Analysis and Semantic Modeling of Modality Preferences in Industrial Human-Robot Interaction

Stefan Profanter[1], Alexander Perzylo[1], Nikhil Somani[1], Markus Rickert[1], Alois Knoll[2]

Abstract— **Intuitive programming of industrial robots is especially important for small and medium-sized enterprises. We evaluated four different input modalities (touch, gesture, speech, 3D tracking device) regarding their preference, usability, and intuitiveness for robot programming.**

A Wizard-of-Oz experiment was conducted with 30 participants and its results show that most users prefer touch and gesture input over 3D tracking device input, whereas speech input was the least preferred input modality. The results also indicate that there are gender specific differences for preferred input modalities.

We show how the results of the user study can be formalized in a semantic description language in such a way that a cognitive robotic workcell can benefit from the additional knowledge of input and output modalities, task parameter types, and preferred combinations of the two.

I. Introduction

Reducing the cost and increasing the efficiency of industrial robot systems for small and medium-sized enterprises (SMEs) can be achieved, among other things, by implementing intuitive programming interfaces. Domain experts from different fields, such as welding or assembly, should be able to teach a robot program by relying on their domain-specific knowledge instead of thinking about which commands are needed to perform the desired action.

To achieve this goal we use a task-based programming approach: the *process* (e.g., assemble gearbox) is divided into multiple *tasks* (e.g., put bearing on axis). This definition is based on object-level and is independent of a specific robot system. Each task is composed of robot *skills* (e.g., move to, close gripper) that directly map to functions of the robot system. The user has to define parameters for each task (e.g., object to pick), while missing parameters on the skill level are inferred automatically.

We conducted a user study to extend our semantic knowledge database by evaluating the usability (efficiency and effectiveness) and intuitiveness (simplicity and ease of learning) of different input modalities (touch, gesture, speech, 3D pen) for programming an industrial robot system. Using questionnaires and experiments, we collected data from 30 participants and analyzed various aspects of using our robotic workcell (see Fig. 1).

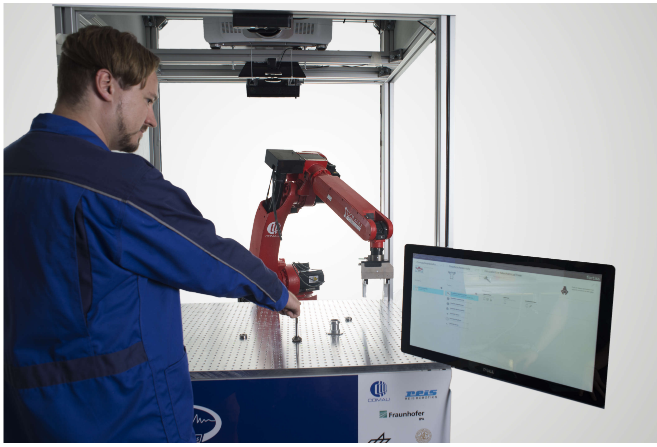


Fig. 1: The cognitive robotic workcell which was used for the user study. It includes a robot, a touchscreen, and different sensors for 3D motion and device tracking. A projector above the table gives visual feedback on the tabletop.

Our hypothesis is that using different multimodal types of input makes robot programming easier and that gesture or 3D pen input will be preferred over touch and speech input. We aim to confirm or contradict this hypothesis with our experiment. The study was built up as a Wizard-of-Oz experiment to provide system and implementation independent results [1].

Based on the results of the user study, we developed a semantic description of input and output modalities, task parameter types, and preferred combinations of the two. This information is used in our cognitive robotic workcell and enables the user to program robot tasks in an intuitive way within the assembly and welding domains using automatically inferred, suitable multimodal input modalities.

II. Related Work

Different studies in the past have shown that multimodal systems are preferred by users over their unimodal alternatives, resulting in higher flexibility and reliability while better meeting the needs of a variety of users [2], [3]. Such systems provide multiple advantages, including improved efficiency, alternating interaction techniques, and accommodation of individual differences, such as permanent or temporary handicaps [4], [5]. Our previously conducted pre-study with one participant has shown that using our intuitive programming approach reduces the required teach-in time by 83% [6].

Using speech as the main input channel to program a robot was evaluated by multiple research groups [7], [9]. They concluded that the bottleneck of using speech input is the availability of easily parameterizable robot skills.

