Feasibility of eye-tracking controlled assistance robots, with two or more degrees of freedom, for paralyzed persons

Fynn Knudsen
Institute of Mechatronics in Mechanics
at Hamburg University of Technology
Email: f.knudsen@tuhh.de

***Abstract*—Eye-tracking systems have evolved into robust data acquisition tools for a variety of use cases. Some cost-effective and verifiably accurate systems are available on the market. The main purpose of such systems still remains the detection of gaze points on surfaces to determine the interest of persons. Very few applications try to use this technology to control devices. In this paper, we analyze the possibilities of controlling robotic devices by eye-tracking in a medical context. In particular, this could be useful as an assistance device for paralyzed persons. It is analyzed that a full 3 DOF tracking is possible, but not recommended, because of the expected quality of measurement. We propose a system using inertial measurement in addition to eye-tracking to measure head movements. The main subfunctions of the system are tested. The difference between tracking and guiding a system was investigated and found to be insignificant, showing that eye-tracking, in general, can be used for controlling robotic devices. Different head movements are shown to be distinguishable and therefore can be used to switch between different functions of the device.**

***Index Terms*—Gaze tracking, Assisted living, Control systems**

I. Introduction

TECHNOLOGICAL progress in medical applications sometimes lags behind, due to high safety requirements, long approval procedures or the lack of financial appeal in case of rare conditions. The technology of eye-tracking is commonly used in many research applications, where human behavior should be studied. Based on this technology, working environments can be optimized and web applications can be build user friendly. In the sector of entertainment, the technology has developed, so that eye-tracking controlled games can be found on the market. In this case, the technology is already being rarely used in the medical field. Eye-tracking devices to control computers are already used by paralyzed people. A remaining problem of these systems is the lack of flexibility in terms of interaction with the environment. Controlling smart home devices is an easy task using an assistance computer. However, tasks like pushing physical buttons, eating or drinking can not be accomplished that way. Particularly environments that are not prepared, can be challenging for paralyzed persons. In the meantime, mature robotic systems are leading to broad automation in many industries. From heavy-duty to high precision applications there are systems for many specific tasks. Highly cooperative robots are more and more able to operate safely close to humans. Moreover new teach-in approaches lead to intuitive controls, which make automation more flexible.

Since the flexibility of human movement is still unequaled, many tasks a healthy person is capable of, seem to be far from being automated soon. However, it may be helpful for persons incapable of easy everyday tasks. Especially for paralyzed persons with cross-sectional syndrome caused by trauma, robotic solutions could increase their ability to live an independent life. Depending on the height of their trauma, many different inputs for controlling an assistance device can be used. Eye-tracking as input system seems to be commonly usable since many paralyzed persons are still able to control eye-movement unrestricted. Furthermore, it is easy to think of an intuitive way of controlling devices by tracking the gaze. A functional system should be able to follow the gaze when needed and stay in place if not used. Additionally, the device should enable the person to grab and release objects in close vicinity, push buttons or move joysticks to interface with the physical environment or other assistance devices. Still, the development of such a device faces major challenges. First and foremost it is imperative to the system to be intrinsically safe. Paralyzed persons are very limited in their abilities to deal with a malfunctioning system. This makes it necessary to ensure reliable and safe functionality. Furthermore, it is questionable whether or not a system can rely on only using eye-tracking. On one hand, this would increase the number of people that are able to use the device. On the other hand, it limits functionality. If additional input devices are necessary it needs to be analyzed, which technology could be used as an easy complement to eye-tracking.

In this paper, we analyze and discuss the usability of eye-tracking for assistance devices. The chances of such an application should be outlined, as well as the limitations. The usability is analyzed especially with regard to using two or more degrees of freedom as input from eye-tracking. Furthermore, an additional input choice as a complement to eye-tracking is presented. The system we propose is shown as a conceptual design.

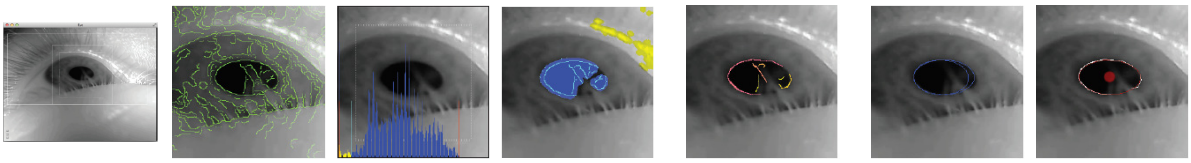
The rest of this paper is organized as follows. Section II presents the related work. Section III discusses the choices for the proposed system. Section IV shows the evaluation. Conclusions and future work are presented in Section V.

II. Previous Related Work

Accurate detection of eye-movement is key to the functionality of an eye-tracking based assistance device. Therefore it is necessary to analyze the existing eye-tracking approaches and asses which system could be used. Furthermore this chapter presents an overview on which technologies can be used as input for degrees of freedom that could not be controlled by eye-tracking.

A. Eye-Tracking Overview

Eye-tracking approaches can be divided into two different techniques of measurement. The eye positions can be detected relatively to the users head or in space, also referred to as Point



(a) (b) (c) (d) (e) (f) (g)

Figure 1: *Pupil Detection Algorithm* [1]

of Regard (POR). In general, it is more precise to track eye positions relative to the head since the measurement is done directly at the eye. One downside of this technique is the need for detecting or assuming the heads position. Some possible techniques according to [2] are:

- Electro-OculoGraphy (EOG)
- Scleral Contact Lens
- Video-Based Combined Pupil/Corneal Reflection

EOG is relative to the heads position. Electrodes around the eyes are used to record the electric potential. The method assumes a constant retinal potential, thus a different potential at the front of the eye, compared to the posterior part, can be observed. Hence after the angle of the eye relative to the head is changed, a potential proportional to this angle can be measured. The biggest disadvantage of an EOG system is the need for new electrodes to be placed properly before every use. According to [2] contact lenses with attached measurement devices are comparatively high in precision. Up to 10 *arcsec* resolution is possible. A wire coil, embedded inside the contact lens is measured while moving through an electromagnetic field. In fact, the range of high precision is limited to about 5°. Moreover, the need of wearing contact lenses could cause discomfort while using the system.

In video-based eye-tracking, a camera picture is analysed in order to detect the gaze direction. This method is affordable and usable for both POR measurements as well as for measurements relative to the head. Commonly the video-based technique is combined with the corneal reflection method. In this case, the *Pupil Detection Algorithm* proposed in [1] is used. The algorithm is used for measurement with eye-tracking glasses. Under perfect conditions, it is able to track fixations with an accuracy of 0.6°. A short latency of 0.0045 *s* makes it suitable for real-time applications. The algorithm takes mainly the following steps to detect the gaze direction:

1) Convert Image to gray-scale
2) Cut image to leave only the area of interest (Fig. 1a)
3) Use a Canny filter (Fig. 1b)
4) Look for user specified dark areas (Fig. 1c)
5) Filter spectral reflections outside the dark areas (Fig. 1d)
6) Extract contours from remaining edges (Fig. 1e)
7) Candidate pupil ellipses are found by least square fitting (Fig. 1f)
8) Search for support contours to candidate ellipse and choose best resulting ellipse (Fig. 1g)

B. Paralysis

Due to the large variety of diseases leading to paralysis, it is important to categorize the conditions, in order to determine for which persons an assistance robot may be suitable.

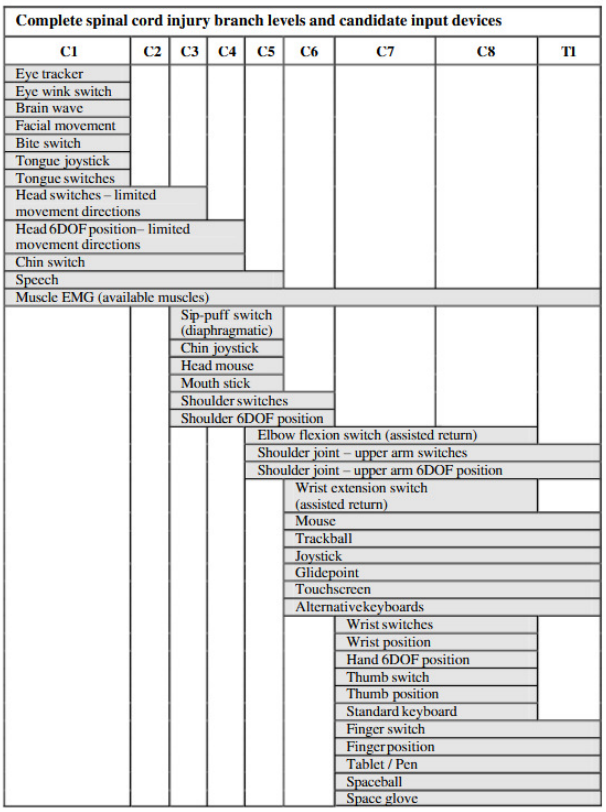


Figure 2: Potential input devices depending on injury [3]

A difference is being made between complete and partial or incomplete paralysis. Partial paralysis restricts the person's ability to feel and move muscles, while a completely paralyzed person is not able to move or feel any of the affected muscle groups. Furthermore, a localized and generalized paralysis can be distinguished. A complete paralysis of skeletal muscle or groups of muscles is called a plegie. A distinction is made according to the number of muscle groups a patient cannot innervate. A tetraplegia is a paralysis of all four extremities. This work will focus on tetraplegia because the usage of eye-tracking is particularly useful here. Paralysis might as well, depending on the cause, be stiff or spastic.

