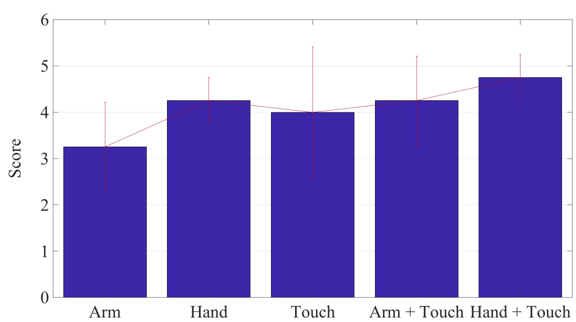

Fig. 9. Users' assessment about interaction's mode using the ERTM framework.

5) The users that were not familiarized with the system (users C, D and E) considered that working with the robot without fences present some degree of danger (they did not felt totally safe);
6) All users reported that the proposed interface allows the human co-worker to abstract from the robot programming, save time in collecting parts and tools for the assembly process, and have better ergonomic conditions by adjusting the robot as desired.

The interactive process proposed in the PRTM framework consumes a significant amount of time. In response to this problem, the PRTM can be setup with the pre-established sequence of tasks so that the human intervention resumes to accept or not the TM suggestions in some critical points of the task being performed. This is the concept used in the ERTM framework. It presents lower flexibility during the task execution and the same flexibility as the PRTM framework taking into account task sequence preparation. Furthermore, the ERTM framework requires fewer gestures to interact with the robot than the PRTM framework which is seen by the users as an advantage. On the same time, the risk of committing mistakes is lower.

The users were invited to assess the interaction mode using the five different interaction modes referred above. The Hand + Touch interaction mode was pointed out by the users as more reliable and intuitive, in some situations is more convenient to use a hand gesture to ask performing a next task to the robot while in other situations is preferable a physical double touch, Fig. 9. In fact, from the five available gestures performable by hand, the one that was better scored by the users in terms of hand configuration as well as not imposing hand movement or psychological constraints on user was the double touch. This hand gesture was chosen to order a next robotic task because a well accept gesture by the user can benefit the utility of the robotic solution and its acceptance.

When the users were questioned if they prefer using other kind of device to interact with a robot, the use of a tablet seems to be a popular option, especially because of the large number of different options/robotic functionalities that can be directly introduced on it. Nevertheless, a tablet also present a lot of disadvantages being one of them the requirement of carrying it all the time. Thus, the use of a tablet as a HRI device is mainly seen by the users as an additional interaction device that is a good solution in some situations (such as to select or start a sequence of tasks) being preferable the use of the proposed gesture interaction in other situations (such as to ask the execution of a next task).

Other question approached was if the users were afraid of performing a gesture that initiate a robotic task without their intention. All the users answer to this question that their afraid could be quantified as low or extremely-low, arguing that when they are interaction with the robot their thoughts are in the robotic task or in performing gestures correctly.

The task completion time was analysed for the presented assembly use case. The task completion time of the collaborative robotic solution (eliminating the parametrization and validation procedures, i.e. using the ERTM framework) is about 1.4 times longer than when performed by the human worker alone. The collaborative robotic solution is not yet attractive from an economic perspective and needs further research. This result is according to similar studies that report that the collaborative robotic solutions are more costly in terms of cycle time than the manual processes [14]. Nevertheless, the system demonstrated to be intuitive to use and with better ergonomics for the human.

V. Conclusions and Future work

Two different frameworks to perform the interface between human interaction mode and robot task execution were proposed as well as five different interaction modes, all of them based on gestures and/or physical contact. The framework preferred by the users, who tested the system, was the ERTM which displays the following advantages:

- It requires low number of gestures to interact with;
- It is relatively fast from the first gesture performance till the robot starts executing a task;
- It is considered by the users as a reliable and safe framework;
- It allows reuse of robot code consequently reprogramming the robot quickly.

A constraint of the ERTM framework is related to flexibility, while the PRTM framework allows to perform any task at any moment during HRI stage, the ERTM framework only allows the same flexibility in an initial stage when the sequences of tasks are defined.

The interaction mode also plays an important role in HRI, an easy to perform, reliable and intuitive interaction is better accept by the users and has influence in user acceptance of new HRI systems. In this study, from the five interaction modes studied, the one preferred by the users was an hybrid solution that interacts with a user by physical contact (a double touch between human and robot) and human hand gestures. In general, when the robot is close to the user, he/she prefers physical contact, otherwise human hand gestures interaction is chosen.