[1] Stefan Profanter, Alexander Perzylo, Nikhil Somani, and Markus Rickert are with fortiss GmbH, An-Institut Technische Universiät München, Munich, Germany

[2] Alois Knoll is with Robotics and Embedded Systems, Department of Informatics, Technische Universiät München, Munich, Germany

Correspondence should be addressed to `profanter@fortiss.org`

Industrial robot programming using markerless gesture recognition and augmented reality is evaluated in [10]. It allows the user to draw poses and trajectories into the workspace. The authors state that such systems show a significant reduction of required teach-in time. A combination of human intention and language recognition for task-based dual-arm robot programming is evaluated in [11]. They conclude that using multimodal input reduces the programming complexity and allows non-experts to move a robot easily.

User group differences in modality preferences for interaction with a smart-home system are presented in [12]. Their study shows that female participants prefer touch and voice over gestures. Male subjects prefer gesture over voice.

A comprehensive overview of programming methods for industrial robots until the year 2010 is presented in [13], including online (sensor guided), offline programming (CAD data), and augmented reality. Hinckley and Oviatt evaluate different input modalities based on various aspects like input speed or user feedback possibilities [14], [15]. By using multiple modalities, the system can make up for the shortcomings of other modalities.

III. User Study Design

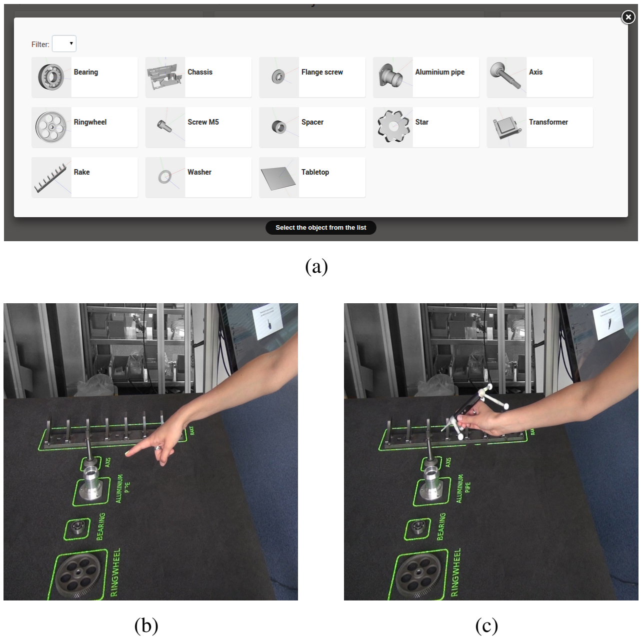
Our user study covered three important use cases for industrial robots used in small and medium-sized enterprises (SMEs): Pick & Place, Assembly and Welding. The main goal was to determine the preferred input modality for different types of parameters and whether using different input modalities is suitable for industrial robot programming. By using the Wizard-of-Oz approach, where a human observer simulates sensor input values without the participant's knowledge, we made sure that the survey evaluates the concept of using multimodal input rather than a specific implementation. The human observer manually set the object to which the participant pointed to (gesture, 3D pen) and also interpreted the speech input by setting the parameters accordingly on the user study administration panel.

The following input modalities were available for defining task parameters:
Touchscreen: displaying a list of objects and 3D visualizations (Fig. 2(a))
Gesture: selecting objects and welding points by hovering over them with the index finger (Fig. 2(b)). [16]
Speech: no special vocabulary was needed due to Wizard-of-Oz. A human supervisor interpreted the spoken commands without the participant's knowledge.
3D Pen input: pen-shaped tool tracked in 3D space with infrared spheres and equipped with a button to confirm the selection (Fig. 2(c)). [17]

The touchscreen was also used to give the user feedback during gesture, speech, and 3D pen input, i.e., it displayed the selected object or value for the parameter. It also allowed the human observer to show an error message to the test subject if the input was unclear.

A. The Four Phases

The study was based on the SUXES evaluation method which allows a comparison of the expected behavior with the experienced behavior of the system [18]. Using this method we try to increase the acceptance of the new system by converging the experience to the expectation during the iterative development process resulting in a more efficient and easy to use system.



(a)

(b)

(c)

Fig. 2: Different input modalities used during the user study: touch input, gesture input, 3D pen input and speech. (a) Touch input modality for object selection, (b) the user points to an object to select it, (c) the user can select objects using the 3D tracked pen.

The questionnaires used during the study are available online[1]. They divide the user study into the following four phases.

1) Introduction to the Evaluation & Background Questionnaire: Gives a short introduction to the different phases and to the main goal of the experiment. The background questionnaire obtained information such as age, gender, expertise in using computers, and knowledge in the field of robotics in general.

2) Introduction to the Application & Expectation Questionnaire: Explains the system, including the different used sensors, which strengthen the impression of a working system. For our Wizard-of-Oz the sensors did not need to be plugged in. The robotic workcell shown in Fig. 1 was used during the study to make the experiment even more realistic. The robot attached to the workcell was not moving during the study, but still it gave the participants a more realistic environment.

