Social Robots 2019-2020

Project Report – Tiago Titanium Sign

Castiglione Cristian, Federico Dalè

26th May 2020

1. Introduction

This project was born from an idea that is inspired by a unique reality of its kind, at least in Italy. It is a very special bar, where those who work there are native and deaf-mute signers. Each member of the staff, whether the owner, a barman or a waiter communicates through sign language and the bar is frequented by both hearing and deaf people. Behind the counter or between the tables, as is more common at the time of writing, given the national health situation, bartenders and waiters must read the lip of hearing people who want to order something to drink or food. Our idea starts from the fact that this communication could be facilitated using a robot, which knows how to interpret the spoken language and translates it into sign language to the barman or the kitchen. This robot should be able to move inside the room, between the tables and do the real job of the waiter, becoming an interpreter. In addition to the advantage of having a translator, thanks to the robot, the waiter is avoided to approach customers and thus minimizes the risk of not maintaining the social distances imposed by the law.

The robot in question is Tiago Titanium, which was chosen because it has an arm and a hand with all five fingers, which is important for signing. This robot is also tall and it does not risk being walked on by people who go to the bar or by a worker. We had to choose a single arm robot because the platform at our disposal for the simulation did not provide a double arm solution with five fingers per hand and we thought it was more appropriate, thanks to the advice of an expert in the sector, to opt for a robot that had all fingers and then to simplify two-handed gestures with one hand, because in our field of application all the signs have symmetry between the two ends. Furthermore, it happens very often that a signer has a busy hand when he wants to communicate with someone and it has to simplify a sign. However, the robot does not have the same joints as human beings and some signs are slightly distorted, but the application context helps the recipient of the communication to understand. To develop and simulate our project we used Webots [4], an open-source application dedicated to simulating robots in physical environments, with the possibility of directly transferring a simulation to the robot. All the material concerning our project can be found on GitHub (https://github.com/crc30/SocialRobots.git).

2. Problem statement and requirements

The problem underlying our project is linked to many factors also dictated by the historical moment in which we live. The idea of using a robot as a communication channel between customers of a bar and deaf-mute staff was born from experience, because in the context of the bar in the centre of Bologna that gave us inspiration it happened several times to see customers talking to staff who could only read the lip. The introduction of a robot would therefore allow the customer to order and to have voice feedback from the robot and the barman to communicate through the signs. In the beginning the project was conceived by imagining a crowded and noisy bar, but given the new provisions of the Italian state regarding public premises used for food and drinks services, the robot adapts perfectly and it helps us, because it allows the bartender and the customer to keep their social distances and to order safely. The robot was therefore designed to automate the entire ordering process and facilitate the bartenders' work who no longer have to be careful about the lip and their work and they can interact with the robot using sign language.

Analysing the entire context, the ideal robot figure to act as an assistant to the bartender is certainly that of a humanoid robot, not too like human because it would generate a sense of fear in certain customers. We would therefore need a robot that looked like as much as possible a "toy" so as to make the customer perceive a sort of feeling between man and machine in an environment that puts him as comfortable as possible. The robot should therefore be able to move, speak and sign with both arms and do everything to seem as natural as possible, so that customers could feel more comfortable by considering the robot more like a game.

The first problem to be faced in such a context is that of the mobility of the robot in an environment shared with the human being and full of obstacles, constantly changing, therefore it must be equipped with wheels and sensors in order to move freely. The second problem is the communication with the customer through verbal language which must be recognized by the robot, which must identify exactly the drink or dish ordered by the customer. The third problem is the interaction with the deaf bartender, to do this you need to sign. Our project focuses on the third problem, defining the signs needed to identify a drink or dish on the bar menu. The use cases of the entire system can therefore be identified as noted in Figure 1.

