# Design, Control, and Evaluation of a Human-Inspired Robotic Eye

**Simon Schulz**

# Design, Control, and Evaluation of a Human-Inspired Robotic Eye

**Simon Schulz**

June 2020

A doctoral thesis presented for the degree of
Doctor of Engineering (Dr.-Ing.) at

Bielefeld University
Cognitive Interaction Technology - Center of Excellence
Inspiration 1
33619 Bielefeld
Germany

### Reviewers

Dr.-Ing. habil. Sven Wachsmuth
Prof. Dr. rer. nat. Axel Schneider

### Examination Board

Prof. Dr.-Ing. Ulrich Rückert
Dr.-Ing. Sebastian Zehe

Printed on permanent paper as per ISO 9706.

# Declaration of Authorship

According to the Bielefeld University's doctoral degree regulations §8(1)g: I hereby declare to acknowledge the current doctoral degree regulations of the Faculty of Technology at Bielefeld University. Furthermore, I certify that this thesis has been composed by me and is based on my own work, unless stated otherwise. Third parties have neither directly nor indirectly received any monetary advantages in relation to mediation advises or activities regarding the content of this thesis. Also, no other person's work has been used without due acknowledgment. All references and verbatim extracts have been quoted, and all sources of information, including graphs and data sets, have been specifically acknowledged. This thesis or parts of it have neither been submitted for any other degree at this university nor elsewhere.

Simon Schulz

Place, Date

# Abstract

The field of human-robot interaction deals with robotic systems that involve humans and robots closely interacting with each other. With these systems getting more complex, users can be easily overburdened by the operation and can fail to infer the internal state of the system or its ”intentions”. A social robot, replicating the human eye region with its familiar features and movement patterns, that are the result of years of evolution, can counter this. However, the replication of these patterns requires hard- and software that is able to compete with the human characteristics and performance. Comparing previous systems found in literature with the human capabilities reveal a mismatch in this regard. Even though individual systems solve single aspects, the successful combination into a complete system remains an open challenge. In contrast to previous work, this thesis targets to close this gap by viewing the system as a whole — optimizing the hard- and software, while focusing on the replication of the human model right from the beginning. This work ultimately provides a set of interlocking building blocks that, taken together, form a complete end-to-end solution for the design, control, and evaluation of a human-inspired robotic eye. Based on the study of the human eye, the key driving factors are identified as the successful combination of aesthetic appeal, sensory capabilities, performance, and functionality. Two hardware prototypes, each based on a different actuation scheme, have been developed in this context. Furthermore, both hardware prototypes are evaluated against each other, a previous prototype, and the human by comparing objective numbers obtained by real-world measurements of the real hardware. In addition, a human-inspired and model-driven control framework is developed out, again, following the predefined criteria and requirements. The quality and human-likeness of the motion, generated by this model, is evaluated by means of a user study. This framework not only allows the replication of human-like motion on the specific eye prototype presented in this thesis, but also promotes the porting and adaption to less equipped humanoid robotic heads. Unlike previous systems found in literature, the presented approach provides a scaling and limiting function that allows intuitive adjustments of the control model, which can be used to reduce the requirements set on the target platform. Even though a reduction of the overall velocities and accelerations will result in a slower motion execution, the human characteristics and the overall composition of the interlocked motion patterns remain unchanged.

# Contents

# List of Figures

# List of Tables

# I

# Research Topic and Related Work

This part of the thesis will give an introduction into the research topic, the underlying motivation, and the contributions of this work, focusing on the implementation of the eye region of a humanoid robotic head with human-like capabilities. In order to create a basis for the following considerations, the human eye and all its properties are established and important characteristics are collected. Based on the combination of this theoretical basis and the insights gained during operation along with the evaluation of a previous robot design, a list of key requirements for the next generation design of a humanoid robotic eye is compiled. Subsequently, the eye regions of different related humanoid robotic heads found in literature are evaluated based on these requirements and the most promising basis for the further development is selected. Finally, the reproduction of human motion by artificial systems and different technical possibilities for the implementation are introduced.

Chapter 1

# Introduction

Often regarded as a 20th century discovery, the history of the automaton that we nowadays would probably call robot, dates back a very long time. Even though the term robot itself was devised less than a century ago by the Czechoslovakian brothers Karel and Josef Čapek in 1920 [1], the history of automaton, apparently moving without direct human intervention, dates back to the Ancient Egypt 1300BC [2]. There are reports on early humanoid robots by the Arabic inventor Al Jazari (1138-1206) [2]. Over the following centuries, from the late Middle Ages to modern times, a variety of early humanoid robots have been developed. One example for such an ancient humanoid robot is the very sophisticated chess playing robot, also known as the *mechanical turk*, that was invented in the 18th century. This machine, remote controlled by a hidden operator, gave the illusion of an autonomous machine playing chess [3].

In the 1950s, the russian-born American writer Isaac Asimov was the first to coin the term robotics [4]. In his science fiction writing *Runaround* (published in 1942) he introduced the famous three laws of robotics that should ensure no human being will ever be harmed by a robotic system. At the time of writing, the term robot was connected with manufacturing industry and autonomous machines taking over repetitive tasks from the human. The first industrial robot, called *Unimate*, was invented by Joseph Engelberger and installed in an assembly line in 1961 [5]. This robot was employed to help factory workers with the unpopular and dangerous job of unloading high temperature parts from a die cast machine [1]. Given the success of the application of robotics in the booming automotive industry, the subsequent generations of those machines found their way into the general industry by the late 70s [6]. In the following decades robots were built for a broader set

of applications to assist and replace more workers in dangerous situations. Those robots were more or less industrial manipulators and were placed in an isolated space with no direct human interaction.

Nowadays the term robot is used for much more than those used in industrial automation. The emerging field of service robotics brought robots in closer proximity to people. This poses new challenges as systems, operated in proximity to humans in their homes or workplaces, are expected to be designed to safely and dependably cohabitate with them [6]. In the 21st century robots are becoming more and more visible in everyone's life and people start to accept these autonomous helpers in their home. One example of this kind of machines is the vacuum cleaning robot *Roomba* manufactured by the company *iRobot*. First introduced in 2002 [7], it has now been sold more than 10 million times worldwide [8]. In 2019, the International Federation of Robotics reported a total number of 16.3 million domestic and service robots that had been sold in 2018. This is an increase of 59% in sales in comparison to 2017. Projections estimate a total number of 61.1 million domestic robots being sold in 2022 [9]. This ongoing inclusion of robotic systems into private domains raised the question of what the best modes for the interaction between robots and humans are [10].

## 1.1. Motivation

With technical systems getting more complex, non-expert users can easily become overburdened operating them. It is also known that people tend to anthropomorphize things (e.g. cars or computers) whose internal mechanisms can not be easily understood, no matter how intelligent or autonomous they are in reality [11], [12]. People will often apply a social model to the interaction with such an autonomous robot as well [12]. The emerging field of sociable robotics tries to exploit this anthropomorphization by transferring rules and conventions familiar to people to robotic systems. Humans are social beings that are highly attuned to human characteristics and thus this transfer facilitates the design of very natural and easy to use interfaces for human robot interaction. For example human faces transfer a variety of information through different channels. Starting with static features such as gender, age, or identity over dynamic facial ex-

pressions, emotions, and communicative feedback. The fact, that humans tend to anthropomorphize things and try to interpret and decode, e.g. the robots intentions by looking at it, can be used to boost human robot interaction. A robot, that features a head, which in turn replicates human features, can benefit from these behavior patterns. For example, it has been shown that humans can successfully interpret gaze and facial expressions of such a robot with regard to communicated intentions and inner states without training [13]. Especially the eye region, consisting of the eyeball, eyebrows, and the eyelids, plays a significant role in this regard and it is in many respects one of the most salient and important parts of a human's or humanoid robot's face. Studies by Looser *et al.* have shown that "human beings are highly attuned to specific facial cues, carried largely in the eyes" [14]. Humans are able to deduce the partner's perceived field of view and especially the focus of attention. Furthermore, humans are able to "read" the partner's intentions, emotions, judgments, and potential cognitive states from looking at the eyes and the eye-related motions. In human-human interaction, both partners benefit from this capabilities. It has been shown that this holds for human-robot interaction as well. Predictable or legible motion in collaborative human robot interaction tasks can lead to higher task efficiency and enhance the workers' perception [15]. These facts show that a humanoid robot, used in an interaction scenario, can benefit from the replication of human features and especially its eye region.

In addition, just like humans use their eyes to perceive their surroundings, a social robot, meant for the direct interaction with a human, will have to sense the environment and specifically its interaction partner as well. This basic requirement can be addressed in two different ways, either using an external camera system or an internal one. Using an external camera is often problematic as its view can be covered or might show the wrong part of the scene. An internal camera, ideally actively orientable, is preferable for multiple reasons. Firstly, the previous statements showed that the replication of the human eye structure can boost the human-robot interaction by expressing cues that the humans are familiar with. Second, despite matching the users' expectation, the replication of human eye motion patterns that are the result of years of evolution often provide further benefits: For example the combination of active eye vision with the motion range of the neck allows to widen the perceivable field of view. Finally, there is another practical implication of placing cameras in the robot's eyes: A human look-

ing towards the robot in an interaction scenario will often look directly into the eyes of its interaction partner during conversations. Therefore, a camera which is included in the robots eyeball can record a front shot of the human face which is very suitable for image processing tasks that e.g. analyze facial expressions.

Beside this rather functional purpose, psychological studies found different social functions of human gaze, e.g. humans use the gaze to send signals or to regulate the flow of conversations [16]. In addition, it is also known that the replication of human motion patterns on a humanoid robot, e.g. mimicking non-verbal cues, and in particular gaze signals, can be interpreted by a human [17]. Therefore, taking advantage and exploiting these communication channels that exist between people allow boosting and enhance human-robot interaction in an unobtrusive way [18] and it is desirable to incorporate these multi-modal features into a robot that closely interacts with humans. Generally, it is assumed that the closer those robots interact with humans, the higher the required level of social skills is [19].

However, this potentially comes at the cost of risking that the robots appearance fails to deliver on its promises because people attribute more into the robot that it is capable to do [10]. Even worse, it is known that robots with human features deviating from the expected behavior or motion tend to be disturbing and can provoke the uncanny valley effect: This hypothesis, coined by Masahiro Mori in 1970, describes a negative emotional response towards a robot that seems almost, but not quite, human. When plotted as a graph with anthropomorphism on the x-axis and the level of familiarity on the y-axis our sense of the familiarity increases up to a certain reversal point as a robot appears more human-like. Right after this point the familiarity level falls steeply to an absolute minimum while forming a valley before increasing again, exceeding the reversal point [20]. Robots that end up in the uncanny valley are experienced more like a dead corpse than a living human. This effect is even more distinct for moving robots where slow-, jerky-, or other unexpected movement patterns of the robot contribute to the uncanny valley effect [21]. There are different opinions on how the level of familiarity behaves after reaching its absolute minimum. Bartneck et al. suppose that the level of familiarity will never exceed the reversal points' familiarity level [22]. Therefore, the design of a social robot and the effective human-robot interaction, can be described as "the balancing of the similarities and differences to ourselves, making robots acceptable and not

feared" [23].

Summing up, it can be said that a robotic head for close human-robot interaction can benefit from the reproduction of human characteristics, but this has to be done very carefully. A pleasant design and construction alone is not sufficient: Without proper and realistic actuation matching the users' expectations, the robot is very likely to pass the tipping point and to nevertheless slip into the uncanny valley effect. In addition, robots with human features that deviate from the expected behavior or motion tend to be disturbing [22] and even minor issues, considered unimportant, such as for example audible motor noise, will be interpreted by the human [24]. This shows the importance of paying attention to all details and the compliance with the users' expectations in terms of e.g. range of motion or motor performance should not be neglected. The requirements for a successful design of a humanoid robotic head can therefore be summed up as:

- The face of a humanoid robot, used in a social context, should replicate human characteristics that can be understood by the human.
- Its outer appearance should match its capabilities, keeping what it promises, by design.
- The design should not, in any event, cause discomfort or other negative reactions.

This list has a variety of implications and different designers of robotic heads place their emphasis somewhat differently. This becomes most apparent with regard to the exterior design of humanoid robotic heads found in literature (see Chapter 4): There are designs that copy the human face at a high level of detail, mimicking the human skin with soft rubber material and even replicate the human hair. Others use a higher level of abstraction by using hard plastic shells to replicate the human facial features in a more comic-like fashion. Some designs go even one step further and exchange the physical representations of e.g. the eyes or the mouth with display devices. Finding the right level of abstraction is still ongoing research and it can vary from one use case to another. However, the use of plastic shells, and a high level of abstraction has the inherent advantage that the replication of the human skin and the anticipated behavior of the moving soft tissue can be omitted. This also holds for the resulting negative impact when failing to

match the users expectation on the skin. Previous work by Hegel focuses exactly on this topic, deducing and explaining the outer appearance of the comic-style humanoid robotic head Flobi [25]. While this humanoid robotic head uses a high level of abstraction by replicating the human facial features by hard plastic shells, it shows a high level of expressiveness and is capable to replicate a variety of human emotions [25].

Based on these findings, this thesis will focus on the class of abstract robotic heads that use plastic shells to hide the mechanics from the user and feature an eye region with active vision. Nonetheless, the final results will not be limited to this class of robots as minor modifications would e.g. allow the addition of soft silicone skin when needed. With regard to characteristics and capabilities of a humanoid robots' eye, studies of the human can be used to define a list of key requirements in terms of outer design and motor capabilities. Interestingly, although they should be one of the key requirements during implementation, comparative numbers such as reachable velocities or the range of motion are hardly ever found in publications on humanoid robotic heads (see Chapter 4). This thesis targets to close this gap by viewing the system as a whole, tackling the optimization problem right from the start by using a predefined set of requirements based on the human model and paying attention to all relevant factors throughout the implementation.

## 1.2. Contribution

Given the importance of the eye region of a humanoid robotic head, this work will focus on the in-depth design, construction, and the evaluation of a robotic eye. Unlike previous work found in literature, this work will be based on a set of predefined strict requirements and the final outcome will be objectively assessed by comparing measurements on the real hardware against those specifications. Therefore, the first mandatory step is to understand the morphology, kinematics, and motion patterns of the human eye. To this end, Chapter 2 will give a detailed introduction into this topic. The combination of the collection of human facts, numbers, and the design aspects, that are based on the studies on the perception of humanoid robots, will allow the compilation of a list of key requirements for the successful

implementation of a humanoid robotic eye.

Before starting with the implementation, further literature research will be conducted and the eye regions of different established robotic heads will be analyzed and evaluated on the basis of the previously compiled list of key requirements. As it will be shown that none of the systems described in literature fulfill the composed list of requirements, two mechanical prototypes, each testing a different actuation principle, will be developed. The first prototype will be based on a human-inspired, tendon-like rope actuation scheme and will facilitate a very large range of motion at the expense of mechanical complexity. The second prototype, using a more classical lever-based approach, will integrate fast eyelids and a high resolution (4K) camera into the design. A subsequent analysis will show that both presented designs can compete or even exceed the human eye in terms of reached maximum accelerations and velocities. The results and the evaluation of the achieved performance were presented at the ICRA conference in 2012 [26] and 2019 [27]. The extended version, featuring a more detailed analysis and a direct comparison of the baseline of a previous robot design, the two prototypes, and the human data will be presented in Chapter 12.

As it turns out, the isolated consideration of the mechanical part without taking the low- and high-level motion control software into account is not sufficient. Therefore, this work will also present a full software stack that includes the full chain from the low-level joint space motion control to a configurable, multi-level control framework for human-like motion generation. The proposed overall system will utilize the presented beneficial effects, as introduced in Chapter 1, to the greatest extend without causing negative effects along the way. The goal of the combined efforts will be the optimization of the eye actuation scheme and the associated motion control in terms of human-like performance while maximizing the user experience. The final evaluation will show that the outcome, based on the proposed combined optimization, will facilitate the user to command the presented mechanical design to move in a human-like way at velocities and accelerations even outperforming the human model to some extends. Furthermore, the presented top-level control interface will facilitate the design of human robot interaction studies by others, without requiring in-depth knowledge of human eye motion patterns. It will allow a study designer to direct the robots gaze to any given point in space while the underlying controller takes care of the human-inspired, model-driven, and synchronized motion generation

for the whole robotic head using the proposed joint level control framework. Unfortunately previous frameworks found in literature tend to be very application specific and often target only a single robotic system. Controlling different robot designs with the same software framework would facilitate comparative studies regarding expressiveness, likability, and liveliness between different robot platforms. Therefore, the proposed frameworks will make extensive use of loose coupling and will possesses a variety of different adjustment options that will allow the transfer and application of the model on a variety of different robotic heads. This portability will be shown by controlling a variety of different robotic heads with this framework (Chapter 10).

Furthermore, the quality and acceptance of the proposed motion generation framework will be evaluated by a user study. First results of this study have been presented at the International Conference on Human Agent Interaction in 2016 under the title "Humotion - A human inspired gaze control framework for anthropomorphic robot heads" [28]. The extended evaluation of this study, presented in this thesis, will reveal valuable indications on the importance of single aspects of the motion control algorithm as well as the need for mechanical considerations concerning e.g. the inclusion of human-like moving eyelids. Further analysis will show that the participants prefer the proposed motion generation algorithm, including model-driven eye-neck coordination and eyelid animation, over a less animated approach (Chapter 11).

## 1.3. Structure

This thesis is divided into four main parts. Part I deals with the introduction into the research topic, the underlying motivation, and its scientific contribution. In order to have a good basis for the subsequent considerations, the human eye and all its relevant properties are introduced and human characteristics, that are important for the design and implementation phase, are collected and form the basis of the compilation of a list of requirements. Subsequently, a literature summary gives an overview of related humanoid robotic heads and their eye region is analyzed in terms of compliance to the previously defined list. Finally, this first part of the thesis concludes with

an overview on the different ways of the reproduction of human motion on technical systems.

Part II deals with the design and construction of a humanoid robotic eye and its control systems. As derived in the previous part, the replication of human motion requires an integrated approach, that takes care of the requirements right from ground up, starting at the low-level control of the individual motors, up to the model-based generation of the complex interplay of human motion. Following this design philosophy, previous work and other systems found in literature laid the foundation for the design, construction, and evaluation of two different actuation principles by means of mechanical prototypes. Subsequently the question on how to actuate such capable systems was addressed by means of a bi-directional motion capture system, which allowed to record and study the complex interplay of different motion patterns first hand. The combination of findings based on datasets of these recordings and the human data found in literature ultimately led to the development of the full motion control framework.

Part III is dedicated to the evaluation of the motion control framework and the proposed mechanical eye prototypes. A user study with participants, ranging from naïves to experts, was conducted in order to assess the quality of the motion generation of the proposed system and the general expectations of the different user groups on how a humanoid robotic head moves. The mechanical eye prototypes were thoroughly measured and evaluated. The joined evaluation of both prototypes, enriched with human data, allowed a direct comparison of both systems to each other and additionally towards the human and a previous prototype.

Finally, Part IV closes this thesis with a discussion of all presented aspects in form of an overall conclusion and an outlook.

Chapter 2

# The Human Eye

As introduced in the previous chapter, this work will focus on the design, construction, and the evaluation of a robotic eye based on the model of the human eye. Before starting with the assessment of humanoid robot eyes found in literature and the realization chapter, it is necessary to characterize the human eye. The following consistent analysis of the human model is mandatory for the formulation of the key requirements which subsequently will guide and facilitate a successful replication in a technical system. This replication aims to mimic human motion patterns, and at the same time benefit from features which humans developed during evolution and their adaption to a humanoid robotic head.

The human eye is the most important sensory organ in the human body with an estimated 80% of all sensory information in humans arise from the eyes' retina [29]. This specialized organ receives the visual input of our surroundings and translates these stimuli into chemical and electrical signals, which are then carried to the brain. Moreover, the eyes make an essential contribution to our facial expressions and facilitate a variety of nonverbal communication patterns. It allows us to get into contact and interact with people before a single word is spoken and plays an important role in social interaction.

The structure and kinematics of the human eye is thoroughly studied and well understood. Detailed descriptions and more background information on this topic can for example be found in the books *Textbook of Ophthalmology* [30], *Strabismus* [31], and *The Neurology of Eye Movements* [32]. The following sections in this chapter are mainly based on these books, supplemented with additions from additional sources where necessary.

## 2.1. Morphology

Before looking closely into the kinematics and motion control of the human eye, a few words about the overall structure are necessary. The adult human eye is approximately spherical with a typical vertical diameter of $\approx 25\,\mathrm{mm}$ and with a minimal larger transversal diameter [30]. The eyeball is located in the eye socket (orbit). The orbits' volume is approximately $30\,\mathrm{mL}$ with the eyeball occupying $6\,\mathrm{mL}$ to $8\,\mathrm{mL}$ [33], [34]. Fat and muscular tissue fill up the remaining space. The average weight of the eyeball is $7.5\,\mathrm{g}$ [35]. A schematic view is depicted in Figure 2.1. It consists of the following parts:

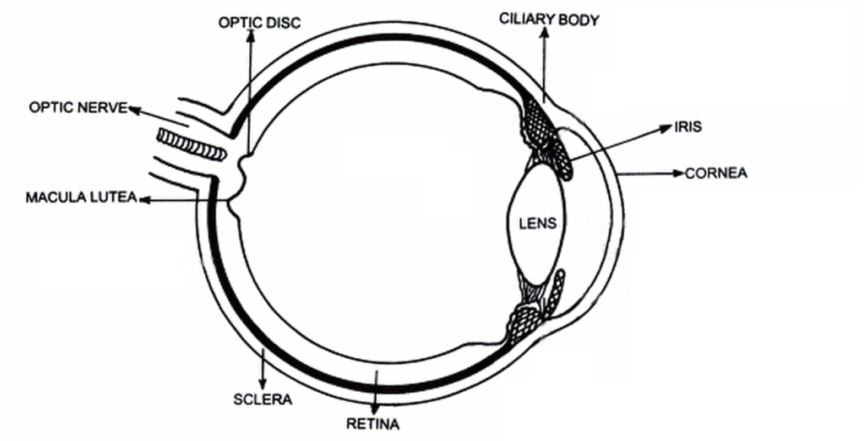



Figure 2.1.: Anatomy of the Human Eye (edited, based on [30]).

**Cornea**
: The transparent outermost part of the eyeball is called cornea. It forms the main refracting surface and is located in front of the iris and lens. The average corneal diameter was found to be $11.71\,\mathrm{mm}$ $\pm 0.42\,\mathrm{mm}$ [36].

**Iris**
: This element performs the function of an aperture to facilitate adaption for varying light conditions. The diameter of the iris is anatomically closely connected to the corneal diameter.

**Pupil**
The term pupil describes the concentric opening of the iris.

**Sclera**
The sclera is the white, outermost layer of the eye. It consists of approximately $1\,\mathrm{mm}$ thick firm tissue and is held in place by a ligamentous apparatus. A complex interlocked network of extra ocular muscles and tendons attach to this tissue in an antagonistic configuration.

**Limbus (cornea)**
The transitional region between the cornea and sclera of the eyeball is called limbus cornea.

**Optic disc**
The point where the optic nerve leaves the eye is called optic disc. It is also known as the blind spot which is caused by the complete absence of photo receptors in this region.

**Retina**
The retina is a circular region with a diameter of $30\,\mathrm{mm}$ to $40\,\mathrm{mm}$ in the eyes back [37]. It encompasses a variety of different regions with non-uniform distribution of photo receptors.

**Macula**
The macula describes the retina region encompassing the whole foveal area. This includes the foveal pit with the highest optical acuity, the foveal slope, the areas called parafovea and perifovea. This region occupies a circular field of approximately $6\,\mathrm{mm}$ around the fovea [38].

**Foveola**
The most central part of the macula is called fovea. It describes a $0.35\,\mathrm{mm}$ wide circular depression with the best visual acuity (also-called foveola). It has the highest density of cone photo receptors [37].

