Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 63 (2017) 396 – 401

www.elsevier.com/locate/procedia

The 50th CIRP Conference on Manufacturing Systems

Human-Robot Collaboration Demonstrator Combining Speech Recognition and Haptic Control

Patrik Gustavsson[a,*], Anna Syberfeldt[a], Rodney Brewster[b], Lihui Wang[c]

[a]University of Skövde, Kanikegränd 3A, Skövde 54134, Sweden
[b]Volvo Car Corporation, Komponentvägen 2, Skövde 54136, Sweden
[c]KTH Royal Institute of Technology, Stockholm 10044, Sweden

* Corresponding author. Tel.: +46-500-448-507; fax: +46-500-416-325. *E-mail address:* patrik.gustavsson@his.se

Abstract

In recent years human-robot collaboration has been an important topic in manufacturing industries. By introducing robots into the same working cell as humans, the advantages of both humans and robots can be utilized. A robot can handle heavy lifting, repetitive and high accuracy tasks while a human can handle tasks that require the flexibility of humans. If a worker is to collaborate with a robot it is important to have an intuitive way of communicating with the robot. Currently, the way of interacting with a robot is through a teaching pendant, where the robot is controlled using buttons or a joystick. However, speech and touch are two communication methods natural to humans, where speech recognition and haptic control technologies can be used to interpret these communication methods. These technologies have been heavily researched in several research areas, including human-robot interaction. However, research of combining these two technologies to achieve a more natural communication in industrial human-robot collaboration is limited. A demonstrator has thus been developed which includes both speech recognition and haptic control technologies to control a collaborative robot from Universal Robots. This demonstrator will function as an experimental platform to further research on how the speech recognition and haptic control can be used in human-robot collaboration. The demonstrator has proven that the two technologies can be integrated with a collaborative industrial robot, where the human and the robot collaborate to assemble a simple car model. The demonstrator has been used in public appearances and a pilot study, which have contributed in further improvements of the demonstrator. Further research will focus on making the communication more intuitive for the human and the demonstrator will be used as the platform for continued research.

Peer-review under responsibility of the scientific committee of The 50th CIRP Conference on Manufacturing Systems

Keywords: Human-robot collaboration; Speech recognition; Haptic control

1. Introduction

The fourth industrial revolution, Industry 4.0, is a top priority for many research institutes, universities and companies [1], because this ultimately shapes the future within the industry. In this revolution human-machine collaboration is one important aspect, and therein Human-Robot Collaboration (HRC). HRC means that a robot and a human work closely together to solve a task related to for example assembling or quality control. By HRC, the unique strengths of a human (such as flexibility and intelligence) can be combined with the unique strengths of a robot (such as strength and the ability to exactly repeat the same a movement an infinite number of times).

Most of the existing industrial robots all over the world require safety fences, because it is not safe to walk close to these robots. However, some of the major industrial robots suppliers, such as ABB and KUKA, have developed new collaborative robots that can be used without a safety fence and thereby make HRC possible. Another supplier is Universal Robots, officially founded in 2005, which focuses on bringing lightweight, flexible industrial robots to the global market. Universal Robots has today three variants of collaborative robots, UR3, UR5, and UR10. HRC is the next step in the development of robots as seen with the prediction of Industry 4.0 and the new collaborative robots.

Peer-review under responsibility of the scientific committee of The 50th CIRP Conference on Manufacturing Systems
doi:10.1016/j.procir.2017.03.126

The common way of interacting with industrial robots is with a teaching pendant. A teaching pendant is a tool connected to the robot which can be used to move and program the robot. However, the teaching pendants way of moving the robot is with either a joystick or buttons, which is both difficult and time consuming for someone not familiar with the controls. The new collaborative robots offer another way to interact with the robot, namely through guidance by hand. This simplifies the way a human can move a robot but is in most cases not enough to achieve an intuitive interaction. To realize a more intuitive way of interacting with the robot, this work attempts to combine haptic control with speech recognition.

There are plenty of research within speech recognition, including some of the largest companies in the world, Google, Apple, and Microsoft. Haptic control have also been thoroughly researched, and there are several focused on robotics, e.g., [2, 3]. However, research on the combination of the two technologies to achieve a more intuitive industrial HRC is limited.

2. Human-Robot Collaboration demonstrator

The research in focus is the combination of speech recognition and haptic control to create an intuitive HRC. A design and creation approach [4] is suitable for this research, because a physical artifact is necessary to evaluate the technologies. Therefore, a demonstrator was planned, because a demonstrator can be used for multiple purposes [5], within and outside the scientific domain. The demonstrator serves as platform for prototyping, and for disseminating the concepts to potential users.

