Head and gaze tracking control for a smart wheelchair with a robot arm

First Stevedan Ogochukwu Omodolor[1†], Second Pablo Mañas Pellejero[1†] and Third Javier Pedrosa Alias[1†]

[1*]Robotics and Automatic Control, UPC, Avinguda Diagonal, 647, Barcelona, 08028, Catalonia, Spain.

Contributing authors:
stevedan.ogochu.omodolor@estudiantat.upc.edu;
pablo.manas@estudiantat.upc.edu;
javier.pedrosa@estudiantat.upc.edu;
[†]These authors contributed equally to this work.

Abstract

For people with severe disabilities who do not suit with conventional wheelchair driving systems, like the manual ones or the electric with joysticks, for example; alternatives and solutions have been researched, leading to a proposal that has been developed based on an interface that simulates the control of a wheel-driven platform, such as a wheelchair, with an integrated robotic arm, both controlled by head gestures and gaze tracking. A webcam and a facial recognition algorithm are used to read the input movements, and a simulation program to perform the actions.

Keywords: human-robot interaction, teleoperation, gaze tracking, head tracking, disability support, wheelchair, robot arm

Introduction

This report includes, on one hand, a research about the current alternative driving systems for wheelchairs, as well as the possibility to integrate a robot arm in them, with the purpose of providing the ability for individuals who cannot move anything other than their head and face, to be able to move around in a wheelchair and manipulate objects or interact with the environment despite their limited movements.

On the other hand, this document also includes the design and development of a proposed solution for the introduced purpose, based on a simulated prototype of a wheel-driven platform, which would imitate the wheelchair, but could eventually be anything else, with a 6-DoF robotic arm above it; which is operated by a facial recognition algorithm to detect the head motions and track the eyes.

The structure of the document is the following: Section 1 corresponds to the the state of art, and contains related literature to the topic, which inspired the prototype development; Section 2 describes the proposed solution, in terms of the main idea, and the implemented software in detail to perform the simulation; Section 3 includes a concise analysis of the achievements and the barriers that have arisen during the development; Section 4 corresponds to the budget, to estimate how much the proposed implementation could cost; and finally, Section 5 includes the conclusions.

1 State of Art

Today there are many possibilities for people who need electric wheelchairs to get around, either because of illness or injury affecting the lower limbs or because of old age and reduced mobility. However, when it comes to people who require special care due to their circumstances, i.e. all those who cannot use their limbs (quadriplegia), or those who suffer from paraplegia accompanied by severe motor disabilities, and therefore require more special systems to be able to perform everyday tasks other than moving around, the existing possibilities are reduced. Thus, this section focuses on explaining the technologies applied to help this group that requires more help in their daily lives.

There is a great deal of research that focuses on alternative ways of controlling a wheelchair, often combined with manual methods (joysticks) to improve them, and, at the same time, provide the user with multiple options for the use of their wheelchair. Based on the search through different articles, it is possible to classify these systems according to the type of technology they use as follows:

- Voice recognition systems [1].
- Inertial Measurement Units (IMU) [2].
- gaze tracking systems [3], [4], [5], and [6].
- Head movement detection and face recognition [3], [7], and [8].
- Brain Computer Interface (BCI) systems [9].

1.1 Voice recognition systems

With the growth of artificial intelligence and the emergence of voice assistant devices such as the popular Amazon Alexa or Google Home, it has become possible to communicate between user and machine in a simple way. Thanks to this, people can communicate through an interface using voice with a computer or adapted device that will transform what the user transmits into specific commands to generate a response to the system to which it is connected, for example, turn on a light or turn off a television.

With this in mind, it is possible to use a device that transforms what the user says into concrete commands to control a wheelchair, as demonstrated in the article [1]. In this text, the authors describe the implementation of a Digital Signal Processor (DSP) in a wheelchair for people with paralysis who can only use voice as a means of communication. They also explain how by analysing the characteristics of the signals produced by the voice: total energy, zero-crossing and standard deviation; they can interpret the 5 selected command words: forward, reverse, stop, left and right. Given that with only these 5 words, the control of a wheelchair is sufficient, it is considered a good approach and a good use of this type of technology.

1.2 Inertial Measurement Units (IMU)

Inertial Measurement Units (IMU) are electronic devices capable of measuring the specific force, angular rate and orientation of a body. Thanks to these devices, it is possible to determine the position of a part of the body such as the head or neck with a certain degree of accuracy and use this signal to generate commands that can be used to control the wheelchair.