The participant had to try out all four input modalities (touch, gesture, speech, 3D pen) to get a first impression of the interaction and was then presented with a questionnaire covering the expectation of the user regarding the different input methods. Participants had to order the input modalities

[1] `https://github.com/fortiss/robotics-mmio-preferences`

according to preference with respect to the following parameters: select an object, set a location where to place the object, set assembly constraints, select a point on the object.

3) User Experiment & Experience Questionnaire: Main evaluation step asking the user to program all proposed tasks using touch input only, then the same tasks with gesture input, followed by speech and finally 3D pen input. The four tasks were:

Pick & Place: Select an object and place it on the center of the table. Parameters: object to pick, location to place

Assemble two objects: Select the bearing and put it concentrically on the axis. Parameters: object to pick, object to place on, assembly constraints/pose

Weld point: Select the rake object and weld a specific point. Parameters: object to weld, point on object

Weld seam: Select the rake object and weld a specific seam. Parameters: object to weld, seam on object

The experience questionnaire asked the user to order the input modalities based on the experienced cognitive load by indicating which input modality was the most demanding. The participant then completed the expectation questionnaire again, allowing us to compare the expectation with the experience which supports the developing of a system which meets the user's expectation. More importantly the results of this questionnaire show which input modalities were preferred for which parameter type.

4) Opinion Questionnaire: The participants had to fill out the opinion questionnaire to state their opinion on the different input modalities and the whole system. The goal of this questionnaire was to understand the overall subject's impression of the system and how useful the usage of different input modalities is.

IV. Evaluation, Results & Discussion

The interaction of each participant was recorded using a video camera and the results of the questionnaire were collected in a spreadsheet[2]. The evaluation of these results is divided into four main categories: participant's background, the expected behavior of the system, preferred input modalities, and opinion about the system. For each participant it took about 30 minutes to complete the questionnaires and experiments.

The following statistical notations are used: M: arithmetic mean, SD: standard deviation, p_{tp}: p-Value using paired t-Test, p_{tu}: p-Value using unpaired t-Test, p_W p-Value using Kruskal-Wallis test, p_{Wd} Kruskal-Wallis test with post-hoc pairwise Dunn's Test.

A. Participants

The data includes answers from 30 participants. A majority of the participants were students and researchers of different technical and nontechnical fields including robotics and embedded systems.

The mean age over all participants was 27 years ($SD = 5$). There were 23 male (77%) and 7 female (23%) participants. 21 participants indicated that they are computer experts, 7 were advanced users, and 2 basic users.

43% of the participants replied to the question "*How much do you know about robotics?*" with "*Not that much (heard or read about it)*". 7% indicated they are hobby roboticists and 50% said they know a lot about robots (studied robotics, computer science, or similar).

60% of the participants did not know what a TeachPad is or how it works. 20% indicated that they know what it is and how it works, but had never used one. The final 20% had already programmed an industrial robot with a TeachPad.

B. Expectation

After actual usage of the input modalities, the participant was asked about his impression on the ease of use. Fig. 3 shows the mean value and standard deviation over all 30 participants for each question. As can be seen, the system outperformed the user's expectation for most of the components in terms of simplicity: Gesture, speech, and 3D pen input were easier to use than expected, mostly due to the fact that these used the Wizard-of-Oz approach and therefore did not rely on a perfectly working recognition. Only touch input behaved as expected since it represented the current state of the art (e.g., as used on smartphones).

Looking at the data separated by gender, there is a notable difference ($p_{tu} = 0.13$) for speech input: Female participants found it easier ($M = 1.7$) than expected to describe tasks using spoken words compared to male participants ($M = 2.3$).

C. Preferred Input Modalities

The questionnaires included questions where the participants had to order the input modalities according to their preferred usage. The same questions were asked during phase two and phase three (before and after the experiment). Figures 4(a) to 4(c) show the results for those questions.

The *Select an object* parameter required the user to set an object model using one of the four available input modalities. Most of the users selected *gesture input* ($M = 1.63$, $p_W < 0.001$) as preferred or second preferred input modality after *touch input* (Fig. 4(a)). The other three modalities lie between $M = 2.5$ and $M = 3$, in which *Speech* has the highest standard deviation ($SD = 1.2$). Speech was often set as the least preferred modality, but on occasion was selected as the most preferred one, resulting in this high standard deviation. Comparing speech input to touch input, the user had to know the name of the object, which was not the case for most of the participants even if it was projected beneath the object on the tabletop.

The graph also shows that before the experiment, users thought they would not like speech input, but after using it, speech gained a better mean than before. This may be due to the fact that most people had a prior experience with speech input or due to the general impression that speech input does not work well enough, and after finishing the experiment they saw that speech recognition is better than they had thought.