The main requirements considered when designing the demonstrator were: (1) it needs to be safe for humans to use, (2) it should be mobile to move around, (3) the task to carry out in collaboration between the human and the robot should be simple yet relevant, and (4) it should involve both haptic control and speech recognition. In the following subchapters, the implementation of the demonstrator is described in further detail.

2.1. Setup of the demonstrator

The following setup was used, as shown in Fig. 1, to meet the requirements of the demonstrator:

- UR3 robot (a) and controller (b) from Universal Robots.
- Flexible 85mm 2 finger tool (c) from Robotiq
- Sennheiser ME 3 EW microphone with Steinberg UR12 USB audio interface (d)
- Computer (e) installed with Microsoft Speech API 11 and EasyModbusTCP, connected to the microphone and the robot controller
- A movable wagon, containing components (a-e)
- A TV as the graphical user interface, mounted on a movable stand

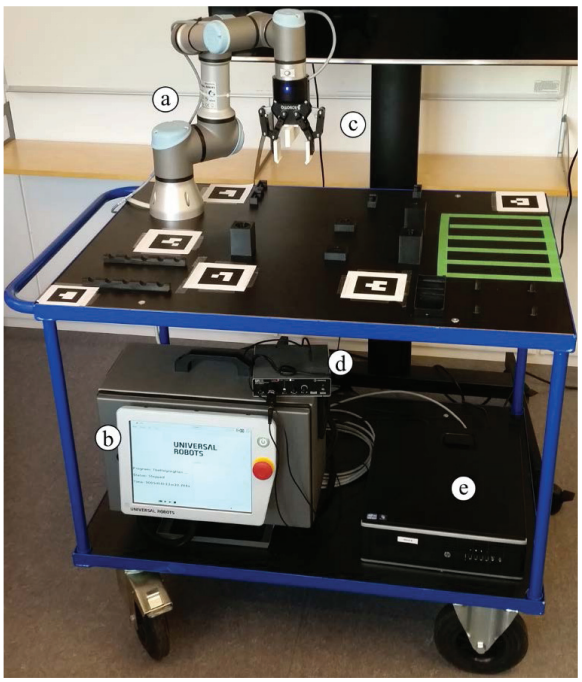

Fig. 1. HRC demonstrator setup, (a) robot, (b) robot controller, (c) robot tool, (d) microphone and USB audio interface, (e) computer. All components placed on a movable wagon.

The UR3 robot was selected because it is certified for working in collaboration with a human, in combination with being one of the cheapest robots for HRC on the market. It is a six axis light weight articulated robot that can lift up to 3 kg. It has joint-by-joint haptic control, called freedrive mode. The freedrive mode uses the impedance/back-drive control which allows a human to move the robot by hand.

The 85mm 2finger tool from Robotiq was selected because it is highly flexible, where the fingers can open 85mm wide and close at 0mm. This tool can also control the speed and force with which it grips an object. In the demonstrator the speed of the tool has been reduced, limiting the possibilities of someone getting stuck.

The computer is the central system controlling what will be displayed on the graphical user-interface, listening to commands by the human and controlling the robot execution. The speech recognition system combines Microsoft Speech API 11, Sennheiser ME 3 EW microphone and Steinberg UR12 USB audio interface. Microsoft Speech API 11 is not cloud based, which is an advantage because depending on the location, Internet access might be unavailable. The Steinberg UR12 USB audio interface connects the microphone to the computer, and this was necessary because the Sennheiser microphone plug is not compatible directly with the computer.

The robot execution is controlled from the computer, through EasyModbusTCP, which acts as a Modbus server. Several signals are defined in the Modbus server, which are: reset, start, next, open, close, and handshake. The handshake signal is used to ensure a good communication between the robot and the computer. The other signals are used for different commands controlling the execution of the robot.

2.2. Task to be carried out in the demonstrator

A simple, yet relevant, task was created where the human and robot collaborate to assemble a toy car. The toy car also has the advantage of having a real world connection of what HRC can be used for. Creating the task was done through two iterations, where the first iteration used a wooden car, Fig 1 to the left, and the second iteration used a 3D printed car, Fig 1 to the right.

