Let’s Develop a Researcher Robot!

Aslı Aydınlar asliaydinlar@sabanciuniv.edu
Computer Science and Engineering - Faculty of Engineering and Natural Sciences, Sophomore

Aysun Öğüt aysuno@sabanciuniv.edu
Computer Science and Engineering - Faculty of Engineering and Natural Sciences, Sophomore

Berna Yıldıran byildiran@sabanciuniv.edu
Computer Science and Engineering - Faculty of Engineering and Natural Sciences, Junior

Göktuğ Gençkaya goktuggenckaya@sabanciuniv.edu
Computer Science and Engineering - Faculty of Engineering and Natural Sciences, Sophomore

Supervised by

Dr. Junko Kanero
Psychology - Faculty of Arts and Social Sciences, Faculty Member

Abstract

Robots have an active and irreplaceable place in almost every aspect of people's daily lives. As the studies in mechatronics and artificial intelligence progress, the areas where robots operate are increasing day by day. Especially with the outbreak of the COVID-19 pandemic, robots have started to be used more in many fields to reduce physical contact between people and to prevent the spread of the pandemic. In this project, it is aimed to develop a “Researcher Robot” which will take the role of research assistant and the robot will greet the incoming participants, chat with them and direct them to the correct study according to their preferences. In the scope of the project, the “Greeting and Ice-breaking” and “Taking Participants to the Research Room” part of the program of the robot had been developed. The implementation of this program focused on robots' interaction with humans and its interaction with the environment through its sensors. The research and programming of these sections carried out through four main functionalities which were *Dialogue*, *Face Detection*, *Speech Recognition* and *Movements*. The SoftBank Robotics’ humanoid robot *Pepper* was used in this project and software development of the Pepper was carried out using Choregraphe 2.5.10.7.

Keywords: Human-Robot Interaction, Social Robots, Social Facilitation, Humanoid Robots, COVID-19

1. Introduction

The relationship between humans and robots continues to advance day by day, and as human-robot interaction projects develop, social robots began to play more roles in different areas such as education and health (Reither et al., 2012, p.41). It can be understood that robots are now not only machines that enable the industry to get heavy work done quickly and easily, but also are agents that can engage in social communication with people.

According to Yang et al (2020), over the years, robots have been used to make hazardous jobs but pandemics in world history have enabled robots to be developed and used socially. The COVID-19 pandemic is one of the epidemics that enabled this development (p.2). With the outbreak of the COVID-19 pandemic, the decrease in communication between people is one of the factors that cause an increase in human-robot interaction. Yang et al (2020) argue that robots take charge to reduce interpersonal interaction during the pandemic. For instance, as they do not carry the risk of infection spread so much in tasks such as delivering basic needs such as food, medicine, and providing controls in different places, their importance has increased. In addition, it is claimed that robots have developed within new opportunities with the life orders changed by the pandemic.

As a result of the quick spread of the pandemic around the world, countries have started to apply quarantine procedures. People's staying at home for a long time also limited their social time with their friends. Yang et al (2020) point out that a decrease in social interaction between people has a negative effect on people's mental health. It is suggested that robots can support people socially in solving this problem and people can socialize without fear that they will be infected.

This project focuses on developing a social robot that welcomes the participant then informing and directing the person about the specific study that the participant will choose. We aim to support the changes that come within the scope of social arrangements brought by COVID-19, such as maintaining social distance at events and reducing interpersonal interaction.

We observed that social robots can help people and make things easier, especially during the pandemic. In light of all these ideas, we argue that understanding the ways to organize human-robot interaction events can contribute in loosening the restrictions. The social robot developed in the scope of this project can also take part in such activities. With the further development of technology, social robots are expected to be seen in more areas.