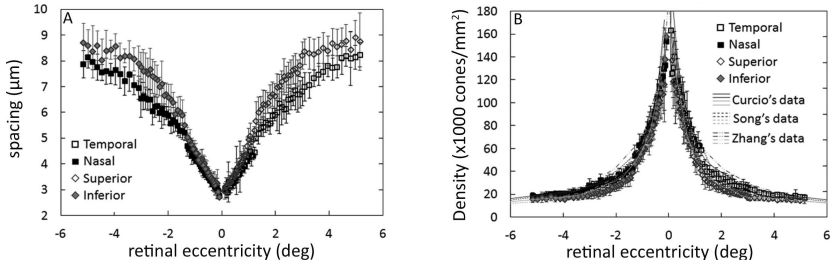
**Fovea**
The fovea describes the central region spanning $1.5\,\mathrm{mm}$ circularly, which equals the inner $5°$ of the visual field.

A recent study by Dodgson found that the average interpupillary distance (IPD) amounts to $64.67\,\mathrm{mm}$ ($\sigma$=$3.7\,\mathrm{mm}$) for men and $62.31\,\mathrm{mm}$ ($\sigma$=$3.6\,\mathrm{mm}$)

for women. The extrema ranged from $52\,\mathrm{mm}$ to $78\,\mathrm{mm}$ for men and $52\,\mathrm{mm}$ to $76\,\mathrm{mm}$ for women. In addition, the breadth of the head was measured as $126\,\mathrm{mm}$ to $173\,\mathrm{mm}$ with an average of $148\,\mathrm{mm}$ ($\sigma = 6.2\,\mathrm{mm}$). This amounts to an average head breadth to IPD ratio of 2.34 [39].

## 2.2. Visual Properties

A closer look at the monocular visual field of view for a steady eye reveals that the inner $30°$ are equal among every healthy human. However, the extended peripheral monocular field of vision varies from one person to another. Typical values found in literature amount to $60°$ up, $75°$ down, $100°$ lateral, and $60°$ medial [40]. The different emphasis for the single axes arise from shadowing by the eyelids or nose.




(a) Cone spacing (b) Cone density

Figure 2.2.: Average cone distribution characteristics in terms of cone density and spacing of the human eye (from [41]).

The single light sensitive cones on the retina are not evenly distributed: The densest region, the fovea, contains up to 160000 cones per square millimeter [41]. This corresponds to a spacing of $\approx 3\,\mu\mathrm{m}$. The density rapidly drops to approximately 20000 cones per square millimeter, or a spacing of $\approx 8\,\mu\mathrm{m}$ outside of the centric $4°$ region (see Figure 2.2). If the central region with the highest visual acuity is defined as 100%, the acuity drops to 15% in a distance of $15°$ from the center. If measured at $45°$, the visual acuity drops to 3% [42]. This allows humans to reach an optical acuity of 1 minute of an arc ($0.0167°$) in the area of the greatest visual acuity [43]. Based on

the lower cone density, this value rapidly drops to the outer regions of the field of view. The non-uniform distribution of photo receptors and the small fovea with the highest visual acuity can be utilized at best when the eye can be moved: A total of seven muscles are located in the orbita with six directly contributing to eye motion. The seventh muscle is responsible for moving the upper eyelid. The six main muscles facilitates the human to center and fixate the area of interest on the fovea which is the region with the highest spatial resolution.

## 2.3. Kinematics

For a better understanding of the eye kinematics one has to investigate the arrangement of connection points and the individual muscles that are depicted in Figure 2.3: A stable eye position is achieved when the sum of antagonistic acting muscles and the passive forces of the orbital tissue zero out. The main forces are applied by the antagonistic muscle pairs, however, in reality things are slightly more complex as the muscles can not be considered separately as they contribute to multiple degrees of freedom (DOF) at the same time. The single muscle attachment points are typically given in relative distance to the limbus.

**Musculus rectus superior**
: This muscle arises from the orbita tip and follows a curved path before attaching to eyeballs' top with an average distance of $7.7\,\mathrm{mm}$ from the limbus and before the equator. By means of this structure and the associated routing this muscle mainly contributes to the adduction of the eyeball. In dependency of lateral eye deflection it also contributes to eye rotation.

**Musculus rectus inferior**
: Similar to the M. rectus superior, this muscle describes a curved path as well but attaches to the bottom of the eyeball. Its insertion point lies slightly closer to the limbus with an average distance of $6.8\,\mathrm{mm}$. Similar to the M. rectus superior its main contribution is to the downward motion (depression) with some influence towards the eye rotation depending on the lateral deflection.

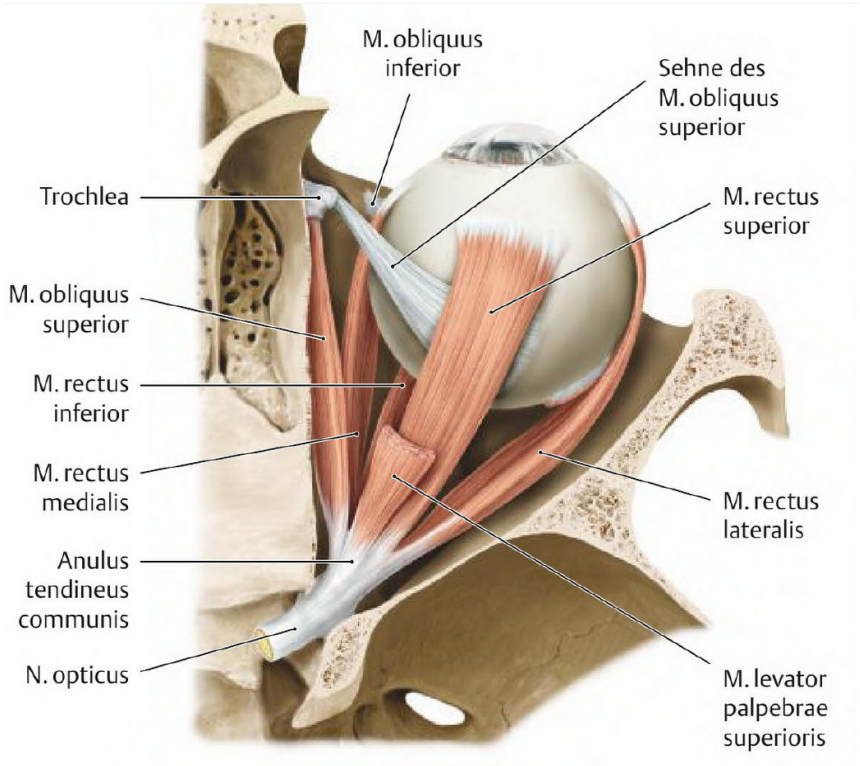



Figure 2.3.: The human eye and its muscles (from [31]).

**Musculus rectus medialis**

Is the most powerful eye muscle. It attaches to the medial side of the eyeball with a distance of $5.7\,\mathrm{mm}$ to the limbus. This muscle is responsible for the eye rotation towards the nose (adduction). It can contribute to a slight elevating or depressing motion on extreme vertical eye deflections.

**Musculus rectus literalis**

This muscle attaches at $7.4\,\mathrm{mm}$ distance to the limbus on the lateral side of the eyeball. This muscles main function is the outward rotation (abduction) of the eyeball. Again, under extreme vertical eye deflections this muscle contributes to elevating and decreasing eye motions.

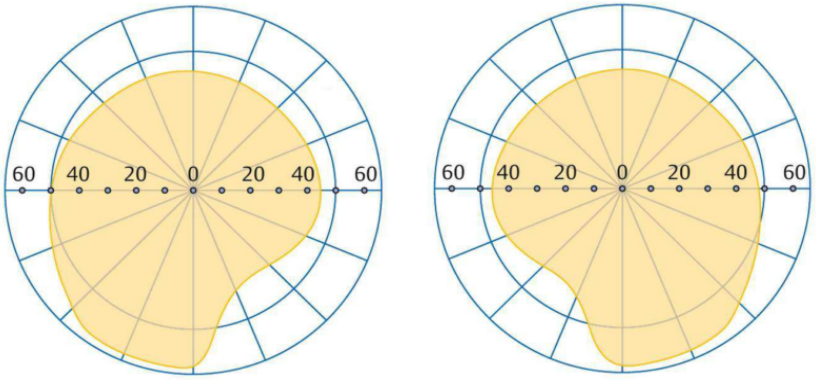
**Musculus obliquus superior**
: This muscle has some unique features: It attaches to the upper back of the eyeball with a comparatively long tendon that is redirected by the trochlea. The trochlea consists of a padded ring of cartilage that redirects the tendon in a sharp crease to the back. Depending on the eyes horizontal deflection the main contribution of this muscle is the inwards rotation (adduction) and the downward motion of the eyeball.

**Musculus obliquus superior**
: This is the only eye muscle that does not arise from the orbitas' tip. Instead, this muscle arises the temporal bone and connects to the lateral back of the eyeball. The main function of this muscle is antagonistic to the M. obliquus superior and responsible for the outward rotation of the eye. It also has some minor contributions towards eye elevation.

**Musculus levator palpebrae superioris**
: This muscle, although located in the orbita, does not contribute to the eyeball motion itself, instead it is responsible for moving the upper eyelid.

## 2.4. Oculomotor Range

For a healthy human oculomotor system the muscle deflections facilitate a maximum adduction and abduction of $50°$, respectively. The elevation capabilities rarely exceed $45°$, whereas the depression can reach nearly $60°$. When combined with the shading effects of the nose this results in a monocular field of view, that is within reach of each individual eye, as depicted in Figure 2.4: Each eye can survey at least approximately $40°$ in every direction with a prominent extension to the lower lateral perimeter. The overall majority of movements take place in the so-called practical field of view ranging between less than $\pm 20°$ vertically, an elevation of less than $10°$, and a depression of less than $-30°$ [31]. According to Guitton *et al.* [44] gaze shifts outside the oculomotor range and the accompanying head movements typically result in a maximum eye displacement of $45°$ before reaching the




(a) left eye (b) right eye

Figure 2.4.: The human monocular field of view (from [31]).

absolute limits. In addition, studies show that up to 86% of naturally occurring human saccades have magnitudes of $15°$ or less [45].

## 2.5. Degrees of Freedom

The human eyeball is suspended in the orbit by a complex interplay of fascia, eye muscles and surrounding fat tissue. The eye can rotate around three axes that lead to horizontal movements (abduction and adduction), vertical movements (elevation and depression) and torsion around the line of sight. At first sight, this gives the impression that the eye will occupy an unlimited number of torsional rotations for a given gaze direction. Interestingly this is not the case: In 1984 the dutch scientist Donders observed that there is only one torsional eye rotation for each combination of horizontal and vertical eye position. In order to understand the implications the following definitions are necessary: The head is held stationary and for now the eye's primary position is initially defined looking straight ahead with the visual axis roughly parallel to the midsagittal plane of the head. The secondary positions are defined as looking solely left, right, up, or down. Tertiary positions of the gaze are then defined as the remaining mixture of horizontal and vertical rotations. Donders' law states that the torsional component is

zero for the primary- and secondary positions. For any other tertiary position there is one specific torsional rotation that depends on the horizontal and vertical elongation [46]. Donders' law does now specify the amount of the torsional component. A more detailed statement is known as Listings' law which states that the eye occupies only those rotations that can be reached from a primary position by a single rotation about an axis in a special plane [47]. This plane is called Listings' plane and it is orthogonal to the line of sight when the eye is in its primary position. The torsional component for a given eye rotation by Euler angles $\alpha$ and $\beta$ of can be approximated by Equation (2.1) [48].

$$tan(\gamma) = \frac{sin(\alpha)sin(\beta)}{cos(\alpha) + cos(\beta)} \tag{2.1}$$

The inverse tangents and thus the angle $\gamma$ can be roughly approximated by Equation (2.2).

$$\gamma(\alpha, \beta) \approx \frac{\alpha\beta}{2} \tag{2.2}$$

Listings definition of the eyes' primary position is a bit different to the clinical definition based on the midsagittal plane of the head. He defines the primary position as the position from which outgoing secondary positions do not involve torsional components. It is important to note that only motions that originate from the primary position can be described this way. Arbitrary start- and end-positions can be described by Listings' half-angle-rule: Any change in position from one gaze direction to a new one can be described by one rotation through a so-called velocity plane. If the origin is the primary position, this velocity plane is orthogonal to the line of sight and thus equal to the previously defined Listings' plane. For a starting point rotated by $\alpha$ degrees from the line of sight, the velocity plane is oriented towards the line sight by only half of that angle ($\frac{\alpha}{2}$). If both eyes are involved, these observations only hold when the gaze focuses on an object at infinity. If vergence is involved, these planes are rotated for both eyes [47]. More recent measurements with more reliable ways to measure eye rotations as in the 19th century have shown that eyes only approximately follow those laws [49].

## 2.6. Motion Patterns

This section will give an overview of typical human eye-, neck-, and eyelid motion patterns, their cause, and interplay expressed in measurable terms in conjunction with a short overview of the underlying control systems. This feature set forms the basis for subsequent implementation of human-inspired eye motion on a humanoid robotic head.

### 2.6.1. The Vestibular and Optokinetic System

The Vestibulo-Ocular Reflex (VOR) is responsible for adapting the eye orientation to fast head and body movements in order to keep a fixated target in the center of the retina. Similar to cameras, the human visual system works the best when the image is held stationary on the sensor or retina, respectively. If this is not the case, the image gets blurred and the visual acuity degrades [50].

Optical, vision-based tracking alone is not always feasible: The optokinetic system alone is simply not fast enough to compensate for fast movements. This is due to the high processing latency of more than 70 ms [50], [51]. And this is where the VOR steps in: This reflex is a direct coupling of stimuli in the inner ears' kinetic labyrinth to associated eye muscles. Depending on the involved eye muscle there are only three to four neural layers involved [31], resulting in a very low latency of around 8-16 ms [50], [52]. During intentional head movements the involved neurons are selectively inhibited to suppress this reflex [31]. Based on how the inner ears receptors work, the vestibular response dies away under sustained head rotation for longer periods without optical feedback (e.g. in darkness) [50]. A similar saturation takes place when the velocities reach and exceed approximately $350\,^{\circ}/\mathrm{s}$.

Most of the time head motion includes linear accelerations to some extent. The otolithic receptors in the human vestibular system compensate for that and contribute to correcting eye motion. In addition, the otolithic receptors are also responsible for the partial correction of head roll motion by torsional counter rotation of the eyeball [50]. On a fixed robot platform with no unintentional or external motion the vestibular response can

be simulated based on measurements of the neck joints and inverse kinematics whereas on a moving, e.g. walking, platform this is typically implemented using an inertial measurement unit with gyroscopes as an additional cue [53].

### 2.6.2. Smooth Pursuit

Smooth pursuit eye motion allows the human to track objects moving at low speeds, typically less than $30\,^{\circ}/\mathrm{s}$, by matching the angular velocities of the eyeball and the target [31]. Angular velocities exceeding this threshold will typically trigger additional correction saccades. Smooth pursuit eye motion allows to keep the area of interest centered and steady on the fovea. In the past, the fixation of stationary objects (visual fixation) has been put on level with the fixation of a "moving" target with zero velocity. However, recent evidence shows that different brain areas and mechanisms are active during visual fixation and smooth pursuit of moving targets. This lead to the assumption that a separate fixation system is being used for stationary targets.

The typical latency of the smooth pursuit control loop is around 150 ms. Interestingly humans can overcome this delay during tracking of a moving object by prediction and adaption to the targets motion [54]. The tracking works best for predictable and consistent target motion. It has been shown that the main input for the smooth pursuit system is the targets' velocity [54]. More recent studies found that the position error play some but less important role and it can not be excluded that the acceleration is being used as well.

The pursuit system is not driven by visual stimuli alone, the internal mental representation of the body state is also relevant: It is known that some subjects can pursue their own fingers in the dark without visual stimuli [55].

### 2.6.3. Eye Saccades

This prominent motion type occurs up to 173000 times a day [56] and is characterized as fast, step-like gaze shifts performed for (re-) fixation of new or lost objects during smooth pursuit or to search for objects out

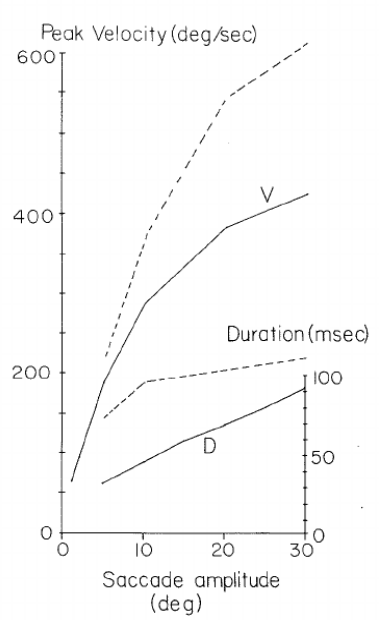
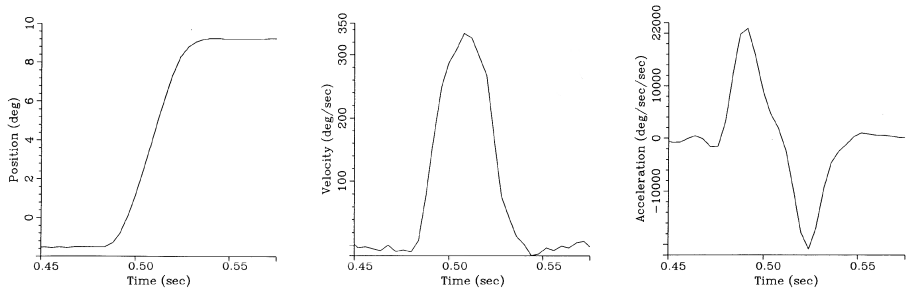



Figure 2.5.: Relationship between saccade amplitude, peak velocity, and duration (from [57]).

of view [31]. Saccades are prevalent amongst foveated animals and the human with the highest visual acuity and resolution in the center of the retina. One differentiates between quick phases and voluntary phases: the former describes unconscious self-centering mechanism for the eye during head rotations [50]. Voluntary saccades are triggered from different stimuli and center targets of interest onto the retina. It takes about $200\,\mathrm{ms}$ between appearance of a target and the actual saccade initiation ($\sigma$ $=25\,\mathrm{ms}$ to $50\,\mathrm{ms}$) [57], [58]. This latency depends on a variety of factors: External conditions (lighting, size, contrast, ...), the type of stimuli (auditory or visual), predictability of the target, the user's age, or motivational and attentional factors influence the recorded values.

Humans can not influence saccade velocities willingly, there is a direct correlation between the total size of the movement and the peak velocity. This is depicted in Figure 2.5: The x-axis represents the total saccade am-




(a) Position (b) Velocity (c) Acceleration

Figure 2.6.: A plot of an exemplary human saccade showing a $10°$ gaze shift (from [57]).

plitude in degrees and the left y-axis represents the peak velocity. The continuous line "V" represents the average correlation between distance and velocity with the standard deviation depicted by dashed lines. Velocities during larger saccades can reach up to $600\,°/\mathrm{s}$ to $800\,°/\mathrm{s}$ with average peak accelerations in the range of $40\,000\,°/\mathrm{s}^2$ for normal individuals [31], [57], [59], [60]. In addition, it has been observed that saccades of up to $50°$ possess a linear relationship between the saccade amplitude and the execution time. This is plotted in Figure 2.5: the right y-axis and the continuous line "D" illustrate this connection. For example a typical $30°$ gaze shift lasts about 100 ms.

An exemplary gaze shift of $10°$ is shown in Figure 2.6: For such smaller saccades the skewness ration, describing the asymmetry of the trajectory, is about 0.5. In that case the acceleration and deceleration phase are equally pronounced as seen in Figure 2.6c. In addition, Figure 2.6b shows that the peak velocity is reached approximately halfway through the saccade. This fact is also supported by research of Richard Abrams [61]. For larger saccade amplitudes the peak velocities are reached earlier resulting in a skewness factor of as low as 0.2. Similar to saccade velocities, this skewness factor depends on saccade start- or end-position, direction and predictability. For non pure horizontal or vertical saccades this causes the trajectories to be curved. It is believed that this effect is caused by neural factors instead of muscle properties. Differences in the velocity profiles of the left and right eye induced by these characteristics causes a transient intersaccadic diver-

gence.

Horizontal saccades often end up with some post saccadic drift. This drift has disjunctive and conjugate components: It has been observed that most saccades slightly undershoot the target. The conjugate component is onward directed and it is assumed that this compensates for the undershoot. The disjunctive component is convergent and is assumed to correct for divergence during the saccade. This effect is called a glissade. It is assumed that this is caused by a mismatch of the neural pulse and step components of the innervational change responsible for the saccade. A saccade that undershoots the target position is corrected by a second correction saccade within $100\,\mathrm{ms}$ to $130\,\mathrm{ms}$. The amplitude of this correction depends on visual and non-visual cues: Undershooting saccades to remembered positions in complete darkness are corrected as well. The basis for the non-visual offset calculation is most likely based on the efferent ocular motor commands.

It was long believed that saccades are ballistic movements that are not stoppable and can not be modified once initiated. Early experiments by Westheimer [62] found that if the target executed a double jump to a new location and returned within $100\,\mathrm{ms}$ the subject would make two saccades: At first to the faked new target position and then back to the start where the second target jump ended. The delay between these two saccades agrees with the saccadic latency of $150\,\mathrm{ms}$ to $200\,\mathrm{ms}$. This behavior was explained by the ”sampled data system” hypothesis by Young and Stark [63]: This model assumes that the visual system samples the scene, calculates the offset and fires a preprogrammed saccadic response. Once completed, the sampling takes place again and the process repeats. More recent research suggests that the saccade calculations can be updated up to a certain point in time: Changes up to $70\,\mathrm{ms}$ before the eyeball starts moving are detected and included in the saccade calculations. This sampled data system hypothesis can not describe all aspects of saccade movements and recent evidence rises doubts on the ballistic nature of saccades. It is believed that based on the short duration and high velocities of the saccade there is not enough time for the acquisition and calculation of a new target during the execution.

In addition, our sense of vision is suppressed during saccades. This phenomenon, called saccadic omission, is caused by two factors: Firstly, saccadic suppression increases the light detecting threshold. However, the second, so-called visual masking effect, has more impact. The stationary and

high contoured visual background before and after the saccade masks up the blurry image during the fast eye movement. This effect is independent of the eye motion itself. It is possible to provoke the same masking effect by moving an external image at saccade speeds while the eyes are not moving at all.

Humans are able to maintain their orientation and sense of straight ahead even though the fast saccadic movements cause the image of the world to move on the retina multiple times a second. The most prominent explanation is that our brain records and shares an efference copy with the sensory systems. This allows the perceptual sense to compensate for these shifted images upon the retina and allows us to ensure spatial consistency.

### 2.6.4. Vergence

When humans shift the visual focus of attention towards objects not located at optical infinity, both eyes rotate inwards. This way the image of the object of interest on the retina is congruent on both eyes. Without this disjunctive positioning offset one would experience diplopia [64]. This oppositely directed position offset is called vergence and typically accompanies saccade motion as changes of fixation points in a three-dimensional space often imply changes of the objects distance as well [50]. Vergence adjustments without an accompanying saccade take place as well. The brain uses the total amount of displacement to assess the subject's distance to the target [64].

It is anticipated that saccadic eye movements are always preceded by convergence and divergence movements of both eyes. Herings' law of equal innervation states that both eyes behave as they have received equal innervation [65]. This is shown in the classic experiment of Müller [66]: The test subject has one eye masked. The seeing Eye changes the focus along its visual axis from a distant to a close target. The covered eye will execute a convergence movement and the seeing eye will execute small vergence movements as well [64].

Vergence adjustments are primarily initiated by two stimuli: Firstly, by disparity between the images on both retinas causing diplopia and secondly by defocused images. The former provokes fusional and the second causes accomodation-linked vergence movements. In addition, changes in size and awareness of the target distance are suspected to have an effect as well.

Vergence movements can be partially deliberately influenced but take place without manual intervention most of the time [64].