[2] `https://github.com/fortiss/robotics-mmio-preferences`

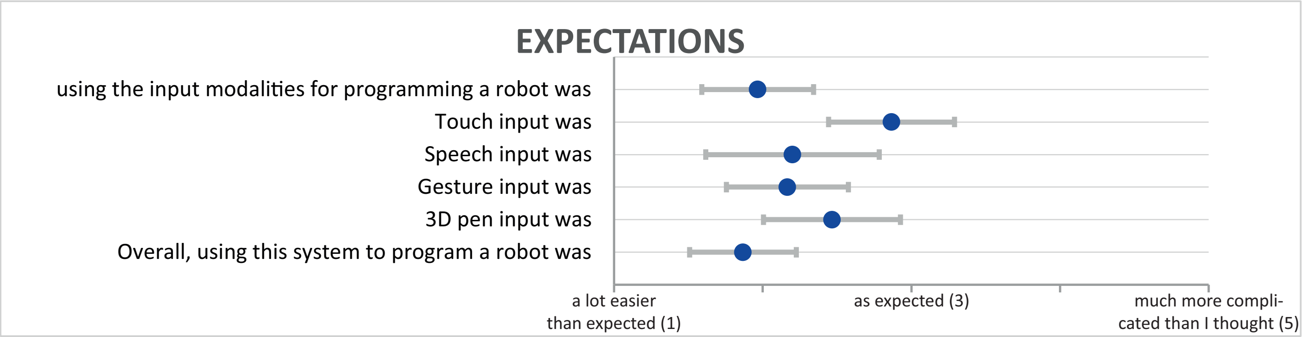


Fig. 3: Diagram showing the mean value and standard deviation for each listed question if it matched the user's expectation.

Separated by gender, there is a more significant difference between male and female participants for touch input ($p_{tu} = 0.098$) and speech input ($p_{tu} = 0.054$). For female participants, touch input was more preferred ($M = 2.1$) compared to the male participants ($M = 2.7$). In contrast to this, most female participants put speech input in the last place. This contradicts the results of [12], where speech was more preferred by female participants than by male participants.

Evaluating the data separately for users who have already used a TeachPad shows a significant difference ($p_{tp} = 0.013$) regarding the expected preferred input modality (answers before the practical component) and the finally preferred input modality. They expected that they would not like speech input ($M = 3.4$). After the practical component the average position for speech input as preferred input modality moved from $M = 3.4$ to $M = 2.4$, which places speech input from least preferred to the second most preferred input modality. This is most likely caused by the fact that those users thought that speech input is not really suited for programming a robot in any way. After using it, they were persuaded and thought that—given the perfect speech recognition in our case—it may be a helpful addition.

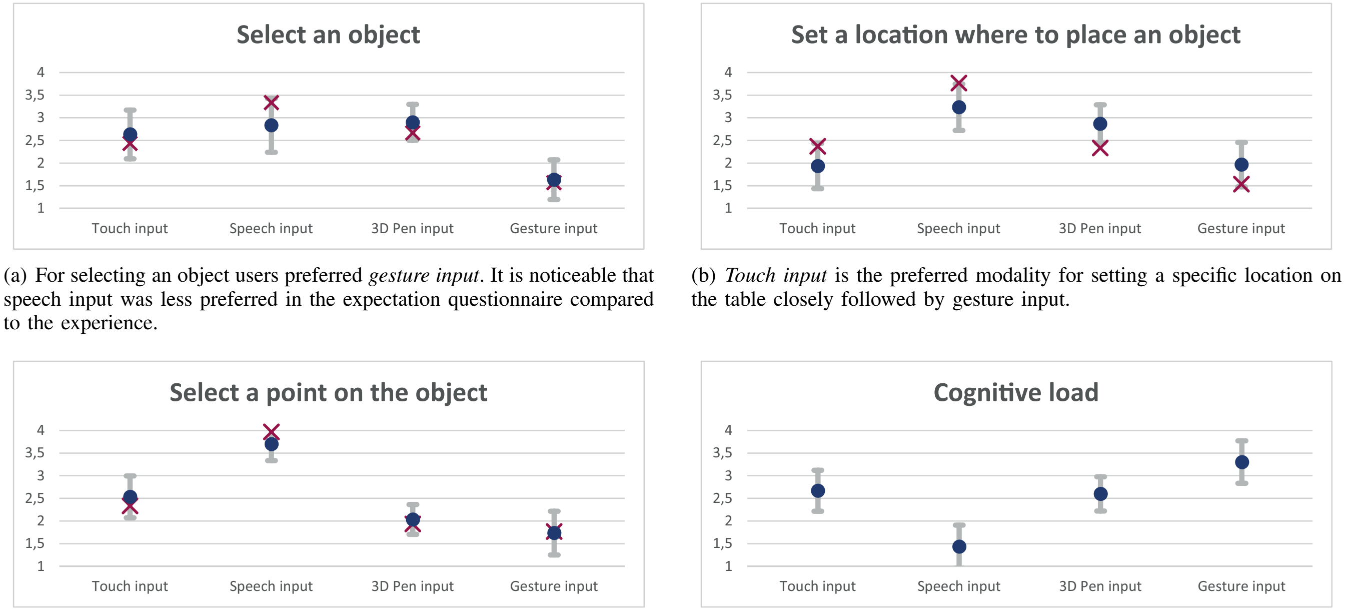
For the *Set location* parameter (Fig. 4(b)), the choice of the participants is quite different. In the first place is *Touch input*, followed closely by *Gesture input*. There is no significant difference between those two modalities ($p_{Wd} = 0.9$), whereas they differ significantly from *3D pen input* and *Speech input* ($p_W < 0.002$). 3D pen input has approximately the same average value as for *Select an object*, speech input has a lower average. During the open discussion at the end of the study, some participants mentioned that for them an influential factor for choosing a specific order was the time needed to switch between touch and other input modalities. This may have played a bigger role for the location parameter, since the visualization of the table on the touchscreen was simply a 2D rectangle, where the user had to touch a specific location. Staying on the touchscreen is much faster than turning to the table and pointing to a specific location. *Speech* was the least popular since it is more complicated to describe a position accurately compared to simply pointing at it.