2. Implementation

2.1. Dialogue

In the program, Pepper was supposed to talk with the respondent till they reach the room where research will take place. To be able to continue the conversation between Pepper and the respondent, adjustments are made to make Pepper talk about movies and songs. This setup was expected to make Pepper more humane and make the Human-Robot Interaction between the respondent and Pepper more intimate. Pepper asks if the respondent wants to hear about the movies: “Can I tell you about a movie that I watched last night? It may cheer you up! Please say yes.” As seen, although Pepper is seen as a robot, speech was selected so that they are able to converse more sincerely. According to the response of the people, the robot decides to speak about the movies or the songs. In the below figure, a box called “Positive Text” is a ‘Text Edit’ box under the Programming, Data Edition of the box libraries section in Choregraphe. With the help of the ‘Say Text’ box found in Speech, Creation; Pepper asks about the respondent’s answer. Speech recognition recognizes the answer of the respondent and sends this output to the ‘Switch Case’ box. Switch Case box directs this stimulation for either movie conversation or song conversation.

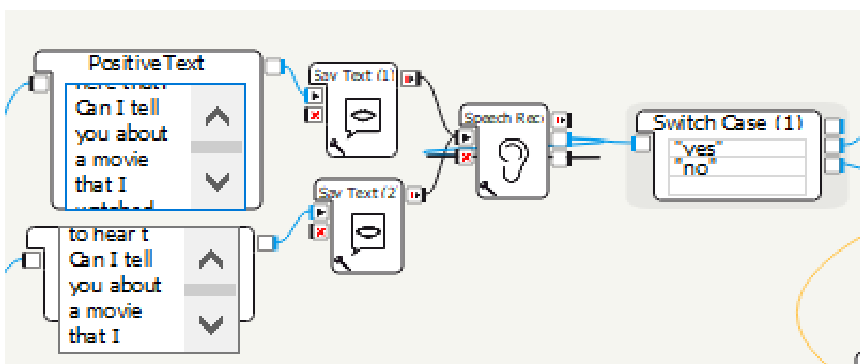

Figure 1. Talking and taking response

Switch Case is designed to separate between different outputs. It classifies the input according to the parameters the box already has. These parameters can be added or removed by right-clicking inside the box and choosing ‘Insert Row’ or ‘Delete Row’ as shown in the figure. Also according to these parameters, by right-clicking the output boxes on the right part of the Switch Case box, it is possible to add, edit or delete outputs as shown in the figure.

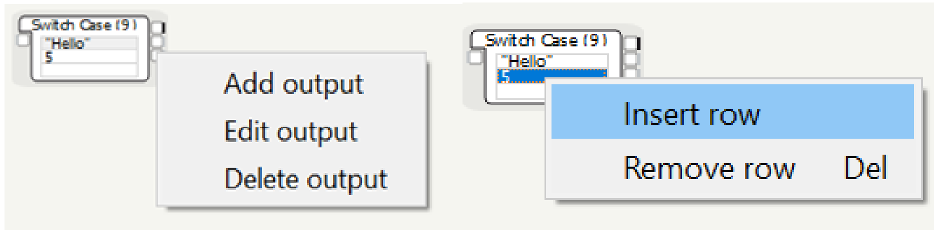

Figure 2 & Figure 3. Managing Switch Case

Thus, according to the mood of the respondent, Pepper asks if the respondent wants to listen to Pepper talk about movies. Recognized answer of the respondent is sent to the switch case box and this box decides to send the signal to talk about the movie part of the program if the speech recognized is 'yes', or sends to the talking about the song part of the program otherwise.

In this part of the program, Pepper is expected to talk about a song. Three songs are chosen for Pepper to be able to talk about and these are 'Hello' by Adele, 'Blurred Lines' by Pharrel Williams and 'I said Hi' by Amy Shark. The simulations that trigger Pepper to wake up are sounds like 'Hi', 'Hey', and 'Hello'. All three of these songs have these triggering words in their lyrics and that's why Pepper has the knowledge about these songs. For making the program choose among these 3 options, 'Random Int' box has been added to the program. This box generates a random integer between the specified range. This range can be chosen by clicking the key-like figure on the left-down side of the box.