An example of the use of this technology applied very specifically to the detection of the 3 degrees of head orientation can be read in article [2]. In this paper, the authors propose the use of an inertial measurement sensor to determine the yaw, pitch and roll angles of the head of quadriplegic users with only upward neck mobility. Once the signal is obtained, applying some formulas and algorithms, the wheelchair can be controlled using the head as if it were a joystick.Finally, they also show the differences between manual joystick control and control using the head and mention the disadvantages of the latter, thus the lack of precision.

1.3 Gaze tracking systems

A different way that can help people with severe mobility problems is the use of systems that recognize where the user focuses their gaze and act accordingly. This type of technology, also known as eye trackers, is applied using cameras, special glasses or adapted devices that capture eye movement. As in any other system, by interpreting this type of data, it is possible to interact or perform a specific action.

For instance, the paper [6] shows a prototype of a wheelchair controlled by gaze tracking. The authors explain that the gaze tracking is performed by capturing the movements by one IR (infrared) camera in each eye; and then processing the obtained patterns by a 2-d (kernel of dimension 2) Convolutional Neural Network (CNN). It also incorporates a safety system to prevent obstacle collisions. In addition, the article describes other types of gaze tracking systems and different existing approaches.

1.4 Head movement detection and face recognition

Apart from the use of inertial sensors (IMUs), it is possible to use cameras together with facial and gesture recognition to determine the orientation and position of the user's head. This different approach is used individually or with the help of sensors to increase precision. In fact, the operation to a large extent is very similar to that described in section 1.2: the yaw, pitch and roll angles of the head are detected and a specific command or action is performed depending on these signals. In some cases, thanks to certain algorithms and with the impulse of neural networks, it is possible to perform facial recognition that allows us to know more in depth some aspects of the user. For example, if the user opens his mouth, the system will be able to detect him and carry out a specific order, which gives more flexibility and allows more possible operations.

In [7], the control of an electric wheelchair is accomplished through recognizing head gestures by using a stereo camera combined with a magnetic sensor attached to the head. Although this magnetic sensor is very accurate, it has to be attached to the user and it may be impossible for the user to reposition the sensor if it slips off. In any case, the combination of these two devices allows people with severe disabilities to operate their wheelchair in any scenario (different environments). The system also considers the user's posture, recognizes involuntary movements and user-dependent head movements or tendencies. Gesture recognition is performed by considering the estimation of the face orientation angle captured with the camera and then, depending on the result, a direction is selected (forward, right, left, down). Other example that can be considered is the article [8] where the control uses an integration of the Adaboost face detection algorithm (Bradski) and the Camshift object tracking algorithm to detect head movements.

1.5 Brain Computer Interface (BCI) systems

Research into the use of devices capable of capturing brain signals, such as electrode helmets that can generate electroencephalograms (EEGs), has been ongoing for some years. With this type of generated signals, which have to be filtered and adapted before being analysed, it is possible to determine the user's intention and therefore generate actions according to what is captured.

Focusing on the issue of wheelchair control, it is possible to find article [9]. In it, it is detailed how using EEG signals together with inertial sensors it is possible to control an electric wheelchair and a robotic arm. In this case, the control of the wheelchair is carried out through an IMU, already explained in previous sections, while the control of the robotic arm is carried out thanks to the EEG signals captured with electrodes. In addition, they provide a graphical interface so that the user can choose the different modes.

1.6 Combined systems

In some of the articles mentioned in the previous sections you can see the use of different technologies in order to control the navigation of the electric chair better or in order to add more elements to the system such as the use of a robotic arm. Therefore, it is necessary to mention that on many occasions, the use of different technologies is very practical to improve the user experience and provide a greater range of possibilities.

In the newspaper article [10], it is detailed how the systems that allowed the prestigious physicist Stephen Hawking to communicate and control his smart chair were adapted and improved. This reading reminds us that it is very important to adapt the systems to the specific user and to calibrate the systems specifically according to their needs. In addition, it highlights the use of multiple systems combined with each other to expand the user's action possibilities, which is very important.

1.7 Interfaces

It is necessary to mention that on many occasions, it is essential to use interfaces between the user and the computer or system that acts accordingly to the received signals. For this, screens are used in the vast majority of applications that allow the user to know the status, the operating mode or the situation of the environment. In case none of this is precise, the user must be able to understand what is happening through what he can feel through the senses.