Select a point is needed to define a welding point. The problem specification was to set the third point from the bottom left as the welding point. For touch input, the user had a 3D visualization of the object and had to touch the correct vertex to select it. Gesture input was the preferred method for setting this parameter followed by 3D pen input. It is more intuitive to use a finger or a pointed object to accurately define a position in 3D. Using speech only is quite difficult and the majority of the participants were unsure how to describe the 3D position accurately. Nearly all participants who indicated they are experts in using computers did not like speech input for setting a point. 20 out of the 21 experts set speech input as their least favorite method ($M = 3.9$).

To indicate the perceived cognitive load for each input modality, the participants had to order the four modalities according to where they had to think the most during the interaction. Fig. 4(d) shows that *speech input* required the most thinking, whereas *Gesture input* was selected by most users as the easiest one. Even if the system was built to understand everything (Wizard-of-Oz experiment), users were in doubt about which vocabulary they could use, despite being told they can talk to the machine as if they were talking to a human. This is also related to the findings by [1], who found out that dialogs between human and human differ significantly from those between human and machine.

Comparing the cognitive load between the seven female and 23 male participants, there is a significant difference between touch input ($p_{tp} = 0.037$) and gesture input ($p_{tp} = 0.065$). The average over all female participant's results show the following order of cognitive load (highest to lowest): Speech input, 3D pen input, gesture input, touch input. Compared to the male participant's (speech, touch, 3D pen, and then gesture input), using touch input was easier for female participants than for men, who selected gesture input as the easiest one.

The overall evaluation, combining all the parameters together, confirms our hypothesis: *Gesture input* is the most preferred modality ($p_{Wd} < 0.0001$). *Touch input* and *3D pen input* can be grouped in second place (no significant difference between touch and 3D pen ($p_{Wd} = 0.64$), whereas *Speech input* is significantly the least preferred modality ($p_{Wd} < 0.0001$).



(a) For selecting an object users preferred *gesture input*. It is noticeable that speech input was less preferred in the expectation questionnaire compared to the experience.

(b) *Touch input* is the preferred modality for setting a specific location on the table closely followed by gesture input.

(c) Selecting a point on the object by pointing with a finger was preferred by most users. *Pen input* is close to *Gesture input*.

(d) Input modalities ordered by cognitive load (where the user did have to think the most, where it required the most attention). *Speech* was perceived as the highest demanding modality.

Fig. 4: Graphs showing the results for each question where the user had to order the input modalities for the different tasks or cognitive load according the subject's experience in the previous practical part. For the cognitive load, 1 means high load and 4 stands for low cognitive load. For all the other graphs 1 is the most and 4 the least preferred modality. The *blue* circle indicates the average over all 30 participants, the *gray* bar shows the standard deviation for each modality. The *red* cross is the average from the expectation questionnaire where the user had to order the modalities according to which he will likely prefer the most.

D. System Opinion

Phase 4 of the SUXES-based User-Study consisted of an opinion questionnaire. It included questions from the After-Scenario Questionnaire (ASQ) [19] and the Software Usability Measurement Inventory (SUMI) [20].