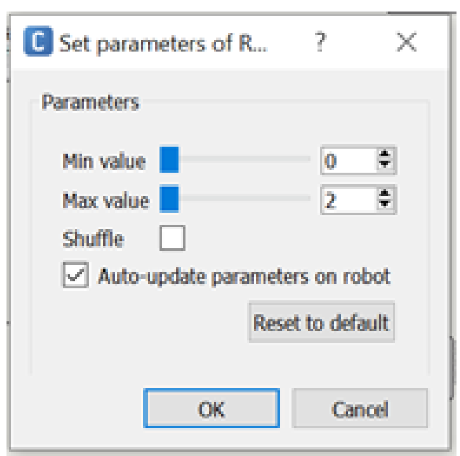

Figure 4. Parameters of Random Int Box

After setting the parameters for maximum and minimum value, this box generates a random integer and sends it as an output to Switch Case. According to the randomly generated number, the switch case directs the signal to one of the song texts. With the help of ‘Text Edit’ and ‘Say Text’ boxes Pepper talks about the songs in its program.

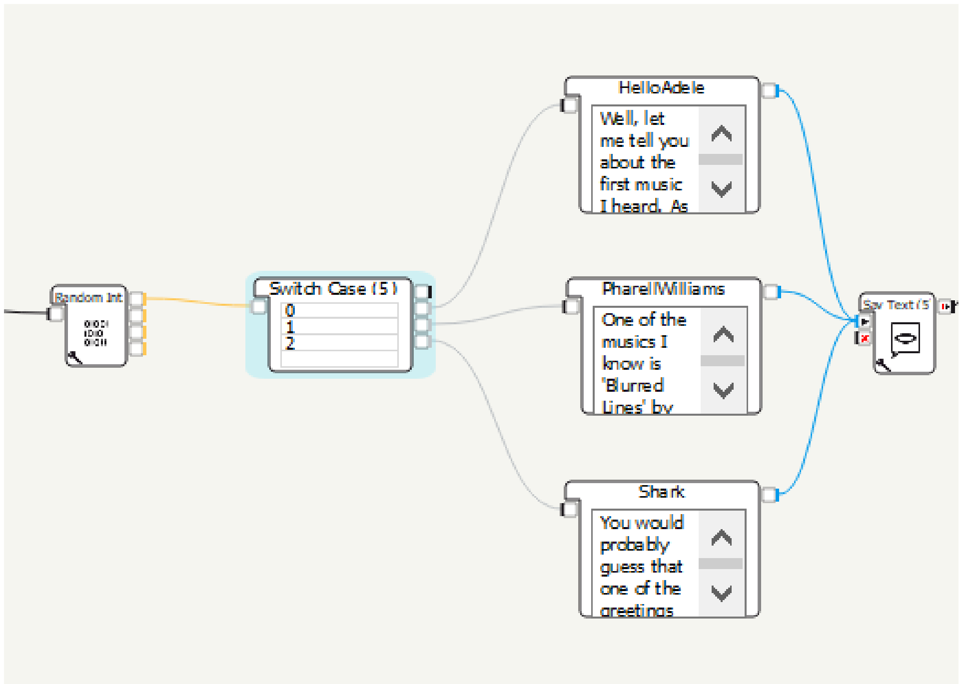

Figure 5. Random songs

In the Choregraphe application, although the English speech of Pepper is adequate, Turkish option is not very useful. Because of that, in the program, if the respondent could not speak in English, Turkish option can be chosen. Instead of using ‘Text’ Edit and ‘Say Text’ boxes in Turkish speaking part, ‘Play Sound’ boxes under the Multimedia, Sound of Box Libraries are used. At the beginning of the program with the ‘Animated Say’ box, Pepper asks if the respondent is able to speak English or not. Response is taken with the help of speech recognition and switch case boxes. In the case that respondents are not able to speak English, with the help of the ‘Set Language’ box under the Speech, Speech settings, Turkish is chosen as the language of the conversation. According to applicants’ choice, Pepper starts to introduce themselves in English (with Text Edit) or in Turkish (with Play Sound).

As mentioned before, Pepper’s Turkish pronunciation is not sufficient and convincing for a native Turkish speaker. Therefore, an alternative solution is pursued through a Turkish Text-to-Speech (TTS) web application called Responsive Voice.[1] MP3 recordings of the

[1] https://responsivevoice.org/

Pepper's Turkish dialogues were obtained from this web application. In order to use these recordings in the Pepper's program, sound files were converted into WAV format and attached to the "Play Sound" box. As a result of this method, Pepper's Turkish pronunciation problem was solved and it communicates with research participants successfully in both Turkish and English.