Therefore, considering the different systems studied together with the combination of some of them, it has been decided to make a prototype that combines the recognition of the movement of the head together with the detection of eye movement to control a wheelchair with a robotic arm.

2 Proposed prototype

As mentioned on the *Introduction*, after assessing the researched literature, we considered that further effort is required in this area to offer more solution options, and if possible, more accessible economically as well as in terms of technology.

That is why we propose a control wheelchair that include a novel component, a robotic arm. Typically, most solution are centered on displacement and there is a lack of focus on the user capability to perform activities of daily life. The goal of this newly added component is to give the targeted users more control and independence in performing activities of daily life that require manipulation of object. In addition, the proposed control system relies on the head, including the gaze, and it is intended to be an easy and intuitive way to interact with the physical devices.

In terms of the developed prototype, it was reduced to a simulation because it would have been too much complicated to design it adequately with the available time we had, and we did not have the necessary resources to develop it. Moreover, a specific application has been selected to narrow down the concept and to have a more defined approach, which corresponds to the use of the robotic platform to push something; like any vertical button panel, such as the one of a lift or doorbell, or even to push an object or door.

2.1 Setup

As mentioned previously, the prototype was developed in a virtual environment, so the hardware we used is simplified in a computer and a webcam. Nevertheless, the configuration required for the prototype, if it were to be carried out in practice, would include the following components:

- **Wheelchair**: it would be the basis of the prototype, where the rest of the components would be integrated, and from where the user would enjoy its use.
- **Robot arm**: it would be an anthropomorphic 6-DOF arm whose location would still need to be studied in order to optimise its performance.
- **Head tracking system**: it would be a camera to be able to capture the head motions.
- **Gaze tracking system**: it could be managed with the same camera as the head tracking; or it could have an specific device to track them exclusively, which would provide more precision.
- **Sensors**: these would be necessary to include, for example, obstacle avoidance systems, or more interesting inputs for the control unit.
- **Computer**: it would be the core of the prototype, the responsible device for processing all input data and managing the different actuators properly.

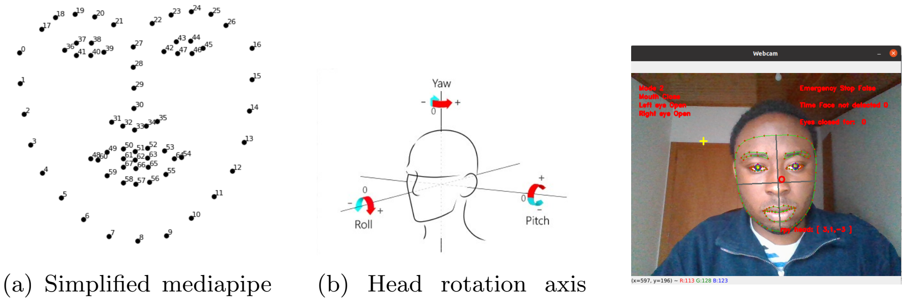
2.2 Interface

For the correct functioning of the prototype, a graphical user interface (GUI) for a computer has been developed, with which the user can drive the robotic platform. It is constituted by two main parts:

- **Head and gaze tracking software**: this part includes the software used to detect the face and track the relevant movements; as well as the GUI handling, all coded in *Python* language. The detection and tracking algorithm is based on a loop that is recording images continuously, and each frame is analysed by the functionalities of two main packages: the *OpenCV*, which is a huge open-source library for computer vision, machine learning, and image processing; and it is able to process images and videos to identify some modeled objects by classical feature detection algorithms (see the documentation for further information[1]). With the help of the a package called *simplified_mediapipe_face_landmarks*[2], which provided a simplified version of the *mediapipe_face_landmarks*[3] face landmark, the gaze and eye tracking system was developed.

  Regarding the implementation, we designed the input value of our robotic system from the previous packages; also defining two operating modes to make the most of these inputs and to have a more controlled operation, so that nothing is disturbed unintentionally. This means that the diverse entries can have different roles depending on the operating

[1] https://docs.opencv.org/4.x/d6/d00/tutorial_py_root.html
[2] https://mediapipe.dev/
[3] https://github.com/bhav09/mastering_mediapipe



(a) Simplified mediapipe face landmarks. (b) Head rotation axis reference. (c) face axes

Fig. 1: Vision system reference

mode. In addition, for safety, this functionality can be disabled by the user. The modes corresponds to:

– **Mode 1**: which focuses the control on the wheeled platform.
– **Mode 2**: which focuses the control in the integrated robotic arm end-effector, to perform a push action more precise.