The goal of this questionnaire was to get a basic impression on what the user thinks about the system and the different input modalities. Fig. 5 shows the distribution of the answers from all 30 participants.

The graph shows that on average, users were satisfied with the ease of completing the tasks and the required time it took to complete them. Speech has the highest standard deviation ($SD = 1.36$) which means that some participants found speech input more intuitive than others, probably due to the fact that non-expert users struggled naming the objects compared to expert users.

Separating the data by the level of expertise shows that experts would prefer if the system preselects the most suited modality, whereas non-expert users would like to select the input modality on their own ($p_{tu} = 0.002$). Experts or daily users of the system tend to rush through the programming steps, thus selecting a specific one would slow them down. Knowing in advance which modality is the default for which parameter would make their interaction with the system much more efficient.

V. Application of Results

A. Self-Adjusting GUI Based on Semantic Descriptions

The findings of the user study have been encoded in a semantic description language to serve as an additional source of information for our cognitive robotic workcell. The software framework used to control the workcell supports processing robot tasks, object models, and workcell setups specified in the Web Ontology Language (OWL). OWL is based on a logical formalism, allows for automatic reasoning, and mitigates the effort of combining and integrating knowledge.

The mentioned workcell offers a human-friendly graphical user interface (GUI) to program the robot in an intuitive way. The GUI adjusts itself based on the domain of the robot task the human operator is currently working on and the types of the corresponding parameters. By defining different types of parameters (e.g., velocities, manipulation objects, 3D or 6D poses) and the preferred modalities for setting that kind of parameter (e.g., pointing gesture, on-screen object libraries), the GUI can automatically choose the preferred modalities and select a suitable appearance. It even supports new task types, as long as they can be parameterized using the known parameter types.

Fig. 6 shows the parameter types and I/O modality taxonomies and how the parameters of a task are specified.

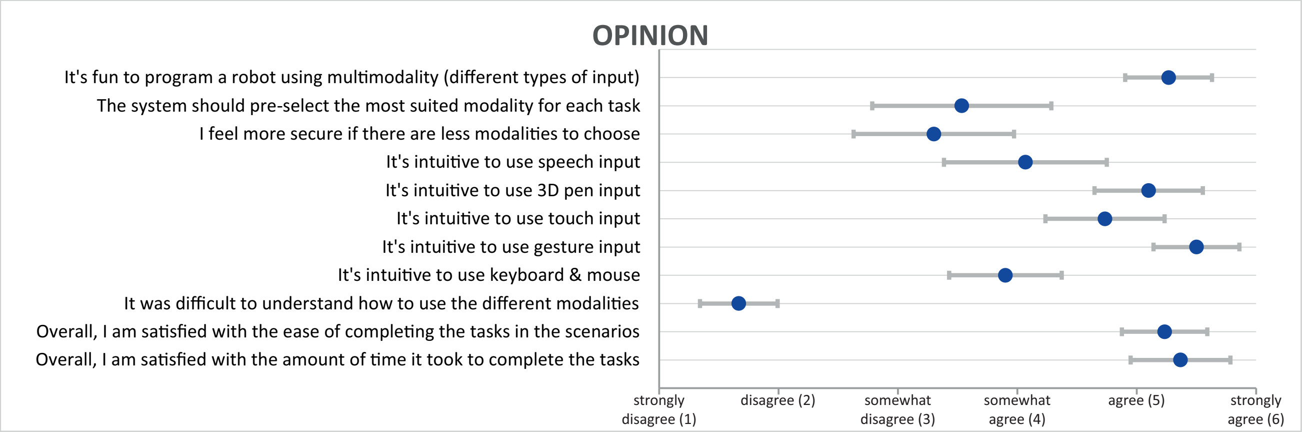


Fig. 5: Distribution of the answers for questions about the intuitiveness of the input modalities and opinions about the whole system. The *blue* dot marks the arithmetic mean of all 30 participants, the *gray* line indicates the standard deviation.

