2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Augmented Reality User Interfaces for Heterogeneous Multirobot Control

Rodrigo Chacon-Quesada and Yiannis Demiris[1]

Abstract— **Recent advances in the design of head-mounted augmented reality (AR) interfaces for assistive human-robot interaction (HRI) have allowed untrained users to rapidly and fluently control single-robot platforms. In this paper, we investigate how such interfaces transfer onto multirobot architectures, as several assistive robotics applications need to be distributed among robots that are different both physically and in terms of software. As part of this investigation, we introduce a novel head-mounted AR interface for heterogeneous multirobot control. This interface generates and displays *dynamic joint-affordance signifiers*, i.e. signifiers that combine and show multiple actions from different robots that can be applied simultaneously to an object. We present a user study with 15 participants analysing the effects of our approach on their perceived fluency. Participants were given the task of filling-out a cup with water making use of a multirobot platform. Our results show a clear improvement in standard HRI fluency metrics when users applied dynamic joint-affordance signifiers, as opposed to a sequence of independent actions.**

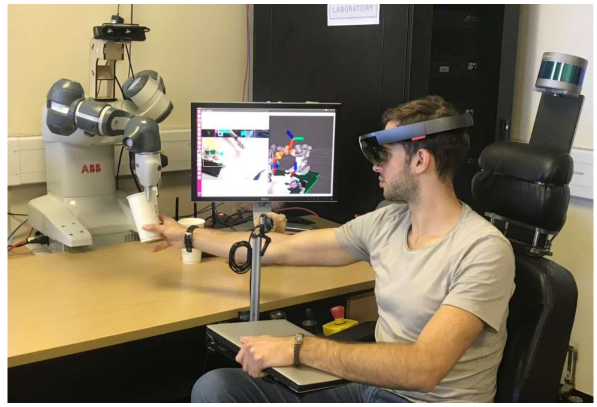

Fig. 1: An assistive multirobot platform composed of a smart wheelchair and a dual arm robot helping a user grasp an object originally out of his reach. The user controls the robots using an AR HMD UI based on the notion of affordances and signifiers.

I. INTRODUCTION

AR user interfaces (UI) for assistive robotics have recently shown great potential towards achieving fluid single-robot control, where additional information, control instructions and affordance sets can be overlaid in the environment using a head-mounted display (HMD) [1], [2], [3]. Benefits of such approaches include more rapid robot instruction [1], [2], or better explanations regarding the internal states or intentions of the robot [4]. Those benefits have motivated us to study the application of AR HMD UI in assistive *multirobot* platforms.

The diverse capabilities required for assistive robotics applications such as mobility assistance (e.g. with smart wheelchairs), household maintenance and meal preparation are difficult, if not impossible to build into a single robot. Therefore, these capabilities can be distributed among different robot types, with each type specialising in different areas [5]. A suitable AR HMD UI must thus be capable of handling heterogeneous multirobot platforms, i.e. platforms made up of robots that are not physically uniform and differ at the software level [6].

Contributions. In section III we present the architecture of our proposed heterogeneous multirobot control platform where *affordance sets*, i.e. representations of the ways in which a robot can interact with an object, are signified using an AR HMD UI (Figure 2). As part of this architecture, we introduce our novel *dynamic joint-affordance signifiers* (Figure 4): AR signifiers that combine actions from multiple robots whenever these actions can be applied simultaneously to an object. These signifiers are overlaid in the environment while considering the relative positions of the user, the objects and the robots (Figure 3). Furthermore, we describe how these signifiers can be updated dynamically, reflecting changes in environmental conditions and multirobot states.

Our next contributions come from the evaluation of our proposed multirobot platform, designed for assistive HRI. First, we present the results from an online survey in subsection IV-A. The survey was aimed at identifying which components of the signifier are more effective towards efficiently presenting the information needed to control such a platform (Figure 5). Influenced by the results from the previous survey, in subsection IV-B we present a user study involving 15 participants that analyses the effects of our approach on standard fluency metrics for HRI [7]. Participants were given the task of filling-out a cup with water making use of the proposed platform (Figures 1 and 6). Our results, presented in section V, show a clear improvement in their perceived fluency when users applied dynamic joint-affordance signifiers, as opposed to a sequence of independent actions. A video abstract is available at `https://youtu.be/owi16TvZYUQ`.