The typical reaction time for blur-driven vergence movements with unexpected target positions is 200 ms. In the past it was believed that general vergence movements are slow, taking up to 1000 ms. However, this is only true when tested in an artificial setup in a laboratory. Real life vergence movements, tested under more natural conditions, are significantly faster [64].

## 2.7. Eye-Neck Coordination

Apart from communicative gestures such as nods, most voluntary head movements are motivated by the need to shift the visual focus of attention. In order to take advantage of the non-uniform distribution of photo receptors on the retina it is inevitable to be able to shift the focus by moving the eye. However, as stated before, the range of motion of each single eye is limited, and interestingly, this is even smaller than the peripheral visual field of view. At this point, the interplay of eye and neck motion comes into play: This complex muscular interplay allows humans to enhance the perceptible field of view [44], [50]. In addition, the interplay between neck and eye muscles allow compensating head perturbations during e.g. locomotion.

Even though reflex mechanisms such as the Vestibulocollic reflex (VCR), which compensates for external disturbances by biasing the muscular tone based on vestibular inputs help to reduce deviation, the dynamic viscoelastic properties of the neck muscles and connecting tissue also play a major role. Perturbations of human locomotion typically result in velocities below $100\,^{\circ}/\mathrm{s}$ with frequencies in the range of $0.5\,\mathrm{Hz}$ to $5.0\,\mathrm{Hz}$. These velocities are small so that these perturbations can be attenuated by the VOR. It is believed that the major contribution of the VCR takes place at the lower excitation frequencies below $1\,\mathrm{Hz}$ [67].

As the head, based on the higher mass, can not accelerate as fast as the eyeball, the head motion lags behind the eyes'. In addition, the actual delay between target selection and the muscle activation measured by electromyograms is different for the eyeball and neck muscles. This delay is shortened by predictable targets [68]. In contrast to the eye saccades, the angular

velocities and deflection of the head can be willingly influenced by humans.

In the following, the eye position stands for the angular deflection of the eyeball, the head position for the angular deflection of the head itself, with the gaze position being the sum of eye- and head position. One application of combined eye and neck motion is to rapidly shift the overall gaze target to a new location. This can take place because of two different reasons: Firstly, the gaze should be redirected to a new target, or secondly, the eye deflection is close to the oculomotor range limits. In the latter case a combined eye neck motion will recenter the eye position to the normal state, which enables future saccades to reach the entire visual scene. In addition, this (re-)centering is also achieved by quick phases of nystagmus during vestibular stimulation [67].

In addition to smooth pursuit tracking using only the eye muscles, humans are able to utilize combined eye neck motion to smoothly track moving objects. People have different preferences which tracking they prefer but in general humans can track equally well using both options [67]. There are different hypotheses how the VOR is nulled during this type of tracking. There are arguments for a direct suppression through gain reduction, cancellation by a direct feedback signal, or a combination of both [67].

An exemplary combined eye neck saccade is depicted in Figure 2.7: In the beginning, the resting gaze position is measured to be at $20°$ to the left. Approximately $200\,\mathrm{ms}$ after the unexpected occurrence of a visual target at $20°$ to the right (A) the eye starts moving by commencing a saccade (B). Once the neck motion starts (C), the VOR drives the eye back to the center position (D). At the end of the motion a small eye saccade (E) corrects for a remaining offset to the target. Note how the summation of the simultaneous movement of the eye and neck contributes to a rapid approach of the target position.

In contrast to this example of a relatively small gaze shift ($\delta = 40°$) with the VOR driving the eye back to the center position during the neck acceleration phase (see Figure 2.7.D), evidence shows that the VOR is inhibited during larger gaze shifts exceeding the oculomotor range. This allows the eyes and head to move together towards the target [50].

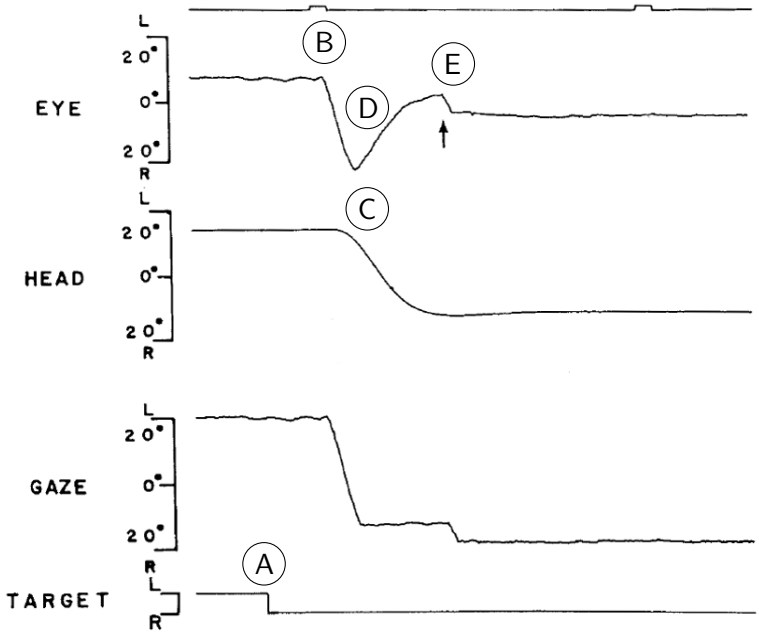



Figure 2.7.: A combined Eye-Neck saccade (edited, based on [67]).

## 2.8. Eyelids

Despite not being part of the oculomotor system as such, eyelid motion is a very salient property of a human face and closely coupled to the eye motion. The muscle that is responsible for the actuation of the upper eyelid is located next to the eyeball muscles, and arise from the orbita as well (M. levator palpebraesuperioris, see Section 2.3). A closer look on the lid-eye interaction during vertical eyeball motion reveals, that the motion and speed of the eyelid matches those of the eyeball during saccades, as well as during smooth-pursuit [69], [70]. The variable clearance between the eyelid and the pupil border is influenced by different factors, e.g. one's

condition or the level of alertness [71].

Another prominent motion type of the eyelids are the different types of blinks. One differentiates between voluntary, spontaneous, and reflexive blinks with distinctive velocity profiles [69]. The frequency of periodic blinks depends on the affective, attentional, and cognitive state [71]. A typical eye blink lasts $100\,\mathrm{ms}$ to $400\,\mathrm{ms}$ [72]. Under resting conditions Cramon *et al.* measured 7.6 eye blinks per minute [73]. Large gaze shifts of more than $33°$ evoke blinks with a probability of 0.97 [74].

All in all, this amounts to that the eyelids are much more than a protective organ. Eyelid motion can and will be recognized and interpreted by humans. The importance for non-verbal communication and expression of the internal state should not be neglected during implementation of a gaze model for an anthropomorphic robotic head.

## 2.9. Discussion

This chapter gave a throughout summary of the features, capabilities, and peculiarities of the human model. This overview is mandatory for a successful replication of the human eye structure by a technical system. In a next step, these single items will be subsequently used to build up a detailed list of key requirements for a mechanical replica. The consecutive inspection of previous technical systems will reveal to what extent present systems deviate from these requirements and thus fail to deliver human-like performance. Multiple reasons might be responsible for this divergence, be it a different focus or technical reasons.

However, as the experience shows, it is important to have exact specifications and target requirements in place before starting the implementation. For example the specification of the total oculomotor range will place specific constraints on the mechanical solution and will thus have a huge impact on the overall design. So it is crucial to have all requirements defined before the implementation takes place. Therefore, the assessment of related work and the consecutive realization will be based and driven by the replication of the subsequently compiled list of requirements, which are in turn based on the human facts and numbers that were introduced in this chapter.

Chapter 3

# Requirement Analysis

As shown in Chapter 1, humans tend to anthropomorphize things and will try to interpret and decode e.g. the robots intentions by looking at its facial cues. In particular the eye region plays a significant role in this regard. It is thus crucial to prepare a detailed list of key requirements based on the human data, that was introduced in Chapter 2, and on the intended application on a humanoid robotic head. This list on human-inspired and human-based requirements alone is not sufficient for a replication. In addition, there are technical limitations and requirements that need to be considered as well. The key driving factor can be identified as the successful combination of aesthetic appeal, sensory capabilities, performance, and functionality.

## 3.1. Appealing Design

The outer appearance of a humanoid robotic head is, in many cases, the first thing a user notices. This consequence directly leads to the first set of requirements - no matter how the details are implemented, the robot should be pleasing to the eye. Covering the robot's face with a mask, hides potentially distracting and deterring parts such as actuators, cabling, and other mechanical parts from the users' view. Based on the positive experience with the Flobi robot [25], the first Requirement $R_{1.1}$ specifies that designed robot eye should not conflict with a closed shell that hides the interior of the humanoid robotic head. At first, this sounds as an easily reachable goal, but the practical implementation often limits e.g. the choice of lever attachment points which will again, limit the range of motion.

In addition to this, the externally visible structure should resemble the human eye morphology as introduced in Section 2.1. In principle, there should be a defined lower and upper limit for the size of the eyeball and its features. However, it is quite common for instance on comic-style robot designs to use different levels of scale and over amplification as stylistic means. Larger eyes, with proportions bigger than normal, can e.g. be used to trigger effects associated to the so-called ”Kindchenschema”: This effect describes the way humans categorize small children, animals, and even mock ups (e.g. dolls) that have a relatively large head, dominant frontal skull, a big and low placed eye, strongly curved forehead, prominent cheekbones, thick and short extremities, firm elastic consistency, and clumsy movement patterns as cute or sweet [75]. A typical example for the use of this kind of over amplification of the eye size is the comic character ”Mickey Mouse”: Over the last decades, the proportion of the eyes in relation to the head length was extended from 27% to 42% [76]. This shows that the limits for an acceptable eyeball size can not be expressed in numbers, instead the proportions have to match the design and a homogeneous overall appearance is desirable. The requirement on the eyeball design can therefore be formulated as follows: The eye should consist of a white eyeball with a structuring element that looks similar to the human cornea and the iris should be indicated as well. Typical implementations use the lens element of the camera as the visual fundamental characteristic but often neglect the replication of the iris.

As derived in Section 1.1, the eye of the robot should resemble the human appearance and capabilities. The fact that the motion of the human eyelid is closely coupled to the eyeball movement is the most evident reason to include eyelids in a robotic eye design. Its proximity close to the eyeball and the expressiveness of the different motion patters are other reasons. In addition, as shown in Section 2.8, eyelid motion is a very salient property of a human, which will be recognized and interpreted by humans. The position and motion of the eyelids relay inner states and signals of the human such as e.g. the emotional state, the level of excitement, or the level of tiredness to a communication partner. The resulting importance for non-verbal communication and ability to express internal states should therefore not be neglected during implementation of a robotic eye. This importance is further supported by the evaluation of the final prototype in Section 11.2.

Summing up, the following list of key requirements can be compiled:

**Requirement 1: Design**

$R_{1.1}$: *Closed shell with no visible mechanics*

$R_{1.2}$: *White eyeball, dark lens, and colored iris*

$R_{1.3}$: *Two independently controllable eyelids*

## 3.2. Camera

As derived in Section 1.1, the eye of the robot should perform in a similar way as the human in order to fulfill the users' expectation. This implies, that the camera should be integrated without adding unnecessary openings into the overall robotic head design and that the visual acuity of the robot eye shows a human-like performance. It is, therefore, necessary to install a state-of-the-art high resolution camera into the eyeball. As introduced in Section 2.1, in contrast to the constant angular acuity given by a typical image sensor, the visual acuity of the human eye is the highest in the centered foveal area and degrades to the extrema of the field of view. As introduced in Section 2.2, the acuity in the center region amounts to one arc minute ($0.0167°$) for a healthy human. In order to match the users' expectation, the combination of the camera sensor and the lens should reach the same angular resolution in this region. This can be achieved by three means: Foveated imaging based on a special image sensor or warping optics can be used to replicate the acuity distribution of the human eye. This option requires a special imaging sensor or complicated lenses and can not be implemented by using commercial components. This is one of the reasons why some robot designs, introduced in Section 4.2, did choose a different option and replicated the combination of the high acuity at the foveal region and the large field of view by a dual camera solution using tele- and wide-angle lenses. This conflicts with Requirement $R_1$ and presents other difficulties as space constraints. Another option is the use of an image sensor that allows to reach the required angular acuity over the full field of view. This would permit to have the same performance as the human in the foveal region and to even outperform the human in the remaining field of view. In the past,

| Sensor Width | Acuity | Performance | Uncovered Region |
|---|---|---|---|
| 320 px | 0.313° | 5% | <40° |
| 640 px | 0.156° | 11% | <20° |
| 1024 px | 0.098° | 17% | <11° |
| 1800 px | 0.056° | 30% | <6° |
| 1920 px | 0.052° | 32% | <5° |
| 2040 px | 0.049° | 34% | <4° |
| 4096 px | 0.024° | 69% | <2° |
| 6000 px | 0.016° | 100% | - |

Table 3.1.: Angular acuity for different sensor resolutions in comparison to the human eye.

this solution was not feasible as typical sensors did not deliver the required resolution at a sufficient framerate, were bulky, or not available at all. The technical progress in the field of image sensors and high performance interface standards makes it possible to build high resolution cameras that can solve this issue. Covering the whole monocular visual field of view of a human (Section 2.2) would require an aperture angle of 100° horizontally and 75° vertically. As the vertical field of view amounts to three fourths of the horizontal field of view. Interestingly this is a good match for typical image sensors as they usually use a 4:3 aspect ratio and thus verifying the horizontal resolution will perfectly match the vertical one as well. Ideally the lens and sensor combination is chosen so that the full sensor is illuminated by the scene. Table 3.1 depicts some exemplary numbers for typical image sensor resolutions. Assuming the sensor has to cover the whole field of view with the same acuity as the human fovea, this table shows that a sensor with a width of 6000 px is required to perform at 100% human level. Sensors with a lower resolution will fail to replicate the acuity of the inner region of the human eye. If one considers the distribution of the cone density of the human eye (Section 2.2), it is possible to compile a rough estimate on which region a given sensor resolution falls short (see Table 3.1): A sensor with a resolution of 1024 px will reach only 17% of the foveal acuity and will outperform the eye beyond the inner 11° region. Approximately doubling the sensor resolution to 1920 px will result in 32% of the foveal acuity. This is good enough to outperform the human eye once the inner region of 5°

around the center is left.

Given that there are no sensors available to reach the required horizontal resolution of $6000\,\mathrm{px}$, not to mention the problems caused by the high data rate, the question arises what is the minimum viable compromise? The World Health Organization classifies people with a visual ability better than 30% as having a mild or no visual impairment, 30% to 10% as classified as moderate, 10% to 5% as severe, and less than 5% as blind [77]. Following this definition, the requirement on the camera resolution can be specified to be better than 30% visual acuity. This corresponds to an angular acuity of better than three arc minutes ($0.05°$), which again, translates to a resolution higher than $1800\,\mathrm{px}$ horizontally and $1350\,\mathrm{px}$ vertically. Cameras with a sensor resolution higher than this number can outperform the human eye when the inner region of $6°$ around the center is left. Please note that these numbers are given for the ideal case. A variety of causes such as the lens quality or the optimum focus can influence the overall performance negatively.

These considerations leads to the following requirements:

**Requirement 2: Camera**

$R_{2.1}$: *The eye integrates a camera sensor*

$R_{2.2}$: *It uses a single lens*

$R_{2.3}$: *The field of view covers* $100° \times 75°$ *(H×V)*

$R_{2.4}$: *The acuity is better than* $0.05°$ *in the center region*

## 3.3. Human Motion

As shown in Section 1.1, it is mandatory to fulfill the expectations of the user of a humanoid robotic head. In summary, it is known that robots with human features, deviating from the expected behavior or motion tend to be disturbing. Therefore, proper, smooth, natural, and anthropomorphic motion is essential in order to fulfill the users' expectations. In order to perform in a human-like way, the robotic eye has to feature human-like properties in terms of total range of motion, reachable velocities and accelerations.

The total range of human eye motion spans approximately $100°$ vertically

and horizontally. This places hard constraints on the actuation system when the pleasant design Requirement $R_1$ should not be violated. For example a lever, attached on the lower half of the eyeball, will be moved $50°$ degrees towards the front. In order to prevent the exposition to the user, the attachment point has to be optimized and shifted to the back, which again, conflicts with the space constraints given by the included camera sensor. This might be one of the reason why most designs found in literature support a smaller range of motion. Another reason might be that humans rarely use the full range of motion : The most used, so-called practical field of view, requires a significantly smaller partition of this region (see Section 2.4). It includes the range of $\pm 20°$ vertically and $-30°$ to $10°$ horizontally.

The total range of motion is not the only important factor when it comes to human motion replication. One prominent and noticeable type of human eye motion is the saccade. The replication of these step-like gaze shifts require high velocities and accelerations. As shown in Section 2.6.3, the human eye can reach velocities between $600\,°/\mathrm{s}$ and $800\,°/\mathrm{s}$ with peak accelerations of up to $40\,000\,°/\mathrm{s}^2$. Another motion pattern, the so-called smooth pursuit (Section 2.6.2), is used to track objects moving at low speeds by matching the angular velocities of the eyeball and the target. The replication of this type of motion requires smooth motion control at low velocities. This can e.g. implemented by a using velocity control.

Given the fact that eyelid motion can follow the eyeball motion and the fact that e.g. eye blinks are very rapid movements, the eyelids need to be agile as well. Typical values found in literature (Section 2.8) indicate a minimal eye blink duration $t_{blink}$ of $100\,\mathrm{ms}$.

Summing up these requirements lead to the following list:

**Requirement 3: Human Motion**

- $R_{3.1}$: *Cover the practical FOV of* $\pm 20°$ *(V) and* $-30°$ *to* $10°$ *(H)*
- $R_{3.2}$: *Cover the full range of* $-60°$ *to* $45°$ *(V) and* $\pm 50°$ *(H)*
- $R_{3.3}$: *Support velocities of at least* $600\,°/\mathrm{s}$ *(*$>40\,000\,°/\mathrm{s}^2$*)*
- $R_{3.4}$: *Smooth motion control*
- $R_{3.5}$: *Agile eyelids (*$t_{blink} < 100ms$*)*

## 3.4. Technical Requirements

Beside the requirements that arise from the human model, there are some aspects for a successful implementation that are not directly driven by human characteristics but are equally important for a technical implementation. One example is the emitted noise during motion, which is caused by the difference in the actuation system: The human eye is suspended in muscular and fat tissues and actuated by pairs of muscles. In contrast to electro motors and other mechanical systems, the human muscles move without making any noise. As shown in Section 1.1, humans will interpret and react to emitted noise. It is thus desirable to reduce the noise level as far as possible by constructive means. Maximum allowable noise levels are not specified here as there are no numbers on noise emission given in literature on humanoid robotic heads.

A technical replication of a humanoid eye, performing with similar velocities and accelerations as the human model, will rarely stand still. This means that the selected motors have to be able to withstand this type of continuous motion and are chosen in a way that they will not be overloaded and will not get damaged by overheating.

A direct consequence of Requirement $R_{3.4}$ is the need for high resolution encoders for the motor control loop. Low velocity motion used during the smooth pursuit motion pattern (Section 2.6.2) requires low velocities of less than $30\,°/s$. Please keep in mind that this is for the highest smooth pursuit speed of $30\,°/s$ and human motion includes lower velocities as well. Assuming a typical control loop update rate of $1000\,Hz$, this results in a change of less than $0.03\,°/s$ per control loop update. A typical absolute encoder with a resolution of less than $0.03°$ will register only one counter increment per control loop update, if any at all. It is quite obvious to see that this will not give sufficient smooth motion control. And indeed, conducted tests have shown, that velocity control at such a low resolution will produce jerky, noisy, and coarse motion. The replication of low velocity motion requires that the encoder, giving the feedback to the control loop, needs a higher resolution in order to allow smooth motion control. As later shown in Sections 7.2 and 8.2, typical absolute encoders can not deliver this performance. A common solution for this problem is the use of high resolution

incremental encoders mounted directly on the motor shaft to facilitate an angular resolution that becomes at least one order of magnitude better.

Even though commonly used on humanoid robotic heads, standard hobby grade RC servo motors can not cope with the Requirement $R_{3.4}$. The internal control loop is typically updated with $50\,\mathrm{Hz}$ and does not feature velocity control. This combination does not allow smooth motion as demanded in Section 3.3.

The use of incremental encoders has the drawback that the system needs some kind of initialization of the angular joint position during power up. This initialization can be obtained in two different ways: The most preferred way is the use of an additional absolute encoder that provide an initialization values directly on start up. The second method uses an end-switch-based, active homing motion in order to obtain a known reference position. The latter method has the drawback that all joints using this method will need to perform this time consuming homing motion during power up.

An additional source of measurement inaccuracy can be caused by a gearbox that is inserted between the motor and the end-effector. The geometry and the clearance between mated gear teeth can introduce a significant backlash that can easily reach $1.5°$ for a standard planetary gearhead [78]. The effect caused by this backlash can be seen best when the direction of the motor movement is reversed: The gearbox output angle will not change as long as the slack is taken up by the counter rotation. This effect can be prevented by special gearbox construction techniques and it is desirable to minimize it to zero.

Depending on the mechanical setup, the mechanical system can introduce eye rotation, which can be accepted in certain applications but can be problematic in others. Applications that use the pictures of the individual cameras on their own might tolerate some rotation. Other applications, such as e.g. the calculation of depth images using disparity information from both cameras can not cope with unconstrained camera rotations. It is thus desirable to reduce the eye rotation to a minimum by design.

## 3.5. Discussion

This chapter combined the research findings from Chapter 1 with the knowledge gained by the study of the human eye (Chapter 2) and condensed the insights into a list of requirements for the successful design and control of a human-inspired robotic eye. These requirements can be grouped into three main parts: First of all, matching the users' expectations gives a list of aesthetic and design constraints. Second, the design should feature an active vision system which, again, has to feature human-like characteristics. Finally, the completed system should be able to replicate human-like motion at human performance. This list is complemented by a set of technical requirements that has to be fulfilled at the same time.

In the recent years, significant progress has been made in subareas that play a role in this regard. Their successful combination in a single system that fulfills all given requirements remains an open challenge. On a stand-alone basis, the individual requirements do not represent unsolvable problems as such. However, the particular problem lies in solving these requirements all at once as they tend to contradict each other. For example hiding all actuation mechanics from the users point of view behind a mask, will restrict the range of motion of any actuating lever that connects to the eyeball. The range of motion, in turn, will require a flexible camera cabling assembly that can follow the full range and, additionally, allows the required high velocities and accelerations. The camera connection needs to support a high bandwidth in order to fulfill the camera requirements. The type of cables that is required for this, is usually not flexible enough and requires a custom solution that narrows down the camera selection as a result. This combination problem becomes apparent in the following section, which gives an overview of typical robotic head designs found in literature.