We are going to focus the entire project on the "Sign" case. The human body parts engaged when a person signs are multiple, starting from all ten fingers of the hands up to the facial expressions, so a robot would be able to replicate all this. The requirements would be: humanoid appearance, with two arms and two hands completed with all fingers, including opposable thumb, face and facial expressions. The goal would be to simulate exactly a human who makes signs, also replicating facial expressions, for example if he smiles or mimics a sad face. On the other hand, this project has various implementation requirements, first the use of the Webots simulator and then the experimentation and the entire test of the project in a virtual environment. Webots offers a series of robots to choose from and unfortunately it is not possible to find one that fully satisfies all the requirements previously indicated. Therefore, it is necessary to select the most suitable robot with features that meet the most important requirements. The first great compromise that we had to accept is that of the facial expressions part, because in these robots the face is completely inexpressive or even absent. However, for what in concerns the arms, many robots have two of them, but none of them have five fingers. However, the only robot with five fingers has only one hand. The robots we selected before making the final choice were Tiago Titanium, ICub and Nao. The first has one hand, one arm and five fingers, it is quite high and has humanoid features, the second instead has two arms, two hands, but only three fingers per hand, it is also really short and this could be a problem inside a room because it could not be seen and walked on. We finally choose Tiago Titanium because by studying the signs which would be needed for our project, we have identified some of them which can be simplified and replicated with one hand, when the signs are symmetrical. This operation is very common for signers because it often happens to have a busy hand and to be able to use only one to communicate, such as for example a real bartender holding the order pad or tablet in his hand.

The gestures replicated by the robot will be simplified and will not present any facial expression, limiting expressiveness. In all this, the context will therefore be very important to identify the signs well, because since we are talking about an order, the signs will certainly be linked to food or a drink in particular and each sign differs from the other in a clear way.

The Tiago robot allows us to fully meet the requirements for mobility, as it is medium size and it has two wheels on the bottom that allow it to perform any movement. In addition, Tiago also has good computing power (thanks to the Intel i5 processor inside) which allows it to be programmed to meet all our needs, from avoiding obstacles to recognizing the human voice.

Although with some compromises, Tiago was our choice because it is very upgradeable and it will be possible to implement new functions later.