II. RELATED WORK

A. Affordance-based robot control

Affordances represent the interaction possibilities in the world; they are relationships between the agent and the

[1]The authors are with the Personal Robotics Laboratory, Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, United Kingdom {`r.chacon-quesada17`, `y.demiris`}`@imperial.ac.uk`.

RCQ is supported by a scholarship from the University of Costa Rica and Ministry of Science Technology and Telecommunications, Costa Rica. YD is supported by a Royal Academy of Engineering Chair in Emerging Technologies.

978-1-7281-6211-9/20/$31.00 ©2020 IEEE

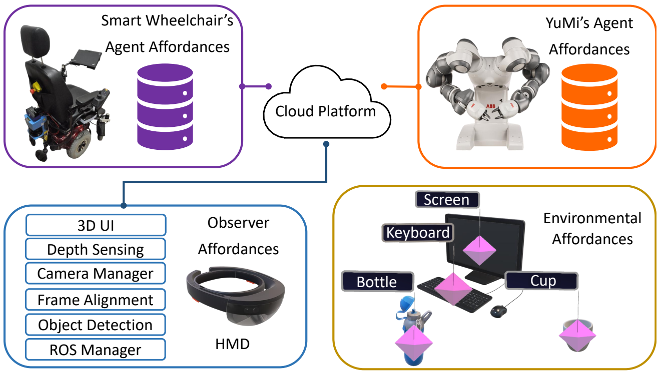

Fig. 2: The HMD handles the interaction with the robots. A cloud platform is used to update a runtime list of observer affordances from the agent affordances reported by the robots. AR signifiers are shown to the user based on the detected objects in the scene.

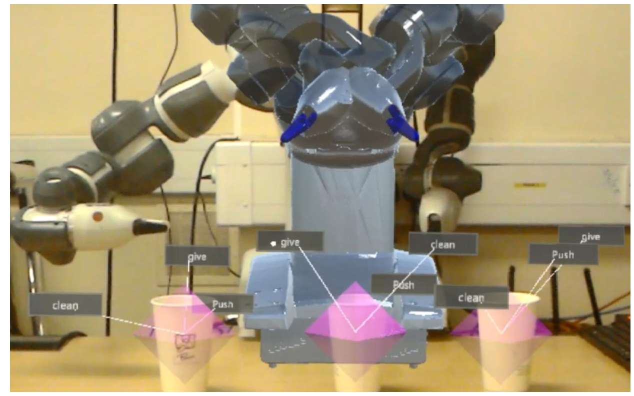

Fig. 3: A real size 3D model of YuMi is overlaid onto the robot. With this, the relative positions of the AR signifiers and the real objects have a 1:1 correspondence relative to the robot's frame.

object, and not stand-alone properties of the object [8]. Signifiers are defined as signals (signs, symbols, static or dynamic icons, or any perceivable indicator) that signify meaningful information to explain what actions are possible, how they should be done and where an interaction should take place [9]. Ideas related to affordances have influenced the design of intelligent robots [10], resulting in applications that can be divided into two main approaches: affordance learning from interaction and geometric information, and visual affordance detection [11].

In the former approach, affordances are learned through direct interaction with the environment. Applications in this domain include learning affordances either from interaction [12] or human demonstration [13]. In visual affordance detection, colour & depth images are used to learn affordances without interaction. Applications in this domain include affordances for HRI [14] and object manipulation [15].

However, most of the previous work associates affordances to single robot platforms, or in the case of multirobot platforms, has been applied to discrete simulated environments [12]. In this paper, we evaluated our approach in a real environment by conducting a user study where participants filled-out a cup with water making use of our proposed affordance-mediated AR HMD UI for heterogeneous multirobot control.

B. User interfaces for multirobot control

Several UI for controlling multirobot platforms have been introduced in the literature. Most of the early works in this domain focused on practical challenges for designing, programming and deploying multirobot platforms, e.g. [16]. However, little attention was given to how the user was engaged with the interaction. In [17], [18], the authors proposed task specific methods for HRI that successfully allowed the user to control the robots, while in [19] a multirobot system learned to extract joint action plans from coordinated demonstrations of multiple tele-operators. More general control algorithms are presented in [20], [21], where either the robots were treated as a unique entity or each robot was engaged independently. However, their methods used traditional graphical UI to interact with the user and were tailored to work on groups of identical robots.