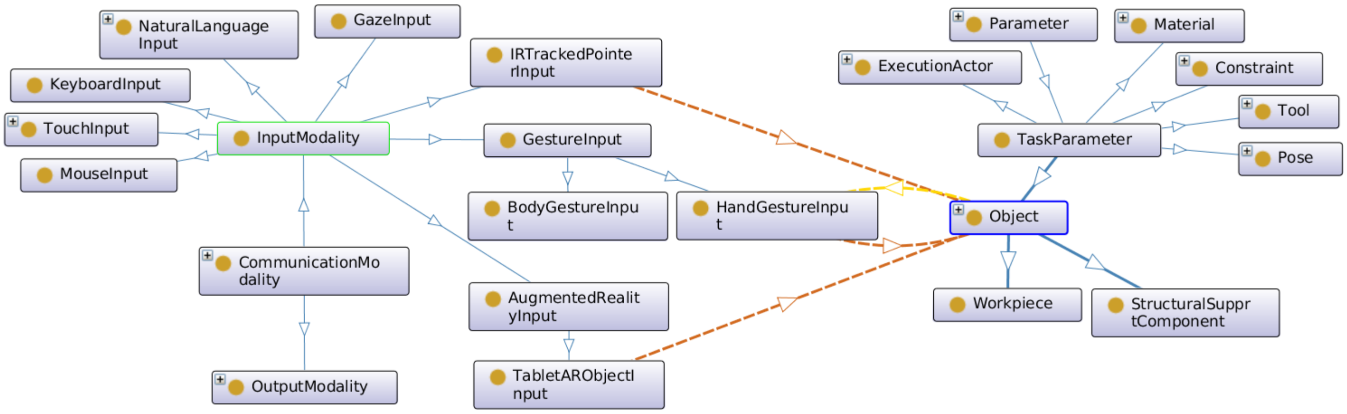


Fig. 6: Visualization of the semantic description of taxonomies for input modalities and task parameter types (*blue* arrows represent subclass relations). The excerpt shows that the *Object* parameter type can be set using the input modalities *IRTrackedPointInput*, *HandGestureInput* and *TabletARObjectInput* (*orange* dashed arrows). The *preferredModality* (*yellow* dashed arrow) for this parameter type is set to be *HandGestureInput*.

The given example shows an excerpt for the *Object* task parameter type. This parameter type can be set using *3D pen input*, *gesture input*, or augmented reality, whereas gesture input is set as the preferred input modality. As the types of parameters are known, the corresponding preferred modalities for setting them can be inferred. The intuitive teaching interface can automatically adapt its appearance accordingly.

B. Robot Cell Design

For the user study we used a robotic workcell without any functioning sensors since we wanted to conduct a Wizard-of-Oz experiment to get system independent results. Using the feedback from the user study, we are now able to build up an improved workcell which better fits the user's needs during robot programming.

We will focus on gesture input, which is one of the preferred input modalities. The touch input user interface will be further improved to minimize the amount of difficulties observed during the study, i.e., a more adapted and straightforward workflow. Better 3D motion tracking systems will be used for 3D pen input which is especially handy for defining accurate positions.

The study also has shown that the participants give system feedback quite a high importance, making the projector mounted above the tabletop an important component of the robotic workcell. It can be used to project additional information and input feedback on the table and to indicate the current system status.

VI. Conclusion and Future Work

The goal of the user study was to analyze and determine the preferred input modalities for task-based robot programming. It was divided into four phases: Background information, expectation, experience, and opinion. These phases evaluated different aspects within multimodal interaction: What the participants expected from the system, how they experienced it, and what their final opinion was.

Evaluating the results of 30 participants we have shown that most of the users prefer gesture input over 3D pen input, touch input, or speech input (Fig. 4). For defining accurate positions in 3D, pen input was slightly more preferred compared to other parameter types. This also confirms the previously published results that users mainly prefer 3D pen input for CAD applications [21]. Additionally, the study also confirms that women feel less secure and comfortable than men when using speech input for technical systems [22].

As expected when stating the hypothesis, the study revealed that speech input on its own is not very suitable for usage in task-based robot programming. The participants struggled to name unknown objects correctly and had problems in describing exact positions. Speech input is also not very accurate: Are position indications (left, right, front, back) defined from the user's point of view or from that of the robot? Minimizing these drawbacks can be achieved by special training for the user, which is not the goal of an intuitive and easy to use interface.

Another conclusion drawn from this user study is to reduce the amount of available input modalities for a specific parameter type: Experts stated that the intuitiveness of the teaching interface would increase if the system preselects the most suitable input modality, whereas non-expert users tend to try out and play with the system, using different input modalities. The presented semantic description of input and output modalities, task parameter types, and preferred combinations of the two will be used to recommend an input modality for a specific parameter type and to adapt the user interface accordingly.

The outcome of the study clearly motivates further research in multimodal interaction for robot programming as it provides the user with a more natural experience, expressive power, and flexibility. The participants stated that using multimodality makes robot programming more intuitive, easier, and faster. Our future research will focus on multimodal fusion by using multiple input modalities at the same time and on improving the detection accuracy followed by another user study using the real system instead of Wizard-of-Oz.

ACKNOWLEDGEMENTS

This research was supported by the European Union's Seventh Framework Programme through the project SME-robotics under grant agreement no. 287787.
We would like to thank all the participants of the user study for their participation.

REFERENCES

[1] N. Dahlbäck, A. Jönsson, and L. Ahrenberg, "Wizard of Oz studies - why and how," *Knowledge-Based Systems*, vol. 6, no. 4, pp. 258–266, Dec. 1993.