Going forward, we aim to further expand the kind of robot tasks covering several industrial fields, understanding and analyzing user needs.

Acknowledgment

This work was supported in part by the Portugal 2020 project POCI-01-0145-FEDER-016418 by UE/FEDER through the program COMPETE2020 and by the ColRobot project, which has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement no 688807.

References

[1] A. Valera, J. Gomez-Moreno, A. Sanchez, C. Ricolfe-Viala, R. Zotovic, and M. Valles, “Industrial robot programming and upnp services orchestration for the automation of factories,” *International Journal of Advanced Robotic Systems*, vol. 9, no. 4, p. 123, 2012.

[2] Z. Salcic, U. Atmojo, H. Park, A. Chen, and K. I-Kai Wang, “Designing Dynamic and Collaborative Automation and Robotics Software Systems,” *IEEE Transactions on Industrial Informatics*, pp. 1–10, 2017.

[3] Y. Maeda, T. Yoshimi, Y. Ando, and M. Mizukawa, “Task management of object delivery service in kukanchi,” in *2012 IEEE/SICE International Symposium on System Integration, SII 2012*, pp. 283–288, IEEE, dec 2012.

[4] T. Ende, S. Haddadin, S. Parusel, T. Wusthoff, M. Hassenzahl, and A. Albu-Schaffer, “A human-centered approach to robot gesture based communication within collaborative working processes,” in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3367–3374, Sept 2011.

[5] M. Simao, P. Neto, and O. Gibaru, “Using data dimensionality reduction for recognition of incomplete dynamic gestures,” *Pattern Recognition Letters*, vol. 99, pp. 32 – 38, 2017.

[6] P. Neto, D. Pereira, J. N. Pires, and A. P. Moreira, “Real-time and continuous hand gesture spotting: An approach based on artificial neural networks,” in *2013 IEEE International Conference on Robotics and Automation*, pp. 178–183, May 2013.

[7] M. R. Pedersen and V. Krüger, “Gesture-based extraction of robot skill parameters for intuitive robot programming,” *Journal of Intelligent & Robotic Systems*, vol. 80, pp. 149–163, Dec 2015.

[8] S. Sheikholeslami, A. Moon, and E. A. Croft, “Cooperative gestures for industry: Exploring the efficacy of robot hand configurations in expression of instructional gestures for human-robot interaction,” *The International Journal of Robotics Research*, vol. 36, no. 5-7, pp. 699–720, 2017.

[9] C. Schou, J. S. Damgaard, S. Bogh, and O. Madsen, “Human-robot interface for instructing industrial tasks using kinesthetic teaching,” in *IEEE ISR 2013*, pp. 1–6, Oct 2013.

[10] P. J. Koch, M. K. van Amstel, P. Debska, M. A. Thormann, A. J. Tetzlaff, S. Bogh, and D. Chrysostomou, “A skill-based robot co-worker for industrial maintenance tasks,” *Procedia Manufacturing*, vol. 11, no. Supplement C, pp. 83 – 90, 2017. 27th International Conference on Flexible Automation and Intelligent Manufacturing, FAIM2017, 27-30 June 2017, Modena, Italy.

[11] R. E. Andersen, E. B. Hansen, D. Cerny, S. Madsen, B. Pulendralingam, S. Bogh, and D. Chrysostomou, “Integration of a skill-based collaborative mobile robot in a smart cyber-physical environment,” *Procedia Manufacturing*, vol. 11, no. Supplement C, pp. 114 – 123, 2017. 27th International Conference on Flexible Automation and Intelligent Manufacturing, FAIM2017, 27-30 June 2017, Modena, Italy.

[12] M. Safeea and P. Neto, “Kuka sunrise toolbox: Interfacing collaborative robots with matlab,” 2018.

[13] M. Safeea, R. Bearee, and P. Neto, *End-Effector Precise Hand-Guiding for Collaborative Robots*, pp. 595–605. Cham: Springer International Publishing, 2018.

[14] O. Madsen, S. Bogh, C. Schou, R. S. Andersen, J. S. Damgaard, M. R. Pedersen, and V. Krüger, “Integration of mobile manipulators in an industrial production,” *Industrial Robot: An International Journal*, vol. 42, no. 1, pp. 11–18, 2015.