AR UI have also demonstrated potential as a mode of communication during HRI in multirobot platforms. In [22], the authors implemented an AR hand-held UI to control a group of identical robots. While AR HMD UI were also proposed in [23], they were mainly used to provide visual aids to the user. In this paper, we expand the state of the art in this domain by introducing an AR HMD UI for the control of heterogeneous multirobot platforms based on the notion of affordances and signifiers.

III. System Description

Our proposed platform is composed of a smart wheelchair, a dual arm collaborative robot (YuMi) from ABB, and the Microsoft HoloLens. The software modules running on the HoloLens are listed in Figure 2 and are used to display AR signifiers based on a transformation from image pixel coordinates to 3D points in the HMD's frame of reference. A complete description of these modules is presented in [2]. For details about the robot navigation module running on the smart wheelchair see [4].

If the robots are to act on the signified objects, the positions of these objects first need to be converted into the robots' own frame of reference. For this, the *Frame Alignment* module determines a correspondence between the HMD and robot reference frames. The strategy followed remains as described in [3] for the smart wheelchair. For YuMi, whose position does not change over time, a 3D model matching its real dimensions is overlaid onto the robot. With this, the relative positions of the AR signifiers and the real objects have a 1:1 correspondence relative to the robot's frame (Figure 3). To obtain the actual position of the objects in the robot's frame, we apply a transformation from the coordinates of the AR signifiers in the HMD's frame to coordinates local to the robot. Furthermore, a spatial anchor is added to the 3D model to make it stay precisely in place and persist that way across multiple application deployments.

A. Agent, environmental and observer affordances

In [24], the authors proposed a formalisation of affordances from three different perspectives with implications for autonomous robot control. These perspectives relate

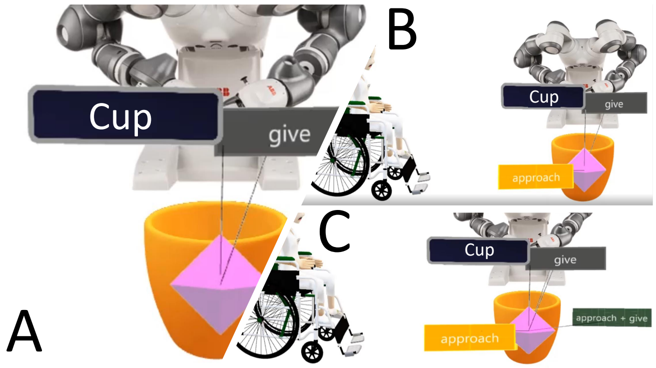

Fig. 4: AR signifiers are dynamically updated based on the surrounding robots. a) only YuMi is present, b) both YuMi and the wheelchair are present, and c) both actions can be jointly-executed.

to that of the agent, the environment and an observer. From an agent's perspective $(effect, (entity, behaviour))$, the affordances reside within the agent that interacts with the environment. From an environmental perspective $(effect, (agent, behaviour))$, the affordances are attached to the objects as extended properties perceived by the agent. Finally, an observer's perspective $(effect, (agent, (entity, behaviour)))$ is used when the interaction between the primary agent and the environment is observed by a third party [24]. In this paper, the term *entity* is used to represent the objects perceived by an *agent*, *behaviour* is used to denote actions that can be applied to these objects, and *effect* represents the consequences of applying such actions.

We applied this formalisation to similarly represent affordances in our platform (Figure 2). To illustrate our approach, for YuMi (*agent*) we developed two different *behaviours* applicable to cups (*entity*): *Give* and *Throw*. *Give* allows YuMi to hand-over (*effect*) cups to the user as shown in Figure 1. *Throw* will discard (*effect*) cups from the table. YuMi's agent affordances are then represented as $(Hand\ Over, (Cup, Give))$ and $(Discard, (Cup, Throw))$.

For the smart wheelchair (*agent*), we developed *Approach*, a *behaviour* applicable to cups, bottles, screens and keyboards (*entities*) that allows the wheelchair to get closer (*effect*) to the positions where these objects are located. These agent affordances are then represented as *(Get-Closer, (Cup, Approach))*, *(Get-Closer, (Bottle, Approach))*, and so forth for the 'Screen' and 'Keyboard' entities.