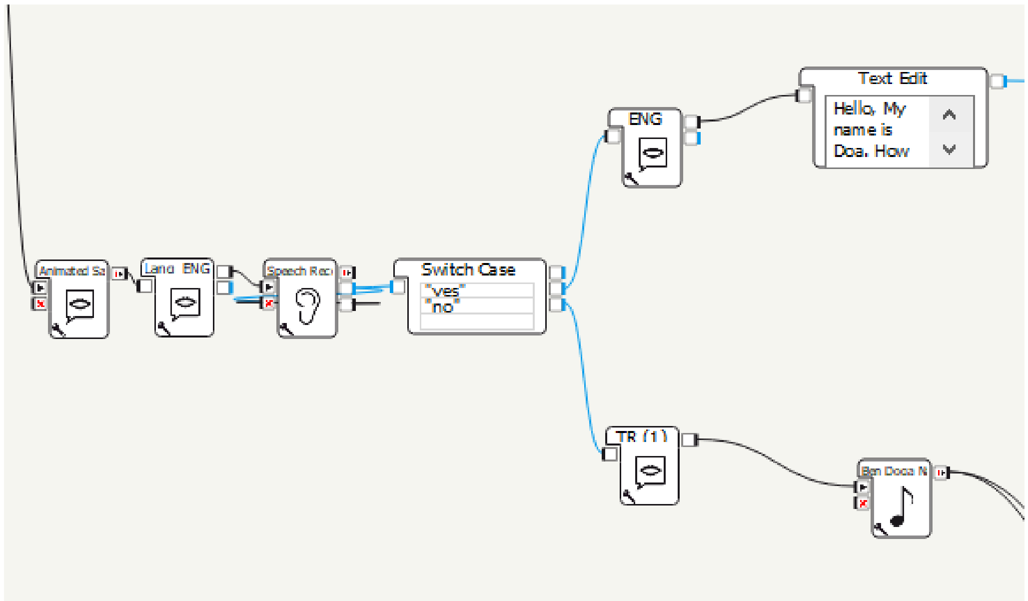

Figure 6. Language selection from Switch Case

In the Turkish conversation part of the program, instead of the 'Text Edit' and 'Say Box' duo to make Pepper speak, Pepper speaks with the sound files recorded from another application. These sound files are used with the 'Play Sound' box. After the speech, the program proceeds just as in the English part of the program. Speech Recognition boxes recognize the words that respondents use to describe their mood and Pepper responds according to the mood of the respondent and asks if the respondent wants Pepper to talk about the movies while they are walking to the research room.

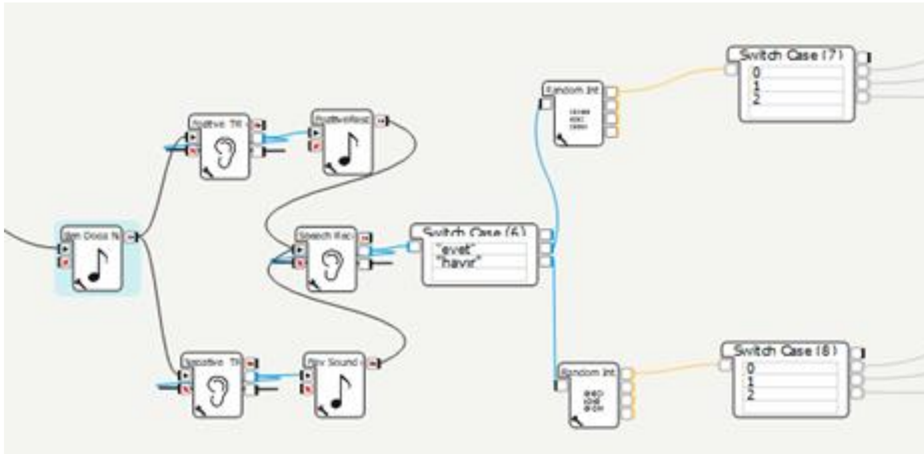

Figure 7. Turkish Dialogue

2.2. Face Detection

One of the most important signals in the initiation of the interaction between human and robot is that the robot understands and reacts when the person comes in front of the robot. It is also important to give this reaction to the user so that he or she understands when communication has started. 'Face Detection' and 'Face Recognition' boxes are used to ensure that the user is detected by the robot.