The causes of paralysis are enormously diverse. The most common ones are strokes and damage to the brain or spinal cord by an accident. Other reasons are various diseases. Demyelinating diseases should be mentioned here. These affect the myelin sheath and decrease the ability of nerve cells to transmit signals. The most common demyelinating disease is

multiple sclerosis (MS) [4]. Other conditions, like motor neuron disease (MND) affect only the motor neurons, which act as a link between brain and muscle. Depending on which part of the transmission path is damaged the disease is categorized as an upper (MND) or a lower (MND). Primary lateral sclerosis is a disease, causing spastic or stiff paralysis by damaging the upper motor neurons. However the incidence rate is relatively low. According to [5] one person per million and year is affected. With upper (MND) muscles tend to be stiff or spastic. An externally driven assistance system might not be applicable. A known lower (MND) is spinal muscular atrophy. This disorder harms the spinal nerves, which connect muscles to the spinal cord. The prevalence is approximately 1-2 per 100.000 persons [6]. The symptoms may differ and not every patient will necessarily be tetraplegic. A relatively common (MND) affecting both upper and lower neurons is the amyotrophic lateral sclerosis (ALS). The incidence rate is approximately 2 per 100.000 and year [7]. (ALS) is a fatal disease, which is causing a tetraplegia at some points of the progression. The main reasons for a tetraplegia are lesions in the cervical spine. According to [8] the incidence rate in Germany is more than 10 per million and year. The exact symptoms may vary depending on the reason for the tetraplegia. Symptoms that could influence the practicability of an eye-tracking system are spastic paralysis, which occurs depending on the injured motoneuron after a lesion [9] and the need for other obstructing medical devices, like it could be the case with artificial respiration, which is the needed for spinal column injuries cranial to the vertebra cervicalis (C4). Furthermore most paralyzed persons who are not able to move their upper limbs are using a wheelchair.

Depending on the height of the cross-sectional lesion, there are different methods of getting input for controlling assistance devices. Fig. 2 shows possible input devices for different heights of spinal cord injuries. Assuming that a person is injured above C3, a major challenge is getting any kind of input from the person. Many devices, dealing with this issue already exist on the market.

One of the most advanced technologies may be voice recognition, which obviously requires the patients to have control over their voices. Full control is only ensured if the lesion is below C5. However, even with limited control speech could be used by recognizing sounds. In this case, many applications e.g. Phone or television can be controlled since most common voice control devices interface with many of them. Another commonly used device is the Sip-and-Puff (SNP) controller. It is a sensory system, detecting changes in air pressure, which the user can induce by using his breath. Most devices can differentiate between sharp and lower-level sips or puffs. That way four degrees of freedom can be controlled. A power wheelchair is an application where SNP devices are used [10]. This kind of controller can not be used for injuries cranial to C3. Patients with less ability are often still able to control their heads position. If so, a head position controller is worth considering. In theory, six degrees of freedom can be controlled by the heads movement. Practically this number reduces, the higher the spinal cord is damaged. However, tilting the head forward/backward and left/right is possible in many cases. Based on that, [11] proposes a head-controlled device, using tilt sensors to control a wheelchair. A similar approach is a chin-switch or chin-joystick. In both cases the neck muscles are used to operate the input device, nevertheless, it is possible to use only jaw muscles, which are innervated by Nervus mandibularis [9]. This nerve originates from a main cranial nerve and is therefore not affected by spinal cord injuries. However, a chin controlled device may be impractical for persons with limited head control. To improve usability and precision different technologies are topics of ongoing research. Electromyography (EMG) is a way of getting input directly from the nervous system, by placing electrodes on or below the skin to measure action potentials. These are rapid changes in electric potential that are transmitted along neurons. Noisy signals are a challenge when using externally measured EMG signals, since action potentials have an amplitude of only about 100 mV. Especially signals closer to the brain are difficult to measure and keep apart from signals, that should not be recorded. Additionally, systems using electrodes on the skin are very sensitive to changes of the electrodes position. Other than that electrodes are expensive and most often not suitable for multiple uses. However it has been shown, that easy tasks like pinching can be done by an EMG controlled exoskeleton if the required signals are available [12]. Since it is impractical and not always possible to measure signals via surface EMG, one approach to improve is using implants. Some results show, that under certain conditions precise input is generated [13]. The same technique, applied to signals in the brain is called Electroencephalography. Signals can be provided for any level of injury, but the assignment of signals to movements is even more difficult than with surface EMG. Another minor field of research is input via tongue control. One approach is to fit the measurement device into a dental retainer. The tongue carries a magnetic tracer, that is pierced into it. The chip inside the device transmits by sending RF signals to a processing microcontroller [14]. This system is very simple and therefore widely accessible to many paralyzed persons.

III. The Proposed Design of an Eye-tracking System

The conceptual development of an eye-tracking device is done according to the methodology proposed in [15]. Therefore the design starts with a process analysis. The goal of this design step is to define system boundaries, interfaces and a functional structure. This structure is visualized in Fig. 3. Each circle indicates a state. The diamond shapes indicate decisions, demanding user input. Processes are shown as boxes. The system should have three modes. The main mode is the tracking mode. While being in this mode, the user shall be able to move the device by fixating the desired point. Permanently the system shall decide, whether it should stop tracking and enter idle mode. This is necessary to prevent the system to follow saccades, which are rapid eye-movements. In idle mode, the user shall be able to control the end effector of the device or start tracking again. The interfaces to the system are the device itself, shown as the actuator interface, as well as the end effector interface. In addition, each input requires an interface to the user, and thus requires information exchange across system boundaries.

The whole process requires user input at different steps. The goal of this chapter is to determine, which inputs can be provided by eye-tracking and what technologies we propose to use for those inputs, where eye-tracking is not suitable. As a requirement in terms of accuracy a reasonable area of up to 1 m distance and 0.5 m laterally shall be controllable with an error well below 10 cm. This covers almost the physiologically range of an arm.

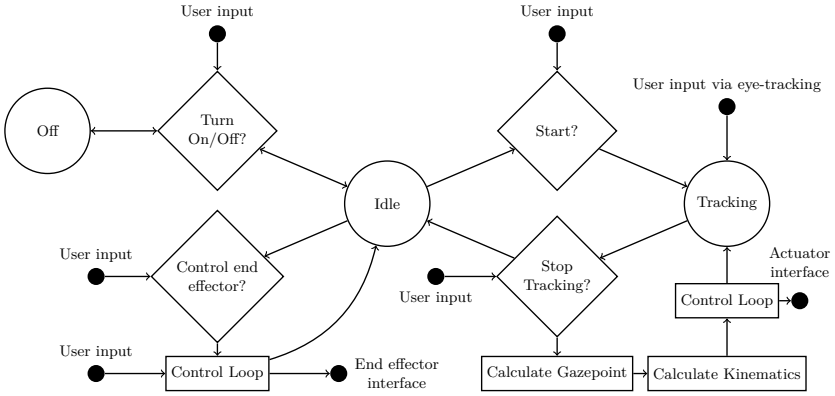


Figure 3: Process analysis

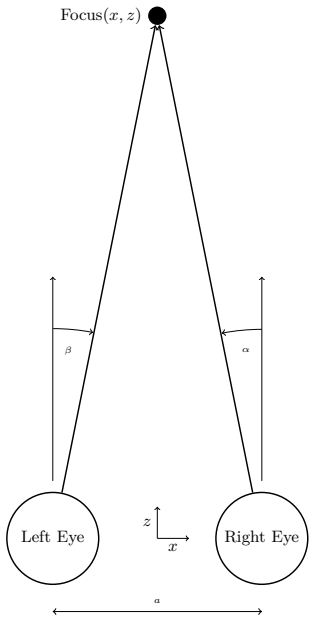


Figure 4: Eye-tracking measurement

A. Subfunctions

1) Eye-tracking input: In order to decide how many degrees of freedom can be controlled, the expected measurement errors of an eye-tracking system are analysed closer.