Chapter 4

# Robotic Heads in Literature

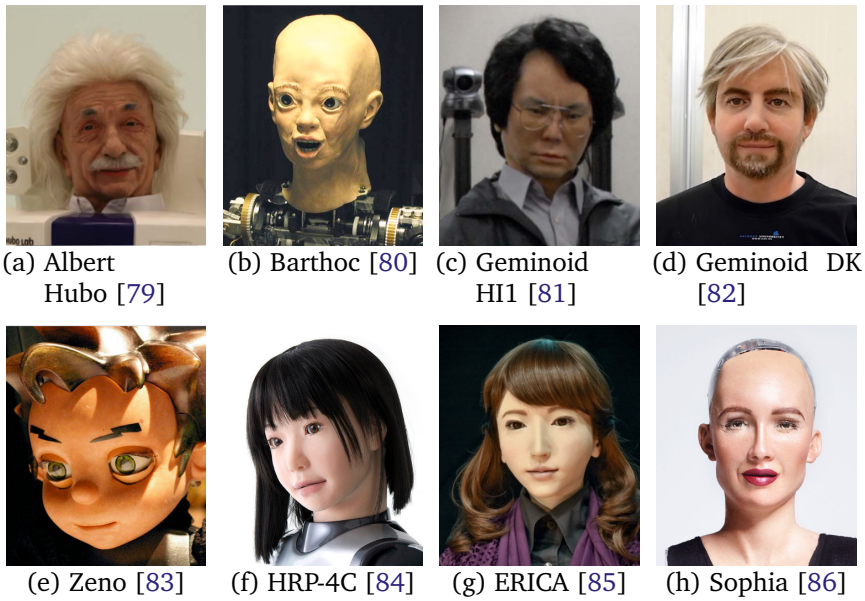
The scientific literature contains numerous designs of humanoid robotic heads. One can differentiate between sensor heads that contain cameras and other input devices and representative heads that have no input capabilities and rely on external capturing devices. But this differentiation is often not that visible from the outside as it is possible to integrate and hide the sensors in a natural way. When it comes to the outer design of a humanoid robotic head, one particularly noticeable difference is the level of abstraction. Some humanoid robotic heads try to replicate the human outer appearance as close as possible by using soft silicon rubber that replicates the structure, flexibility, and tactile properties of skin and muscles. Others use a varying degree of abstraction and, similar to cartoons, maintain only the most apparent and most expressive features. An additional distinctive feature of robotic heads is the way how the facial features are physically represented. In addition to a mechanical representation, there are also robots that use a translucent outer skin or openings in combination with a display device in order to show animated facial features. A further simplification leads to a group of robots that do not feature a closed outer shell and expose mechanical parts to the user. The following list of robotic heads will therefore be divided into four classes: life-like, mechanical, display-based, and comic-like robots. The following list of robots is compiled from the relevant most-cited publications from the following four important robotic conferences ICRA, IROS, Humanoids, and Ro-Man.

## 4.1. Life-like Robotic Heads

Over the years a variety of different life-like robotic heads have been presented. This class of robots maximizes the level of reproduction accuracy by using a low-level of abstraction. A common characteristic is the use of soft materials that replicate the touch and appearance of the human skin.

**Albert Hubo**: An early example of this class of robots is the Albert Hubo robot which was published in 2005 [79] (Figure 4.1a). It features a total 31 RC servo motors for facial and neck actuation. This robotic head does not include cameras in the eyes.

**Barthoc**: Approximately at the same time a similar approach, also using RC servo motors and a silicon-based skin, has been published in [87]. This robot, depicted in Figure 4.1b, was named Barthoc. In contrast to the Albert



(a) Albert Hubo [79] (b) Barthoc [80] (c) Geminoid HI1 [81] (d) Geminoid DK [82]

(e) Zeno [83] (f) HRP-4C [84] (g) ERICA [85] (h) Sophia [86]

Figure 4.1.: A compilation of life-like humanoid robotic heads. Please note that the dimensions of each head are not to scale. Photo credits are given in brackets.

Hubo, this robot featured eyeball cameras.

**Geminoid**: Another prominent example for this type of robots is the Geminoid series by Ishiguro [81]. This line of robots, dating back to 2006, uses a different actuation approach than the previously-mentioned robots. Instead of using RC servo motors, these robots use pneumatic actuators in combination with a silicon-based material that mimics the human skin. These robots were mostly tele-operated where an operator triggered a limited set of predefined motion patterns via a graphical user interface and the operators speech and lip motion was transferred to the robot in real time [88]. In 2011, the same technology was used to produce the Geminoid DK [82], forming an alter ego of the danish professor Henrik Scharfe (Figure 4.1d).

The separation between comic-style and realistic reproduction is not strictly black and white, mixed forms do exist as well:

**Zeno**: The Zeno robot (Figure 4.1e) by Hanson Robotics is another remarkable robot in this category. Again, it uses a silicon-based skin and includes cameras in its eyes [89], [90]. The head features a total of 9 degrees of freedom and uses a mixture of RC servomotors and Dynamixel robot actuators [91]. This robot is particularly interesting as it can be seen as a mixture between comic-style and life-like robots. On the one hand, this robot possesses a silicon-based soft skin, but on the other hand uses abstraction to make it look comic-like.

**HRP-4C**: The HRP-4C robot (Figure 4.1f) is mixing a comic-style body with a realistic human-like head [92]. This robot includes a camera in one of its eyeballs while maintaining the realistic outer appearance.

**ERICA**: In 2016, Glas *et al.* presented ERICA. This robot, shown in Figure 4.1g, is a successor of the Geminoid robot series. It uses silicon skin and pneumatic actuators but does not resemble a living human. Instead, this robot's facial features were artificially modeled according to principles of beauty theory. Its facial features were optimized to be neutral while resembling an attractive robot that people would feel comfortable to interact with. The face and neck feature a total of 16 pneumatic actuators for motion control with an update rate of $20\,\mathrm{Hz}$. Both eyeballs contain an integrated camera with a resolution of $1280\,\mathrm{px} \times 1024\,\mathrm{px}$ $(30\,\mathrm{fps})$.

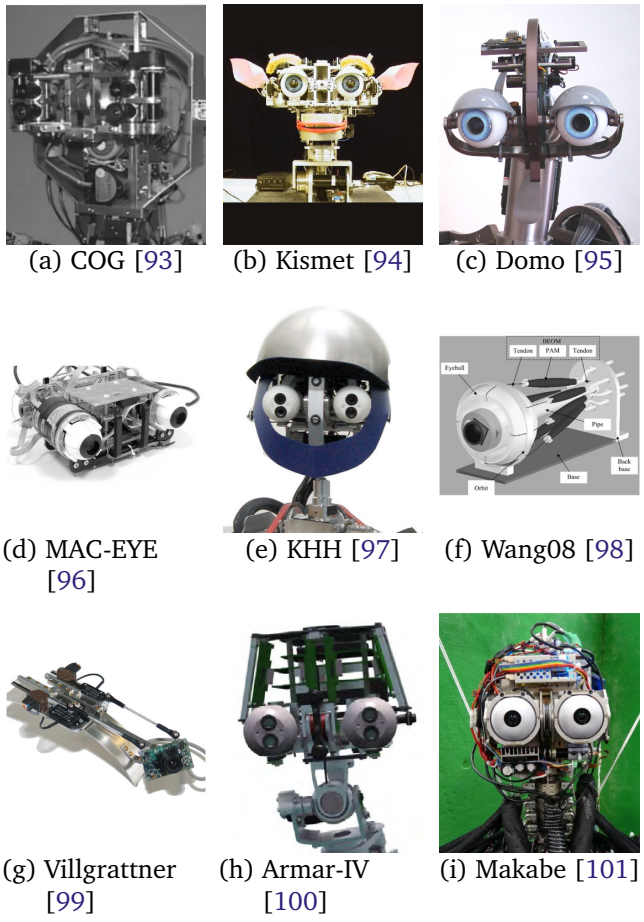
**Sophia**: In the same year Hanson robotics revealed another sophisticated robotic head called Sophia [86]. This version, depicted in Figure 4.1h, features a larger number of servo motors that facilitate a rich set of facial ex-

pressions. This robot advertised for its artificial intelligence that it allows fully autonomous operation and human robot interaction.

## 4.2. Mechanical Robotic Heads

**COG**: One of the early examples in this category is the robot from the Cog project (see Figure 4.2a) [102], [103]. This upper torso humanoid robot and its binocular foveated vision system was presented in 1999. Scassellati designed its visual system by following nature's example, trying to mimic the sensory and sensory-motor capabilities of the human. The eye core structure measures approximately $100\,\mathrm{mm} \times 260\,\mathrm{mm} \times 120\,\mathrm{mm}$ (LxWxH). In order to replicate the high visual acuity of the human retina, this system relied on two different camera and lens subsystems per eyeball. First single camera implementations did exist at that time but the authors state that their experimental nature and the resulting cost, reliability, and availability drove their decision for a multi-camera setup. Therefore, this system combined the view of a $3\,\mathrm{mm}$ wide angle- ($115.8° \times 88.6°$) with a $15\,\mathrm{mm}$ tele-lens ($24.4° \times 18.4°$). The angular acuity of this system reaches human performance ($0.033°$). Based on human values during eye saccades, their system was designed to reach accelerations of up to $79\,640\,°/\mathrm{s}^2$ and resulting high velocities. The original design did not feature any face masks, exposing the bare mechanics. Later work by Edsinger *et al.* added a 3D printed shell [93]. One of the drawbacks of this system is its size: Especially its width, measuring $260\,\mathrm{mm}$, would require a head design that would measure twice that much as found on a typical human. In addition, the dual lens system causes the outer appearance of the eyeball to visually deviate from the human archetype. As shown in Section 1.1, deviations from the expected norm can be disturbing and might trigger hostile reactions. Nevertheless, this system was one of the first used to study and explore a variety of human-like visual motor routines.

**Kismet**: The Kismet robot, presented in the year 1999, is another early example for humanoid robotic heads (see Figure 4.2b). Its design incorporates anthropomorphic features such as neck and eye kinetics in a cute animalistic appearance. Even though this robotic head is not strictly anthropomorphic, it can successfully enhance communication through facial expressions and



(a) COG [93] (b) Kismet [94] (c) Domo [95]

(d) MAC-EYE [96] (e) KHH [97] (f) Wang08 [98]

(g) Villgrattner [99] (h) Armar-IV [100] (i) Makabe [101]

Figure 4.2.: A compilation of mechanical humanoid robotic heads. Please note that the dimensions of each head are not to scale. Photo credits are given in brackets.

social cues [104].

**Domo**: Based on the experience with the design of the Cog robot, Edsinger worked on the design of a new robot called Domo [105], [106] (see Figure 4.2c). This robot, introduced in 2004, featured a new head and eye actuation scheme based on a one by one copy of the design used on the Mertz robot [107]. This head featured a total of seven degrees of freedom, and a pair of Point Gray firewire cameras. According to Edsinger, this head was designed to be able to execute human-like eye movements in terms of the different motion patterns and competitive velocities. The addition of a single degree of freedom eyelid improved the expressiveness of the overall system. The size of this head has been reduced to approximately $157\,\mathrm{mm}$, resembling a more human-like size in comparison to the older cog robot.

**MAC-EYE**: In 2005, Biamino *et al.* presented an interesting eye concept called the MAC-EYE system [96]. This prototype is actuated by four motors in an antagonistic setup using strings. This setup is an interesting sample as it mimics the muscles and tendons of the human oculomotor system. The drawbacks of this system are that it lacks a high resolution camera and the actuating tendons are running visibly over a large portion of the eyeball. The publication does neither include numbers on the available range of motion nor the reachable velocities.

**Karlsruhe Humanoid Head**: Similar to the ”dual-camera per eyeball” approach of the COG robot, the Karlsruhe Humanoid Head (KHH), depicted in Figure 4.2e, was presented by Asfour *et al.* in 2008 [97]. It comes as a realistic human sized head without an outer shell measuring approximately $180\,\mathrm{mm} \times 165\,\mathrm{mm} \times 180\,\mathrm{mm}$ (LxWxH). This head was developed with the goal to resemble human-like characteristics in terms of motion, response, and high visual acuity. The capabilities of the human retina are emulated by a dual camera setup featuring a wide- and narrow-angle lens.

**Wang08**: Besides traditional motor and piezo-based systems there are also systems using pneumatic actuators: For example the three DOF system presented by Wang *et al.* [98] is based on pneumatic air muscles that attach to the eyeball (Figure 4.2f). These artificial muscles are made by the Shadow Robot Company [108] and have a similar contraction-ratio and force displacement as human muscles. This eye prototype concept is quite interesting in terms of actuation principle, however there are no numbers given concerning the eyeball diameter, attainable velocities, or reachable acceler-

ations.

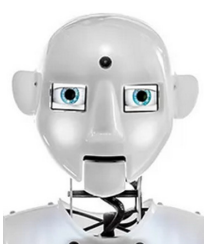
**Villgrattner**: Another interesting concept has been published in 2009 by Villgrattner *et al.*: The initial prototype [99] has two degrees of freedom and has been extended to three DOF in 2010 [109]. Both designs use a complex piezoactuator-driven parallel kinematic system to move the on-board camera at speeds way beyond the human capabilities of up to $1500\,°/s$. Incorporating this into an eye visually matching a humanoid robot would result in a very capable and agile system, however matching the center of rotation of an eyeball with the current bearing point of the parallel system would result in an impractically large eye diameter.

**ARMAR-IV**: In 2013 Asfour *et al.* published the design of the Armar-IV Robot [100] (Figure 4.2h). This design is based on previous work on Armar-III and the Karlsruhe Humanoid Head (Figure 4.2e). Both eyes feature foveated vision-based on a dual Point Grey Dragonfly 2 camera setup. Each eye has a separate pan and tilt actuator. There is no information given concerning the attainable range, velocities, or acceleration.

**Makabe18**: In 2018, Makabe *et al.* presented their "movable binocular highres eye camera unit" [101]. This design (Figure 4.2i) is small enough to be integrated into the head of their humanoid robot Kengoro [110]. Both eyes include a 4K camera running at $30\,\mathrm{fps}$ and share a common tilt axis with individual pan axes. Each eye can reach $-20°$ to $6°$ and $-22°$ to $21°$ vertically and horizontally, respectively. The horizontal and vertical resolution is specified at $0.15°$. The maximum reachable velocity amounts to $300\,°/s$.

## 4.3. Display-Based Robotic Heads

This category of robotic heads exchange some or all actuated parts of the human face by a projection system or a build-in screen. The aspect, that a screen or projection represents the robots' eyes, prevents the integration of movable camera eyes at the same location. With movable cameras being one of the key requirements of this thesis, this overview is just given as a reference. In the following, this robot class will not be analyzed in more detail.



(a) Robothespian [111]

(b) SociBot [112]

(c) ISR [113]

(d) R1 [114]

(e) Furhat [115]

Figure 4.3.: A compilation of projection- and display-based humanoid robotic heads. Please note that the dimensions of each head are not to scale. Photo credits are given in brackets.

**Robothespian**: One example of this class is the head of the Robothespian robot [111]. This head features a combination of a mechanical mouth and two LC-Displays as the robots' eyes (Figure 4.3a).

**SociBot**: There is also a second iteration of the head design that uses a back-illuminated diffuse mask that is called SociBot [112] (Figure 4.3b).

**ISR**: The ISR robotic head uses a total of three LC-Displays, one for the mouth and two for the eyes [113]. The displays are used to show animated facial expressions (Figure 4.3c).

**R1**: One very futuristic looking example is the R1 robotic head by Parmiggiani *et al.* [114]. It uses a custom made 3D-shaped LED matrix to represent facial expressions (Figure 4.3d).

**Furhat**: The Furhat robot (Figure 4.3e) makes the maximum use from this concept. The face mask is made from a diffuse material and a rear projection system projects an animated face onto the mask [115]. The head

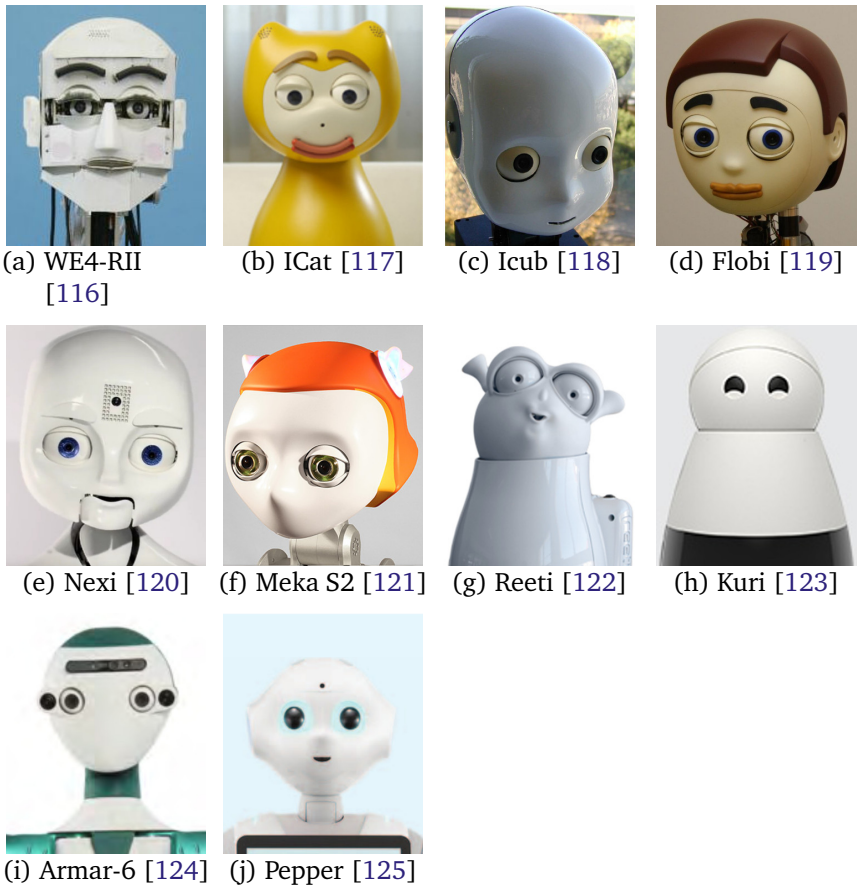
is $16\,\mathrm{cm}$ wide and $22\,\mathrm{cm}$ high, which resembles a typical human head size. A fur hat is used to cover and hide the mechanical and technical structure of this setup. The head base is articulated by a two degree of freedom pan-tilt unit.

## 4.4. Comic-Style Robotic Heads

This class of robots uses a more pronounced level of abstraction than the life-like robots and tries to hide the internal mechanics. Typically, a selection of key features of the human face are adopted. Single aspects of a human face are shown in an attenuated or reduced form and others are left out. As with the previously introduced classes of robots, many publications on this type of robots lack technical specifications of the eye region.

**WE-4RII**: The Humanoid Robot WE-4RII (Figure 4.4a), built by the Waseda University, features a remarkable robotic head in terms of degrees of freedom [126]: This head was originally presented in 2002 and upgraded in 2003. Its eyeballs can reach angular velocities of up to $600\,°/\mathrm{s}$ and have three degrees of freedom with a common tilt axis shared by both eyeballs. Each eye contains a CCD camera with a resolution of $320\,\mathrm{px} \times 240\,\mathrm{px}$. A special feature is the expressive eyelid construction of this robot, which has a total of six degrees of freedom, with two degrees of freedom allowing the upper eyelid to rotate. The velocity of opening and closing the eyelid reaches up to $900\,°/\mathrm{s}$. The lower eyelids are mechanically coupled to the eyeball tilt axis, which facilitates coordinated eyeball following eyelid motion without active control. The overall dimensions of the eye region amount to a width of $180\,\mathrm{mm}$ at a height of $140\,\mathrm{mm}$. The total head breadth can be estimated to $\approx$200 mm. Despite its size and the fact that there is only a partially covering shell that does not hide all the internal mechanical parts from the user, this robot is remarkable for its time.

**iCat**: The Philips iCat, shown in Figure 4.4b, resembles a combination of animalistic and human features and was built to study various aspects of human robot interaction [127]. This robot, published in 2005, is probably one of the earliest comic-like models featuring a completely self-contained, comic-like outer appearance that does not expose any mechanical components to the user. Unfortunately this robot does not feature camera eyes.



(a) WE4-RII [116] (b) ICat [117] (c) Icub [118] (d) Flobi [119]

(e) Nexi [120] (f) Meka S2 [121] (g) Reeti [122] (h) Kuri [123]

(i) Armar-6 [124] (j) Pepper [125]

Figure 4.4.: A compilation of comic-style robotic heads. Please note that the dimensions of each head are not to scale. Photo credits are given in brackets.

**iCub**: Another remarkable anthropomorphic robotic head was developed for the iCub robot in 2004 [128] (see Figure 4.4c). It possesses anthropomorphic neck- and eye-kinetics combined with additional expressive features. The eyes were designed to span an oculomotor range of $90°$ and $80°$ and to reach a maximum velocity of $180\,°/s$ and $160\,°/s$ for the pan- and tilt-actuator, respectively [128]. More than 30 iCub robots have been built so far and were distributed to research institutes in Europe, the US, Korea, and

Japan [118]. The fact that the complete design of this robot is distributed under an opensource license is one of its unique features. Later iterations feature a set of four eyelids that are actuated by a single servo motor. Similar to the Cog robot, this single degree of freedom reduced the expressiveness of the eyelids as e.g. eyeball following motion could not be executed. The eyebrow and mouth implementation is based on LEDs shining through the facial masks. An upgraded mechanically actuated mouth construction has recently been presented [129].

**Flobi**: The Flobi robot [119], invented and built in-house at Bielefeld University, follows a similar approach with its comic-style outer appearance (Figure 4.4d). This head comes at the size of a typical adult human head and its large number of actuators facilitate the playback of a wide range of emotional expressions. The head breadth amounts to $16.5\,\mathrm{cm}$. Please refer to Chapter 6 for a detailed description of this robotic head.

**Nexi MDS**: Another expressive and capable robot is the Nexi MDS robot that was presented in 2008 (Figure 4.4e) by the MIT Media lab [120]. Published videos that show the robot replicating human motion, based on motion capture recordings show a wide range of emotions and very natural and smooth motion. Published facts and numbers are sparse on this head, but based on the known size of the built in Swissranger 3D camera this head can be approximated to have a breadth of around $240\,\mathrm{mm}$ or 1.6 times of a typical human head.

**Meka S2**: The Meka Robotics S2 head (see Figure 4.4f) follows a similar design concept. It features a comic-like, self-contained exterior shell and the eye mechanism was marketed to have very low inertia and high dynamic performance levels optimized for smooth pursuit and saccade motion [121]. The anthropomorphic design is complemented by a pair of movable fairy ears that can light up in different colors. The information on this robot is sparse since Meka robotics has been acquired by Google and shut down its operations.

**Reeti**: Similar to this, the Reeti robot [130] mixes fantasy and human elements. This robot, released in 2011, consists of a human-like head and a non-functional torso (Figure 4.4g). It features a camera and an upper eyelid per eye. It was developed for emotion feedback and human robot communication. There is no scientific publication on this robot yet, considering the promotional video it looks like this robot most likely uses RC servo motors

that offer no velocity control for actuation. The motion pattern playback on the manufacturer's video shows a lot of jerk and shaky motions.

**Kuri**: A different level of abstraction is used on the Kuri robot that was presented in 2017 [123]. The head of this robot, shown in Figure 4.4h, is reduced to the absolute minimum, resembling a nearly perfect sphere with small openings as the eyes. These openings contain fixed high resolution cameras and actuated eyelids. Promotional videos indicate the expressiveness of the implementation [131].

**Armar-6**: Unlike its predecessors such as the Armar-IV (Figure 4.2h), the latest generation of the Armar robot series features a closed outer facial mask. The Armar-6 (Figure 4.4i), presented in 2018, features a pair of Flea3 stereo cameras integrated into eyeballs that are mounted 27 cm apart [124]. The eyeballs are not articulated in the current design.

**Pepper**: Another recent example of a humanoid robotic head is used on the Pepper robot [125]. This robot, shown in Figure 4.4j, is a commercial product made by the company Softbank [125]. It features a matching outer appearance and a three degrees of freedom neck. Unfortunately the eyes can not move, which is why this robot will not be further elaborated in this thesis.

## 4.5. Discussion

As stated in the introduction, this work will focus on the design of a human-inspired eye that can be used in a humanoid robotic head. Based on the numerous advantages of an abstract comic-style design that have been deduced in Section 1.1, this design is targeted, but not limited to the application on a robot of this class. Minor modifications, for example by using soft tissue for the eyelid, would allow this eye to be used on life-like robots as well. As the previous list of robots suggests, publications on humanoid robotic heads rarely include technical specifications of e.g. the oculomotor range or the eyes' reachable velocities and accelerations. However, these numbers are particularly important when it comes to the perception by the user and thus play a significant role for the implementation of a system that fulfills the given requirements as introduced in Section 1.1.