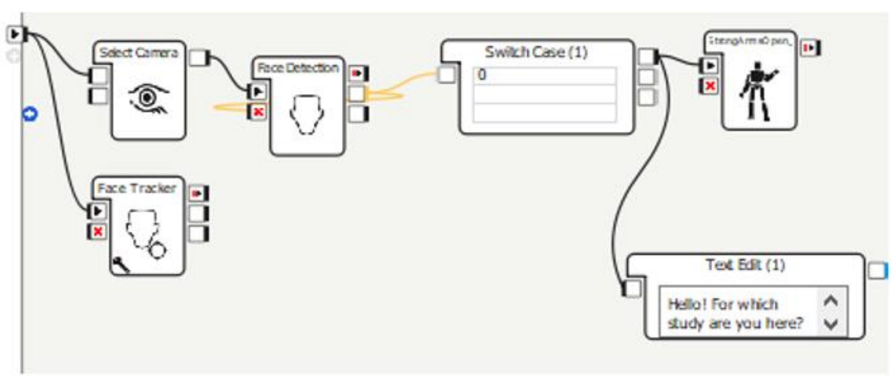

Figure 8. Detecting people

Figure 8 shows the correct usage of the Face Detection feature. This allows the robot to take action when it detects someone. First of all, the ‘Select Camera’ box should be used to activate the robot’s camera. With the help of this box, it can see its surroundings and detect objects around it. After the camera is turned on, the ‘Face Detection’ box is activated. The output of the Face Detection box represents the number of faces recognized. When the output from this box is connected to the switch case, the program will not continue to run as long as the number of faces is 0. If it is different from 0, it will understand that the person has been detected and the program will continue to run.

If it is thought that there is always a human in front of the Robot during the interaction, the Face Detection box will work continuously and thus the number of faces will always show a value different from 0 as an infinite loop. But in fact, the moment it sees a face, it has to finish that part now. That's why the box must be killed. Thus, a break command should be given. Figure 9 shows how the face detection box is killed and does not enter the infinite loop.

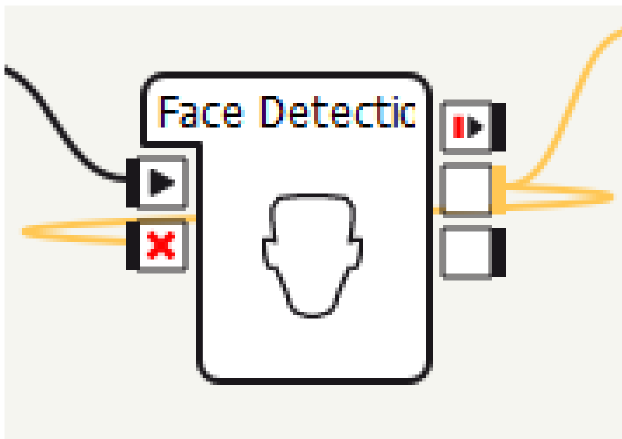

Figure 9. Killing the box

When the switch case catches any input, a connection is made to the place it wants to go from the onDefault output. Thus, the face detection part is finished. While all these are happening, if the ‘Face Tracker’ box is used in parallel, it is ensured that the robot follows the user with its head throughout the actions. As a result, the robot detects the person standing in front of it and reacts accordingly. In addition, it makes eye contact with the user. It can be said that this is a core part of human-robot interaction.

2.3. Speech Recognition

Taking input from the user and taking action accordingly is also an important part of interaction. Guiding correctly according to the user's answers gives the user the feeling that it is understood by the robot and thus strengthens the interaction between user and robot. One way to understand user responses is to use the ‘Speech Recognition’ box. This box takes some strings as input and tries to find the correct path by sending the detected string to the switch case.