a) Two DOF: Assuming the distance of the focussed object is known, the measurement of two degrees of freedom can be done by observing only one eye. Let x_r be the x coordinate of the focus point, calculated by the medial/lateral angle α (right eye) as shown in Figure 4. One can denote that

$$\begin{aligned} x_r &= a/2 - z \cdot tan(\alpha) \\ x_l &= z \cdot tan(\beta) - a/2, \end{aligned} \tag{1}$$

where a is the distance between both eyes. The y coordinate can be calculated in the same way using the elevation angle. If we allow for a measurement error $\Delta\alpha$ and assume, that $\alpha + \Delta\alpha$ is small so that $tan(x) \approx x$ the error will be

$$\begin{aligned} \Delta x_r &= z \cdot \Delta\alpha \\ \Delta x_l &= z \cdot \Delta\beta. \end{aligned} \tag{2}$$

This shows, that the accuracy decreases proportionally with the distance. Statistically, the error could be reduced by using both measurements and calculating the means. In general, the assumption that has been made does not hold. For large angles, Δx_r depends non-linearly on $\Delta\alpha$ and causes even larger errors. Nevertheless, (2) gives a rough approximation of how the accuracy of a two DOF eye-tracking system will be. Using 1 one can calculate, that the expected error stays below 3 cm for the whole area, defined as requirement, if we assume having an measurement accuracy of 1°.

b) Three DOF: If both angles α and β are observed, theoretically the distance of the focus point can be calculated. The problem simplifies if the measured lines of fixation intersect in space. In general this is not the case. Assuming that they do intersect, we can solve the planar problem shown in Fig. 4. The measured coordinates are

$$z = \frac{a}{tan(\alpha) + tan(\beta)} \tag{3}$$

and

$$x = a \cdot \left(\frac{1}{2} - \frac{tan(\alpha)}{tan(\alpha) + tan(\beta)} \right). \tag{4}$$

From (3) the measurement error of the distance can be calculated if we allow for errors in angle measurement. One can observe

$$\Delta z = a \left(\frac{1}{tan(\alpha) + tan(\beta)} - \frac{1}{tan(\alpha + \Delta\alpha) + tan(\beta + \Delta\beta)} \right). \tag{5}$$

To analyze further how the accuracy of the eye angle measurement translates into the accuracy of distance calculation, we assume an object moving on the z-axis, should be tracked. Therefore, $\alpha = \beta$. Furthermore, we assume that the accuracies are the same for both angles, thus $\Delta\alpha = \Delta\beta$. This assumption is reasonable because the same system should be used to detect both angles. With these simplifications (5) reduces to

$$\Delta z = a \left(\frac{1}{2 \cdot tan(\alpha)} - \frac{1}{2 \cdot tan(\alpha + \Delta\alpha)} \right). \tag{6}$$

In Fig. 5 this function is evaluated for different accuracies and different angles. One can observe that for large angles, corresponding to close distances, the accuracy is relatively high. Below 10°, which is approximately $z = 0.3\ m$, the quality of measurement starts to drop rapidly. Even with an angle measurement quality of $\Delta\alpha = 1°$, Δz will be roughly 0.5 m for a distance $z = 1\ m$, which is a relative error of 50 %.

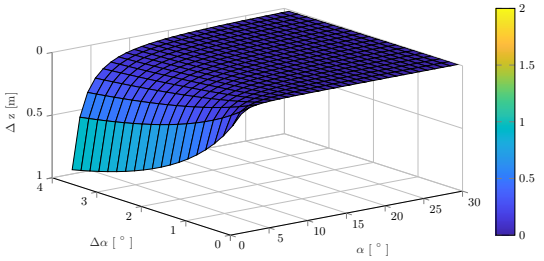


Figure 5: Accuracy in distance measurement by eye-tracking

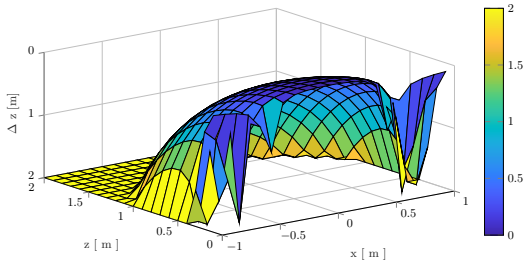


Figure 6: Distance measurement accuracy for $\Delta\alpha = \pm 1°$ in the field of view

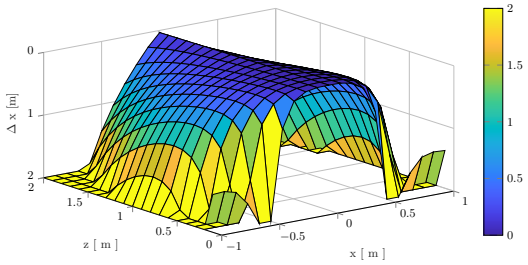


Figure 7: Horizontal measurement accuracy for $\Delta\alpha = \pm 1°$ in the field of view

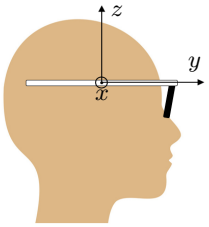


Figure 8: Coordinate system of the sensor

The expected accuracy of an eye-tracking system in the field of view is shown in Fig. 6. Within close proximity to the eyes, the measurement is relatively accurate. For $z = 30\ cm$, $\Delta z = 0.03\ m$. In one meter distance, the accuracy has already dropped to $\Delta z = 0.53\ m$, which is unacceptable. Laterally the divergence between true and measured value will be even worse. In general one could state, that distance measurement with eye-tracking is not suitable for precise control tasks, requiring an area of operation, which is larger than 40 cm x 40 cm.

However, not only the distance measurement is a matter of concern. The lateral accuracy should be investigated as well. Therefore Eq.4 needs to be evaluated with diverging angles and subtracted from the true measurement. This leads to the expression

$$\Delta x = a \cdot \left(\frac{tan(\alpha + \Delta\alpha)}{tan(\alpha + \Delta\alpha) + tan(\beta + \Delta\beta)} - \frac{tan(\alpha)}{tan(\alpha) + tan(\beta)} \right). \tag{7}$$

Fig.7 displays how the accuracy is distributed. Overall it should be noticed, that the area with high accuracy is larger and extended in the z-direction, compared to the highly accurate areas of the distance measurement. Still, the lateral roll-off in accuracy is progressively steeper for higher distances. While Δx at $x = 0.3$ is still below 10 cm for small distances $z < 0.6\ m$, the accuracy drops to $\Delta x \approx 0.19\ m$ for a distance of $z = 1\ m$.

This result reveals the weakness of eye-tracking for both, large distances and laterally deviating focus points. The requirements can not be fulfilled by using binocular eye-tracking and calculating the distance this way. Furthermore this shows, that accurate eye-angle measurement is essential in order to determine the exact position of the focus point. An alternative control of the z-direction could improve the accuracy of a system using eye-tracking. Therefore, we propose a 2D tracking, which is possible by observing one eye.

2) Receiving input for the third DOF: Since the process analysis revealed, that not only the z-direction, but also switching modes and controlling the end effector demands user input, the candidate should provide as many DOF as possible. Furthermore, it should be easy to implement the device additionally to eye-tracking glasses. Both these requirements are leading to the use of head movement, since it provides at least three rotations. Also, the muscles causing the heads rotational movement are mainly innervated by nerves originating above C3 which increases the chances, that persons suffering from a cross sectional trauma can still use the device.

The movement can easily be detected by acceleration sensors, which can be integrated into eye-tracking glasses. Microelectromechanical systems (MEMS), like the acceleration sensor used in this case, offer fast and precise measurements. Additionally the sensors are small enough to be integrated into glasses. The experiments were performed, using the MPU-6050, which is a cheap capacitive MEMS. The sensor provides an *Inter-Integrated Circuit* (I2C) Bus interface. As a master device, an Arduino Nano is used. This way three axes of translational acceleration and three axes of angular acceleration data can be measured and transferred via USB-Serial interface in 84 ms on average. The sensor is attached to the glasses with its x-axis pointing to the right, its y-axis pointing forward and its z-axes upwards as shown in Fig. 8. Different types of movement, that represent possible candidates for controlling the device are measured and analyzed. Possible candidates are rotations in the sagittal plane, causing a pitching movement. Both, fast and slow movements are investigated. The fast movement is furthermore split up into an initial flexion, which is a nodding move, leading to a negative rotation around the the x-axis of the sensor and an initial extension which is causing a positive rotation around the same axis. The last candidate analyzed is lateral flexion which is a rotation of the head in the frontal plane. This movement causes the sensor to rotate mainly in its y-axis. To qualify as a proper user input, the movements must be distinguishable.