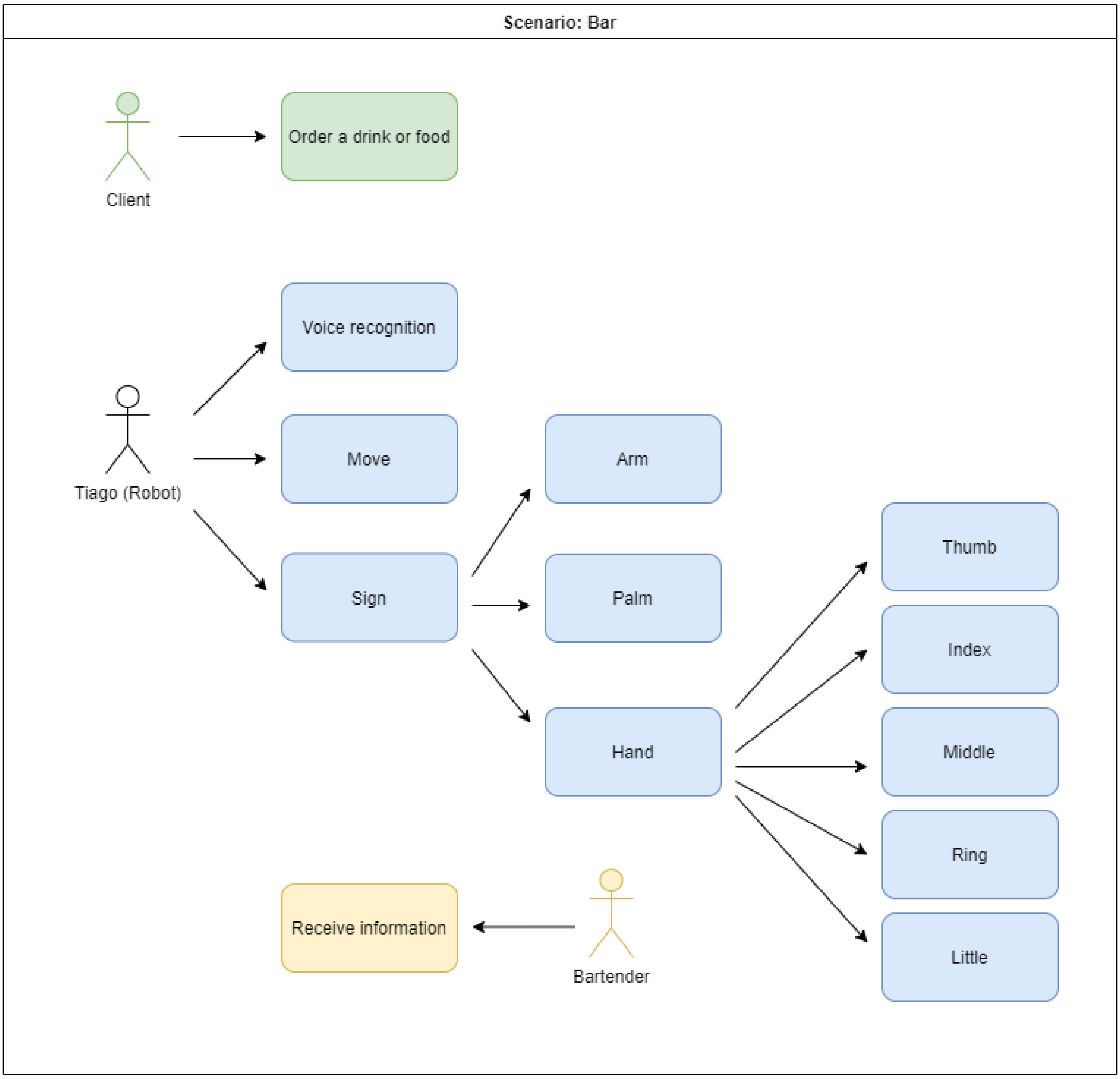


Figure 1: Use Case Diagram

3. Related Work

The problem we deal with is very specific because it is based on a truly revolutionary idea in the catering world. This bar is entirely run by deaf and dumb people and the ordering problem relates only to such a scenario, which is the first ever bar in Italy run in this way. What characterizes our scenario is the fact of using a signing robot in a working environment and not just as a linguistic teaching support to a human user, this means that the robot has no margin of error, otherwise there would be errors in the order.

The general problem we focused on in this project is to program a robot to sign, in order to be understood by a natural signing person. There are several projects that have studied sign language reproduced by humanoid subjects, in different fields, but still related to ours.

The first study [1] presents an interactive game assisted by a humanoid robot for teaching sign language. The game is specially designed for children with communication problems. The children play with a humanoid robot that expresses a series of words chosen in sign language using hand movements, body gestures and facial expressions. The robot can communicate with children by recognizing coloured cards through a camera-based system and generating a selected subset of signs. As in our case, the set of signs is predefined and it must be updated by a developer. The application is very similar, because it generates an output in terms of signs from a user input, which in this case are for teaching purposes only. This type of approach is widely used in the field of research and teaching of children because it leads children to imitate the robot signs, making them their own.

Another type of study [2] takes the problem to a completely different level because it aims to create not only a translator from words to signs but also vice versa. This field of study is much more complex than the one we approach because it aims to build a learning model through Machine Learning that allows it to extend the robot's signing ability through extensive training and training sets so it can be more precise in translation and can be improved over time. This could be useful for our purpose for future development that includes teaching the robot new signs for expanding the menu of a bar.

Another signs recognition method [3] tested was to transform the acquisition of the sign into 2D images to be then passed to a recognizer who classifies them and leads to a result, all through a neural network.

These two projects, even if they are disconnected from ours, are interesting because they can be a starting point to expand the range of signs available to us by adopting a different technique in a future development.

4. Approach

As previously described, Webots was chosen for the development of the entire project, an obligatory choice given the historical moment in which we find ourselves. The software provides a predefined list of robots and each of them is distinguished from the other by shape, size, purpose, type and scenario of use. The choice of the robot on which we have developed the entire project was the first step to be taken, since in order to mark a robot it needs humanoid features we have identified the following models: ICub (Figure 2), Nao (Figure 3) and Tiago Titanium (Figure 4).

The first is developed at IIT within the EU RobotCub project. It has 53 motors that move the head, arms and the hands, the waist and the legs making the movements very fluid and like that of a human being. We had to discard the possibility of using it because it does not have fingers, a fundamental element in the world of signs and is also not sufficiently high.
The second robot we examined was Nao, developed by SoftBank Robotics, like ICub, it is a robot with three fingers per hand and this would not have allowed us to reproduce many words in sign language, it is also very short and for the scenario in which it has to operate is not good because it would risk not being seen especially in situations where the place is crowded.
Finally, we found Tiago Titanium, developed by PAL Robotics, this robot that has not entirely humanoid features, has a bust, a head and a single arm.