The designed input sources correspond to:

1. **The head rotation movements depicted in the Figure 1a**. Their interpretation is possible thanks to a face feature detection *simplified_mediapipe_face_landmarks* model, which provides the key-points shown in the Figure 1b. Then, a reference axes are defined with four of the available landmarks, in such a way they halves the face horizontally and vertically, finally dividing it into four quadrants (as shown in the Figure 1c). When the user moves the head, these quadrants dimensions change, and these changes can be used to scale variables as entries for the control system. In the implementation, each of the possible rotations drive different parts of the robotic platform. Their roles coincide between modes, but the Yaw is only available in Mode 1. The roles are specifically:
2. **The gaze direction**. With the help of the landmarks of the eye position, we are able to determines where the user is looking by a comparison of the pupils location with the eyes bounding. With this technique, we are able to determine whether the user is looking the following directions: up, down, left, right. In the implementation, this is only available on Mode 2, and it is treated in a joystick-like manner to move a cursor shown in the GUI 1c. It represents the position the end-effector has to position it-self in order to perform the push action.
3. **The state of the mouth**, in a binary form, distinguishing between open and closed. It is interpreted by the same face feature detection *simplified_mediapipe_face_landmarks* model, which also has landmarks corresponding to the mouth. Then, we chose two of them so that they form a vertical line in the middle of the mouth; and this

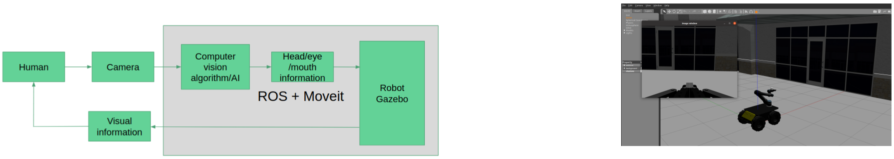
line length is thresholded to indicate, by a bound Boolean variable, its state. In the implementation, it is available from both operating modes indistinctly, and it is used as the enabling flag for the robotic platform actions (Pitch and Yaw, as well as the end-effector positioning by gaze), meaning that if it is not enabled, the platform should stop; and should quit the program instantaneously in case an emergency mode is triggered by turning the head in the Roll axis ($> 15°$ in any direction).

4. **The eyes detection**. It is interpreted by using the same technique as mentioned previously in the state of the mouth. With this, the state of the eyes can be known(closed or open) or when lost. In the current implementation, the eye detecttion is available in all mode and has different functionality mentioned as follows:
   – Right eye wink (lost about 1 s) + Roll rotation ($> 15°$ in any direction) $\rightarrow$ enable operating Mode 1 if mode 2 is enabled.
   – Left eye wink (lost about 1 s) + Roll rotation ($> 15°$ in any direction) $\rightarrow$ enable operating Mode 2 if mode 1 is enabled.
   – Closed eyes (both lost more than 10 s) $\rightarrow$ it is a safety check that is triggered when they have fallen asleep; or when they become unconscious.
5. **Emergency trigger**, In order to ensure safety of the system, multiple emergency protocol were placed that can be trigger intentionally by the user or by the system itself, this can be seen in the right part of the image depicted in 1c. A summary of the triggers that causes the emergency flag to be set to true:
   – The user eyes are closed for a set period of time as mentioned in the previous item.
   – If the user rotates the head in the roll axes past a fixed angle in any direction with the mouth open
   – When the user face is not detected in camera's range for a set period of time.

- **Wheelchair and robotic arm System**: As mentioned previously, in order to test the proposed solution, a simulation is used to represent the real system. The simulation was done using Gazebo, which is an advance 3D simulator. In order to control and manage the control of the whole system, ROS was used. Figure 2a shows the design of the whole system.

  In order to simulate the wheelchair, a husky UGV attached to a UR5 arm was used. A camera was placed on the robotics arm to allows the user to perform the gaze control mentioned previously. Figure 2b shows simulation environment.

  The robot controller works differently based on the mode the user is in and the actions the user wants to perform.
  – **Mode 1**: The angular data from the user head are converted to translation/rotation speed, with movement activated only when the mouth is open.