For the robots presented in Section 4.1, there are no such numbers given

in literature. Robots of the second class, using display devices to show facial features, can neither move the eyes nor include cameras inside the eyeball. This is why the following assessment does not include these two types of robots. In general, further evaluation will only include robot eyes that can move. A summarizing tabular comparison can be found in Table 4.1: The single systems have been analyzed by means of the requirements that were introduced in Chapter 3. As no publication mention the noise level produced by the eye actuation system, this table will not show any data on that. However, it can be assumed that RC servo based implementations will produce an audible and disturbing gearbox noise by design. Due to the lack of published data, this requirement can not be verified at this point. In addition, some of the head breaths of the individual systems had to be estimated based on referencing known object dimensions in the published pictures. These values can be identified by the prefixed approximation symbol. A complete fulfillment of the requirement is denoted by a green tick, a partial compliance with an orange tick in parentheses, a non-fulfillment is shown by an 'x', and missing data is denoted by a question mark. The score row represents the sum of all partially ($+0.5$) and completely fulfilled ($+1$) requirements.

Individual aspects and single numbers in terms of performance are striking if considered separately. Systems such as the iCub, the iCat, and the Flobi do well in terms of outer design and size. Others, such as the WE4-RII and the Villgrattner prototype show suitable numbers in terms of reachable velocities. The dual camera approach of the Cog, the KHH, and the Armar IV robot deliver a suitable field of view and acuity — however the eyeball design fail to deliver a human-like outer design. Considering the overall table, it becomes quite obvious that publications rarely include exact numbers in terms of the reachable range of motion or velocities. And even if there are numbers given, they are often based on the design specifications and not on actual measurements taken on the final hardware.

| Robot | Year | Head Breadth | $R_1$: Design | | | $R_2$: Camera | | | | $R_3$: Human Motion | | | | | Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $R_{1.1}$: Shell | $R_{1.2}$: Eyeball | $R_{1.3}$: Eyelids | $R_{2.1}$: Included | $R_{2.2}$: One Lens | $R_{2.3}$: FOV | $R_{2.4}$: Acuity | $R_{3.1}$: Practical | $R_{3.2}$: Full | $R_{3.3}$: Velocity | $R_{3.4}$: Smooth | $R_{3.5}$: Lids | |
| Cog | 1999 | 26 cm | ✗ | ✗ | ✗ | ✓ | ✗ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | ✗ | 6 |
| Kismet | 1999 | ≈25 cm | ✗ | ✓ | (✓) | ✓ | ✓ | ✗ | ✗ | ? | ? | ? | ? | ? | 3.5 |
| WE4-RII | 2003 | ≈20 cm | ✗ | ✓ | ✓ | ✓ | ✓ | ✗ | ✗ | ? | ? | ✓ | ✓ | ✓ | 7 |
| Domo | 2004 | 15.7 cm | (✓) | ✓ | (✓) | ✓ | ✓ | ✗ | ✗ | ? | ? | ✓ | ✓ | ? | 6 |
| ICub | 2004 | 15.2 cm | ✓ | ✓ | (✓) | ✓ | ✓ | ✗ | ✗ | ✓ | ✗ | ✗ | ✓ | ✗ | 6.5 |
| iCat | 2005 | 18 cm | ✓ | ✓ | (✓) | ✗ | ✗ | ✗ | ✗ | ? | ? | ? | ? | ✗ | 2.5 |
| MAC-Eye | 2005 | - | ✗ | (✓) | ✗ | ✓ | ✓ | ✗ | ✗ | ? | ? | ✗ | ? | ✗ | 2.5 |
| KHH | 2008 | ≈18 cm | ✗ | ✗ | ✗ | ✓ | ✗ | ✓ | ✓ | ? | ? | ✗ | ? | ✗ | 3 |
| Meka S2 | 2008 | ≈20 cm | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✗ | ? | ? | ? | ? | ? | 5 |
| Nexi | 2008 | ≈24 cm | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✗ | ? | ? | ? | ? | ✓ | 6 |
| Wang | 2008 | - | ✗ | (✓) | ✗ | ✓ | ✓ | ✗ | ✗ | ? | ? | ? | ? | ✗ | 2.5 |
| Villgrattner | 2009 | - | ✗ | ✗ | ✗ | ✓ | ✓ | ✗ | ✗ | ? | ? | ✓ | ✓ | ✗ | 4 |
| Flobi ('09) | 2009 | 16.5 cm | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✗ | (✓) | ✗ | ✓ | ✓ | ✓ | 8.5 |
| Reeti | 2011 | ≈15 cm | ✓ | ✓ | (✓) | ✓ | ✓ | ✗ | ✗ | ? | ? | ✗ | ✗ | ? | 4.5 |
| Armar IV | 2013 | ≈18 cm | ✗ | ✗ | ✗ | ✓ | ✗ | ✓ | ✓ | ? | ? | ? | ? | ✗ | 3 |
| Armar VI | 2018 | ≈45 cm | ✓ | ✓ | ✗ | ✓ | ✓ | ✗ | ✗ | ? | ? | ? | ? | ✗ | 4 |
| Makabe | 2018 | 15.2 cm | ✗ | ✗ | ✗ | ✓ | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | ? | ✗ | 2 |

Table 4.1.: A Requirement analysis for the related robotic heads and eye prototypes.

Summarizing it can be said that none of these systems gets to fulfill all the requirements laid down. Under the prerequisite of a closed outer shell, the robots that are most likely close to the given requirements are the iCub and the Flobi robot. Both robots form a good basis for the future development as they fulfill several requirements and the CAD data is accessible. The iCub is an open hardware project with all its design data being available for free and the Flobi is an in-house development. In order not to limit the integration of the presented prototypes into any given robot platform, the following prototypes will be designed in a way that they do not rely on any specific internal structure. The single eye actuation systems for the left and the right eye will be individually place- and mountable in any given head platform, given that there is enough free space. The disadvantage of the iCub robotic head is that it includes the on-board computer and there is a very limited amount of free space available for the eye and eyelid actuation system. The integration into the iCub would require an extensive redesign of the core computing unit that occupies the head. Based on these considerations, the final integration of the final prototype presented in this thesis will therefore take place on a Flobi based design.

Chapter 5

# Reproduction of Human Motion

The best humanoid robotic head hardware is worth nothing without proper motion control. One possible way to generate human-like motion is to track and record human motion for live display or even time shifted playback. However, typical use case requires a more dynamic approach which allows the robot to e.g. rapidly shift the gaze direction based on scene input. In order to cope with this issue, different modes of tele-operation with a varying degree of scene feedback and various levels of interactivity have been described in the literature. A second class of control systems generate artificial human-like motion based on a model. One advantage is that these systems do not rely on a on-site human operator and can be very reactive. The separation between recorded and generated motion is not strictly black and white, mixed forms, that combine on-the-fly customized pre-recorded datasets and automatically generated motions, can be used as well.

## 5.1. Motion Capture

Tracking the human face is a crucial step for the understanding of human motion and the derivation of a model. Facial expression motion capture solutions can be differentiated into two major groups: marker-based or marker-less systems, with both groups have their pros and cons. marker-based systems rely on artificial markers that are attached to the actors face. A prominent example in this category is Vicon motion capturing system which uses retro-reflective spheres that are attached to various facial and other anatomic features using a skin friendly adhesive [132]. marker-based systems require a tedious and time consuming preparation of the actor. This

disadvantage is eliminated by the use of marker-less systems that are based on a captured 2D or 3D image of the actors face. One example of a 3D based facial motion capture system is the work by Weise *et al.* that was later commercialized under the name faceshift [133]. In addition, recent developments in the field of machine learning allow the tracking of an operators face by e.g. using a single 2D color camera [134]. Data recorded by one of these systems can be either used study the human motion in order to compile a model or use the more direct approach of playing back recorded datasets.

## 5.2. Tele-Operation

Pre-recorded human motion can be played back smoothly and does indeed look very real with an impressive level of detail. However, the plain reproduction of a pre-recorded dataset is of limited use when it comes to live interaction between a human and the robot. Placing the operator in a direct loop by using immediate scene feedback can facilitate a more realistic operation. With the final objective being a completely computer driven robot control system, one might think why is tele-operation needed at all? Tele-operated systems date back long in history of robotics, as already mentioned in the thesis introduction, one of the earliest known tele-operated robots has been developed in 1769 by Wolfgang von Kempelen. This robot, nowadays also known as the mechanical turk, gave the illusion of an autonomous chess playing robot. The actual mode operation remained a secret until the year 1820 when Robert Willis unveiled that this robot was in fact tele-operated by a person hiding in the cabinet below [135]. A more recent example are the (partially) tele-operated Geminoid humanoid robot series by Hiroshi Ishiguro. Their autonomous operation can be augmented by partial tele-operation by a human operator: A limited set of predefined motion patterns (idle, speaking, ...) can be triggered via a graphical user interface. In parallel, the operators speech and lip motion can be recorded and played back on the robot in real time. This system uses an external motion tracking systems and face markers [81], [88].

These systems can be used to gain experience on how a robotic system should behave and on how different users react on specific motion and be-

havior patterns. Different operational modes, movement-, and behavior-patterns can be tested without actually building an automated control system and thus assessing beforehand where to set priorities in the implementation of such systems. The second use case is the recording of human motion for playback at a later stage. Even though the playback of recorded motion is of limited use in an interactive scenario where the motion and especially the robots' gaze has to be dynamically adjusted to the scene, playback of recorded motion gives valuable first insights. One field of application is the assessment of likability and acceptance of a robot prototype that is playing back a recorded dataset in a user study.

## 5.3. Motion Transfer and Replication

The replication of human motion can be divided into two major groups, either using pre-recorded datasets that are modified to be played back on the target character or model-driven approaches. The first approach, using the playback of recorded human motion, requires an actors face to be tracked and the motion data is extracted, processed, and stored. During playback, the facial expressions have to be transferred to the target character. Different methodologies exist in literature for this step: When the physical differences between target character and executing actor are small enough, a relatively direct mapping technique, such as presented in [136] or [137], can be applied. Recent work by Ribera *et al.* tries to solve this mapping problem by using a facial prior and an automatic range of motion alignment. However, for computer generated animations with a major morphological disparity between actuator and character this mapping becomes inappropriate and a different approach might give better results: Mapping the captured data in an expression space (e.g. using the facial action coding system [139]) from different facial muscle group actuations to the corresponding muscle groups on the virtual or robotic character gives much better results [140]. Similar concepts exist for the generation of the human gaze: Gaze models can be differentiated between two classes of gaze models: Data-driven, based on human recorded datasets, and the procedural approach. One example for a data-driven approach is the system presented in [141]: Emotionally expressive gaze datasets were recorded beforehand

and are then warped and over blended before being superimposed with a model-driven gaze. There are also mixed models, e.g. the ”Eyes Alive” model by Lee et al [142] that is based on an empirical model of saccades combined with a statistical model-based on eye tracking data.

Unfortunately most of the models found in literature do not take full eyelid motion into account. If present at all, eyelid animation is often limited to blinking, covering only a subset of human eyelid motion patterns. As stated in Section 1.1, due to the anthropomorphic outer appearance, such a deviation from the users' expectations, can cause disturbances and degrade the overall perception. A system proposed by Deng *et al.* tries to overcome this limitation by recreating combined eye and lid motions based on recorded datasets using texture synthesis [143].

The application of data-driven models, based on recordings from humans, for the animation of anthropomorphic virtual characters with slightly different proportions, kinetics, or geometries is not straightforward. These variations, if not handled properly by complex morphing and adaption of the recorded data, cause deviations from the norm, which again, degrade the overall perception. In order to address these issues, Pejsa *et al.* proposes a procedural system to overcome those issues by adaption and remapping to their target character [144].

One example for a model-based control framework applied on anthropomorphic robotic heads is the emotion and gaze control architecture that was developed for the iCat [117]. This framework incorporates blinking but does not include eyeball pursuing motion. Another example for a procedural control framework is the iKinGazeCtrl module developed for the iCub [145]: This approach features an anthropomorphic, model-based combined actuation of neck- and eye joints — but again, no eyelid animation at all. Unfortunately the iCubs' eyelid system is limited by a single actuator to move all four eyelids at once. This structure does not allow the variety of motions required for the full eyelid animation, however plain blinking can be added on top using a separate blink controller [146].

## 5.4. Discussion

Even though the topic is widely researched and every robotic head with human kinetics will benefit from model-based actuation systems, the accessibility of actual implementations or frameworks for gaze control is sparse. The published implementations are tightly coupled to a single robot platform and not customizeable to the constraints of a different robot. For example iKinGazeCtrl does not feature eyelid animation as the iCubs' kinematics and its limited degrees of freedom concerning the eyes do not allow smooth eyelid animation.

This fixation to a single target robot is common and might be one of the reasons why there are no comparative studies of different robotic heads that are controlled by the same gaze control software. Further developments on robotic heads would benefit from comparative studies regarding expressiveness, likability, and liveliness between different robot platforms. In order to change that, the proposed motion control library, introduced in Section 10.6, uses a high level of abstraction. In addition, it is kept configurable wherever possible in order to allow the application of this library on any robot platform. A unique feature of the proposed library is the possibility to reduce the maximum accelerations and velocities used during the path planning in order to adept the framework on less agile robot platforms without producing jerky or incorrect motion patterns.

# II

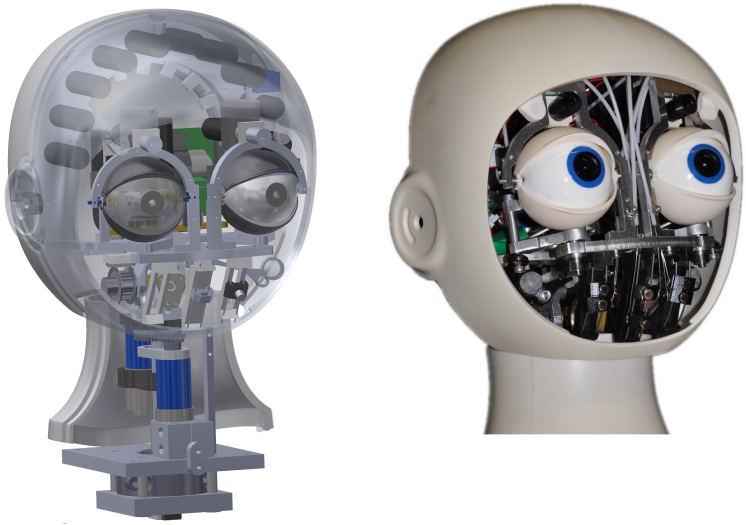
# Realization

With the previous part of this thesis working out the human capabilities, deriving the key requirements for the next generation of humanoid robotic eyes, and analyzing different humanoid robotic heads and control methods found in literature in that regard, this part will describe the realization of an overall system that is based on- and combines all this accumulated knowledge. It is pointless to have an isolated robotic eye without being integrated into a suitable humanoid robotic head. In order not to start completely from scratch, a suitable robotic humanoid head is chosen as a base and its current state and areas that need improvement are recorded in Chapter 6. Subsequently, a total of two different mechanical eye prototypes, each evaluating a different actuation approach, are introduced in Chapters 7 and 8. These prototypes were presented at the IROS conference in 2012 [26] and the ICRA conference in 2019 [27], respectively.
With this improved hardware in place, capable to replicate human performance and motion, the question arises, what is the correct way to control a technical replica? In order to address this question, Chapter 9 introduces a bi-directional motion capture system. This system allowed it to gain first insights on how humans move their eyes in general and, moreover, it allowed to observe the complex interplay of eye- and neck-muscles at first hand. The combination of insights of these experiments and the operation of the prototypes finally led to Chapter 10, which presents the developed motion control hardware and software stack.

Chapter 6

# The Initial Flobi Design

As introduced in Section 1.1, it is not expedient to show the mechanics of a human-like robotic eye prototype without a matching shell. The prototype has to be integrated into a robotic head. Based on the analysis of the requirement compliance in Section 4.5, the most logic starting point was identified as the Flobi robot. Before starting with the actual realization of an improvement, it is mandatory to document and analyze the design state that marks the starting point for this work in full detail: The original design of the outer appearance of the Flobi robot has been presented in Frank Hegels doctorate thesis [25]. This publication marks the starting point of the work presented here: Hegels design of a sophisticated outer appearance based on a list of key requirements for a pleasant robotic head laid the foundation for further enhancements. The mechanical and actuated design was initially developed by Mathias Hackel from the company Mabotic. It combined Hegels outer design of the shells and enhanced it by actuators, the inner core structure, and enriched it by state-of-the-art sensing functionality. This robotic head has a breadth of $16.5\,\mathrm{cm}$, a height of $19\,\mathrm{cm}$, and a depth of $18\,\mathrm{cm}$. It features a total 18 degrees of freedom and it was initially meant to use standard, analog RC servo motors for actuation. Early tests showed significant noise emissions and, as those motors allowed no speed control, jerky motion patterns. As shown in Section 1.1, these unwanted artifacts tend to be disturbing to the user and adversely impact human robot interaction. This initial design was never published, instead it was re-iterated and these flaws were fixed by reworking the actuation concept. Due to limited space, tightly integrated motion control hardware was necessary. This was implemented by means the custom motion control circuit boards that are presented in Section 10.1. This initial, fully actuated Flobi design was presented in [119].



(a) CAD front view (b) Real Robot

Figure 6.1.: The first iteration of the Flobi design shown without the face mask (2009).

With the improved performance and smooth motion control in place, a new weak spot was identified: The original design was based on bowden cables for eyelid and the mouth actuation. A CAD rendering and a real world picture of this iteration is shown in Figure 6.1a and 6.1b, respectively. The bowden cables, depicted in Figure 6.1b, run through white PTFE sleeves and are actuated by levers that are attached to associated DC motors. The length of these bowden cables and their disadvantageous routing introduced a lot of friction and moving the actuators required a lot of power. Beside limiting the overall performance in terms of reachable velocities, this put a lot of strain on the motors. Which, in turn, caused jerky motion and overheating issues of the eyelid and mouth motors.

The eye actuation scheme of this early prototype was using a traditional four bar linkage configuration. Exemplary for the whole system, the relevant mechanical parts are depicted in Figure 6.2, on the basis of the right eye:

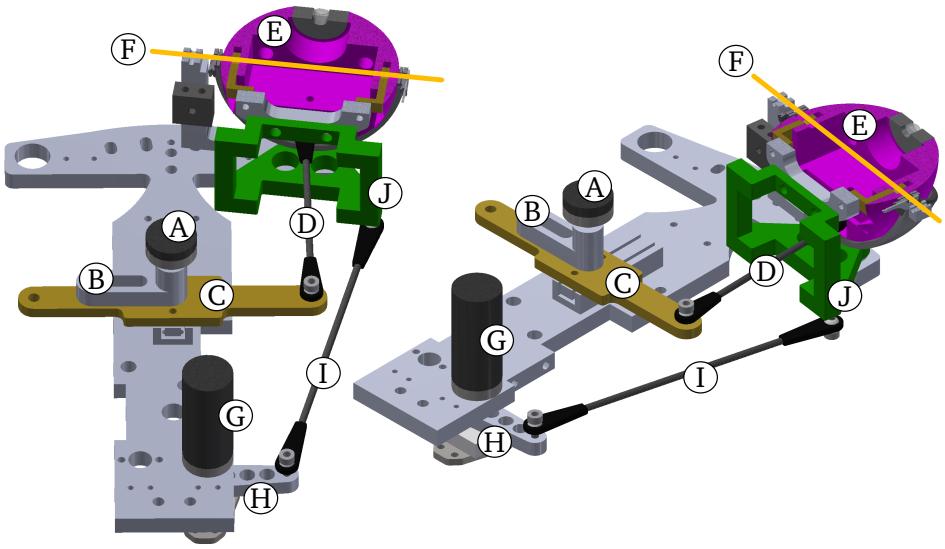



Figure 6.2.: CAD rendering of the actuation principle of the eyes as present on the first iteration of Flobi (2009).

Both eyes share the common motor (A), lever (B), and slider (C) for the eye tilt actuation. A separate rod, depicted as (D), is connected to the main eyeball sphere (E) and transfers the sliding motion to the tilt rotation around the axis (F). There is an additional rod connected to the left eyeball (not shown here). Each eyes' pan rotation is driven by a single motor, here exemplary depicted as (G) for the right eyeball. The motors rotation is directly coupled to the eyeball by using lever (H), rod (I), and the structural part (J) that acts as the final lever.

The choice of this motor-gearbox combination and especially the lever-slot mechanism (B) introduced some slack and backlash during operation. In addition, the lever dimensions, the choice of attachment points, and occurring collisions with other structural parts including the eye lids, limited the range of motion for the eye pan- and tilt-angle to $\pm 30°$ and $\pm 20°$, respectively. This is substantively lower than typical human figures introduced in Section 2.4. Despite this, other human characteristics such as the achievable velocity, acceleration, and the human-like appearance was matched quite well. Please refer to Chapter 12 for a throughout evaluation of the

characteristics, accuracy, and performance of this prototype.

This thesis has its focus on the replication of human eye motion and thus the mechanical properties of the eye region are the most important properties at a first glance. However, at a second, more detailed look, this is only half of the story. As shown in Section 2.7, eye motion is often accompanied by matching neck motion and therefore, has to be considered as well. The initial prototype of the neck actuation mechanism, specifically the geared transmission and the belt driven part depicted in Figure 6.3, both caused jerky motion and introduced mechanical play. In that design, the neck pan

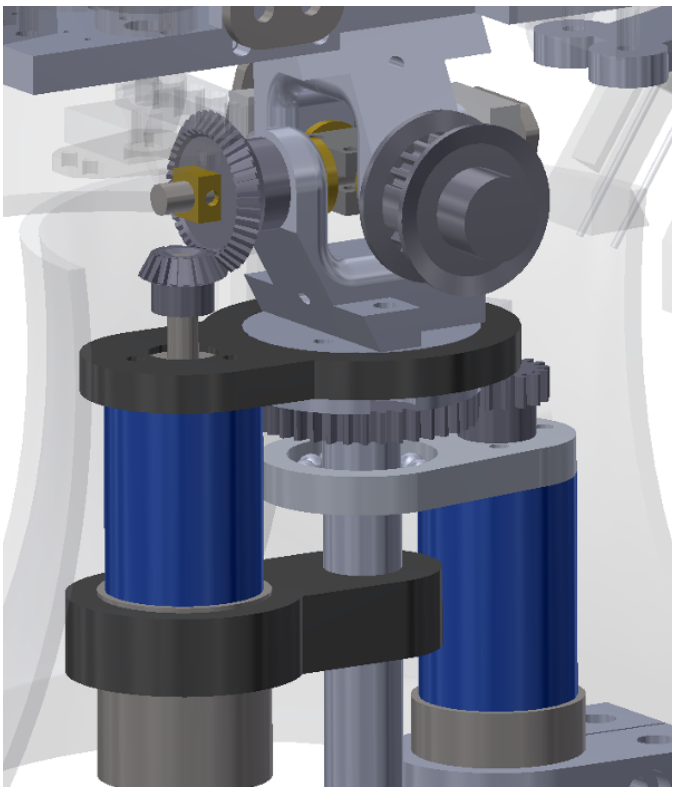


Figure 6.3.: CAD rendering of the first iteration of the Flobi neck design.

motor-gearbox combination drives a second, single stage gearbox that consists of two coarse gears. The same applies for the roll motor with the last, single-stage gearbox redirecting the axis of rotation by 90 degrees. The neck

tilt motor is coupled via a toothed belt to the actuator. This design uses Faulhaber 2224 series [147] motors in combination with matching 20/1 [148] series gearboxes. The combination of the gearbox, specified to possess ”less than one degree” [148] of backlash with the additional backlash of the external single stage gear- and belt- transmission led to a significant overall backlash and gave a wobbly impression. Different actuation techniques

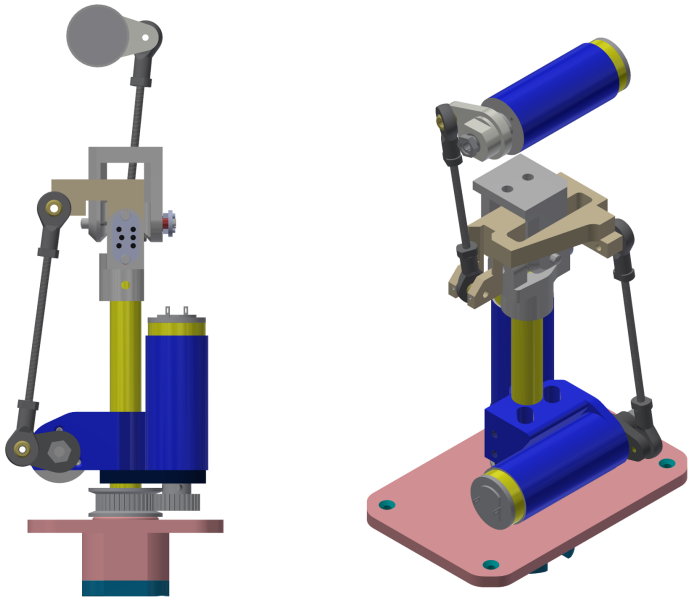


Figure 6.4.: CAD rendering of final iteration of the Flobi neck design.

have been tested and evaluated by the means of their mechanical stability and noise emissions. For example a second iteration, based on a lead-screw linear actuator, had to be discarded due to non-natural and irritating noise emissions.