Initially we were perplexed about the compromises that we would have to accept because if for height and size and for mobility and power it would have been the perfect choice, a single arm would have excluded many signs that we had seen until then. Studying the opinions of some experts in the sector, it emerged that many times people have a busy hand while they sign and simplify the gestures by performing them only with the other hand. We therefore thought it appropriate to choose this robot for the characteristics described above and simplify the two-hand gestures following the symmetry. Several two-handed signs, including all those which interest us, perform the same gesture with both the right and left hands symmetrically. This simplification seemed the easiest to put into practice so as not to affect the comprehension of the signs. Sign language also involves the use of facial expressions to communicate with others and it would have been appropriate to insert them, unfortunately the robot does not have any control on the face and therefore it is not possible to integrate them.

Our choice was finally Tiago, because any compromise we had to accept did not result in the loss of information of necessary importance. The strong point of this robot is that it is equipped with 44 motors that allow it to move all parts of the body. The arm and hand which are the parts that interest us the most are equipped with 39 motors that allow us to reproduce in detail all the signs provided in our project. Another interesting aspect concerns the fingers as the robot can manage all the phalanges which allows to faithfully reproduce any position that the fingers of a human being can assume. For the development of the project we initially drafted a very simple, but complete menu. For each dish or drink on the menu we studied the relative sign in L.I.S thanks to the online vocabulary "SpreadTheSign [5]" which provides a video of the sign performed by a human. After the translation of our menu through the corresponding signs video, we identified the main details and the differences between one sign and another. We then described each sign in its entirety, distinguishing four main characteristics: the position on the body, the orientation of the palm, the position of the fingers and the movement of the arm or hand.

Below is an example of coca cola:

- Finger position: joined upwards.
- Palm orientation: vertical.
- Location on the body: to the chin.
- Movement: rotation of the wrist.

After describing all the signs and identifying the details that characterize each of them, we thought about how to compose a sign starting from the details and then how to compose it using other signs described above. The first approach, identified as vertical composition, bases the process of creating a sign on the four points described above and codifies them as samples by defining the various positions of the arm, the orientations of the palm, the position of the fingers and the movement, so as to create a sign by composing the usable samples. For example, the sign that identifies the Coca-Cola and that of the beer have in common the arm which from the rest position is raised to the chin, therefore it will be used the same sample in both situation. This type of vertical compositionality has allowed us to identify within our menu signs that have common movements with others and therefore we went to create samples ad hoc for our domain that allowed us to create all the necessary gestures. Some items on the menu have compound signs such as first courses. The gestures in question are all created from the general "pasta" gesture, then specifying the sauce in question. In this case we have the possibility to recall a sign previously created and then executing another immediately after. This approach is called horizontal compositionality, very useful in creating signs of compound dishes.

Development therefore concentrated first on the creation of the various samples, then on the vertical composition of each sign and then on the horizontal composition of more complex signs, covering the entire menu. We initially imported the robot supplied by Webots on a minimal world that provides a flat surface delimited by barriers around the perimeter, then we initialized the controller. The project is entirely developed in C language. The Tiago Titanium robot integrated in Webots provides a default controller that uses an array containing double values that specify the position that each robot motor must take. We took care of creating a function that would take our samples as input and from these create the array in question and then pass it to the robot in order to create a movement towards that configuration. As for the various samples, they are small double arrays that contain information on the position of each motor that concerns the part of the gesture in question.

For each sign a function is created which calls, based on the number of movements, the function setTiagoPositionCompos. This function takes as input the different samples and a delay in seconds which indicates after how long the movement must activate. This function allows you to set the positions of the robot motors through the values that are passed to it in input and the repeated calls to this function allow the robot to create the desired sign by moving from one configuration to the next. Each function that identifies a dish from the menu calls the setTiagoPositionCompos several times with different values and as previously said this sequence of calls generates the desired sign. The strength of this approach is the methodology that allows future developers to create their own signs in a simple way thanks to the compositions, also defining new samples.

For simplicity, the robot is controlled by keyboard and depending on the key pressed, the sign corresponding to a menu item is executed. This is because we did not deal with the part of interaction between the human and the robot but only with the part of execution of the signs and this was the easiest way to give commands to the robot.