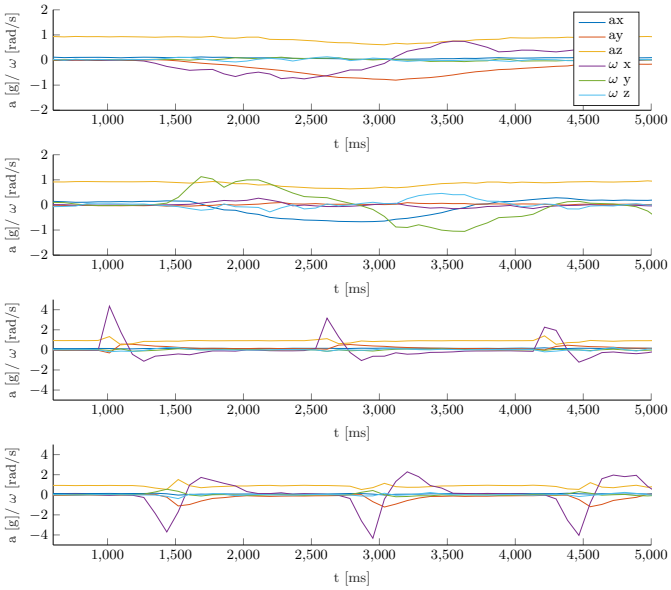


Figure 9: Candidates for control movements (From top to bottom: slow pitching, slow lateral flexion, three fast pitching moves with initial extension, three fast pitching moves with initial flexion)

Fig. 9 shows the six sensor outputs for each candidate. The acceleration data is expressed in g, since the slow movements are expected to represent only the gravitational acceleration. For reasons of better visualization, the ordinate is scaled differently for the fast movements. As expected, in steady-state the reading for acceleration in z-direction is almost one, while all other variables equal approximately zero. Due to the slight misalignment of the sensor or a tilted resting position of the users head, the reading does not exactly equal 1 g in z-direction. This issue should be taken care of when designing a calibration algorithm.

The data reveals some significant characteristics for each movement, that could be used to detect them. Slow pitching movement can be easily sensed by observing the acceleration in z- and y-direction. We can continuously measure the pitch angle, which allows for controlling a continuous DOF. The same measurement can be done for the lateral flexion movement. From the acceleration in z and x- direction a roll angle can be calculated and used for controlling another DOF freedom. The fast pitching movements can be identified by observing the angular velocity in x-direction. It can be observed, that both inputs can be triggered by a threshold, since $|\omega_x| \leq 1.5\ s^{-1}$ for all slow movements. The difference between the fast movements is the direction of the initial deflection of ω_x. The angular velocity, when turning the head back to resting position, sometimes exceeds reasonable thresholds as well. To avoid triggering the opposite control signal, a time of approximately 600 ms after exceeding the threshold should be ignored. The data was collected from a healthy person without limited movement possibilities. A data-set of actually paralyzed persons may vary in terms of intensity of each movement. Therefore a mature system shall not have the thresholds hard coded but calculated from data gathered in a calibration process.

Each decision shown in the process analysis in Fig. 3 can be controlled by the fast pitching movement. In idle mode the system could output to the user whether it is going to start tracking, control the end effector or turn off. By fast pitching in one direction the user could switch to the next option. By fast pitching in the opposite direction the user could invoke starting and stop the selected option. This way other user modes could be added later.

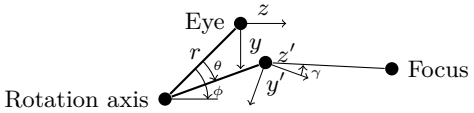


Figure 10: New mapping

Slow pitching movements are suitable for providing a third DOF. If we assume only head rotations in the sagittal plane, the mapping of eye-angles to world coordinates is only slightly more complicated. The axis in which a person rotates its head around when nodding runs transversally between external auditory meatus and mastoid process [9]. Therefore, the distance in the sagittal plane between the center of the eyeball and the axis of rotation can be measured with an error of less than 1 cm. Fig. 10 shows the new mapping problem. The angles γ and θ are measured. Since the calibration is done with the original coordinate system, we are interested in the y coordinate of the gaze point. In order to solve this, the angle ϕ has to be known which can be approximated from the anatomical position of the rotation axis. From [9] we measure a value of $\phi \approx 28°$. In a calibration process this angle could be forced by fixating the head for its duration. For now on we assume this angle as given. It can be shown, that the eye moved in the z-direction for

$$\Delta z = r \cdot (\cos(\phi - \theta) - cos(\theta)) \tag{8}$$

and in y-direction for

$$\Delta y = -r \cdot (sin(\theta) - \sin(\phi - \theta))\,. \tag{9}$$

It is straight forward to calculate the coordinates of the gaze point to be

$$y_{focus} = \tan(\gamma - \phi) \cdot (z - \Delta z) + \Delta y. \tag{10}$$

The calculation 1 for the x-coordinate simply changes to

$$x_{focus} = a/2 - (z - \Delta Z) \cdot \tan(\alpha). \tag{11}$$

3) Attachment to the carrier: The attachment can be done directly by using the body as a base for the device, as shown in Fig. 11a. Alternatively, the device could be attached to the wheelchair of a person (Fig. 11c). Fig. 11b shows, that assistant devices combining both properties exist and could represent a possible solution. In order to decide which attachment should be used, each approach should be analyzed in terms of usability for this project. The body-based device has an obvious advantage for persons without wheelchair, since this is the only possible solution, that is meeting the requirement of mobility in this case. However, the condition of two paralyzed arms and fully functional legs is extremely rare. On the other hand, the designing effort for the body-based system will be significantly greater. The design needs to mimic the kinematics of the physiological arm precisely in order to ensure that no harm is done to the wearer. Furthermore, the attachment to the body could leave bruises, especially if the device is not built as lightweight as possible. Also, the attachment to the

(a) Body based device

(b) Wheelchair and arm attachment

(c) Wheelchair based system

Figure 11: Different attachment approaches

person wearing the device, is more complicated compared to a wheelchair based approach and possibly requires redesigning for different users. In terms of safety, the wheelchair based system must primarily prevent the event of a collision with the person which is obviously an easier task than causing no harm to the person while moving its arm. The interface to the wheelchair is comparably easy to design. Even movements outside of the space a human arm could reach are possible, as long as accurate enough eye-tracking data is supplied to the controls. For gripping tasks advanced robotic end effectors already exist which could be used in case of a wheelchair based solution. Therefore a wheelchair based solution offers the most benefit.

4) Software: The eye-tracking shall be done by *Pupil capture* [1] which is using the algorithm presented in section II. The program can be interfaced with a *msgpack* based API. Both eye-angles are being published by pupil capture. The system shall be implemented using the *Robot Operating System* (ROS) framework. The benefits of using ROS are easy and flexible interfacing between different parts of the program which makes building a prototype easier. New parts of the application can be integrated quickly. ROS offers simulation tools like RVIZ which is been used to test the first version of the system.

Fig. 12 shows the architecture of the proposed system. Systems using the ROS framework consist of nodes that run independently. The ROS Master node acts as a message broker that allows communication between the nodes. The eye-tracking data is acquired by the gaze point publisher that subscribes to the eye angles from pupil capture. The acceleration data processing node is reading the serial port, the *Arduino Nano* is writing to. It is receiving all three accelerations and three angular velocities. The acceleration data processing node is publishing the heads pitch angle, as well as the roll angle. Furthermore, it is publishing the raw reading of ω_x to the control signal detector node which is searching for fast movements of the head. The gaze point publisher subscribes to the heads pitch angle and calculates the three-dimensional gaze point. The state machine acts as a control unit of the process. When the corresponding control signal is being detected, it forwards the gaze point coordinates or the end-effector set point. The control nodes interface with the hardware. The node RViz is a tool for visualizing the system which can be helpful for debugging or testing the system before using it on physical hardware. This system architecture offers high flexibility for future work since it is easy to integrate new nodes in the data pipeline.

IV. Evaluation

For evaluation purposes, each subfunction is implemented individually to verify a good functionality.

Table I: DH parameters

Joint	θ_i	d_i	a_i	α_i
1	θ_1	s	0	90°
2	θ_2	l_u	0	−90°
3	θ_3	0	l_f	0°

A. Eye-tracking

In order to compare a physical representation of a robotic device with the simulated one, based on eye-tracking, the task for evaluation should be to track and display the movement of a functional human arm in a simulation. The focus should be on whether the system is fast and robust. Furthermore, it should be observed, if it is possible, to guide the arm in the simulation to a point in space without moving the physical arm first. This ability is an essential requirement for the success of the project. The system is implemented in ROS with two degrees of freedom from eye-tracking. The gaze point publisher node is assuming fixed values for the distance z.