(a) System design scheme

(b) Simulation environment

Fig. 2: System design

- **Mode 2**: In this mode, the position of the cursor obtained from the user's gaze is used to determine where the user wants the robotic arm to push upon. When activated by the mouth, the robotic arm perform a sequence of actions in other to push the object of interest and the necessary action to return to its initial position. To improve manipulation and ensure success, the robotic base is programmed to autonomously move closer to or further from the object so as to be in the ideal work space for the robotic arm.

It is worth noting that independently of the move, when the emergency stop is activated, the whole robotic system shuts down and is not able to receive any command only when the whole system is restarted. The full code implementation of the system can be found in [4]

3 Achievements and barriers

After evaluating the prototype by performing multiple simulations for the specific application of moving the wheelchair and pressing a button using the robotic arm, it is possible to identify the achievements made and the barriers or impediments observed. This section therefore tries to explain what objectives the current prototype fulfils, what elements should be incorporated to improve the system, and what difficulties users might encounter when using the developed system. Demos of the system working can be seen in [5]

First of all, it is important to highlight that the prototype was designed for the application specified above, and then, if other tasks were to be performed with the robotic arm, probably, new instructions, commands and modes of use would have to be added to the model.

Secondly, it has been observed that the gaze tracking system, although working well with a webcam, and even in different users, lighting and environment conditions; could be improved by using more sophisticated hardware systems, such as those described in the Section 1.3, which are quite more accurate. Also, the joystick-like control approach should be replaced by a direct gaze cursor control, i.e. display the cursor exactly where the user is looking.

---

[4] https://github.com/stevedanomodolor/Head_and_gaze_tracking_control_for_a_smart_wheelchair_with_a_robot_arm.git

[5] https://drive.google.com/drive/folders/1Af1GDXXvdE6rkpkNgx9cCG_ovlDowA7C?usp=sharing

Thirdly, the use of different modes that work independently, is proved to be a correct and appropriate decision, since it allows the user a more precise handling of the tools at his disposal. If the two defined modes were to operate at the same time, the user could experience discomfort when moving the arm or the wheelchair, as the head movement would cause the point at which they are looking to vary, and the result of the relevant operation would be less precise and more unstable.

Furthermore, it has been observed that depending on the user using the GUI, some adaptations are necessary in certain aspects. This happens due to the fact that each person has different physiognomy characteristics and, in many cases, it is necessary to previously study and modify some factors, such as redefining the thresholds for the rotation angles, or for the gaze dealing. Related to this issue, a calibration method could be implemented, i.e. when the user starts the system for the first time, by performing a series of specific gestures and movements, the program would be able to redefine some key values, in such a way that the solution would be adapted for each user.

Finally, it is necessary to mention that the prototype has been made in a simulated environment and designed by users who have full mobility, therefore, some of the barriers that could appear in a real case may not have been taken into account. With this in mind, it should be noted that, if the implementation was finally carried out, the specific user would have to be analyzed in the first instance, that is, their basic functionalities and motor skills; since perhaps some of the operating modes developed could be modified to enhance the user's experience. For example, if the person using the wheelchair still has some hand mobility, the use of a supplementary joystick to steer the chair could fit in the model.

4 Budget analysis

Although the prototype has been performed in a virtual environment, and the cost of materials is practically zero; it is important to answer the question of what the cost would be if the system were implemented in practice. Then, accounting the components mentioned in section 2.1, and analyzing the different market possibilities for each of the parts, Table 1 details the hardware budget.

Regarding these choices, they could vary depending on the available funds or the user preferences. In this case, the selection includes an electric wheelchair with a built-in joystick, so that it can be a more versatile solution for any user; a robotic arm that was designed to fit with wheelchairs; and finally, a complete device that incorporates a computer, a camera and an eye tracking system in the same product. All the useful sensors to avoid collisions and to determine the position of the arm, for example, are already included within the selected elements, so it is not required to add more. However, it should be noted that the final price would be around 40,000 \$, depending mainly on the robotic arm model chosen, because this element is the one with more economic clout.

Table 1: Budget

Description	Model	Quantity	Price
Wheelchair[1]	BC-6011 ComfyGO Electric Wheelchair[2]	1	1,449 $
Robot arm[1]	JACO Robot Arm Kinova[3]	1	35,000 $ - 50,000 $
Eye tracker System + + Computer + Camera	MyTobii P10[4]	1	1,100 $
		Total	37,549 $ - 52,549 $

[1]Sensors included.