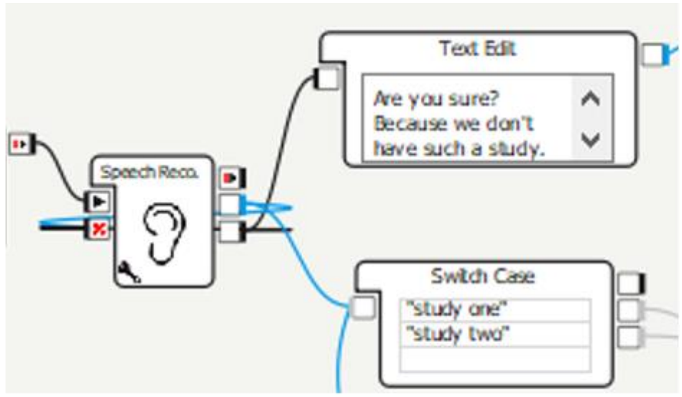

Figure 10. Speech recognition

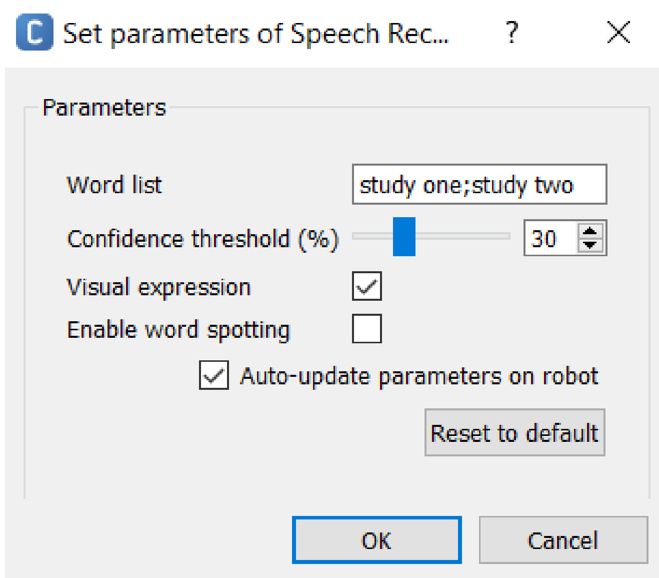

Figure 11. Parameters of Speech Recognition

Speech recognition requires a switch case, as seen in figure 10 because when looking at the parameters of speech recognition, it is seen that strings should be introduced to the robot by the programmer. These strings are separated by semicolons and each represents a case (Figure 11) The point to note here is that no spaces should be placed after the semicolon. As a result of the tests, it was determined that when a space is inserted after the semicolon, it does not detect the second case.

The middle output in the Speech Recognition box (Figure 10) is activated when the recognized string is one of the inputs of the box and it directly goes to the switch case to understand which case it is. Like other boxes that require input from the environment, the speech recognition box must be killed, otherwise the box may be stuck in the infinite loop and prevent the program from progressing. If the user says a word other than the parameters the box has, the robot will not understand it and will activate the onNothing output. When looking at the manual dialog screen on the Choregraphe, it was observed that the onNothing output was used because it perceives another word as one of its parameters but its confidence threshold is low. So selecting a parameter of the box, 'confidence threshold' too low can cause it to count other words as correct. The default confidence threshold (30%) is suggested to use.

The Speech recognition box does not require any other boxes before it, but it waits for an input from the moment it is run. Therefore, before Speech Recognition is run, the user may be asked a question so that the user understands that he or she has to give an input. Then, according to the input person gives, other ways are created with the help of the switch case and thus an interaction is created between the user and the robot.

2.4. Movements

Pepper is used in a variety of places such as banks and medical facilities in Japan.[2] Similarly, in this project we wanted to use Pepper in order to greet people who are coming to participate in different studies in our school facility. Since it is invented in order “to make people enjoy life” we wanted to fulfill this premise as much as possible.[3] To fulfill this premise we try to make Pepper look as friendly as possible and consequently make participants more comfortable in the study environment. To do this, there were certain steps we needed to achieve along the process. First, we needed to make sure it walks properly and second, it looked friendly.