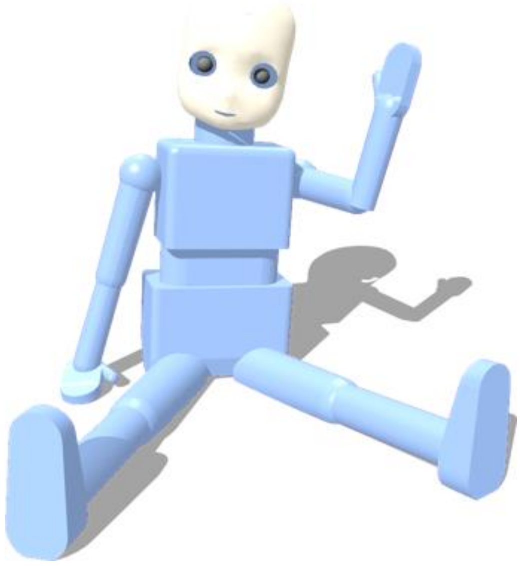


Figura 2: ICub

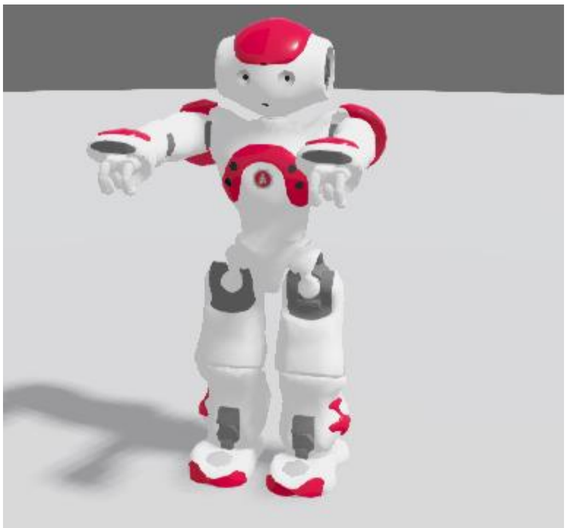


Figura 3: Nao

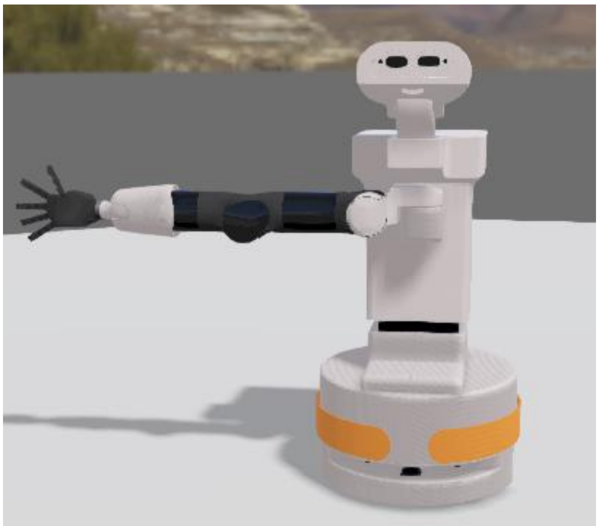


Figura 4: Tiago Titanium

5. Results and evaluation

After the initial development phase of the robot we moved on to the test phase of the signs created and we refined some of them which were not entirely precise. Overall, the signs created are very similar to those of SpreadTheSign's online vocabulary, with the small difference that the robot's movements are less fluid than those of a human being who has much more mobile joints. By following the well-structured procedures described in the approach phase in which each sign was described in detail with the help of the videos, we obtained good results. The approach we adopted is the result of a study on sign language and on the process of realization of a sign by the human being, this approach has allowed us to create an algorithm that makes the process of developing sign intuitive and faithful.

After completing the development part, we moved on to the actual test part. To evaluate the work we have done, we decided to have a test done by two people who have never had to deal with the world of signs and who do not even know a sign in the Italian sign language, to understand if they can guess the meaning of some of those reproduced by the robot. Initially we showed him mini videos in which signs are performed by a human being and then we have shown others in which the same signs are performed by the robot. We then asked testers specific questions about the signs of both the robot and the human. Finally, we averaged the ratings and starting from this we analysed the results by splitting them into positive and negative.