[2]More information: https://www.comfygomobility.com/product/bc-6011-comfy-go-electric-wheelchair/

[3]More information: https://assistive.kinovarobotics.com/product/jaco-robotic-arm

[4]More information: http://www.spectronics.com.au/product/24271

In addition to the final budget, it would be necessary to count the price of the software development, installation, and the final platform mounting; which should have to meet part of the developers' and assemblers' salaries, as well as part of the investment in maintenance and improvements. Nevertheless, as we do not have available numbers about that, the estimation of this extra cost was not possible.

5 Conclusions

To sum up, research in systems for people who have lost motor or mobility capabilities for any reason is a topical issue, but more efforts should be made to broaden the horizons for the same purpose; because, as mentioned above, there is a lot of focus on assisted displacement. Although, there are lot more capabilities that are to be explored in this field. That is what we have wanted to achieve with the incorporation of the the robotic arm together with the mobile platform, in order to expand the manipulation capability of the end-user. In addition, the purpose also includes the aim of proposing solutions that are more accessible to everyone.

Regarding the engineered prototype, it has resulted to be very interesting with the innovating mobility combination and the precise control of the end-effector with the gaze. Even so, there are some outstanding aspects to improve in future researches, like ones explained in Section 3; in order to make the system more robust and efficient.

References

[1] Qadri, M.T., Ahmed, S.A.: Voice Controlled Wheelchair Using DSK TMS320C6711. 2009 International Conference on Signal Acquisition and

Processing, 217–220 (2009). https://doi.org/10.1109/ICSAP.2009.48

[2] Mandel, C., Rofer, T., Frese, U.: Applying a 3DOF Orientation Tracker as a Human-Robot Interface for Autonomous Wheelchairs. 2007 IEEE 10th International Conference on Rehabilitation Robotics, 52–59 (2007). https://doi.org/10.1109/ICORR.2007.4428406

[3] Matsumotot, Y., Ino, T., Ogsawara, T.: Development of intelligent wheelchair system with face and gaze based interface. Proceedings 10th IEEE International Workshop on Robot and Human Interactive Communication. ROMAN 2001 (Cat. No.01TH8591), 262–267 (2001). https://doi.org/10.1109/ROMAN.2001.981912

[4] Ju, J.S., Shin, Y., Kim, E.Y.: Vision based interface system for hands free control of an intelligent wheelchair. J NeuroEngineering Rehabil **6**, 33 (2009). https://doi.org/10.1186/1743-0003-6-33

[5] Berenguer, E., Soria, C., Nasisi, O., Bastos, T., Mut, V.: Robotic wheelchair controlled through a vision-based interface. Robotica **30** (2012). https://doi.org/10.1017/S0263574711000919

[6] Dahmani, M., Chowdhury, M.E.H., Khandakar, A., Rahman, T., Al-Jayyousi, K., Hefny, A., Kiranyaz, S.: An Intelligent and Low-Cost Eye-Tracking System for Motorized Wheelchair Control. Sensors **20**(14) (2020). https://doi.org/10.3390/s20143936

[7] Yoda, I., Sakaue, K., Inoue, T.: Development of head gesture interface for electric wheelchair. i-CREATe '07: Proceedings of the 1st international convention on Rehabilitation engineering assistive technology: in conjunction with 1st Tan Tock Seng Hospital Neurorehabilitation Meeting, 77–80 (2007). https://doi.org/10.1145/1328491.1328511

[8] Lu, T., Yuan, K., Jia, P., Hu, H.: Head gesture recognition for hands-free control of an intelligent wheelchair. Industrial Robot **34**(1), 60–68 (2007). https://doi.org/10.1108/01439910710718469

[9] Achic, F., Montero, J., Penaloza, C., Cuellar, F.: Hybrid BCI system to operate an electric wheelchair and a robotic arm for navigation and manipulation tasks. 2016 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO), 249–254 (2016). https://doi.org/10.1109/ARSO.2016.7736290

[10] Redacción Barcelona: Así funcionaba el sistema de comunicación de Stephen Hawking. La Vanguardia (2018). https://www.lavanguardia.com/ciencia/20180314/441518656736/sistema-comunicacion-stephen-hawking-acat.html (visited: 2 of October, 2021)