2.4.1. Walking

One of the problems we faced is that Pepper does not start moving when it perceives an obstacle. Lengths, which are not issues and solved simply by moving just a little in order to avoid the perceived obstacles for an average human, makes Pepper not want to move at all. Though we programmed the Pepper for a designated area and seems to work fine, one should not undermine the effects of other obstacles that might pop up any time. For example; even slight changes in the place of a chair/or a desk would probably result in Pepper stopping and remaining there indefinitely. Also, our current program does not solve problems like moving objects, for example in a room full of people there will be many unstill movements as expected. In a sitıation like this, Pepper would stop and remain like that as well. Since this project ran remotely, these things remained unsolved. One might expect to solve these problems given proper conditions such as hands-on experience with Pepper and more time.[4]

[2] https://www.softbankrobotics.com/emea/en/pepper
[3] https://en.wikipedia.org/wiki/Pepper_(robot)#cite_note-17
[4] http://doc.aldebaran.com/2-4/family/pepper_technical/pepper_dcm/actuator_sensor_names.html

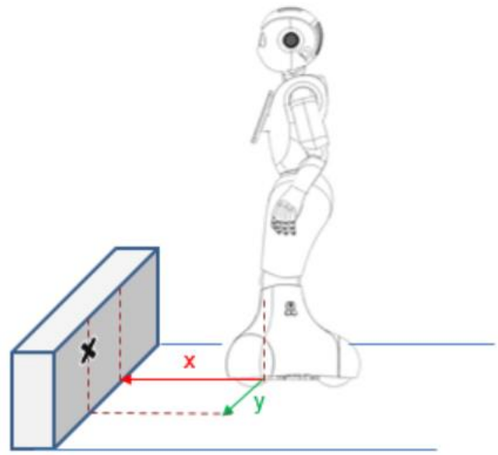

Figure 12. Obstacles

2.4.2. Achieving Anthropomorphism

We used many properties of Pepper such as it’s 20 degrees of freedom for natural and expressive movements; it’s touch sensors, LEDs and microphones for multimodal interactions; it’s infrared sensors, bumpers, an inertial unit, 2D and 3D cameras, and sonars for omnidirectional and autonomous navigation.[5] We can say that the properties mentioned above were both necessary and sufficient for our project.

Pepper allows users to create their own movements in its environment using timelines. A Timeline enables users to synchronize boxes with movements, movements with each other and/or Boxes with each other. And it’s constituted from a Motion layer which contains movements and one or several Behavior layers which contain Flow diagrams.[6]

Though we had some experiments in order to create our own movements, we predominantly used premade movement boxes from the platform since they satisfied our needs in the creation of a program that greets participants.

[5] https://www.softbankrobotics.com/emea/en/pepper
[6] http://doc.aldebaran.com/2-4/software/choregraphe/panels/timeline_panel.html

2.4.3. Premade Boxes

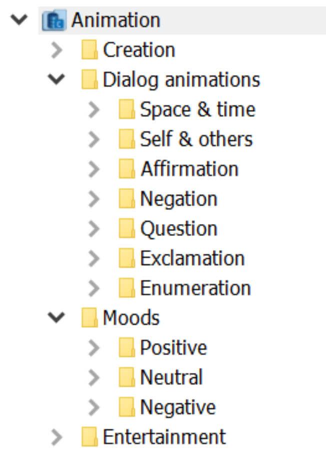

Figure 13. Premade Boxes

Pepper does not come back to its previous state after a user-made movement so one might note that it needs an additional box, or a movement within the same timeline, for Pepper to go back to its previous position; since humans do not continue a movement indefinitely.

We used movements while Pepper is talking with the participant since it feels more natural. Some examples can be seen in Figure 14 and Figure 15.