Fig.14 shows the model used, that represents a simplified right arm, using the Denavit-Hartenberg (DH) convention. The shoulder, which is a ball joint, is modeled by two rotary joints. The connecting body has neither length nor mass. The elbow is modeled by a hinge joint. It is connected to the shoulder by the upper arm with the length l_u. The origin of the world coordinate system is placed cranial to the center of the body so that the coordinate system of the shoulder joint is shifted only in the z-direction. The origin of the third coordinate system is placed at the hands position. Using this model, the kinematics can be calculated easily. Table I shows the 12 parameters, that describe the model completely. θ_1 and θ_2 are the variable joint position of the shoulder joints 1 and 2 respectively. Both angles are zero in Fig.14. Positive angles of θ_1 correspond to anteversion, negative angles to retroversion. A medial rotation equals negative angles of θ_2, also lateral rotation is expressed by positive angles. θ_3 is the angle of the elbow. Positive angles correspond to flexion, negative angles to extension.

For calculating the inverse kinematics, a geometric approach is used. The first thing to notice is that the distance of the hand, relative to the shoulder only depends on θ_3. Therefore one can calculate

$$\theta_3 = 90^\circ - \arccos\left(\frac{\left\|\mathbf{p}^{(0)} - \mathbf{s}^{(0)}\right\|^2 - l_u^2 - l_f^2}{-2 \cdot l_u \cdot l_f}\right), \qquad (12)$$

where $\mathbf{p}^{(0)}$ denotes the point which the hand should reach expressed in the world coordinate system. $\mathbf{s}^{(0)}$ represents the position of the shoulder. By subtracting the inverse cosine from

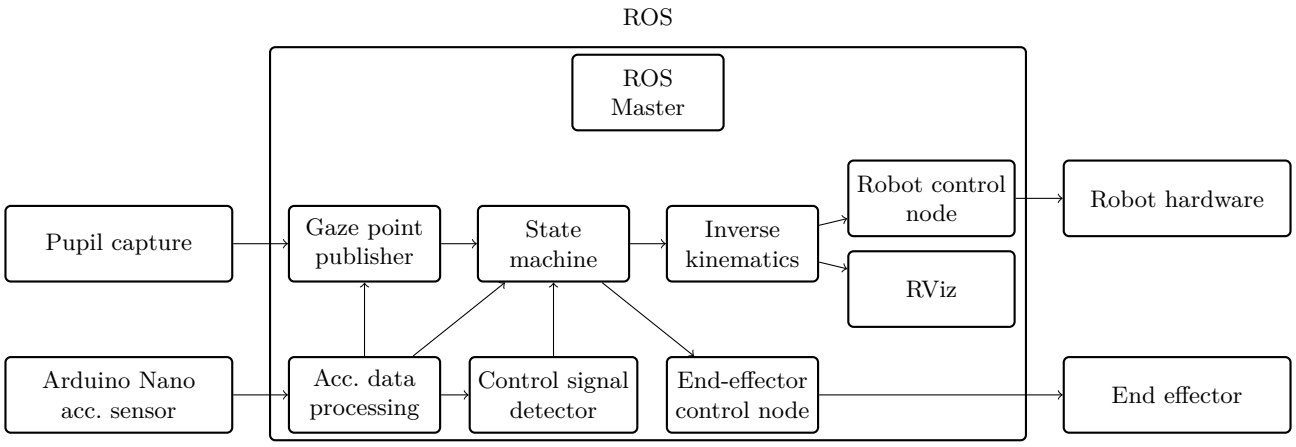


Figure 12: Service oriented architecture

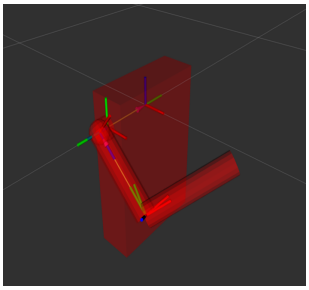


Figure 13: Visualization

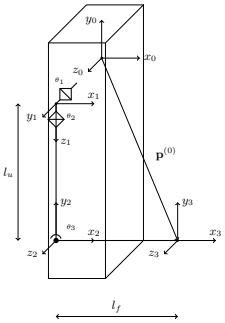


Figure 14: Kinematic model

90°, the range of values will be $-90° < \theta_3 < 90°$, which is anatomically reasonable. This way the solution is unique. Using the forward kinematics it can be shown that

$$p_z^{(0)} = s + l_f \cdot \cos(\theta_3) \cdot \sin(\theta_2). \tag{13}$$

This way θ_2 can be calculated as

$$\theta_2 = \arcsin\left(\frac{p_z^{(0)} - s}{l_f \cdot \cos(\theta_3)}\right). \tag{14}$$

Now only θ_1 remains unknown. One can calculate $\mathbf{p}_0^{(0)}$ which denotes the hands position if $\theta_1 = 0$. After that calculation one can state, that the angle θ_1 is the difference between the angle of $\mathbf{p}^{(0)}$ and $\mathbf{p}_0^{(0)}$, being projected onto the x_0, y_0 plane. Therefore,

$$\theta_1 = \text{arctan2}\left(p_x^{(0)}, p_y^{(0)}\right) - \text{arctan2}\left(p_{x0}^{(0)}, p_{y0}^{(0)}\right). \tag{15}$$

By using the function arctan2 it is not necessary to take care of the correct quadrant. After calculating the joint coordinates, they are passed to RViz in order to visualize the movement (Fig. 13).

For calibration, the users head was placed on a chin rest and a point directly in front of the left eye should be focussed. The measured angles were taken as zero. In order to achieve the accuracy as stated in [1], where the accuracy is already proven, one should calibrate by using at least three points. In this case only the tracking behavior should be analyzed specifically for the proposed use case, therefore the easier calibration process is reasonable. The tests were performed at a constant distance of $z = 0.35\ m$.

For a qualitative observation, the user was asked to focus his hand and move it in a constant distance. The representation of the physical arm in the simulation showed the actual movements in good approximation. It could be noticed, that a greater divergence from the model to the physical arm appeared when it was moved further laterally. This is in line with expectations, as the tracking error increased and the user utilized DOF in his shoulder which were not modeled in the visualization. When passing raw eye-tracking data to the visualization in steady-state, some rapid movements due to eye-movements or imprecise measurement can be observed.

In order to determine, whether the system can be used to control, instead of just emulating physical movement, further tests have been carried out. Each compared the difference between eye-movements which follow an object and eye-movements that try to follow the same trajectory, without anything to focus. The user was asked to guide his gaze from one point to another twice. In only one case there was an object to follow. Fig. 16 shows the residuals after subtracting the position of the trajectory from the measured position. The detected gaze position stays within 2 cm from the trajectory for both cases. The standard deviation with an object to focus is 0.58 cm while being 0.77 cm for the tracking without an object. Both are acceptable values and one could not observe a significant difference.

Furthermore, more complex trajectories are analyzed. The tested shapes were a triangle with visible edges, a triangle with only vertices, and a circle. For all geometries, a slight offset is

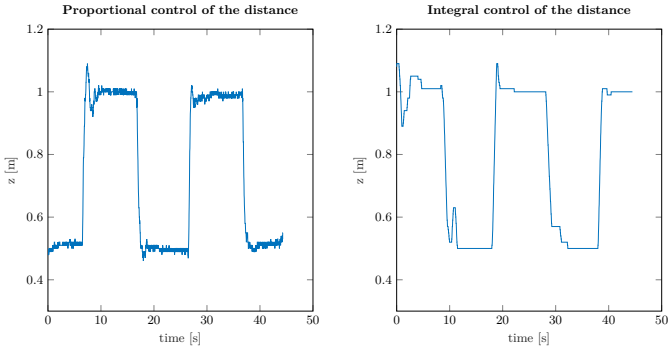


Figure 15: Distance control

visible which is due to the described calibration process. It is assumed to be removable by using a robust calibration algorithm. For each test, the calculated gaze trajectory is displayed with the true shape, as well as both γ and β, which are elevation and lateral angle respectively. Outliers have been filtered by using a median filter with the order of three. Additionally, a moving average filter with a window size of three was applied to remove measurement noise, caused by the pupil detection algorithm. Fig. 17 shows a slightly less smooth signals in the data-set, where no object moved along the edges of the triangle. The eye was unable to move along the edges of the triangle in a clean fixation. This observation does not hold in general for every trajectory with no object. Fig. 18 shows relatively precise movement of the eye with no object to focus. The data-set with an object to focus seems comparatively imprecise which is also caused by a blink at 19 s. This reveals the need for blink detection that was not mentioned yet. In Fig. 19 a circular shape is investigated, revealing a similar result for both cases. However, speed seems to have an impact on the quality of data that can be generated from eye-tracking. Fig. 20 shows how a slight increased speed leads to a worse ability to follow the trajectory smoothly.