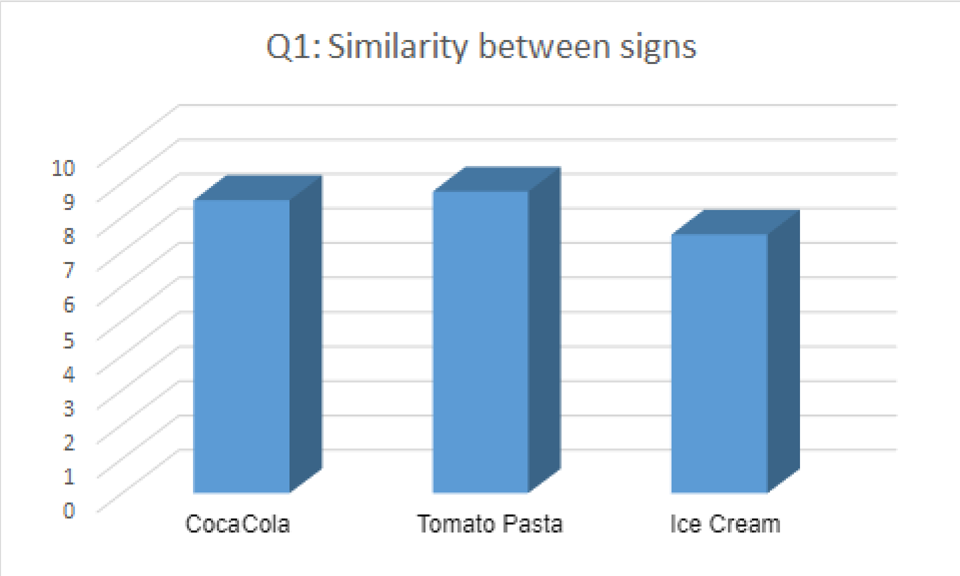
The test was prepared in the following way:

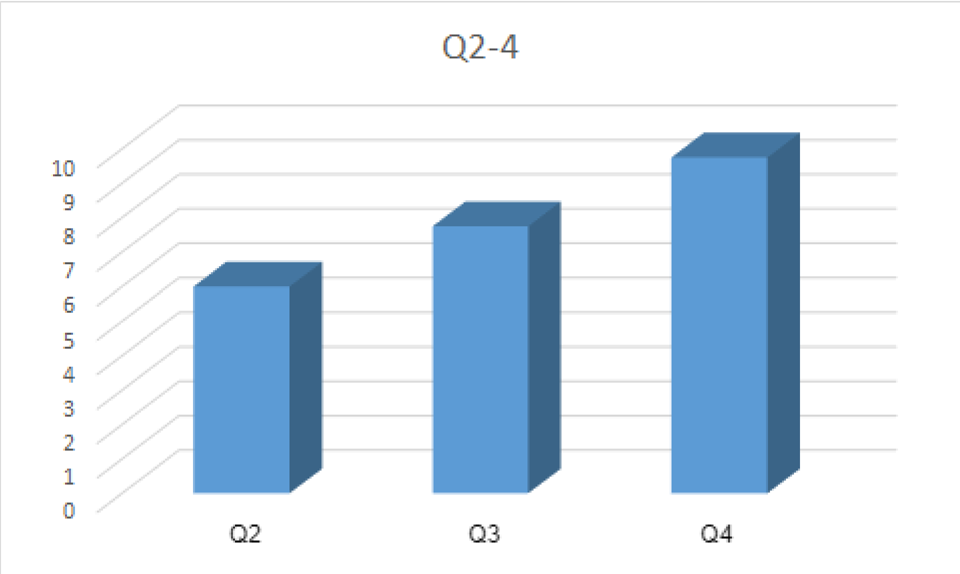
1. We showed three videos of three different signs: coca cola, tomato pasta and ice cream.
2. We showed them the same signs of the three previous videos, however performed by the robot.
3. We asked the following questions after watching the six videos:
   i) Is the sign performed by the robot the same as the one performed by the person? (one question for each pair of signs)
   ii) Do you think that facial expressions are totally necessary to better understand the signs performed by the robot, despite knowing the scenario of use?
   iii) Do you think that the use of both hands is totally necessary to better understand the signs performed by the robot, despite knowing the scenario of use?
   iv) Do you think this robot is useful to replace a waiter who takes orders in a bar during this covid-19 period?