The key functionality to enable heterogeneous multirobot control is a stand-alone application we developed to report all previous agent affordances to the HMD, which is represented by the dataset symbols in Figure 2. Through this application we can include, edit and delete agent affordances separately for each robot. When the agent affordances are reported by the application, it adds a unique *agent* identifier, information used by the HMD to build a real-time-updated observer affordances run-time list.

This run-time list of observer affordances allows the *3D UI* module to dynamically add or remove AR signifiers if the conditions of the environment change. For example, behaviours currently being performed by a robot will not be shown to the user, or if more robots are added to the platform, their behaviours will be shown when applicable. Affordances selected by the user broadcast a message that includes the selected *behaviour*, the object class (*entity*), the position of the associated object in the robot's frame, and the *agent* identifier of the robot who reported the agent affordance. This way, only the robot with this *agent* identifier will execute the *behaviour* that matches the $(entity, behaviour)$ tuple from its set of agent affordances.

B. Dynamic joint-affordance signifiers

In several cases an object affords multiple actions to the user of an assistive multirobot platform, e.g. when multiple behaviours that originate from one or more robots apply to the same object. Additionally, it is often desirable or even necessary to apply some of these actions simultaneously to the object, i.e. they are sub-actions of another action. For example, a drinking action is composed of a sequence of grasp, pour and drink sub-actions. A suitable AR signifier must thus be capable of simultaneously signifying all of these possible actions. To this end, we developed an AR signifier that dynamically shows multiple behaviours coming from one or more robots while considering if these actions can be jointly applied to an object. We term these novel AR signifiers as *dynamic joint-affordance signifiers* (Figure 4).

Our proposed dynamic joint-affordance signifiers are composed of a *core 3D shape* (the purple diamond in Figure 4), a salient *object class label* and a set of *affordance labels* to separately signify each applicable action. The *object class labels* are added by the *3D UI* module when the dynamic joint-affordance signifiers are created. These signifiers then compare the object class to the *entity* information of all entries in the observer affordances run-time list. All matching entries are added to a list of potential affordances using an environmental perspective representation. This list is handled separately for each dynamic joint-affordance signifier.

A behaviour tree is used to add *affordance labels* to these proposed signifiers. For every entry in the list of potential affordances (if any), this behaviour tree first tests if there is any environmental restriction that makes the behaviour not affordable and therefore should not be shown. If this test is successful, an affordance label is shown to signify the corresponding behaviour and the affordance is added to an environmental affordances list. This test is based on the position of the robot relative to the signifier. In the case of YuMi, the test is successful if the signifier is within its workspace for the right arm. The workspace dimensions were empirically determined by running YuMi's behaviours multiple times while changing the object's position. In the case of the wheelchair, the test succeeds if the signifier is less than 5 metres away from the robot, which was experimentally found to be suitable for the wheelchair's navigation module.

Based on the entries in the environmental affordances list, the behaviour tree runs a second test responsible for creating affordance labels that combine multiple behaviours that can be applied simultaneously to an object. For example, if they are associated with different robots or relies on different

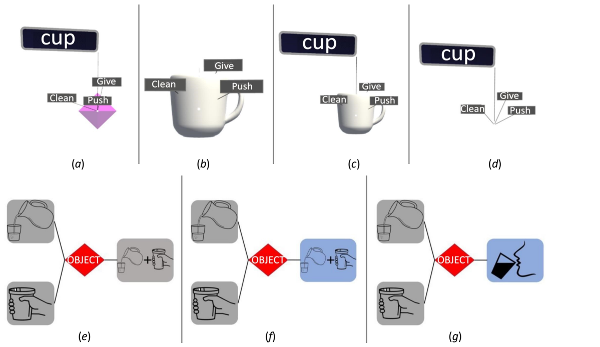

Fig. 5: Survey. (a)-(d) were used to ask about the preferred option to represent affordance sets, while (e)-(g) to ask about the preferred option to represent the joint-execution of two actions over an object.