B. Distance control

For evaluating the distance control, two types of strategies were evaluated. The first strategy is calculating the distance by a proportional gain from the pitch angle θ. The second strategy is feeding the angle to an integrator to calculate the desired control distance. The proportional gain was calculated in a way, that within $-30° < \theta < 30°$ a reasonable range of $0\ m < z < 1\ m$ could be covered. The integrator input was chosen to be only angles with $|\theta| > 5°$ in order to ease maintaining a steady-state value. With both types of control, a user was instructed to hold z values of 1 m and 0.5 m repeatedly. Every 10 s a sound signal indicated the time of a reference change. Fig. 15 shows the result of both strategies. The proportional design shows a faster response as expected. In this particular case a time constant of $\tau = 0.32\ s$ was measured for the proportional control, while the integral approach had an average of $\tau = 0.68\ s$. With the proportional design on the other hand, it is not possible to maintain a steady state precisely. The data shows fluctuation around the steady state value with an amplitude of $\pm 1\ cm$, while the integral approach does not vary at all after reaching the steady state, since angles smaller than 5° are ignored. The integral control tends to swing slightly more than the proportional approach. Nevertheless, the fact that integral control allows for an upright head position while being in steady state, makes it superior to the proportional control as a complement to eye-tracking. Also, it seems to be easier to avoid overshooting, as long as the steady state is approached slower.

C. Mode switching signal detector

In order to test the switch signal detector, the proposed algorithm is implemented. The algorithm is detecting whether the angular velocity ω_x is exceeding a threshold of 1.5 s^{-1}. It reports the kind of movement, depending on the sign of the spike and can, therefore, distinguish between two types of switching movement. After an event 10 samples are ignored to avoid triggering the opposite signal when the head turns back. This way the test showed a reliable detection with 50 out of 50 proper classifications. It was not triggered when moving the head slowly. To prevent fast movements from affecting the accuracy of the eye-tracking and z-direction control, a buffer should be implemented, so that after a rapid movement appears to stop tracking, the coordinate can be taken from before the event.

V. Conclusions and Future Work

Using eye-tracking for assistance robots could lead to large improvements in terms of flexibility for paralyzed persons. Two DOF can be controlled by an eye-tracking system accurately and precisely provided that a calibration procedure is implemented with the system which should be part of the future work on this project. The data shows no significant difference in accuracy or precision whether the eye controls or just follows an object. A third DOF must be controlled differently because the expected errors of an eye tracking system are unacceptably large. Furthermore, it is necessary to study the differences between saccades and fixations more deeply since an algorithm needs to be designed that tells the difference in a robust way, so that tracking can be stopped in case the user scans his surrounding without any further input. Moreover, the data collected shows that the raw data needs to be filtered. Future work needs to be done to raise the quality of input to the robotic device.

As a complementing technology, an inertial measurement unit suits the requirements. It is easy to integrate into eye-tracking glasses and offers the ability to use the device for every paralyzed person with cross-sectional lesions below C4. Head movements are proven to be able to control a state machine changing modes reliably, as well as accurate control of another DOF can be done. As control strategy, an integral type should be used, instead of a proportional for this DOF because it offers the ability to keep the head in the initial position in steady state.

The framework ROS is proposed because it offers great flexibility to extend the project continuously. The structure and good visualization options make it easy to develop and debug the system. The future work shall implement the state machine and integrate each subfunction shown in this paper. In particular, the new mapping, taking pitching angles into account, needs to be validated. As a next step, the system should be tested with robotic hardware to validate the performance of the overall system.

A major challenge for future works will be implementing safety features to ensure that no harm is done to the user or others. Since, paralyzed persons are the target group for the device, approaches to this problem might be different than the usual robotic safety precautions.

References

[1] M. Kassner, W. Patera, and A. Bulling, "Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction." [Online]. Available: http://arxiv.org/pdf/1405.0006v1 [Accessed: 11.05.2020]

[2] A. T. Duchowski, *Eye tracking methodology: Theory and practice*, 2nd ed. London: Springer, 2007. [Online]. Available: https://www.springer.com/de/book/9781846286087 [Accessed: 07.05.2020]

[3] Richard Bates, "A computer input device selection methodology for users with high-level spinal cord injuries," 2002.

[4] "Types of paralysis," 2019. [Online]. Available: https://www.webmd.com/brain/paralysis-types [Accessed: 10.02.2020]

[5] "Lateralsklerose, primäre," 2004. [Online]. Available: https://www.orpha.net/consor/cgi-bin/OC_Exp.php?lng=DE&Expert=35689 [Accessed: 11.02.2020]

[6] Ingrid E. C. Verhaart, "Prevalence, incidence and carrier frequency of 5q–linked spinal muscular atrophy: a literature review," 2017. [Online]. Available: https://ojrd.biomedcentral.com/track/pdf/10.1186/s13023-017-0671-8 [Accessed: 11.02.2020]

[7] G. Logroscino, B. J. Traynor, O. Hardiman, A. Chiò, D. Mitchell, R. J. Swingler, A. Millul, E. Benn, and E. Beghi, "Incidence of amyotrophic lateral sclerosis in europe," *Journal of neurology, neurosurgery, and psychiatry*, vol. 81, no. 4, pp. 385–390, 2010.

[8] K. Röhl, "Halswirbelsäulenverletzungen mit tetraplegie," *Trauma und Berufskrankheit*, vol. 5, no. 2, pp. 231–243, 2003.

[9] G. Aumüller, *Anatomie*, 3rd ed., ser. Duale Reihe. Stuttgart: Thieme, 2014.

[10] Torsten Felzer, "Alternative wheelchair control," 2007.

[11] S.-H. Chen, Y.-L. Chen, Y.-H. Chiou, J.-C. Tsai, and T.-S. Kuo, "Head-controlled device with m3s-based for people with disabilities," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439)*. IEEE, 17-21 Sept. 2003, pp. 1587–1589.

[12] L. Lucas, M. DiCicco, and Y. Matsuoka, "An emg-controlled hand exoskeleton for natural pinching," *Journal of Robotics and Mechatronics*, vol. 16, pp. 482–488, 2004.

[13] P. M. Rossini, S. Micera, A. Benvenuto, J. Carpaneto, G. Cavallo, L. Citi, C. Cipriani, L. Denaro, V. Denaro, G. Di Pino *et al.*, "Double nerve intraneural interface implant on a human amputee for robotic hand control," *Clinical neurophysiology*, vol. 121, no. 5, pp. 777–783, 2010.

[14] H. Park, B. Gosselin, M. Kiani, H.-M. Lee, J. Kim, X. Huo, and M. Ghovanloo, "A wireless magnetoresistive sensing system for an intra-oral tongue-computer interface," in *IEEE International Solid-State Circuits Conference digest of technical papers (ISSCC), 2012: 19 - 23 Feb. 2012, San Francisco, CA, USA ; volume 55*, L. Fujino, Ed. Piscataway, NJ: IEEE, 2012, pp. 124–126.

[15] C. Backhaus, "Methodisches vorgehen in der praxis," vol. 3, pp. 65–79.