Test users could answer questions by indicating a value from 0 to 10, with 0 totally disagree and 10 totally agree. The people tested are 4 of both sexes and between the ages of 26 and 35 and none of them had ever talked to a deaf and dumb person. The results obtained from this test provided us a positive feedback that allowed us to understand how similar the signs made by the robot are to those performed by the human. Most people who did the test think that the use of both hands is essential to fully understand the meaning of the sign. A very positive result was found in the last question relating to social distancing due to covid-19, a robot can be essential in this period to maintain the necessary safety distances from waiters and customers.

For what it concerns the question related to facial expressions, the answers have been discordant, as some people think they are important while others do not. However, the tests had a positive result.

The graphs below show in detail the average of the answers to the questions of the various test users.





As for future developments, in addition to the expansion of the vocabulary known by the robot thanks to the development of new signs and new samples, it would be interesting to complete the project with the voice recognition functions, so that the robot can speak with the customer. Furthermore, the function of moving and avoiding obstacles could be integrated, so that the robot can move independently in the room. In addition, it would be interesting to generate vocabulary translations in various languages in order to communicate with foreign signers. Another very interesting future development, that certainly requires more effort, is the integration of sign recognition through the robot's eyes. In this way, through a recognizer and machine learning techniques, it could be created a new robot which recognizes new signs, learns them and increases its vocabulary, becoming almost a simultaneous translator. Of course, all this would require a lot of work and would be something truly innovative.

For what it concerns the real robot, it would be interesting to replace Tiago with its more advanced version that is not present on Webots, as it has two arms and two hands which can reproduce all the possible signs.

6. Conclusion

This elaborate focused on the part of sign generation through a robot, managing to imitate, even with some small differences, the real human signs. The most complicated problem that arose was that of the joints. Tiago has a wrist that works in a different way than the human one and performs less fluid and wider movements to turn the palm of the hand in one direction rather than the other. All this means that the sign is less faithful to the original, precisely because of a physical difference between the human arm and the Tiago's. Another problem arose when dealing with the vertical compositionality described above. Our study is based on a previously defined park of signs (the menu) and the various compositional parts of the signs have been designed starting from a well-defined domain, this means that it is possible to adapt them but that certainly do not cover all the signs, not even in the restaurant business. Furthermore, another problem arose when a rotation by a robot motor was needed. We used a sine function that allows us to simulate a circular movement by updating the position of the motor in question at each step of the simulation.

Surely one of the most important things we have learned is that in the world of robots it is even more important to create applications and projects rigorously and that they can be used by third parties to improve them and bring them to a higher level, or they can serve other people in bigger projects. In addition to all this, working with robots makes you enter in a completely different perspective compared to classic programming, because you go to use your knowledge about coding and you apply it on something completely new.

Our project would like to form the basis of a much larger study, in which the robot should be able to move autonomously in the room, avoid obstacles and make the shortest possible journey, as well as understand speech and therefore interface between bartender and hearing customer. A certainly very interesting future development could be to create the opposite, that is teaching the robot to understand the signs and translate them into speech and maybe at the same time learn them and replicate them through machine learning algorithms. This would be a very powerful development because it would be able to create a robot which knows the spoken and signed languages, almost becoming a simultaneous translator and an interpreter.

Our solution is a first approach to the use of this kind of a robot, with all its limitations, aimed to creating a link between the conception of human sign and the conception of position by a robot.

Bibliography and Sitography

[1] P. Uluer, N. Akalin and H. Köse, "Learning sign language from a social robot peer by playing an interactive game," 2014 IEEE-RAS International Conference on Humanoid Robots, Madrid, 2014, pp. 437-437, doi: 10.1109/HUMANOIDS.2014.7041397.

[2] A. Balayn, H. Brock and K. Nakadai, "Data-driven development of Virtual Sign Language Communication Agents," 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Nanjing, 2018, pp. 370-377, doi: 10.1109/ROMAN.2018.8525717.

[3] Paulraj M P, S. Yaacob, M. S. Zanar Azalan and R. Palaniappan, "A phoneme based sign language recognition system using 2D moment invariant interleaving feature and Neural Network," 2011 IEEE Student Conference on Research and Development, Cyberjaya, 2011, pp. 111-116, doi: 10.1109/SCOReD.2011.6148718.

[4] *Webots: robot simulator*. https://cyberbotics.com/.

[5] SpreadTheSign: https://www.spreadthesign.com/it.it/search/