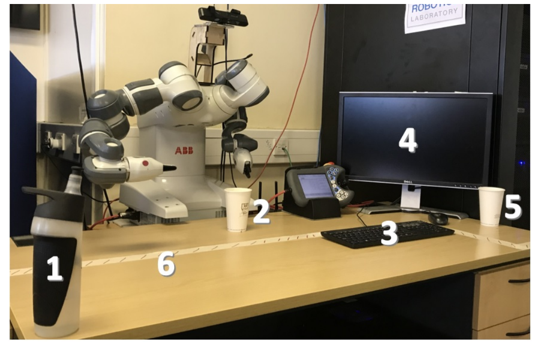

Fig. 6: A bottle (1), two cups (2 and 5), a keyboard (3), and a screen (4) were part of the experimental setup. Participants were not allowed to cross the white line (6) with their hands.

actuators from the same robot. In Figure 4 we illustrate how *affordance labels* are added to a dynamic joint-affordance signifier when $a)$ only YuMi is present, $b)$ both YuMi and the smart wheelchair are present, and $c)$ the two individual robot actions can be simultaneously applied to the cup.

IV. Methodology

A. Survey

To represent dynamic joint-affordance signifiers in our platform, we propose the use of AR signifiers such as the ones shown in Figure 5a-d. Yet, we wanted *affordance labels* combining multiple behaviours to be distinct from the ones used to signify individual behaviours. We turn to the users for identifying the features needed to make this distinction evident and present the results from an online survey aimed at investigating which components of the signifier are more effective towards this goal.

Participants first answered some demographic questions. Then, they chose their preferred representation for an affordance set among four different options. They also had the possibility to choose none of these options and propose an alternative instead. Finally, participants had to choose their preferred representation for the joint-execution of two actions over the same object. Figure 5 shows the representations used in the survey. Table I shows the demographics of the participants and a summary of the results.

The results obtained suggest that an AR signifier representing the overlaid object, e.g. a 3D model of a mug to represent a real mug, is important for participants as Figure 5b and Figure 5c were the two options accumulating more selections. Furthermore, a label describing the object class was important for most participants. In regards to the joint-action representation, most participants chose to explicitly show the two given actions in a single icon whilst this icon is also of a different colour. These results, except for the inclusion of a 3D model for each overlaid object, influenced the design of the AR signifiers presented in this paper as described in subsection III-B and shown in Figure 4.

B. Experimental procedure

We performed a user study to analyse the effects of our proposed dynamic joint-affordance signifiers on standard objective and subjective fluency metrics for HRI [7]. We gave participants the task of filling-out a cup with water making use of our multirobot platform. Participants were only allowed to use signified actions to complete the task. They were not allowed to make use of the smart wheelchair's joystick or get-off from it. They were also not allowed to cross a line marked on the desk with their hands (Figure 6).

To start, participants gave their informed written consent and filled-out demographics and baseline questionnaires. Then, participants started the experiment using the voice command **Start experiment**. AR signifiers were not visible before this time. Participants then needed to decide their strategy to complete the task based on the provided AR signifiers. When the task was completed, participants finished the experiment using the voice command **Stop experiment**.

1) Experimental setup: To avoid variations in the execution time of each robot behaviour between participants, the start position was always the same, approximately 2m from the desk. The smart wheelchair's *Approach* was available for all objects from this point. The positions of the objects shown in Figure 6 were also kept constant. Cup 2 was behind the line marked on the desk but within YuMi's workspace and therefore YuMi's *Give* was also signified for this cup. Furthermore, an *affordance label* combining *Approach* and *Give* was also available for this object. Cup 5 was outside YuMi's workspace and behind the line, it could not be used to complete the task. The bottle was in front of the line, participants could use it just after approaching it. The keyboard and screen were not useful to complete the task.

2) Evaluation: The objectives metrics used to compare between groups were: the percentage of the total task time that the user is not active (H-IDLE), the percentage of the total task time that the robot is not perceivably active (R-IDLE), and the accumulated time, as a ratio of the total task time, between the completion of one agent's action and the beginning of the other agent's action (F-DEL) [7]. Subjective indicators were rated on a seven-point Likert scale from *Very strongly disagree* to *Very Strongly Agree* using the following statements: *(1):* The platform and I worked fluently together; *(2):* The platform contributed to the fluency in the task accomplishment; *(3):* The platform was committed to the success of the task; and *(4):* The platform had an important contribution to the success of the task.

TABLE I: Demographics of the survey participants and the number of times each representation was selected ($N = 37$).