The final design, depicted in Figure 6.4, uses special zero backlash gearboxes in combination with levers that possess very little play in order to overcome previous issues caused by mechanical play and noise emissions. This design is based on the same 2224 series motors that were initially used but combines them with the Faulhaber 22/5 zero backlash spur gear head [147], [149]. The pan axis is driven by the same motor gearbox com-

bination using a pre-tensioned belt transmission with a ratio of 36:28. All three axes uses the same gearbox transmission ratio of 69.2:1. These gearboxes use a special technique to reduce the backlash to a minimum, the datasheet specifies zero degrees play at no load. Even though no numbers on noise emissions are given, Faulhaber recommends this gearbox for low noise applications. And indeed, this motor gearbox combination proved to be very quiet during operation and the emitted noise is pleasing to the ear. From all tested combinations it emitted subjectively the most pleasant motor noise of all designs. An IE2-128 series [150] relative encoder with 128 lines per rotation delivers 512 ticks per full motor shaft rotation or 35430.4 ticks per gearbox output shaft rotation. In combination with the belt transmission ratio the final pan axis resolution amounts to $0.0086°$ per encoder tick or 116 ticks per degree of end-effector rotation. The transmission ratio of the different lengths of the neck tilt actuation levers result in a final resolution of $0.0054°$ per encoder tick or 185 ticks per degree of tilt axis rotation. Finally, the neck roll angle resolution amounts to $0.0065°$ per encoder tick or 153 ticks per degree of final neck roll rotation. At this point the neck noise emissions were drastically reduced and smooth and accurate velocity and position control was established for the neck joints. However, the performance of the eye and eyelid actuation system was still not satisfactory. Their capabilities diverged to far from the human parameters and a throughout analysis and redesign was considered mandatory.

Chapter 7

# The Tendon Prototype

The limitations and restrictions of the initial prototype (see Chapter 6) led to the development of a more human-inspired eye actuation system. This prototype has been presented at the IROS conference with the title "An Affordable, 3D-printable Camera Eye with Two Active Degrees of Freedom for an Anthropomorphic Robot" [26]. The aims of this prototype were twofold: One goal was the creation of a system with a more human-like oculomotor range as specified in Chapter 3. And, in addition, this prototype was used to explore and evaluate emerging technologies in the field of rapid prototyping: Instead of using a classical actuation system by using e.g. a four bar linkage or driving belts, this system used a tendon-like system that combines tear-resistant strings and 3D printed parts to achieve a highly-mobile, low-cost robotic eye for humanoid robots. One of the main issues with the previous, lever-based construction presented in Chapter 6, was the fact that it was not possible to reach the necessary angular eyeball displacement values because of collisions between the actuating lever and e.g. the face masks. This problem has been eliminated by deploying a combination of a tendon-based actuation scheme and intelligent selection of attachment points inspired by the human oculomotor system as introduced in Section 2.1. Instead of using actuator pairs in an antagonistic setup as present in the human oculomotor system (see Section 2.3), this prototype has only one single actuator for each degree of freedom. On the one hand this simplifies the system in terms of construction complexity, weight, and part count, but on the other hand it introduces some non-linearities and possible unconstrained torsion (see Section 12.5).

## 7.1. Mechanical Design

An overview of the full system is shown in Figure 7.1: Beside the expressive function of a moving eyeball the main purpose of it is to capture the surrounding scene. In this case, a small CCD image sensor is embedded in a small camera PCB (K) and a lens (J) projects the scene onto it. The

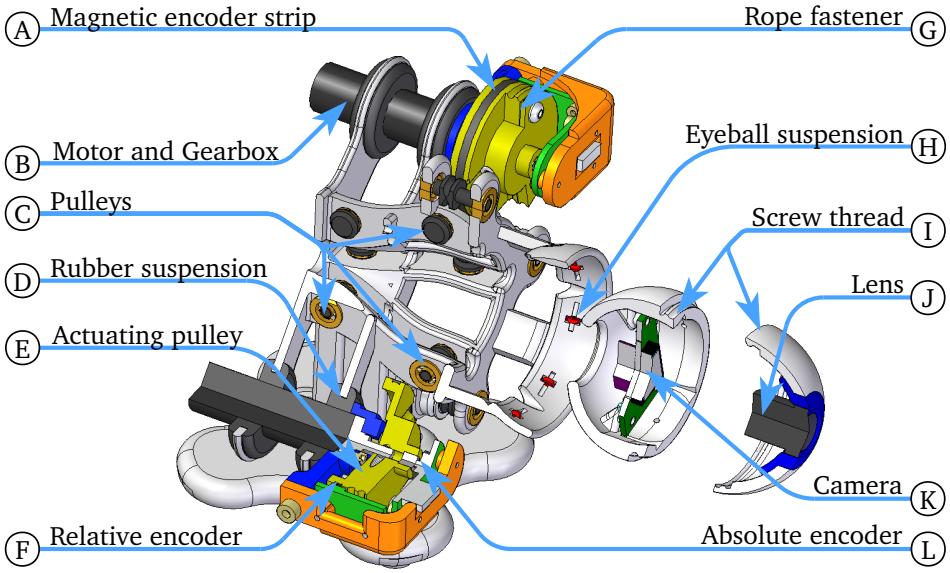



Figure 7.1.: Dissected CAD view of the tendon eye prototype.

eyeball (I) measures 45 mm in diameter and consists of a blue iris part, a white front- and back- part. These two parts contain 3D printed threads and can be screwed together, encapsulating the camera in a visually clean way. The deviation from the human eye diameter is founded in the overall design of the Flobi robot, which in turn was driven by the dimension of the camera. As the name of this prototype suggests, pairs of artificial tendons made of tear resistant strings are used to actuate the eyeball. Before going into the details of the overall actuation system please refer to Figure 7.2 for a more detailed overview of the attachment points, the redirection, and tendon routing system. Similar to humans, the actuating tendons are con-

nected to the eyeball, but located at a slightly different position, namely $-13°$ behind the equator. In comparison, the human muscles attach much further to the face of the eye, even in front of the equator (Section 2.3). The reason for changing the attachment point further to the back is that the design of the targeting robot makes use of baby face cues like the relatively big eyes which leads to large portions of the eye being exposed to the user. Without any soft tissues covering the tendon insertion points as found on humans, moving these points further to the back was the best way to hide all mechanics as defined by the requirement analysis. This position was not selected by random guessing but instead the position of the pulley and the tendon insertion points have been numerically optimized to allow the intended movement of $\pm 55°$, while at the same time minimizing the exposure of the tendon running on the front of the eyeball. As a result of moving the insertion points to the back, the tendon path is switching between a roll-on phase and a free pulling movement which results in non linear behavior. In

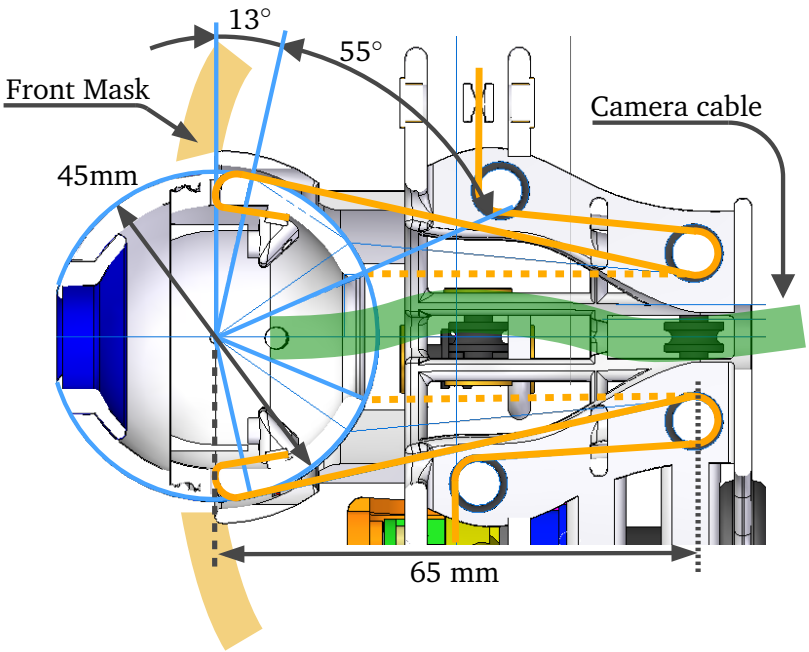



Figure 7.2.: Overview on the actuation scheme of the Tendon eye prototype.

combination with having only one motor per degree of freedom and rolling up while unwinding the other side of the string on the eyeball results in a changing total tendon length during movement. This has been calculated to be less than $0.5\,\mathrm{mm}$ and will be compensated by the rope stretching. The actuation string is made out of a special fishing line which consists of a neutral colored, braided $0.1\,\mathrm{mm}$ dyneema fibers with a total breaking force of $6.0\,\mathrm{kg}$ and less than 2% elongation.

An initial overview of the complex system of redirection pulleys is shown in Figure 7.2: The first pulley is $65\,\mathrm{mm}$ away from the eyeball center and inverts the pull direction. Again, the exact position was calculated to allow a maximum eye rotation of $\pm 55°$ on the one hand and to leave enough space for the escaping camera connection cable on the other hand. A second, different angled, view of the whole setup is shown in Figure 7.1: The pulley system (C) guides the tendons through a system of pulleys. The tendons are finally pairwise attached to a driving pulley (E) with a diameter of $34.6\,\mathrm{mm}$. A 3D printed string fastener (G) secures the tendon after final assembly. The angular displacement of the driving pulley is measured directly in place using a combination of a magnetic strip (A) glued onto the actuator and a relative linear hall encoder (F). In addition, a second, absolute hall encoder (L) is mounted directly on the pulley axis and is used to establish a reference position during initialization. Please refer to Section 7.2 for more details on the sensor setup.

The power train is implemented by a combination of a 12 V brushed DC motor (B) and 67:1 planetary gear head with a diameter of $13\,\mathrm{mm}$ (Maxon RE-max [78]). This motor-gearbox combination is held in place by a flexible rubber suspension (D) with the only additional connection to the core structure being the driving tendon. This elimination of all stiff mechanical contact surfaces is deployed in order to reduce body borne noise.

It is not feasible to reproduce all aspects of the human oculomotor system: Whereas the human eye is held in place by a combination of soft tissues and the eye muscles, in our case, a system of eight $3\,\mathrm{mm} \times 1\,\mathrm{mm} \times 1\,\mathrm{mm}$ miniature ball bearings form the artificial eyeball suspension system. These bearings (H), are equally spaced in a circular arrangement behind the eyeball. Comparable to the human, the tendons that drive the eyeball are constantly applying pressure to the eyeball and holding it in place. Similar to the human visual nerve the outlet port for the camera signal cable is on the back of the eyeball. The connecting cable, shown in Figure 7.4, is custom

made out of a strand of highly flexible wires in order to reduce the bending resistance.

## 7.2. Sensors and Electronics

Similar to the previous prototype presented in Chapter 6, this system uses a combination of a high-resolution relative encoder and a lower resolution absolute encoder. The higher resolution of the relative encoder facilitates smooth motion control whereas the absolute encoder allow to restore a defined, calibrated position during power up. The initialization procedure uses an AS5045 [151] absolute magnetic encoder from Austria Microsystems together with a simple 2-pole magnet mounted on-axis at the actuating pulley for every degree of freedom (see Figure 7.1 (E)+(L)). This sensor connects to the motor control board using a SPI connection and is read out during boot-up to determine the absolute eye position with a resolution of $\approx 0.068°$. This translates to 4096 encoder ticks per full $360°$ rotation of the magnet axis, which again, translates to no more than ≈14.78 ticks per degree of eyeball rotation. At a first glance this might look quite high, but this is not sufficient for smooth velocity control as it exceeds the requirements introduced in Section 3.4. This is where the additional AS5311 [152] magnetic linear displacement sensor can score with a resolution which is an order of magnitude better than the absolute encoder: A fine pitched, multi pole magnetic strip is mounted onto the driving pulley (see Figure 7.1 (A)). Based on the encoder resolution of 512 ticks per mm and the transmission ratio from the pulley to the eyeball, the angular eye positioning resolution can be calculated as $<0.005°$. This translates to ≈212 encoder ticks per degree rotation of the eyeball, which is again, an order of magnitude better than the absolute encoder and facilitates very accurate and smooth eye movements. In addition, measuring the angular displacement directly at the actuator pulley, eliminates any interfering gearbox backlash and allows the use of standard motor-gearbox combinations. The magnetic multi pole strip is flexible and comes in a $300\,\mathrm{mm}$ long, $2\,\mathrm{mm}$ wide, and $0.5\,\mathrm{mm}$ thick package. The magnetic pole pair length matches the AS5311 specification and amounts to $2\,\mathrm{mm}$. It was cut to length and glued into a designated gap on the pulley. This specific magnetic strip was supplied by Bomatec Switzerland.

Just as in the case of the first prototype presented in Chapter 6, this prototype uses a Point Grey Research Dragonfly 2[153] Firewire camera. It is

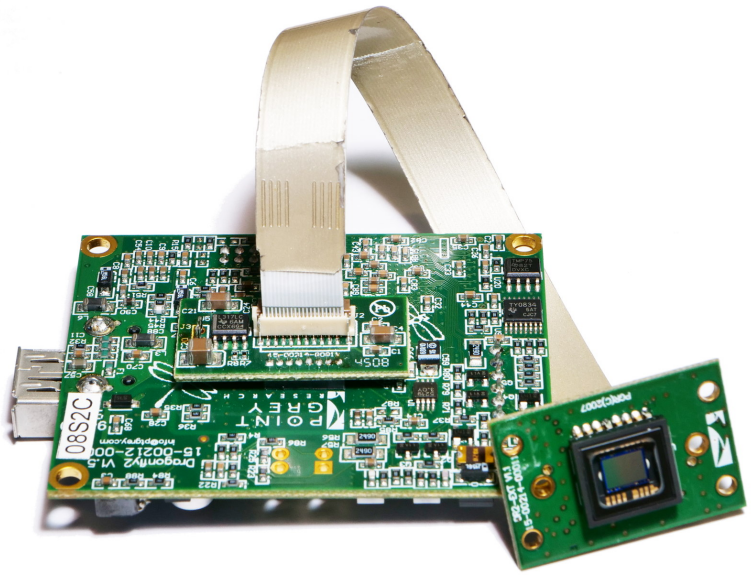



Figure 7.3.: Point Grey Dragonfly 2 Camera in remote head configuration with stock FFC cable (Model 08S2C).

available in a variety of different configurations. The 08S2C variant that is used for this eye prototype is depicted in Figure 7.3 and comes in a remote head configuration that outsources the CCD sensor to a separate and lightweight sensor board. This configuration is the best suiting option and comes with a $1/3$" color Sony ICX204[154] CCD sensor. It can deliver up to $30\,\mathrm{fps}$ at a resolution of $1032\,\mathrm{px} \times 776\,\mathrm{px}$. The final link to the outside world is realized by means of a M12 Lens with a focal length of $4\,\mathrm{mm}$. In combination with the $1/3$" sensor, this results in a visual field of $68° \times 48°$. The stock cable that connects sensor- and main-board is a flexible flat cable (FFC). Similar to a strip of thick paper, this type of cable is flexible in one direction and can e.g. be rolled up, but on the other hand it does not allow lateral movements. In order to allow both types of movements, this cable was replaced with a flexible, custom made adapter cable that fits into the standard FFC connectors on both PCBs. The prototype of this cable is

shown in Figure 7.4: It consists of two flex PCB based adapter boards that converts between the original FFC connector and solder joints on the other side. In order to protect the signals from external interferences, the signal lines were identified and were pairwise grouped with matching ground wires and twisted together before being soldered to the adapter boards.

Figure 7.4.: The custom-made highly flexible replacement camera cable.

## 7.3. Manufacturing and Costs

The main structural component consists of one complex shaped main geometry and additional small pulleys which can be printed in different synthetic materials. The first prototype, depicted in Figure 7.5, was printed in ABS using fused deposition modeling on a Dimension SST1200 series printer. Initial endurance testing showed that the ABS pulleys tended to break under high load. This was caused by the print lines introduced by printing layer by layer. This effect has been circumvented by using a selective laser sintering printing process, which uses a polyamid based material (PA2200). It comes with a higher strength and, in addition, possesses self-lubricating properties. The higher resolution and accuracy of this process resulted in a better fit for the bearings. A fully assembled prototype is shown in Figure 7.5. The individual indices shown in this graphic match those given in Figure 7.1.

After printing, this eyeball was sanded to a shiny and very smooth finish leaving no traces of any unevenness from the layered printing process. This step is important for the bearing based suspension as well as for appearance

reasons. The main geometry has a volume of $36.7\,\mathrm{cm}^3$ and, when printed in ABS, a mass of $\approx 40\,\mathrm{g}$. The full assembly, including pulleys, bearings, motors, encoders, and the camera has a total weight of about $258\,\mathrm{g}$. The

| Count | Item | Total |
|---|---|---|
| 2 | Motor & Gearbox | 212 € |
| 1 | Motor Control Board | 90 € |
| 1 | Set of 3D printed parts | 60 € |
| 4 | Encoder Boards | 46 € |
| 34 | Bearings | 32 € |
| 1 | Miscellaneous small items | 20 € |
| | Total Sum: | 460 € |

Table 7.1.: The total cost of the 3D printed prototype including all installed parts.

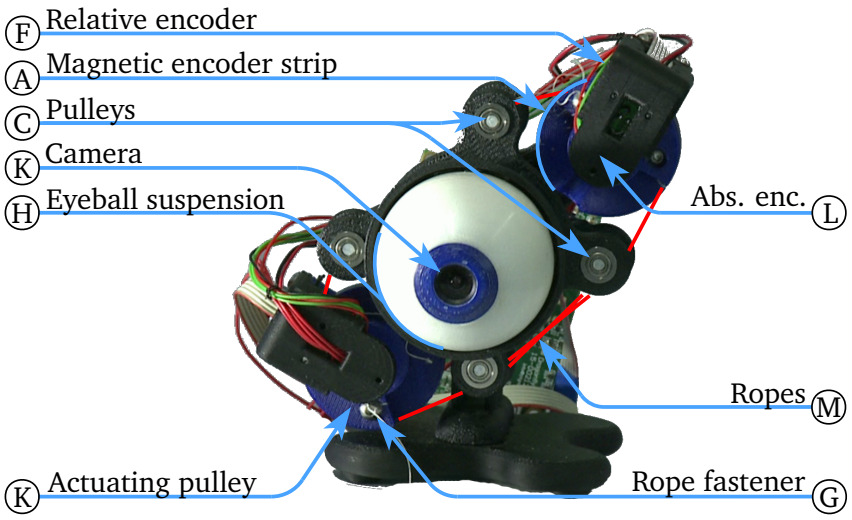



Figure 7.5.: The 3D printed prototype fully assembled. Figure indices equal to those used in Figure 7.1.

use of rapid prototyping significantly reduced the costs of manufacturing: There are no expensive CNC or workshop costs plus all parts can be made and assembled with standard and affordable tools in the lab. A detailed cost breakdown is shown in Table 7.1: The whole setup without the camera costs around 460 € where the main portion of the cost are for the motor-gearbox combination at 212 €. Beside the costs for the 3D printed structure (60 €), the other parts are cheap and easy to source.

## 7.4. Discussion

The presented prototype successfully implemented a mainly 3D-printable robotic eye with human characteristics. There are some design ideas that worked out quite well whereas some aspects will need some more work and improvements. First of all crafting the main structure and the redirecting pulleys using a 3D printer has been a success: Arbitrary shapes, short lead time and uncomplicated assembly combined with good overall stiffness can be seen as a positive outcome. Reduced cost and manufacturing time speeds up the development process and results in a faster evolution of our robotic head.

The tendon-based approach allows a huge range of motion of up to $55°$ combined with maximum velocities and accelerations that met the requirements given in Chapter 3. The total backlash error of the whole system has been reduced from $1.5°$ to less than $0.5°$ by using magnetic linear encoders combined with a magnetic strip glued onto the actuating pulley. However, there is some remaining distance dependent positioning error and some unconstrained eye torsion. As stated in Section 3.4, eye torsion can be accepted in certain applications but might interfere with others. If proven necessary, further work in this direction would be needed to minimize this error. Another aspect that needs some minor adjustments is the tendon tensioning system: The contact pressure of the screw based tensioning system is not high enough, after long periods of use or end position collisions the tendon might get loose and requires a re fixation and re calibration. Additionally, in order to have better control over the unconstrained eye torsion, future designs should use some additional bearings that guide the tendon at a position closer to the eyeball. Please refer to Chapter 12 for a detailed

survey and an in depth analysis of the performance.

In sum, while there is still room for improvement, this prototype exceeds the previous and most other robot eyes in all respects, and, at the time of publication, was the only one in the literature to fulfill both the appearance and the motion requirements that were set up.

Chapter 8

# The Floka Prototype

The lessons learned during the development and evaluation of the tendon-based prototype paved the way to the second eye prototype which emerged during the planning and design of the successor of the Flobi robotic head. This new head, called Floka, combines all lessons learned from the previous iterations of the eye-, neck-, and additional mouth- prototypes into a new self-contained design. The eye prototype and an associated analysis have been presented at the ICRA 2019 conference under the title "See and Be Seen – Rapid and Likeable High-Definition Camera-Eye for Anthropomorphic Robots" [27].

## 8.1. Mechanical Design

This prototype was meant to combine the large range of motion of the tendon-based approach while avoiding the associated drawbacks such as the unconstrained camera rotation, integration complexity issues, and the low resolution camera. Instead of artificial tendons, this iteration uses a more classical, lever-based approach. Each eyeball has two degrees of freedom, actuated by a two separate electro motors in a four bar linkage configuration.

The eyeball consists of a hollow partial sphere with a diameter of $45\,\mathrm{mm}$ and houses the image sensor and the lens system. The internal structure is depicted in Figure 8.1 by means of a cross section rendering: The sensor (M) is mounted to the camera circuit board (N). In order to keep the assembly as small and lightweight as possible, a custom M12 lens holder (O) acts as

the main structural part. It keeps the lens in place, connects the assembly to the eyeball shell (Q), and attaches to the actuation lever at position (F). Finally, a visually appealing thin colored vinyl sticker (R) mimics the iris and optically closes the gaps between the lens (P) and the eyeball shell. The current prototype with the structural parts made from aluminum weights a total of 32 g which is less than the previous design at 40 g (Section 7.3). A more detailed rendering of the actuation system is shown in Figure 8.2: The motor (A) actuates a four bar linkage using the lever (B). The four bar linkage transfers the motion to lever (C) which finally actuates the eyeball pan axis. A second Motor (D) drives the eyeball tilt axis via a rod that is driven via lever (E). The structure that holds the camera PCB and the eyeball is finally driven at point (F). Similar to the previous two prototypes, magnetic absolute encoders (G) and (H) are used for machine homing and initialization of a reference position during startup (Section 3.4).