Attachments

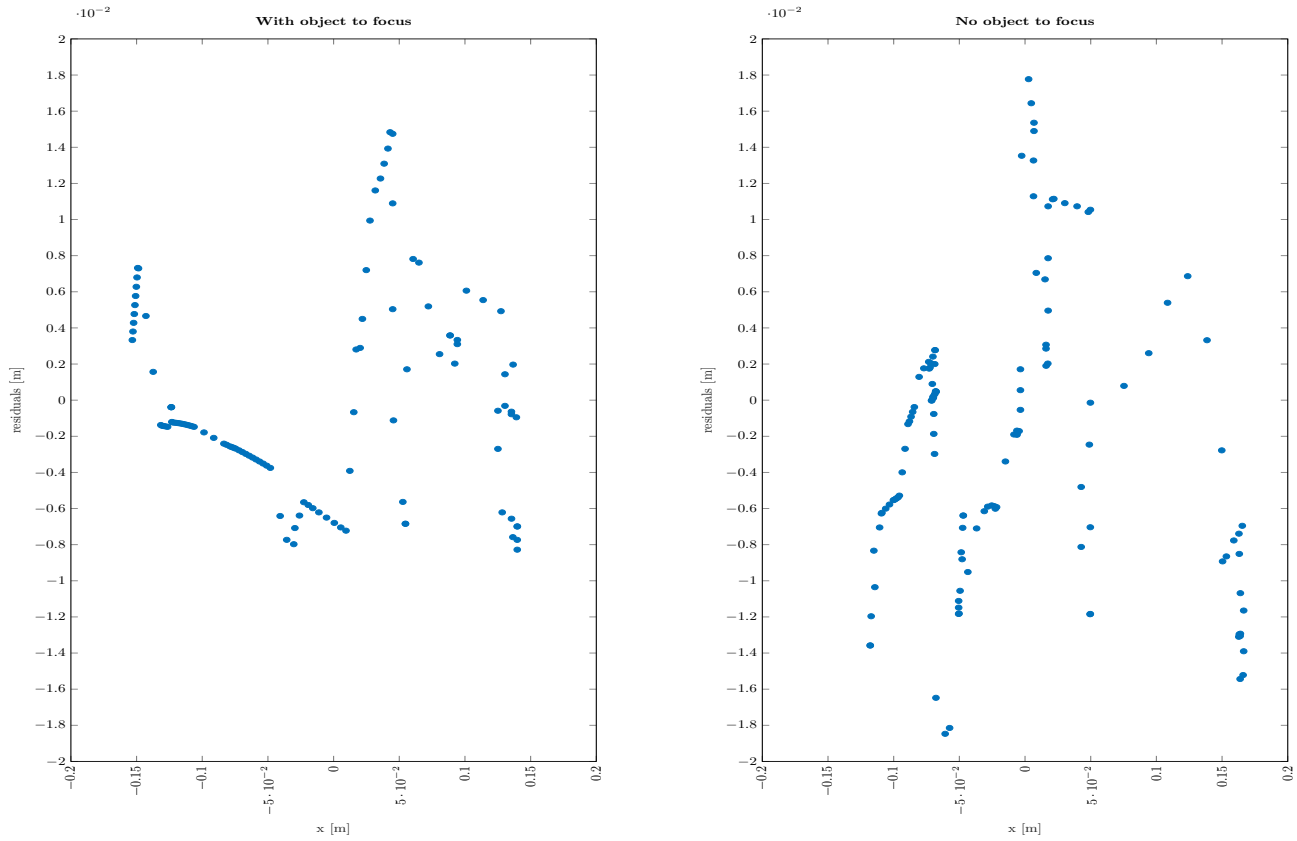


Figure 16: Eye-tracking measurements with and without object to focus (line)

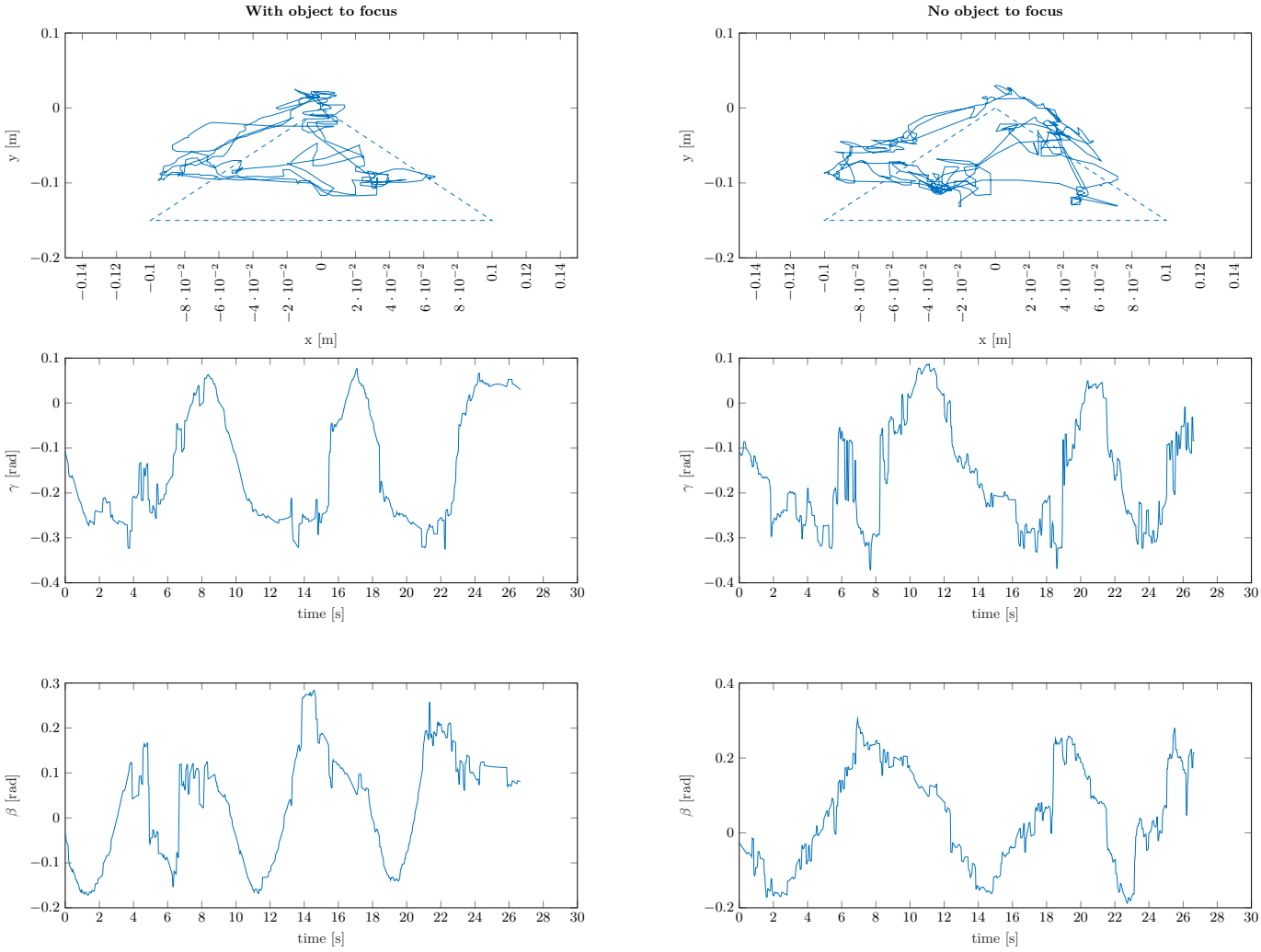


Figure 17: Eye-tracking measurements with and without object to focus (triangle)

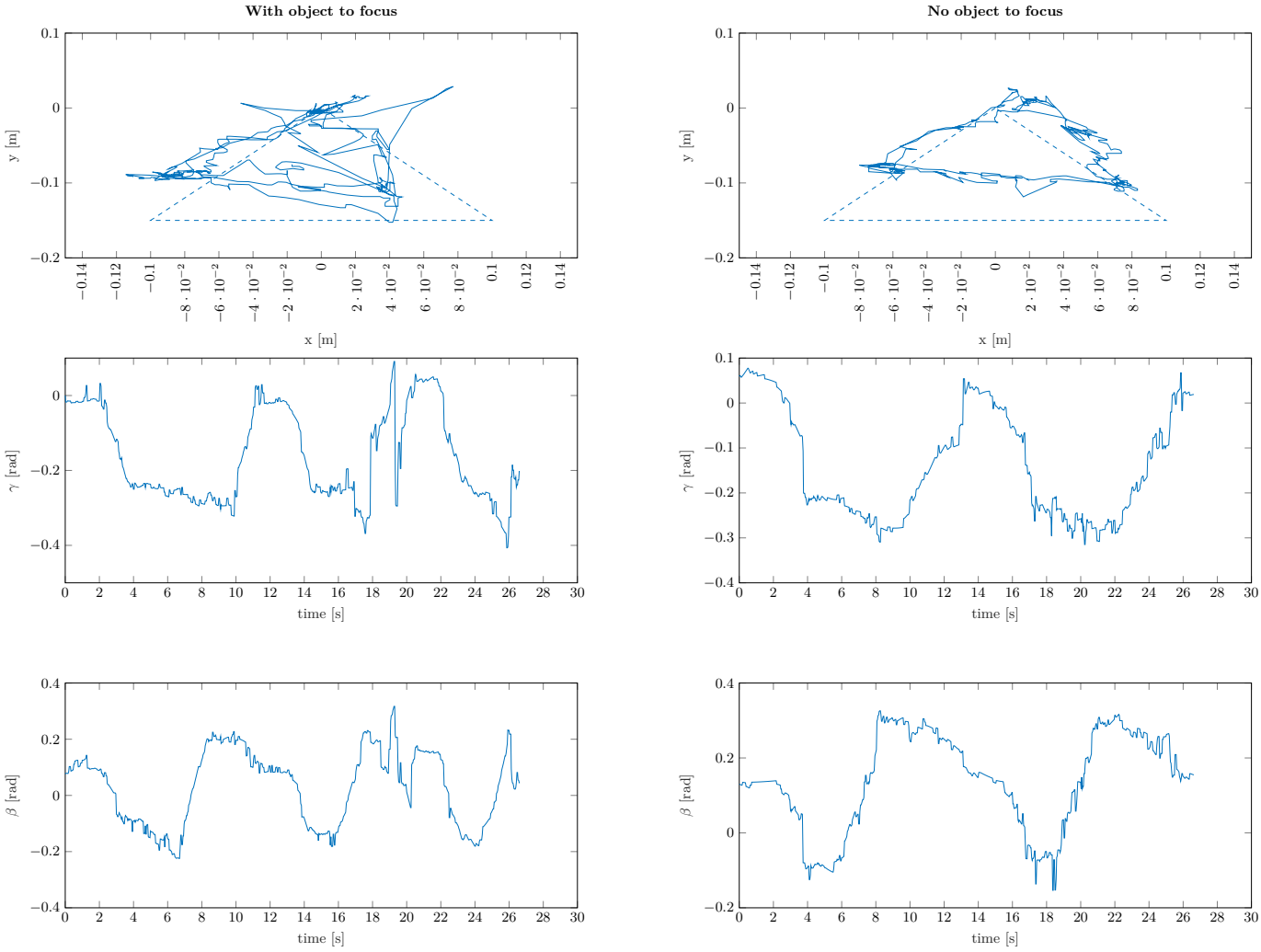


Figure 18: Eye-tracking measurements with and without object to focus (triangle with only corner points)