Factor	n	%
Gender		
Female	12	32.4
Male	25	67.6
Age (years)		
18–24	2	5.4
25–34	24	64.9
35–44	8	21.6
45–54	3	8.1

Signifier	n	%
Affordance set		
Fig. 5a	1	2.7
Fig. 5b	13	35.1
Fig. 5c	15	40.5
Fig. 5d	6	16.2
Other	2	5.4
Joint-action		
Fig. 5e	3	8.1
Fig. 5f	20	54.1
Fig. 5g	14	37.8

3) Hypotheses: The use of dynamic joint-affordance signifiers, which combine multiple robot actions that can be jointly-applied to an object, will increase the median score of subjective and objective indicators of fluency in HRI.

V. Results

Participants were classified into two groups based on the strategy they followed to complete the task. Participants in Group 1 triggered *Approach* and *Give* simultaneously. Participants in Group 2 triggered both behaviours separately instead. Either way, participants were then able to proceed filling-out the cup with water using the bottle. A total of 15 participants completed the experiment. Five of these were classified as part of Group 1 (33%) and 10 as part of Group 2 (66%). Group 2 was further divided depending on whether participants first triggered *Approach* (Group 2 A) or *Give* (Group 2 B). In Table II, we show the demographics and a summary of participants' reported experience with VR, AR, computer games and smart wheelchair technologies.

In Figure 7, we show a time diagram representation of the strategies followed by participants to complete the task. Therein, $T_A = 9.61s$ and $T_G = 26.01s$ represents the execution times of *Approach* and *Give* respectively, t_{TT} is the total task time, t_{FD} is the time between the completion of the first and second behaviour, and $T_R = 21.26s$ is the time in which YuMi releases the cup. T_R is considered separately as the cup can be used immediately after being released by YuMi, even though *Give* has not finished by this time. From these diagrams, we derived the equations used to calculate the objective indicators described in Section IV-B.2:

$$H\quad IDLE_{G1} = H\quad IDLE_{G2A} = T_R/t_{TT} \tag{1}$$

$$H\quad IDLE_{G2B} = T_A/t_{TT} \tag{2}$$

$$R\quad IDLE_{SW} = (t_{TT}\quad T_A)/t_{TT} \tag{3}$$

$$R\quad IDLE_{YM} = R\quad IDLE_{N\quad G1} = (t_{TT}\quad T_G)/t_{TT} \tag{4}$$

$$R\quad IDLE_{N\quad G2} = (t_{TT}\quad (T_A + T_G))/t_{TT} \tag{5}$$

$$F\quad DEL_{G1} = (T_A\quad T_G)/t_{TT} \tag{6}$$

$$F\quad DEL_{G2} = t_{BS}/t_{TT} \tag{7}$$

For Group 2, Equation 1 and Equation 2 treat participants as active (whether scanning the scene or selecting a behaviour) up to the point of second action selection. In regards

TABLE II: Demographics of the experiment participants and their reported experience with other technologies ($N = 15$).

Factor	n	%
Gender		
Female	4	26.7
Male	11	73.3
Age (years)		
18–24	2	13.3
25–34	13	86.7

Baseline	n	%
VR	8	53.3
AR	8	53.3
Computer games	12	80
SW visual feedback	2	13.3
None	13	86.7

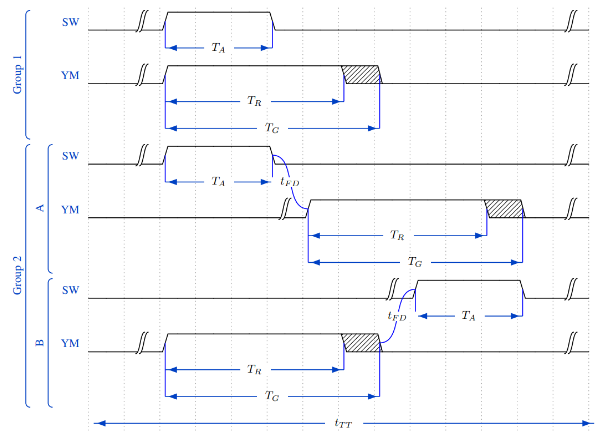

Fig. 7: Time diagram representation of the strategies followed by participants to complete the task.

to R-IDLE, there are two possible ways of measuring it in our experiment, one approach is to measure it separately for each robot (Equation 3 and Equation 4), the second possibility is to measure it as the time when none of the robots were active (Equation 4 and Equation 5). Equation 5 assumes no overlapping during the execution of the behaviours. The values obtained with Equation 6 will be negative as the behaviours overlap during their execution.