The motors (A) and (D) are out of the Faulhaber 1524 motor series [155] in combination with a special, zero backlash gear head of the 15/8 series [156] (76:1). This special type of gear head reduces the backlash to a minimum by using counter-rotated individual gear passes that are locked in place on the motor pinion gear. A relative position encoder of the IEH2-512 series [157]

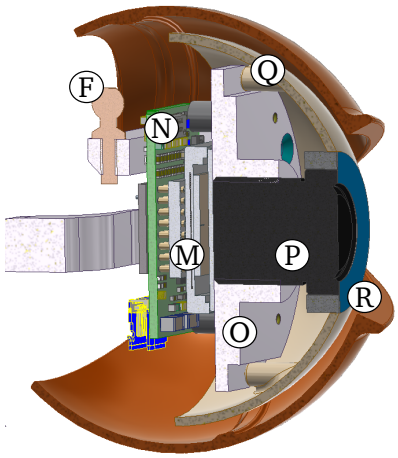



Figure 8.1.: A CAD rendering of a cross-section of the proposed eyeball design.

with 512 lines per revolution is mounted directly on the motor shaft.

As requested by Requirement $R_{1.3}$ (see Chapter 3), the final robotic head should include eyelids. In this design, each eye encompasses two independently movable eyelids (I) and (J). The eyelids are spring-loaded to be closed in the idle state and a cable pulley system, driven by motors (K) and (L), is used to pull the eyelid open. Both motors are from the Faulhaber 1331-series [158] and use a gearbox of the 15/5 series [159] with a transmission ratio of 11.8:1. Again, positional feedback is provided by relative encoders mounted directly to the motor shaft. The encoder (IE2-400 series [160]) delivers 400 lines per revolution, which translates to 1600 Encoder ticks per motor shaft rotation.

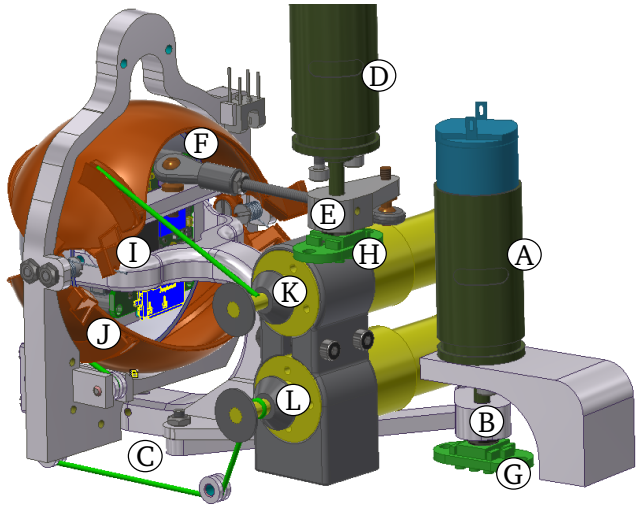



Figure 8.2.: A CAD rendering of the actuation principle of the prototype.

## 8.2. Sensors and Electronics

As already introduced in Section 3.4, smooth motion control during runtime relies on high resolution encoder feedback and a proper initialization on startup. As implemented on the previous prototypes (see Chapters 6 and 7), this prototype relies on a combination of a high resolution relative position encoder and a proper initialization during startup as well. For the eyeball this initialization is based on an absolute reference position that is acquired by using a 14-bit magnetic position encoder (AS5048 [161]), which is calibrated once during assembly. This sensor delivers 16383 encoder ticks per $360°$ rotation, which corresponds to a resolution of $0.022°$ per encoder tick on the lever. In combination with the lever to end-effector transmission ratio this amounts to a resolution of $0.019°$ and $0.015°$ of measured eyeball deflection per absolute encoder tick for the pan and tilt axis, respectively. As introduced in Section 3.4, this is not sufficient for smooth motion at low velocities. Therefore, once initialized from the absolute encoder value, the main position and velocity control loop for the eyeball uses positional feedback from an additional relative position encoder with 512 lines per motor shaft revolution for the pan- and tilt axis, respectively. This translates to 2048 single quadrature encoder ticks per motor rotation, or, in combination with the 76:1 gear head, 155648 encoder ticks per gearbox output shaft rotation. Throughout eye calibration of one prototype showed a resolution of 495 and 616 encoder ticks per degree or $0.0023°$ and $0.0016°$ of end-effector rotation for the pan- and tilt- axis, respectively. This is approximately one order of magnitude better than the absolute encoder resolution and allows very accurate, smooth, and silent eye movements which are mandatory for this application (see Section 3.4).

The eyelid position is initialized by a similar homing motion. Due to the lack of mechanical space, in this case, the homing is based on a calibrated end switch position. The drawback of end-switch-based homing is that it requires an active calibration or reference drive during every power up whereas an absolute encoder based initialization can be used straight away. Once initialized the motion controller switches to the relative encoder as the input source for the position and velocity control loop. The amount of rope that is spooled onto the actuating pulley during construction influences

to total transmission ratio between motor and end-effector. The presented prototype was calibrated at 470 and 530 encoder ticks per degree for the upper and the lower eyelid actuation angle, respectively. This translates to a resolution of $0.0021°$ and $0.0019°$ per encoder tick for the upper and lower eyelid, respectively.

As stated before, a nice looking exterior and smooth motion control is only half of the story, the integration of a decent camera system into the eye is equally important. Our proposed system integrates the off-the-shelf USB3 board level camera MQ042CG-CM-BRD (see Figure 8.3) made by Ximea [162]. It measures $26\,\mathrm{mm} \times 26\,\mathrm{mm}$, weighs $8\,\mathrm{g}$, and consumes $1.6\,\mathrm{W}$ of power. This module features an $1$" Austria Microsystems CMV4000 color CMOS sensor with a global shutter that can stream images in a resolution of $2048\,\mathrm{px} \times 2048\,\mathrm{px}$ at $90\,\mathrm{fps}$ or up to $180\,\mathrm{fps}$ at a reduced image resolution [163]. Due to the relatively large $1$" sensor and the resulting large pixel size of $5.5\,\mu\mathrm{m}$ this sensor features a very good low light performance. As stated in Section 3.2, the sensors' resolution has the potential to fulfill the Requirement $\mathrm{R}_{2.4}$ if this sensor is combined with the correct lens. At the time of planning and component procurement in 2015 this was the only board level module with a USB3 flat cable adapter available. As of writing this, a large variety of modules, based on different image sensors, are

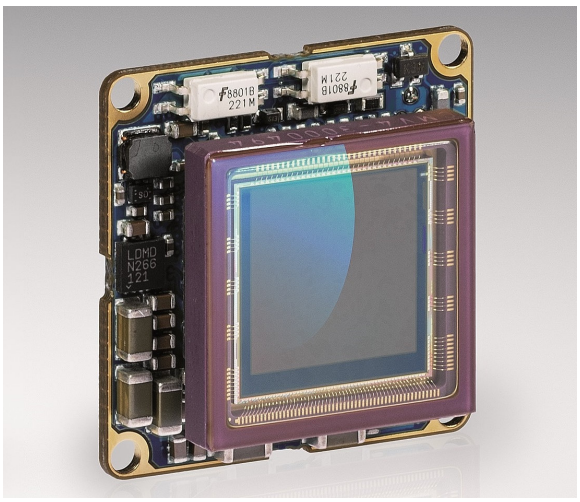


Figure 8.3.: The Ximea MQ042CG-CM-BRD camera module (from [162])

now available. Future developments and designs can thus benefit from a wider range of choices. It may, for instance, be feasible to build a price reduced version of this prototype where the image resolution is not the most important factor. The enabling factor of this module is the flat flex cable similar to the remote head configuration of the previously used Dragonfly2 camera (see Figure 7.3). The reason that makes this camera special is the

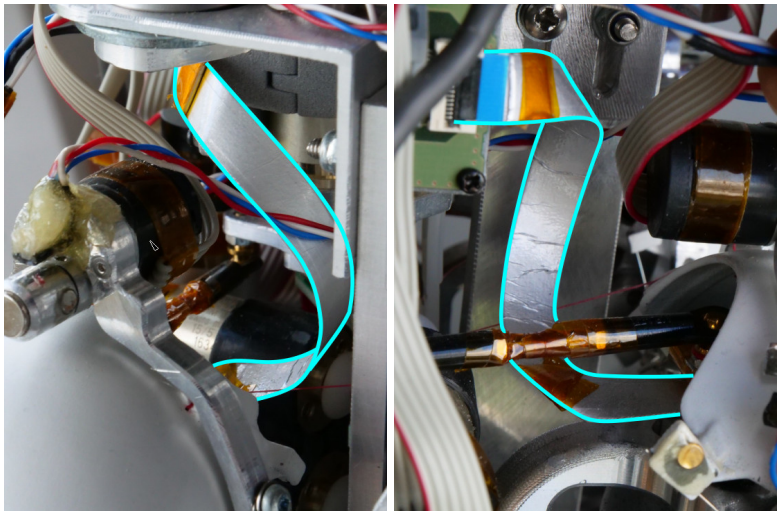


Figure 8.4.: Front- and back view of the camera's flex cable routing inside the Floka robot.

use of a special flexible flat cable for the complete USB3 connection: Typical consumer grade USB3 cables are too thick and their bending resistance would introduce to much counter force during fast eye motion sequences whereas the flat flex cable can be routed in a tight loop (see Figure 8.4). If properly planned and executed, it is possible to route this cable in a special way without causing creases that could sooner or later introduce fatigue fractures during repeated motion. Replacing the FPC cable in the same way as on the tendon prototype was not possible due to signaling and shielding constraints of the high speed USB3 link.

Different lens combinations have been evaluated and the best trade-off in terms of opening angle, image quality, and image distortion was chosen. Based on the geometry of the eyeball and the camera sensor, the sum

of the back focal length and the total lens length can not exceed $19.5\,\mathrm{mm}$ due to mechanical constraints. The opening angle should be similar to the human monocular visual field of view as introduced in Requirement $R_{2.3}$ (Section 2.2). On the one hand a big lens is preferable in terms of quality and light transmission ratio, but on the other hand the lens diameter is limited due to mechanical constraints and aesthetic reasons. The S-Mount or M12 industry lens standard, which was used on all previous prototypes as well, is a good fit. This standard uses a M12 thread with a diameter of $12\,\mathrm{mm}$ for mounting. Due to size constraints these lenses can not be built as sophisticated as e.g. C-Mount lenses and come with a worse aperture value. As stated before the $1$" sensor was the only available option at time of construction. Unfortunately M12 lenses are usually specified for $2/3$" or

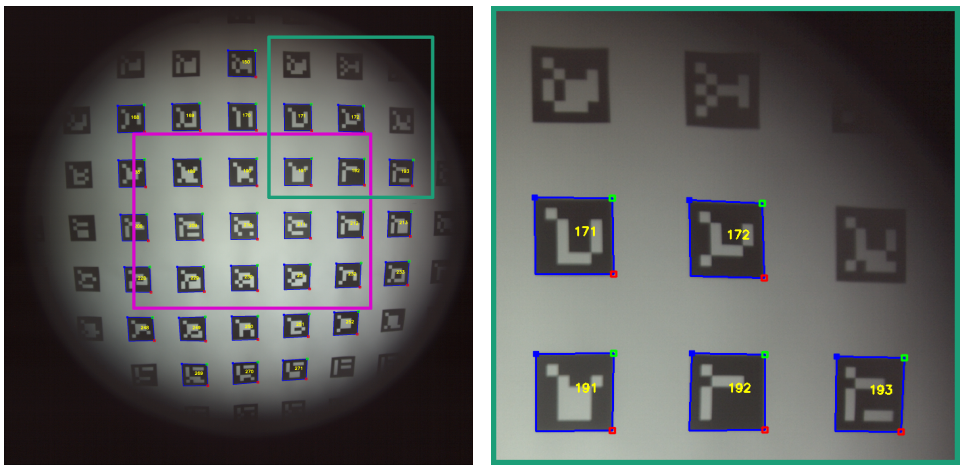
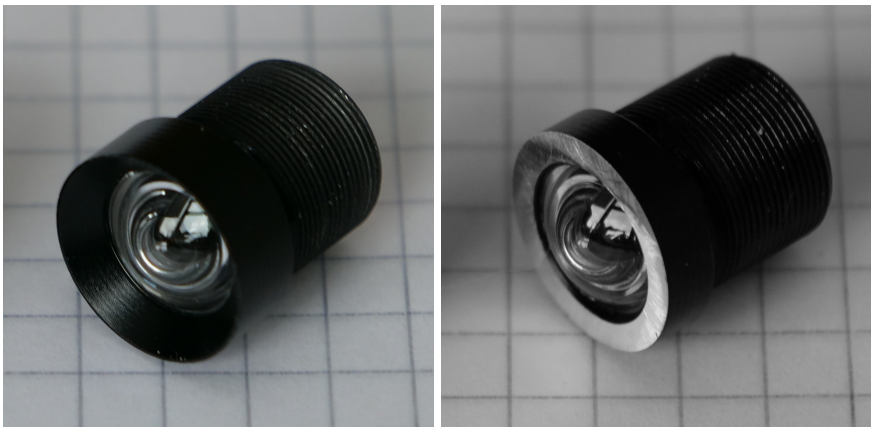



Figure 8.5.: Uncropped- next to magnified view of an image acquired by the camera-lens combination in front of a marker map with an overlay of the Aruco [164] marker detection results.

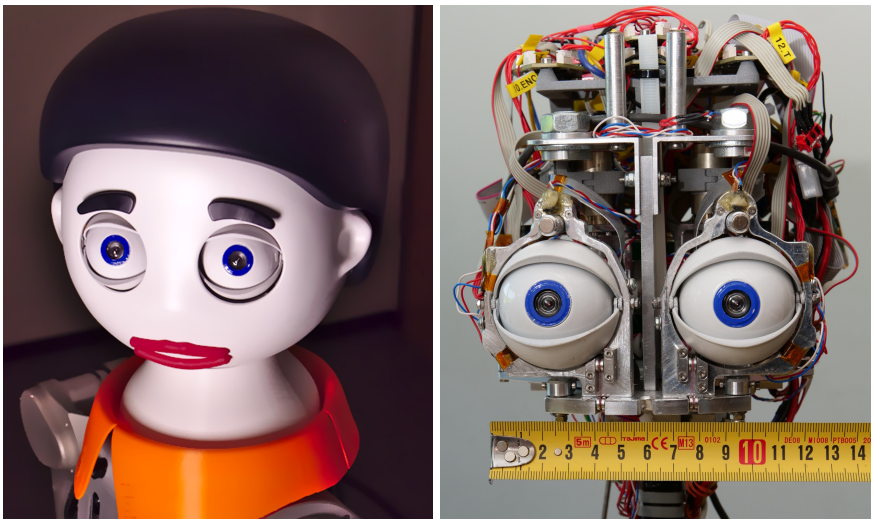
smaller sensors as it is not possible to properly illuminate a bigger sensor with a wide angle lens in this form factor. This causes an uneven circular illumination pattern with vignetting and black border. The combination of the $1$" sensor and a lens made for a $1/2.5$" sensor is shown in Figure 8.5. The most limiting factor for the lens selection turned out to be limited available space measured from the sensor plane to the planned iris. Most of the lenses that fulfill the opening angle Requirement $R_{2.3}$ require more space



(a) Before modification (b) After modification

Figure 8.6.: JSD3428 lens before and after modification of the lens shade.

than available. However, there was one lens that fits all given criteria: The $3.6\,\mathrm{mm}$ f/3.0 lens JSD3428 made by JSD-Optical [165]. It is specified to provide a field of view of horizontally $80°$ and vertically $65°$ in combination with a $1/2.5$" sensor. When combined with the installed $1$" sensor, it delivers an image with a circular field of view of $\approx 105°$. The lens is marketed as a low distortion type with less than 0.35% distortion. An uncropped example image from the used sensor lens combination is depicted in Figure 8.5 (left). The purple marked rectangle shows the field of view of the $1/2.5$" sensor size that the lens manufacturer specifies. A zoomed view of the green rectangle shows vignetting effect in more detail (right). The front lens element holder is $14\,\mathrm{mm}$ wide. The total length from the front element to the image plane is $19.0\,\mathrm{mm}$, which fits to the space constraints. Due to aesthetic reasons the outer part of the lens shade has to be cut off. This creates an even surface for placing a thin blue vinyl sticker that visually acts as the iris. Figure 8.6 shows the lens before (left) and after this modification (right).



(a) Floka with shells (b) Internal mechanics

Figure 8.7.: The final eye prototype mounted on the Floka robotic head shown with and without the final outer shell.

## 8.3. Discussion

Similar to the tendon-based prototype introduced in Chapter 7, this prototype of an anthropomorphic robot eye features human-like velocities and acceleration. The special feature of this prototype is the integration of fast responding eyelids, that even exceed the human performance, and the successful combination into an overall system. All mechanical parts have been successfully hidden from the users' view by using matching exterior shells. Its enclosed and uncoupled nature facilitates integration into any arbitrary humanoid robotic head design as the proposed system does not rely on any coupled joints and therefore each individual eyeball with its eyelids and actuators can be placed and mounted on a freely chosen position without restrictions. For a standard binocular setup, two mirrored copies of this system have to be installed next to each other (see Figure 8.7).

The aged low-resolution firewire cameras that have been previously used, have been replaced by a state-of-the-art high resolution cameras with a

modern, high-performance USB3 host interface. A throughout evaluation verified the human-like performance in terms of required velocities, accelerations, range of motion, and accuracy. Please refer to Chapter 12 for a detailed survey and an in depth analysis of the performance.

The presented prototype has been successfully integrated into the new humanoid robotic head research platform Floka that is based on the previous Flobi design (Figure 8.7). The single eye prototype requires approximately $55\,\mathrm{mm}$ of horizontal space. This small form factor facilitates the integration into human sized robotic heads as specified in Section 1.1. For example on the Floka head, with two eyes are mounted next to each other, the total width of the eye region is similar to a human and amounts to $120\,\mathrm{mm}$ with an interpupillary distance of $70\,\mathrm{mm}$, which is similar to the human values given in Section 2.1.

Chapter 9

# From Humans to Robots

The mechanical completion of the different humanoid eye replica raises a variety of new questions: How should such a humanoid robotic eye be actuated? Basic human motion patterns have been derived from the human data in Chapter 2. But how does this motion look on a mechanical replica and what are the key elements needed for a successful implementation? Even though the final objective is a fully computer driven control system, first answers to these questions can be obtained by using a tele-operated system or by playing back pre-recorded human motion. As shown in Section 5.3, the transfer of recorded human motion is not always straightforward and might not be suitable for all areas of application, however, some valuable insights can be gained by the application of a tele-operation system.

This chapter introduces a novel full-duplex motion capture system that allows a human operator to take the role of an anthropomorphic robotic head by wearing a motion capture helmet, which not only provides direct and natural control of the robot through the operator's own head- and facial-movement, but also includes a projection of the scene perceived by the robot. This system was presented at the ICRA conference under the title "Robot Reality - A Motion Capture System that makes Robots Become Human and Vice Versa" [166]. It allows the operator to control the robot in the most natural way possible: The robot's gaze is controlled by visually fixating an object on a projection screen, while, at the same time, the facial expressions of the operator are transferred to the robot.

## 9.1. Concept

One of the main conceptual ideas of the proposed system is the immediate feedback of the scene as perceived by the robot to the operator. This is implemented by means of a scene camera that is mounted to the robotic head and thus follows all its head movements. The robotic head motion itself is in turn directly controlled by the operators head movement. This requires the scene display to be physically connected to the operators head. At the time of construction, there were no suitable head mounted displays available: Beside low resolution and high latencies, systems from this era did not allow eye tracking or blocked large portions of the face, rendering facial expression tracking impossible. As a workaround, a wearable projection screen was developed. Ideally, this screen needs to be relatively wide in order to cover the whole field of view of the operator. However, the fact that the screen is physically connected to the operators head, poses some limits on the total dimensions and, associated therewith, the system weight. A lightweight system will be less demanding in terms of inertia and the required muscular forces. The second task of the proposed system is to transfer the operators facial expressions to the robot. One drawback of the relatively large screen in front of the operators face is that it blocks external motion cameras as the operator moves in space. Therefore, the proposed system uses a single, head stationary, face tracking camera that is mounted in the center of the screen. This guarantees an unobstructed view of the face at the expense of a small circle on the screen where not projection is possible. This black spot can be clearly seen in Figure 9.1. As introduced in Section 5.1, facial expression motion capture solutions can be differentiated into two major groups: marker-based or marker-less systems. At the time of construction marker-less systems either used heavy and bulky 3D sensors such as the Kinect [167], were not mature enough when based on 2D sensors, or were missing key elements such as eyelid tracking. marker-based systems on the other hand, are stable, easy to implement, and work subject independent but require tedious and time-consuming preparation of the actor.

Before going into details of the actual implementation, it is necessary to give a short overview of the overall structure. The composition of the

system is depicted in Figure 9.2. It is necessary to differentiate between the operator, whose motion, gaze, and emotion controls the robot, and the user, who is placed in front of the tele-operated robot and e.g. takes part in an interaction study. The robot and the participating user are placed in the same room while the operator and the computer equipment is placed in an adjoining room invisible to the user. The complete scene perceived by the robot, including audio and video, is transferred to the system worn by the operator. At the same time, the operator's gaze, facial expressions, head movement and audio is played back on the robot. An exemplary area of application is shown in Figure 9.3: The robot and a study participant are engaged in a card game. The study participant interacts solely with the tele-operated robot. The operator wearing the motion capturing helmet is placed out of sight.

In contrast to systems based on operator input on graphical computer interfaces, the approach presented here is intended to be less demanding for the operator and requires no training or experience in the operation of a robot. This is achieved by a twofold approach: On the one hand, this system exploits the natural human motion as the main input source and

Figure 9.1.: An operator wearing the proposed bi-directional motion capture and feedback system.

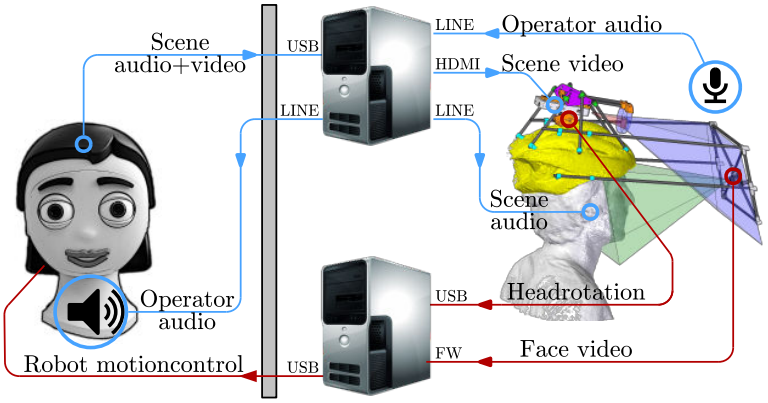



Figure 9.2.: Overall structure of the proposed system as to be used in a ”Wizard-of-Oz” study.


Figure 9.3.: Exemplary ”Wizard-of-Oz” study showing the application of the proposed bi-directional motion capture and feedback system.

requires no buttons or other user interfaces to be used. On the other hand it gives the operator more situation awareness and maximizes the level of immersion by transferring the view perceived by the robot to the operator. This combination makes this system as transparent and intuitive as possible which allows the extraction of undisturbed and natural motion.

In addition, the overall construction philosophy was to use no special,

expensive, or hard to obtain parts wherever possible in order to facilitate the system reproduction by others. This was accomplished by using 3D printed parts for the fastening structure in combination with off-the-shelf parts. Furthermore, all design files are available free of charge and the necessary software is available under an opensource license as well [168]. The software framework has to solve two different problems: First, streaming the video from the robot to the projector requires a configurable distortion correction while retaining minimum latency. And Secondly, the facial image processing, the motor control command generation, and the control of the robotic head.