[2] P. Cohen, M. Johnston, D. McGee, and S. Oviatt, "The efficiency of multimodal interaction: a case study," in *International Conference on Spoken Language Processing (ICSLP)*, Sydney, Australia, 1998.

[3] S. Oviatt, R. Lunsford, and R. Coulston, "Individual Differences in Multimodal Integration Patterns : What Are They and Why Do They Exist?" in *Conference on Human Factors in Computing Systems (CHI)*, New York, USA, 2005.

[4] N. Ruiz, F. Chen, and S. Oviatt, "Multimodal Input," in *Multimodal Signal Processing*, J.-P. Thiran, F. Marqués, and H. Bourlard, Eds., 2010, ch. 12, pp. 231–255.

[5] M. Turk, "Multimodal interaction: A review," *Pattern Recognition Letters*, vol. 36, pp. 189–195, 2014.

[6] A. Perzylo, N. Somani, S. Profanter, M. Rickert, and A. Knoll, "Toward Efficient Robot Teach-In and Semantic Process Descriptions for Small Lot Sizes," in *Robotics: Science and Systems (RSS), Workshop on Combining AI Reasoning and Cognitive Science with Robotics*, Rome, Italy, 2015.

[7] B. Akan, A. Ameri, B. Çürüklü, and L. Asplund, "Intuitive industrial robot programming through incremental multimodal language and augmented reality," in *International Conference on Robotics and Automation (ICRA)*, Shanghai, China, 2011.

[8] M. Stenmark and P. Nugues, "Natural Language Programming of Industrial Robots," in *International Symposium on Robotics (ISR)*, Seoul, Korea, 2013.

[9] M. Stenmark and J. Malec, "Describing constraint-based assembly tasks in unstructured natural language," in *World Congress of the International Federation of Automatic Control (IFAC)*, 2014.

[10] J. Lambrecht and J. Krüger, "Spatial Programming for Industrial Robots: Efficient, Effective and User-Optimised through Natural Communication and Augmented Reality," *Advanced Materials Research*, vol. 1018, pp. 39–46, Sept. 2014.

[11] S. Makris, P. Tsarouchi, D. Surdilovic, and J. Krüger, "Intuitive dual arm robot programming for assembly operations," *CIRP Annals - Manufacturing Technology*, vol. 63, no. 1, pp. 13–16, 2014.

[12] B. Weiss, S. Möller, and M. Schulz, "Modality preferences of different user groups," in *International Conference on Advances in Computer-Human Interactions (ACHI)*, Valencia, Spain, 2012.

[13] Z. Pan, J. Polden, N. Larkin, S. V. Duin, and J. Norrish, "Recent Progress on Programming Methods for Industrial Robots," in *International Symposium on Robotics (ISR)*, Munich, Germany, 2010.

[14] K. Hinckley and D. Wigdor, "Input technologies and techniques," in *Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*, 3rd ed., J. A. Jacko, Ed. CRC Press, 2012, ch. 9, pp. 151–168.

[15] S. Oviatt, "Multimodal interfaces," in *Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*, 3rd ed., J. A. Jacko, Ed. CRC Press, 2012, ch. 18, pp. 405–429.

[16] N. Somani, E. Dean-Leon, C. Cai, and A. Knoll, "Scene Perception and Recognition for Human-Robot Co-Operation," in *First International Workshop on Assistive Computer Vision and Robotics (ACVR 2013). The 17th International Conference on Image Analysis and Processing*, Naples, Italy, 2013.

[17] A. Gaschler, M. Springer, M. Rickert, and A. Knoll, "Intuitive Robot Tasks with Augmented Reality and Virtual Obstacles," in *International Conference on Robotics and Automation (ICRA)*, Hong Kong, China, 2014.

[18] M. Turunen and J. Hakulinen, "SUXES - user experience evaluation method for spoken and multimodal interaction." in *Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Brighton, United Kingdom, 2009.

[19] J. Lewis, "IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use," *International Journal of Human/Computer Interaction*, vol. 7, no. 1, pp. 57–78, 1995.

[20] J. Kirakowski and M. Corbett, "SUMI: the Software Usability Measurement Inventory," *British Journal of Educational Technology*, vol. 24, no. 3, pp. 210–212, 1993.

[21] X. Ren, G. Zhang, and G. Dai, "An experimental study of input modes for multimodal human-computer interaction," *Advances in Multimodal Interfaces - ICMI*, vol. 1948, pp. 49–56, 2000.

[22] R. Wasinger and A. Krüger, "Modality preferences in mobile and instrumented environments," in *International Conference on Intelligent User Interfaces (IUI)*, New York, USA, 2006.