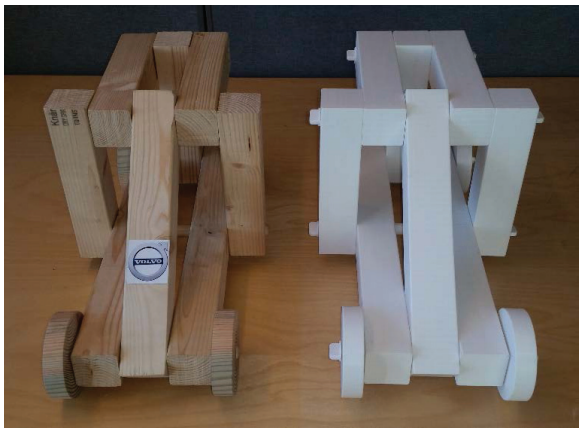

Fig. 2. The two iterations of car models, to the left the wooden car, and to the right the 3D printed car.

The graphical user-interface displayed information of the task and speech commands. The task was described with both plain text and augmented reality by highlighting both the pick and assembly position. The available speech commands were displayed with plain text and when the user spoke the interpreted command were displayed along with the speech-recognition confidence.

The speech recognition system filtered out commands that did not have at least 85% confidence. This was to make sure that the system did not misinterpret the spoken words. The speech recognition in the first iteration used “Start”, to begin with the demonstration, and “Next” to continue on each step. On the second iteration the speech recognition used words and sentences connected to the task at hand, e.g., “Rotate car”, “Open”, and “Next step”.

Friction was used to hold together the structure of the wooden car. This resulted in difficulties when assembling the parts, because some parts required the human to apply more force to assemble. Because friction holds the car together no fasteners were used, which is unrealistic in a real-world assembly task. For these reasons the 3D printed car was developed. This car used approximately the same model and measurement as the wooden car, but instead used locking rings and thumbscrews to hold the car together. Because of these changes, different fixtures and custom tool parts were created to work with the car model.

There were also another major difference between the tasks created in the two iterations, in the first iteration the whole car was assembled, while in the second iteration only parts of the car were assembled. The second iteration focused more on a realistic work station, where parts of a product are assembled, not the whole product. Both iterations of the tasks have been separated in several steps, and each step could be categorized into three different levels of HRC, direct, indirect, and no HRC.

- Direct HRC refers to steps when both the human and the robot actively work together on the same part.
- Indirect HRC refers to steps when the human or the robot support each other but only one of them is actively working on the part.
- No HRC refers to steps when the human and the robot can work without support from each other.

Fig. 3 illustrates direct HRC and indirect HRC steps used in the demonstrator. Direct HRC (a) has been used when the human guide the robot using haptic control. Indirect HRC (b) has been used when the robot holds the car while the human assembles parts onto the car. In these cases the robot has been stiff to ensure that the user have no problem assembling the parts.

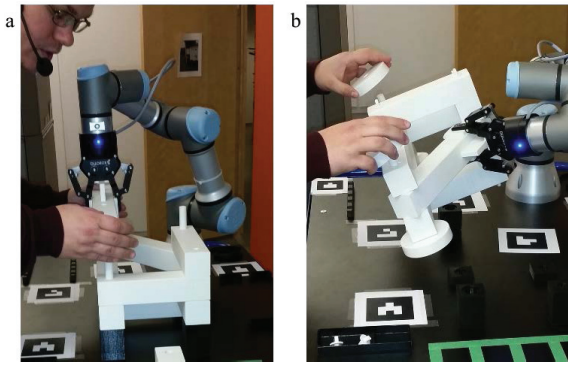

Fig. 3. Steps with (a) direct HRC and (b) indirect HRC.

3. Testing the demonstrator publically

To test the functionality of the demonstrator it has been publically exposed at several occasions. One of the main purposes of the demonstrator is also to disseminate the concepts and the research to potential users and stakeholders, which goes hand-in-hand with exposing it publically. It has been quite popular in these occasions, and it has been especially useful to show the industry what is possible using HRC.

During these public appearances the first iteration of the task was used, with the wooden car model. Anyone at these occasions was allowed to test the demonstrator. At least one instructor was always available to help them get started and to help them when problems occurred. The instructor was also necessary to ensure that there were no safety risks during the demonstration. The first time a person reached a step with direct HRC, when parts were assembled in collaboration with the robot, then the instructor guided the person on how they should execute that step.