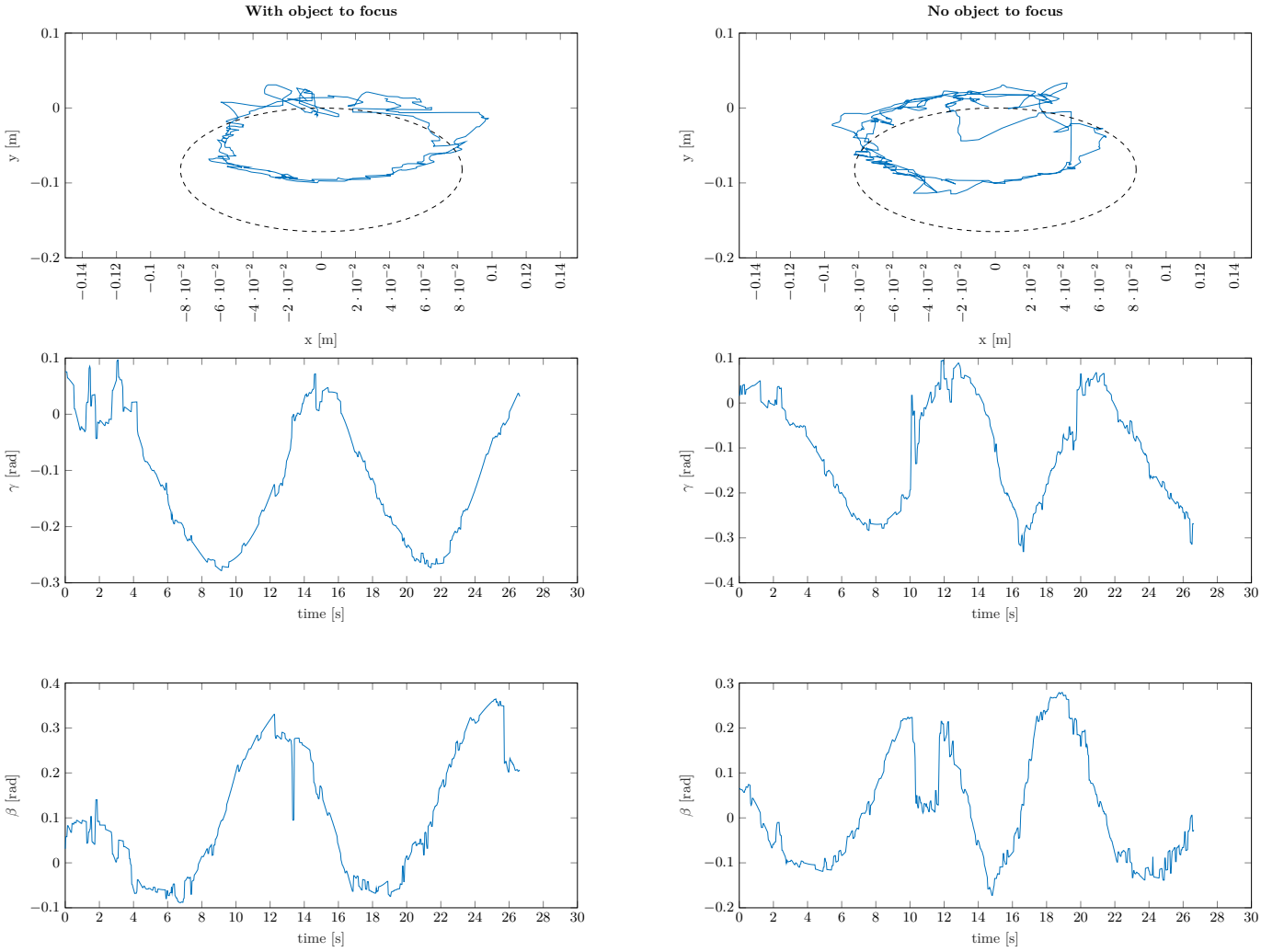


Figure 19: Eye-tracking measurements with and without object to focus (circle)

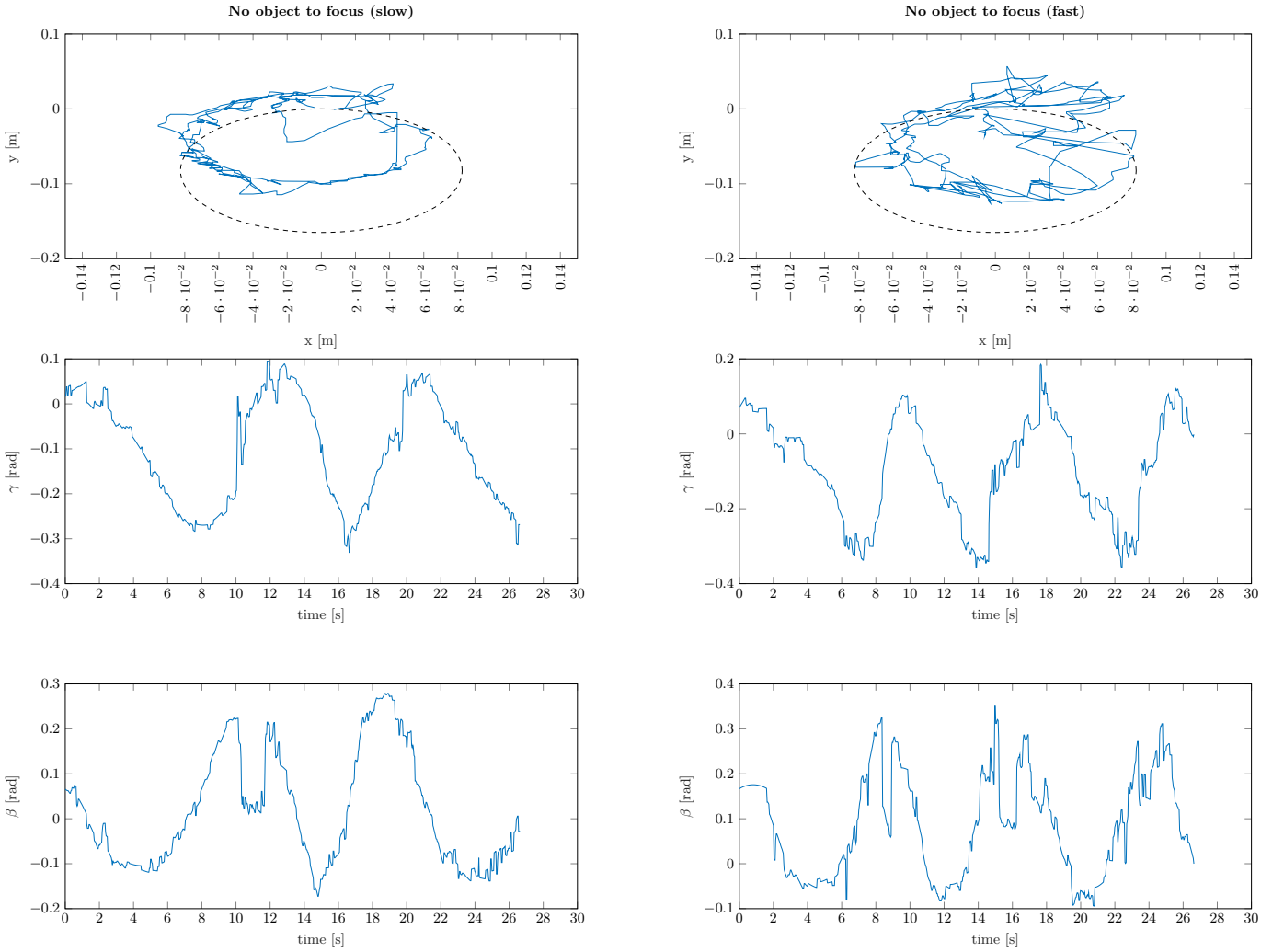


Figure 20: Eye-tracking measurements with different speeds (circle)