Subjective metrics (Figure 9) were rated higher in all questions by participants in Group 1, giving a mean score of 6.7, higher than the 5.98 reported by Group 2. We do not see an effect of the dynamic joint-affordance signifiers over H-IDLE but note a tendency towards higher values in Figure 8a where results positively correlate with fluency [7]. Figure 8b-c show the results of R-IDLE separately for each robot. The mean scores obtained are lower for Group 1 in both cases, R-IDLE is reported to be consistently inversely correlated with fluency [7]. The outliers shown in these two plots helps to reinforce this last argument as they are both from participant 2, who rated all subjective metrics as 7.

If R-IDLE is measured when both robots are inactive, the mean value obtained for each group is similar. However, the upper outlier shown in Figure 8d corresponds to participant 6, which was the slowest amongst the participants. Interestingly, contrary to what was expected, this participant rated their subjective metrics highly (average score of 6.75). Finally, Figure 8e shows the results for F-DEL. The mean values obtained are higher for Group 2, suggesting a clear improvement. Furthermore, the outliers shown in this plot are associated with participants who on average rated 6.25 (below Group 1's average) and 5.25 (lowest score obtained)

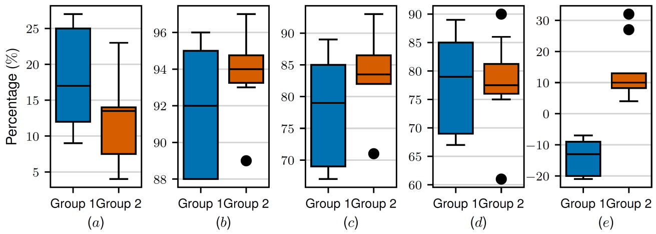

Fig. 8: Objective fluency metrics results. (a) H-IDLE. (b) $R{-}IDLE_{SW}$. (c) $R{-}IDLE_{YuMi}$. (d) $R{-}IDLE_{N}$. (e) F-DEL.

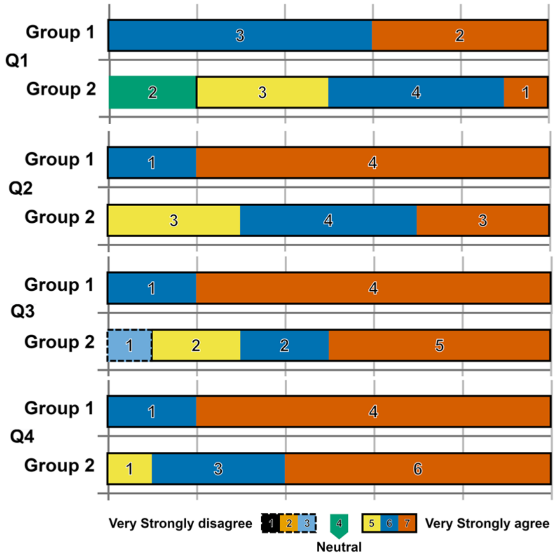

Fig. 9: Likert chart summarising the answers of the participants for each question in the post-experiment survey (best seen in color).

their subjective metrics. This means that our results are in line with [7], who demonstrated F-DEL to have the strongest correlation with subjective fluency perception in HRI.

VI. Conclusions

User Interfaces (UI) for assistive robotics that enhance interaction fluency have the potential to improve the quality and efficiency of HRI. In this paper, we presented a novel AR HMD affordance-mediated UI for controlling a heterogeneous multirobot platform, and detailed our architecture for creating, updating, displaying, combining and selecting affordances. Furthermore, we demonstrated how our interface positively impacts subjective and objective metrics of fluency in HRI, especially when AR signifiers that automatically combine multiple robot actions are encapsulated in the UI. In addition, we provided remarks on how such signifiers should include the corresponding 3D model and object class for each object, and use distinct features, e.g. colour, to signify joint-affordances. Future work will further probe visualisation methodologies for presenting higher level options to the user, which require multirobot affordances that are more complex in terms of sequence length and action space.