The knowledge gained from these occasions helped to develop the second iteration of the task. Some of the problems learned from these public appearances were:

- Limited speech recognition usage, in the first iteration only the word “Next” was used to step through the program.
- Too much force required to assemble some parts, because friction was used to hold the car together.
- Too many steps of the task did not include HRC, because the whole car was assembled.
- The steps with direct HRC were difficult to move in a straight line, because the freedrive mode of UR3 is limited to joint-by-joint control, see Fig. 4.

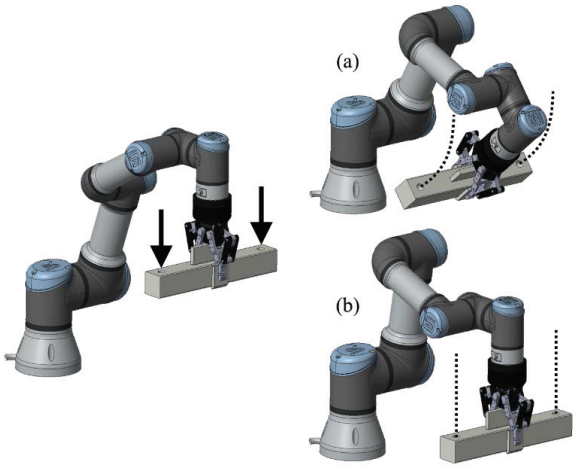

Fig. 4. Illustration of (a) UR3 freedrive mode and (b) linear motion, when applying force onto the part held by the tool.

This demonstrator has also been shown in both local and regional newspaper, to inform the general public about ongoing research at the university.

4. Pilot Study

During the previously mentioned public appearances the instructor guided the users completely in order for them to use the demonstrator. However, the main purpose of the research is to create a more intuitive interaction using speech recognition and haptic control. Therefore, the demonstrator has also been used for a pilot study, where the aim was to study the performance of the system. This pilot study invited students from technical high schools between ages 16-19. These students make good test-subjects, because they are likely to work in the future in the manufacturing industry. The pilot study had three goals:

- Comparing the accuracy of the speech recognition system when using one-word and multiple-word commands.
  - By comparing the accuracy when using one and multiple words, it is possible to determine whether the current speech recognition system is more suited for one or the other.
- Gaining insights on how the demonstrator is working and the test subjects are performing without interference from instructors.
  - With these insights, it is easier to find which problems the system has, which ultimately leads to future improvements of the demonstrator.
- Gather the interests of technical high school students toward human-robot collaboration.
  - The interests of the students, is mainly an indicator whether the students would in the future want to work with HRC. This knowledge is useful for both academia and industries, because academia want to attract new students to study HRC, while industries want to attract new employees.

4.1. Structure of the pilot study

In the pilot study the second iteration of the task was used, with the 3D printed car. The study took place in a classroom, and the demonstrator was placed on the floor in front of the seats so that all participants could see what the test-person was doing.

There were three instructors in total in this study. One instructor was tasked to handle all questionnaires, giving the correct questionnaire to each person. One instructor was tasked to introduce the students to the experiment, select the test-persons, and observe the test-person while filling in an experiment protocol. One instructor was tasked to help the test-persons getting started, intervene when a problem occurred, answer questions from the test-person, and switch programs between the groups.

Four programs were prepared. The programs were created to test two variants of the speech recognition system, and two variants of the graphical user-interface. The speech recognition variants tested one-word commands, and multiple-word commands, to study which one is more suitable. The graphical user-interface variants tested non-blinking and blinking highlights, to study how it affects the user performance.

The following programs were prepared:

1. One-word commands and blinking highlight
2. One-word commands and non-blinking highlight
3. Multiple-word commands and blinking highlight
4. Multiple-word commands and non-blinking highlight

4.2. Execution of the pilot study

Four groups, ranging between 26-31 technical high school students, participated in the experiment separately. For each group, three students were selected as test-persons by asking for volunteers. Two of the selected test-persons were asked to leave the classroom, to ensure they did not learn by watching the other test-persons. The remaining test-person was asked to stand in front of the demonstrator. The instructor gave information on how to put on the microphone and about the graphical user interface, containing all instructions. After the first test-person finished, that person received a user questionnaire, and then the audience received a public interest questionnaire. Then the second test-person was brought in to start the demonstration, the same information was given to this person. After the second test-person finished, that person received a user questionnaire. This was repeated with the third test-person. When all questionnaires were completed they were

collected and the group left the room. The next prepared program was loaded in the demonstrator and then the next group was brought in.