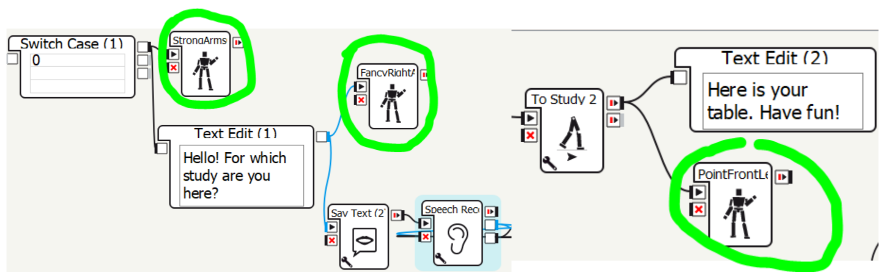

Figure 14 & Figure 15. Humanoid behaviors

3. Discussion and Conclusion

In order to comply with the new regulations related to COVID-19 the need for social robots increased and in this project Pepper, a humanoid robot that can engage in social communication with people, is used. In this research, focus was on developing a social robot which welcomes the participant and then to inform and direct the very person to the study they came for. With the help of this reception robot, reducing interpersonal interaction time and maintaining social distance at the designated study area is aimed.

Researchers created the Choregraphe environment to be able to give and regulate commands to Pepper robot. Dialogues between the robot and the human are organized in a way that makes the Pepper robot appear more humane and intimate.In this project this property has been used many times as Pepper was programmed to talk sincerely with the participant on the way to the designated study place. Likewise, the movements were adjusted with other properties of Pepper to make this experience feel more natural and humane. Moreover, these dialogues can both be in Turkish and in English. Although Pepper's own language and voice is used in the implementation for English speaking, for Turkish speech, an alternative solution is pursued by using a web application and recording MP3 voices. Furthermore, continuation of the dialogues are controlled with the speech recognition done by Pepper. In both languages, Pepper was able to recognize what the respondent said and responded back to the participant and/or Pepper got the information about which study the participant will participate in according to this recognized speech input.

Initiation of the interaction between the research participant and Pepper starts when a person comes in front of the robot or when Pepper is called. Though Pepper can also recognize faces, it was decided to not be used in the final programme since recognizing participants faces was not necessary and it could cause more issues in a place where there is a large circulation of people.In the research, implementations were made to establish a dialogue with the participant and walk them to the selected study area. Unfortunately, the robot is error-prone and a tiny error in any part of interaction cannot be reverted. For example, because of the ambient voice, speech recognition may not work properly and can recognize an input that is not the original response of the participant. Furthermore, in the case of a changed environment, Pepper would stop walking since it won't be able to understand the changes in its environment. Future research may focus on solving these problems when an error occurred.

Division of Labour

Abstract- *Berna Yıldıran*
Introduction - *Aysun Öğüt*
Dialogue - *Berna Yıldıran, Göktuğ Gençkaya*
Face Detection - *Aysun Öğüt*
Speech Recognition - *Aysun Öğüt*
Movements - *Aslı Aydınlar*
Discussion and Conclusion - *Aslı Aydınlar, Göktuğ Gençkaya*
Layout Design - *Aysun Öğüt, Berna Yıldıran*

References

Brin, S., Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. In: Seventh International World-Wide Web Conference (WWW 1998), April 14-18, 1998, Brisbane, Australia.

Einstein, A., Podolsky, B., Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete? Physical Review, 47 (10), 777-780. doi: 10.1103/PhysRev.47.777.

G. Z. Yang, B. J. Nelson, R. R. Murphy, H. Choset, H. Christensen, S. H. Collins, P. Dario, K. Goldberg, K. Ikuta, N. Jacobstein, D. Kragic, R. H. Taylor, M. McNutt. (2020). Combating COVID-19—The role of robotics in managing public health and infectious diseases. *Science Robotics*, 5. doi:10.1126/scirobotics.abb5589

Hofstadter, D. R. (1979). Gödel, Escher, Bach: An Eternal Golden Braid. New York, NY: Basic Books Inc.

N. Riether, F. Hegel, B. Wrede and G. Horstmann. (2012).”Social facilitation with social robots?”, 2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 41-47, doi: 10.1145/2157689.2157697.

Pepper the humanoid and programmable robot: SoftBank Robotics. SoftBank Robotics - Group. (n.d.). https://www.softbankrobotics.com/emea/en/pepper.

SoftBank Robotics. SoftBank Robotics Documentation. (n.d.). http://doc.aldebaran.com/.

Wikimedia Foundation. (2021, May 19). *Pepper (robot)*. Wikipedia. https://en.wikipedia.org/wiki/Pepper_(robot)#cite_note-17.