References

[1] R. Chacon and Y. Demiris, "Augmented reality control of smart wheelchair using eye-gaze–enabled selection of affordances," in *IROS 2018 Workshop on Robots for Assisted Living*, 2018. [Online]. Available: https://spiral.imperial.ac.uk:8443/handle/10044/1/71986

[2] ——, "Augmented reality controlled smart wheelchair using dynamic signifiers for affordance representation," *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2019)*, pp. 4812–4818, 2019.

[3] M. Zolotas, J. Elsdon, and Y. Demiris, "Head-mounted augmented reality for explainable robotic wheelchair assistance," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct 2018, pp. 1823–1829.

[4] M. Zolotas and Y. Demiris, "Towards explainable shared control using augmented reality," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 3020–3026.

[5] L. E. Parker, "Alliance: An architecture for fault tolerant multirobot cooperation," *IEEE transactions on robotics and automation*, vol. 14, no. 2, pp. 220–240, 1998.

[6] T. Balch and L. E. Parker, *Robot teams: from diversity to polymorphism*. AK Peters/CRC Press, 2002.

[7] G. Hoffman, "Evaluating fluency in human–robot collaboration," *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 3, pp. 209–218, June 2019.

[8] J. J. Gibson, *The senses considered as perceptual systems.* Boston, MA, USA: Houghton Mifflin, 1966.

[9] D. Norman, *The design of everyday things: Revised and expanded edition*. Constellation, 2013.

[10] P. Ardón, È. Pairet, K. S. Lohan, S. Ramamoorthy, and R. Petrick, "Affordances in robotic tasks–a survey," *arXiv preprint arXiv:2004.07400*, 2020.

[11] S. McMahon, N. Sünderhauf, B. Upcroft, and M. Milford, "Multi-modal trip hazard affordance detection on construction sites," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 1–8, 2017.

[12] H. Min, R. Luo, *et al.*, "Affordance matching from the shared information in multi-robot," in *2015 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2015, pp. 66–71.

[13] H. S. Koppula and A. Saxena, "Anticipating human activities using object affordances for reactive robotic response," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 1, pp. 14–29, 2016.

[14] C.-Y. Chuang, J. Li, A. Torralba, and S. Fidler, "Learning to act properly: Predicting and explaining affordances from images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 975–983.

[15] T. Nagarajan, C. Feichtenhofer, and K. Grauman, "Grounded human-object interaction hotspots from video," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 8688–8697.

[16] M. Li, A. Alvarez, F. De Pellegrini, B. Prabhakaran, and I. Chlamtac, "Robotrak: a centralized real-time monitoring, control, and coordination system for robot swarms," in *Proceedings of the 1st international conference on Robot communication and coordination*, 2007, pp. 1–4.

[17] S. Bashyal and G. K. Venayagamoorthy, "Human swarm interaction for radiation source search and localization," in *2008 IEEE Swarm Intelligence Symposium*. IEEE, 2008, pp. 1–8.

[18] X. C. Ding, M. Powers, M. Egerstedt, and R. Young, "An optimal timing approach to controlling multiple uavs," in *2009 American Control Conference*. IEEE, 2009, pp. 5374–5379.

[19] M. F. Martins and Y. Demiris, "Learning multirobot joint action plans from simultaneous task execution demonstrations." in *AAMAS*, 2010, pp. 931–938.

[20] Z. Kira and M. A. Potter, "Exerting human control over decentralized robot swarms," in *2009 4th International Conference on Autonomous Robots and Agents*. IEEE, 2009, pp. 566–571.

[21] A. Kolling, S. Nunnally, and M. Lewis, "Towards human control of robot swarms," in *Proceedings of the seventh annual ACM/IEEE international conference on human-robot interaction*. ACM, 2012, pp. 89–96.

[22] J. Patel, Y. Xu, and C. Pinciroli, "Mixed-granularity human-swarm interaction," in *2019 International Conference on Robotics and Automation (ICRA)*, May 2019, pp. 1059–1065.

[23] M. Daily, Y. Cho, K. Martin, and D. Payton, "World embedded interfaces for human-robot interaction," in *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the*. IEEE, 2003, pp. 6–pp.

[24] E. Şahin, M. Çakmak, M. R. Doğar, E. Uğur, and G. Üçoluk, "To afford or not to afford: A new formalization of affordances toward affordance-based robot control," *Adaptive Behavior*, vol. 15, no. 4, pp. 447–472, 2007.