The experiments became hectic, because some groups required more time, leaving less time to prepare for the next group. Therefore a mistake was made where two groups tested the same program. This resulted in group 1 using program 1, group 2 and 3 using program 2, and group 4 using program 3.

4.3. Experiment protocol

For this experiment certain type of events were of interest, therefore a systematic observation [4] was used to count these events. An experiment protocol was therefore created; the protocol was used to study the graphical user-interface, haptic control, speech recognition, and combination thereof. This protocol logged for each step:

- Number of errors, when following the instructions, including missing parts, untightened fasteners, and not fully executed steps.
- Dropped parts, all parts that were dropped onto the wagon or the floor.
- Number of questions from the test-person.
- Misinterpreted commands, including commands that fell below the 85% threshold and commands that were interpreted to a different phrase.

Table 1 lists the average result from each group. The results from the protocol may have some errors, because at some occasions the instructor, responsible for the protocol, needed to tell the audience to stop laughing or to lower their voices. However, these results can still give indications to what needs improvement.

Table 1. Average and standard deviation results from the experiment protocol rounded to one decimal, each group had three test-persons.

Group	Program	Errors	Dropped parts	Questions	Misinterpreted commands
1	1	5.7	0.3	1.7	6.3
2	2	11.7	0.7	2.0	3.7
3	2	11.3	0.0	1.0	15.7
4	3	9.3	0.3	1.7	2.7

From the results, it is clear that the system is not yet intuitive enough to work with. There were in total 11 steps for this demonstration, each test-person in group 2 and 3 did in average one error per step. This clearly indicates that the system needs improvements. The speech recognition had more difficulties interpreting the test-persons in group 3. However, it is important to mention that group 3 also had the most noise in the background, i.e., chatter and laughter. Program 1 and 2 tested one word commands, but the number of misinterpreted words were mostly connected to the background noise. The implemented speech recognition is clearly not ready for industrial use, because the word error rate is too high.

4.4. User questionnaire

The System-Usability-Scale (SUS) developed by [6], was used to get an indication of the usability of the demonstrator. This questionnaire is a simple yet efficient tool for assessing a system's usability [7]. It is divided into ten questions, each question uses a five level Likert scale; from strongly agree to strongly disagree. Every odd numbered question has a positive point of view while every even numbered question has a negative point of view. Each question was translated to Swedish, and focused on the haptic control and speech recognition. The result of the SUS is a score between 0 and 100 that correlates to the usability of a system. A score above ~73 is a good system, while a score above ~85 is an excellent system [7].

The average score per question from each group is illustrated in Fig. 5. The score from the SUS varied from 30.8 to 72.5, between the groups. The sample size of each group was three, and therefore the results cannot be statistically proven, but can be used as an indication of usability.

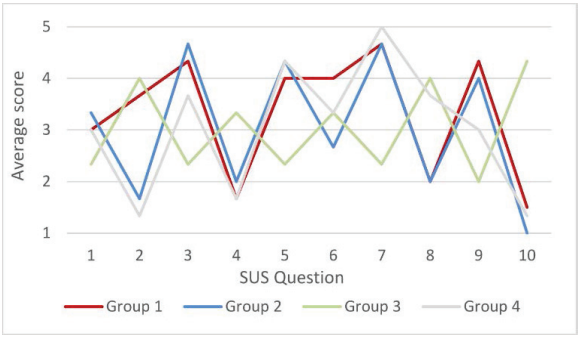

Fig. 5. Results from the SUS, average score per question for each group.

There is a clear difference between the score of group 3 and that of the rest of the groups. From observations, there were a lot of external disturbances for group 3, where the students in the back talked and laughed. This is believed to be the reason for the large differences in the SUS-scores. In the groups 1, 2, and 4 the SUS scores vary between 65.8 and 72.5. This indicates that without external disturbances, the system is close to being good, but needs clarification from a larger sample size.

4.5. Public interest questionnaire

A questionnaire was developed for the audience, to get indications of the public interest. A questionnaire was selected because it provides an efficient way of collecting data from many people [4]. This questionnaire had 11 questions, each question using a five level Likert scale; from strongly agree to strongly disagree. The following questions were used:

1. I am interested in technology
2. I am interested in robotics
3. It would be interesting to work with this robot
4. It would be interesting to develop systems with this robot
5. I thought the system seemed practical to work with

6. The system seems to be safe to work with
7. I use speech recognition in my daily life
8. I think I would like to use speech recognition for work
9. I thought the speech recognition seemed practical
10. I think I would like to use the control by hand for robot programming
11. I thought the control by hand seemed practical

Question 5 had an average result between 4 and 5, question 7 had an average result between 2 and 3, and all other questions had an average result between 3 and 4. These results indicates that the kind of technology used in the study is interesting for students in a technical high school, and that they might be interested in working with this technology in the future. The results also show an indication that speech recognition is not used in the daily life, but that the technology itself might be interesting to work with.

5. Conclusions

This study aimed to investigate the combined use of haptic control and speech recognition for human-robot collaboration. A demonstrator has been developed that combines speech recognition and haptic control and serves as a platform for prototyping and experimentation. The demonstrator has been used both for public dissemination of the research concepts and for undertaking a pilot study of the concepts. During the public appearances anyone was allowed to test the system, which helped improving the demonstrator.

From the pilot study, important knowledge of the system was gathered. It was shown that the system has great potential, but that too many errors and misinterpretation occurred which indicates that the system needs improvements. A SUS was used to measure the usability of the system and the results showed a clear indication that external sources, in this case chatter and laughter, affected the user experience heavily. An important finding from the study is therefore that external disturbance can largely affect the results from user experiment to render unusable and careful measures should be taken to avoid this. To draw further conclusions from the study, more participants need to be involved and doing this is included in the plans of the near future.

Even though the system has some problems, there seems to be an interest of working with this kind of system amongst technical high school students. This is important knowledge for both academia and industries, where the academia wants to attract more students, and the industry wants to attract more workers.

6. Future work

Future experiments require more participants testing the demonstrator because three participants are not enough to make any statistical proofs, although it gave valuable insights. The pilot study had two controlled variables, variation of speech recognition and graphical user-interface. However, the results might be affected depending on the combination, therefore future experiments should focus on one controlled variable. Future experiments should also have a more controlled environment, isolating the user with the HRC demonstrator to avoid disturbances like chatter and laughter.

For the pilot study 16-19 years old high school students were selected because they are potential future workers within an industrial manufacturing setting. However, future studies need to consider using actual workers within a manufacturing industry. Their perspectives could provide insight on applications where HRC could be implemented. They also make good test-subjects for future experiments when evaluating improvements with HRC, because they have experience working in the manufacturing industry.

The speech recognition in its current form is not good enough for industrial use, therefore different speech recognition engines and microphones needs to be tested. Further experiments also needs to apply controlled noise, since within certain industries noise is quite common. The misinterpretation results from the experiment protocol in the pilot study were not perfectly accurate. The reason was because the instructor got distracted, and also it was difficult to keep track of everything that happened. Therefore, in the future, an automatic way of logging the accuracy or word error rate needs to be developed.

All direct HRC steps, that included some form of haptic control, were inconvenient because the freedrive mode of the UR3 robot moves joint by joint. Therefore, future work should look into implementing technologies where haptic control can be used to move the robot with linear motions. One such technology is the ActiveDrive developed by Robotiq, which allows a human to control the robot with, translation movements, tool orientation, etc.

References

[1] M. Hermann, T. Pentek, and B. Otto, "Design Principles for Industrie 4.0 Scenarios," in *2016 49th Hawaii International Conference on System Sciences (HICSS)*, 2016, pp. 3928-3937.

[2] V. Chu, I. McMahon, L. Riano, C. G. McDonald, Q. He, J. Martinez Perez-Tejada*, et al.*, "Robotic learning of haptic adjectives through physical interaction," *Robotics and Autonomous Systems,* vol. 63, Part 3, pp. 279-292, 1// 2015.

[3] D. Surdilovic and J. Radojicic, "Robust Control of Interaction with Haptic Interfaces," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, 2007, pp. 3237-3244.

[4] B. J. Oates, *Researching information systems and computing*. London ; Thousand Oaks, Calif.: SAGE Publications, 2006.

[5] J. Moultrie, "Understanding and classifying the role of design demonstrators in scientific exploration," *Technovation,* vol. 43–44, pp. 1-16, 9// 2015.

[6] J. Brooke, "SUS - A quick and dirty usability scale," in *Usability Evaluation in Industry*, P. W. Jordan, B. Thomas, B. A. Weerdmeester, Ed., ed London: Taylor and Francis., 1996.

[7] A. Bangor, P. Kortum, and J. Miller, "Determining what individual SUS scores mean: adding an adjective rating scale," *J. Usability Studies,* vol. 4, pp. 114-123, 2009.