## 9.2. System Design

The proposed system was completely planned and constructed using computer aided design tools. The camera and projection systems were modeled inside a CAD program and the field of view and the projection path clearance was evaluated and verified using this digital model. In order to keep the costs low and to allow reproduction by others, a commercially available bicycle helmet was chosen as the base of construction. The complete design is depicted in Figure 9.4. A three-dimensional model of an operator wearing the helmet was acquired with the help of a Kinect 3D camera [167] in combination with a 3D reconstruction Software. The acquired mesh was imported into the CAD software system and checked for scale accuracy. This approach allowed it to design the whole system around the holding structure. The 3D model, based on the scan of the real world object, has been used to verify that the tracking camera can capture the whole face by virtually testing different camera opening angles. The finally chosen lens opening angle is depicted as a green cone in Figure 9.4.

### 9.2.1. Feedback System

The robot's surrounding are projected onto the screen (F) in front of the operator. This way the operator experiences the world through the robot's eyes which helps to enhance the perceived level of immersion. After comparison and evaluation of different projector systems, an Optoma PK-201

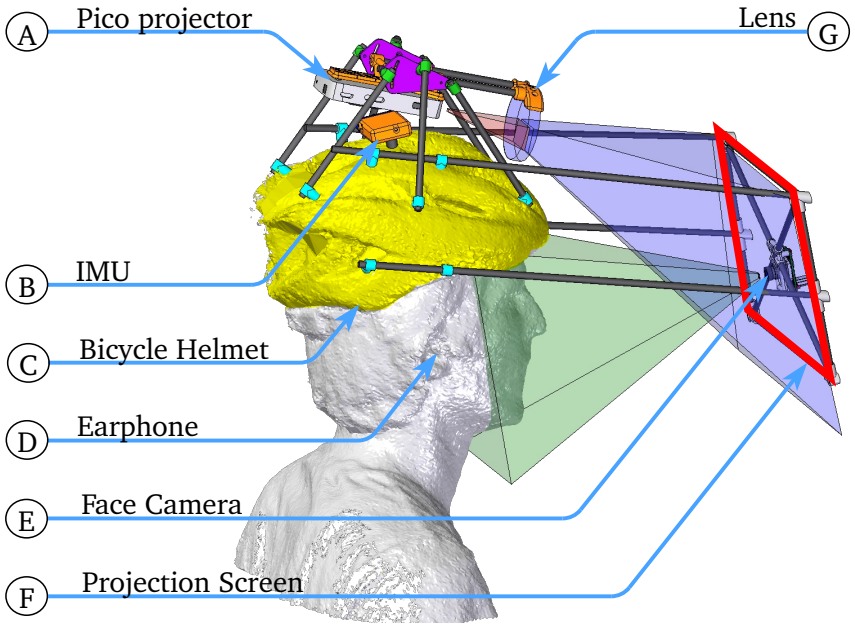



Figure 9.4.: The CAD model of the proposed system, modeled around the 3D-scanned operator wearing the bicycle helmet base.

pico led projector (A) was selected. The key factors were the low weight of $160\,\mathrm{g}$, battery operation, high brightness (20 ANSI Lumen), and a comparable high native resolution of $848\,\mathrm{px} \times 480\,\mathrm{px}$. This device uses a combination of LED illumination and a DLP chip. Laser based projectors also looked quite promising, however they were eliminated from the selection list due to safety concerns based on the low distance to the operators eye and multiple reflecting surfaces in the direct projection path. Unfortunately there was no pico projector in a short-throw configuration available. This required to correct the narrow aperture angle by inserting an additional lens element (G) into the beam path. For the proposed setup a single lens element with a focal length of $f = -100\,\mathrm{mm}$ was calculated to allow the full illumination of the screen. The corresponding beam path, shown as the blue cone in Figure 9.4, was verified on the real setup. In order to minimize the reflective brightness loss, a coated eyeglass lens blank with a diameter of $d = 70\,\mathrm{mm}$ was used. This type of lens is very easy to procure and cheap but, on the other hand,

this comes at the cost of chromatic aberration and image distortion. Both can be dealt with by using a software based distortion correction on each of the three color channels separately. Please refer to Section 9.3.1 for a detailed description of the effect and the proposed algorithmic solution. The projection surface is held in place by the camera holding frame (F) and is based on a highly reflective paper sheet. Its dimensions are calculated to reproduce the field of view of the robot's scene camera of approximately $48° \times 36°$.

Utilizing the cameras inside the robot's eyes for scene feedback is complicated because the proposed system requires a head-fixed camera view. The robots eye view would need to be re-projected and transformed in order to be shown to the operator. For reasons of simplification, a different approach was chosen: An additional fixed, front-facing pinhole camera was added to the robot by drilling a $1.5\,\mathrm{mm}$ hole into the hair mask and invisibly installing the camera module inside this shell without influencing the overall impression of the robot. The disadvantage of this solution is that this camera has a very low resolution and all benefits of a camera included inside a moving eyeball, as mentioned in Section 1.1, are lost. These drawbacks can be neglected as the camera is only used as the operators' scene feedback camera as the resolution of the projection system is very limited as well. The specific camera module is a replacement module for a Dell Inspiron Mini 10 notebook. It is cheap and easy to procure as a spare part. The used CMOS sensor provides a video stream at up to $640\,\mathrm{px} \times 480\,\mathrm{px}$ at $30\,\mathrm{fps}$. The internal connection uses standard USB signaling on a proprietary flat cable. Fabricating a custom cable that connects a standard USB plug to the module circuit board is straightforward. Please refer to the associated opensource release and the wiki [168] for detailed cabling instructions and the module pinout.

### 9.2.2. Tracking Hardware

The main tracking camera (E) is mounted approximately $30\,\mathrm{cm}$ in front of the operators face. Due to the laws of the lever and its mounting position, a heavier camera will require that the operator applies a bigger counter-rotational force to the helmet. Depending on the weight, this might be tolerable, become uncomfortable after longer periods of time, or might be

unusable at all. The whole camera selection process was driven by the resulting requirement to keep the weight down, while at the same time, getting a sufficient image quality and high framerate. Different camera alternatives have been evaluated and assessed by means of these requirements and additional latency measurements have been carried out (see Section 9.4.1). The chosen compromise was to use a Point Gray Firefly MV [169] camera. This camera features a firewire bus and delivers up to $752\,\mathrm{px} \times 480\,\mathrm{px}$ at $60\,\mathrm{fps}$. In order to further optimize the total weight of the system, the standard CS lens mount has been removed and was replaced by a small and lightweight M12 lens including an appropriate mount. Based on the CAD model and preceding calculations, it was estimated that the cameras focal length needs to be approximately $4\,\mathrm{mm}$ in order to maximize the exploitation of the spatial resolution while capturing the operators face.

The angular head orientation data is provided by an additional inertial measurement device (B) that is mounted directly on the helmet (x-IMU [170]). This device consists of a three-axis gyroscope, accelerometer, and magnetometer. The sensor fusion and state estimation is calculated on the device itself and the final orientation vector is transferred to the PC at a rate of up to $512\,\mathrm{Hz}$. Both, bluetooth and USB connectivity is available, but the latter is used here because of the lower latency. The sensor housing was removed and the circuit board was directly mounted in a smaller 3D printed case that was mounted directly on the helmet base.

### 9.2.3. Structural Design

The overall system consists of a variety of different elements that needs to be fixed to the helmet. A steady and defined camera position is advantageous for stable tracking results and lower post-processing labor. This is achieved by the combination of printed parts and bending resistant lightweight hollow carbon fiber rods. This construction fixates the camera (E) and the projection screen (F) in place. Even though the camera was selected for low weight, the mounting position and the resulting off-axis weight applied to the system, turned out to be uncomfortable after longer periods of use. This issue was resolved by adding a counterweight with a total weight of $250\,\mathrm{g}$ to the back of the helmet. This shifted the center of mass towards the center and resulted in a balanced weight distribution. The overall system

weight including all proposed components amounts to $980\,\mathrm{g}$. In order to maximize the wearing comfort and the freedom of movement, ultra thin and highly flexible cables were chosen for the HDMI-, Firewire-, USB-, and audio-connection to the host PC.

## 9.3. Software Stack

For the robot control stack, smooth and low latency motion control is a key requirement for this operation. This section will deal with the data preprocessing and transfer between human and robot. The underlying chain from low-level sensory in- and output towards smooth motion control is covered in Chapter 10. In order to facilitate the best possible immersion and smooth (e)motion transfer between human and robot, a set of software components have to work seamlessly hand in hand with minimal latency. Similar to the non availability of an all-in-one hardware solution there was no readily available software for this specific task. The proposed overall system combines existing software components, where available, with new developed components by using proven technology where applicable. The complete task of performed by the proposed system can be split in two parts: Firstly, the scene perceived by the robot has to be transferred to the operator under the constraints to maximize the immersion effects. And secondly, at the same time, the operators emotion, head motion, and gaze has to be transferred back to the robot with minimal latency.

### 9.3.1. Scene Feedback

The first part, providing the operator with direct visual feedback of the scene perceived by the robot, requires a low latency transfer of the robotic head camera image stream to the wearable projection screen. Although at first glance this sounds like an easy task, the elegance lies in the detail: Due to the unavoidable beveled mounting angle of the projector and the associated external lens element, the projected image is distorted by keystone- respectively pincushion-distortion and in addition chromatic aberration effects are visible. These effects can be seen in Figure 9.5 which shows the operators views on the scene depicted in Figure 9.3: The rectangular input

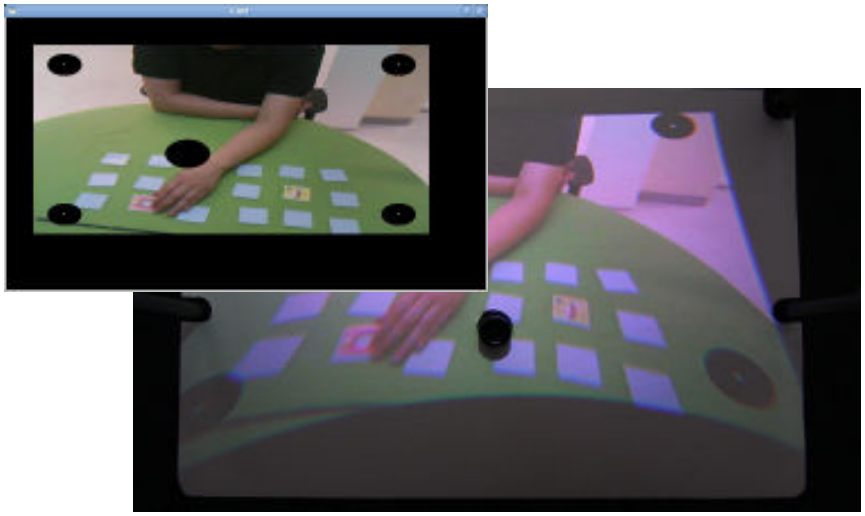


Figure 9.5.: Input image and distorted projection on screen without any distortion correction applied.

image on the top left is warped and heavily distorted and chromatic aberration blurs the image and smears the color information. When the inverse of the distortion function is known, those effects can be corrected in software by applying it to the image before projecting it on the screen. Once projected, the inverse- and the distortion function will cancel out and the result will be an accurate and rectangular image. Assuming this correction matrix has been determined and there is no chromatic aberration, one inverse distortion correction matrix for all three color channels is sufficient. The result is shown in Figure 9.6: The keystone- and pincushion-distortion has been successfully eliminated. The borders are straight and there is no circular distortion visible. However, in reality there is chromatic aberration which results in the projected image being blurry and washed out. This can be seen best in the magnified view of the game card section on the left in Figure 9.6. Before projection, the image looks sharp and the red, green, and blue color pixel match. The magnified view after the image is projected show the shift on the different color channels to each other caused by the chromatic aberration of the lens elements. Chromatic aberration depends on the light wavelength and the angle of impact on the glass surface.

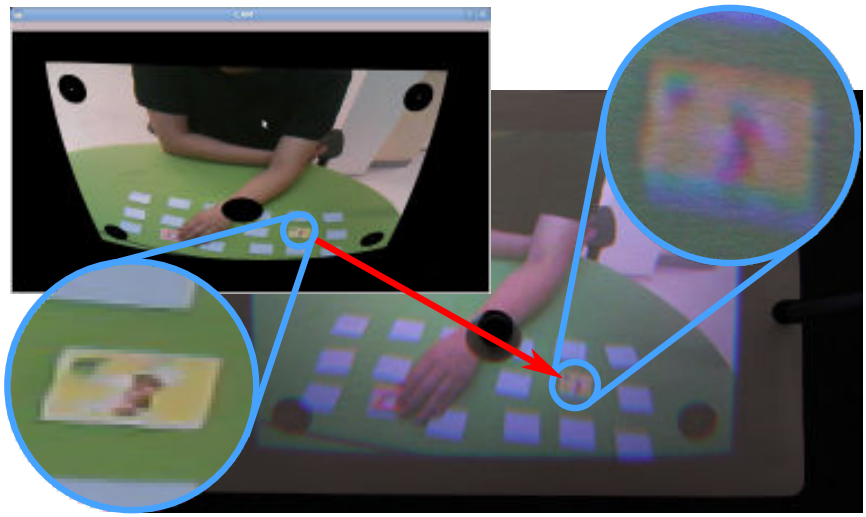


Figure 9.6.: Input image and corrected projection with a one dimensional distortion correction applied to all three color channels at once.

The RGB based DLP projector uses three separate light emitting diodes and subsequently projects the single color channels of the original image to the screen. This allows to apply a 3-dimensional distortion correction with three separate matrices for the red, green, and blue channel. The final result is shown in Figure 9.7: The left zoomed view shows the projector HDMI input image as shown on an external monitor with the calculated ”inverse” chromatic aberration- and distortion-function applied. The final image on the right, projected through the lens and onto the screen, looks much clearer as in Figure 9.6 and does not possess any distortion or chromatic aberration.

In order to calculate the inverse of the distortion- and chromatic aberration-function it is necessary to obtain the visual parameters of the projection setup from HDMI input up to the final projected image on the screen. These parameters can be estimated by placing a printout of an equally spaced 8x6 grid of points on the projection screen surface. With these on screen coordinates known, it is necessary to obtain the corresponding projector x- and y pixel coordinates. These coordinates are determined by manually and suc-

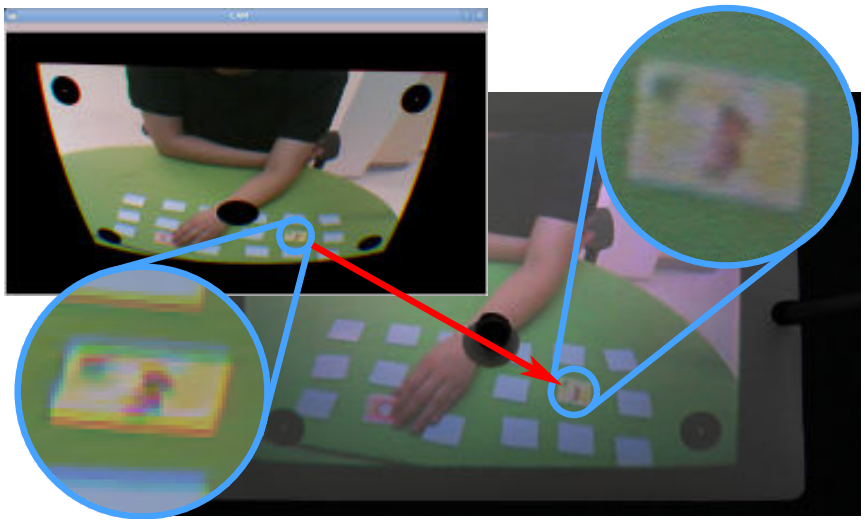


Figure 9.7.: Input image and corrected projection with a three-dimensional distortion correction separately applied to all three color channels.

cessively placing a red, green, and blue crosshatch on the screen, matching the points on the printed grid. The software toolchain assists during this calibration, the crosshatches are displayed one by one and simple mouse clicks store the associated x- and y-coordinates. The standard OpenCV camera calibration routines are then used to calculate the corresponding inverse transformation matrices from screen- to real world-coordinates. For speed and efficiency the application finally stores these matrices in a separate translation lookup tables for each color channel. Once successfully calibrated, the processing pipeline can utilize these lookup tables to remove the distortion and chromatic aberration effects by utilizing the standard OpenCV remap() function.

A high priority was placed on latency reduction in order to maximize the level of immersion felt by the operator. The lowest display latency was achieved by disabling triple- and double-buffering and executing the remapping based distortion correction and the display routines on the graphics card using CUDA. This successfully reduced the image latency from

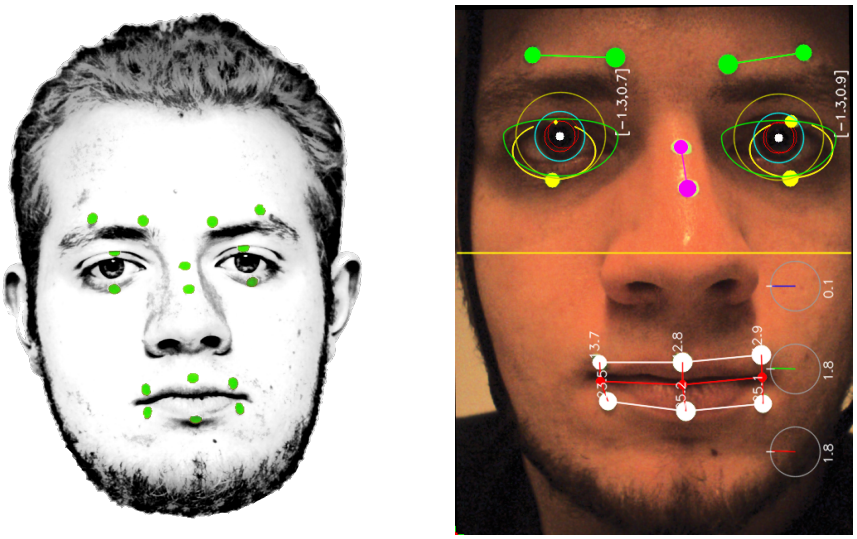
≈22.1 ms to ≈2.5 ms in the case of the one dimensional correction. The latency of the optimal, three dimensional remapping for the individual color channels, was reduced from ≈25.3 ms to ≈4.9 ms. Please refer to Section 9.4.2 for a throughout latency analysis of the projection step and the overall system. Some latency based effects such as the lag introduced by the combination of e.g. a delayed head rotation acquisition and the final image output could be mitigated to some extent by using head position prediction and image deflection algorithms as presented in [171]. However, in this application, other effects such as the delayed perception of scene changes in an interaction scenario can not be removed by such pre-calculations. Therefore, the current system is not using any motion prediction and future projection algorithms.

### 9.3.2. Tracking Gaze, Head, and Facial Expressions

With the scene feedback in place, it is now necessary to capture the operator's motion and replicate it on the robot. As stated in Section 5.2, motion capture for character animation is an extensively researched field. The application presented here is quite similar, even though there are some differences such as the robot's limited number of actuators and the fact that our robot does not allow mesh deformations that can simplify the actual data transformation and mapping step. Due to low processing needs and the ease of implementation a marker-based tracking system was implemented. It facilitates minimal latency and a tracking at full camera resolution at a high framerate. The operator's face is tracked using a low count of 16 green marker dots placed at certain positions on the operators face (see Figure 9.8a). In order to minimize the effort for the translation and mapping step (Section 5.2), these marker positions are chosen to resemble the effector position on the robot. The markers are made of thin paper sheets of fluorescent green color with a human skin friendly adhesive on the back. The color was chosen for three reasons: First, it will appear brighter because of its fluorescent characteristics. And second, the bayer pattern of the camera CCD sensor has twice as much green as red or blue pixels which results in a higher spatial resolution for green. Last but not least the large distance between skin colors and green in a normalized red-green color space allows a very simple and efficient color based segmentation. After blob detection, fil-

tering, and motion estimation using a Kalman filter, the 2D marker positions are known. Marker occlusion and permutation is handled by a combination of finding the optimal assignment using the Kuhn-Munkres algorithm [172] and model-based verification. The final output of the segmentation, detection and tracking algorithm is depicted in Figure 9.8b.

In every social interaction, be it human-human or human-robot the gaze direction of the communication partner is a very important factor: Directed gaze can direct the attention of a human, express a system state and regulate a conversion [173]. Therefore, tracking the operator's gaze is mandatory for our system and was implemented by tracking the pupil movements of both eyes. The search area for the pupil can be restricted to a small region which can be derived from the known coordinates of the eyelid marker positions. After segmentation and tracking the 2D pixel coordinates for each eye are translated to individual pan and tilt angles based on a sphere model of the human eyeball. Even though the face tracking as such is stable and does require minimal configuration, unfortunately the current pupil tracking al-




(a) Marker placement (b) Tracking result

Figure 9.8.: Marker placement and the result of the marker-based tracking system.

gorithm does not work reliable for brown eyes because of the simple segmentation algorithm. This issue could be solved by using a more elaborate algorithm or the use of infrared scene illumination as found in commercial eye tracking solutions. An additional potential weakness of this system is caused by the camera distance to the eye. The spatial resolution at the eye is typically limited to less than 100x60 px. However, for our application the results are satisfying and the proposed system obtains smooth and natural looking eye movements. Once facial features are tracked, the facial expressions have to be transferred to the character. Please refer to Section 5.2 for an overview of tracking systems and the different motion transfer and mapping approaches. For simplification and because of the limited set of actuators a direct mapping approach was chosen here: The marker placement, depicted in Figure 9.8b, resembles the positions and motion axes of the single actuators. The mapping of the marker positions in camera coordinates to the metric space allows a direct calculation and mapping of eye- and eyelid-rotation values to the robot's actuators. As our robot offers only one degree of freedom (rotation) for each eyebrow, an approximate conversion based on the pixel locations of two marker positions is used. Every single of the six lip actuators requires a target position in mm. The necessary conversion is straightforward once scaling and offset correction for the rest position have been applied. The additional two markers on the nose are used for helmet-head dislocation compensation in terms of shift and rotation. This was used on previous helmet prototypes as they did not properly fixate the cameras orientation. By using carbon fiber rods and a total of eight fixation points on this prototype helmet this issue has been dramatically reduced.

The final missing link in the total motion reproduction chain is the ability to track the operators three dimensional head rotation. A x-IMU is used as the inertial measurement device used on this prototype (see Section 9.2.2). It handles all processing and the sensor fusion on board, the incoming angular data can be directly converted to the required neck rotation values. Within the scope of developing this prototype a C++ library to access the incoming data was developed and released under an opensource license [174].

In the course of the development of the motion capture and playback toolchain the actual controller code running on the robot's microcontroller boards was enhanced with a control module that allows velocity-based smooth

and quiet motion control for fixed rate position data.

## 9.4. Evaluation

The proposed system is meant to place the operator in a direct and closed loop. It is highly susceptible to delays and latencies that can possibly occur in many places and thus minimizing these effects is one of the key requirements for a good performance. This is why one of the central points during the realization of this system was to keep the individual latencies, and finally the total sum, as low as possible. This is even more important for future systems that might e.g. be used in combination with a robotic torso that include arms. If the sum of all delays in the control loop exceeds a certain threshold, natural robot control, e.g. hand-eye coordination during grasping, would face serious issues. This section will present the tests that were carried out to objectively measure the performance of the system. Many of these test measurements were performed during component selection and the initial implementation, leading to the optimal selection of components and implementation details. All tests were performed on a off-the-shelf PC with an Intel Core i7-870 CPU, 8 GB of main memory, and a Nvidia GT430 graphics card. The operating system was a standard Ubuntu 10.04 installation with a 2.6.38-13-generic-pae kernel.

### 9.4.1. Camera Latency

The first step in a series of tests was the characterization of the camera latency measured from a change in the visual scene to the detection on the host machine. For the evaluation of the face tracking camera, two candidates, the Playstation 3 Eye camera known for its high framerate modes, and a PointGrey Firefly MV camera were tested. In addition, the robot's scene camera, which captures its view on the surroundings has been evaluated as well. In order to measure the total latency from scene change to the arrival of the corresponding image on the PC the following test setup was constructed: A test program was flashing a light emitting diode by using a control signal of a hardware serial port of the host PC at non-periodic

intervals. It is important to note that only plain, non-USB based serial and parallel ports, offer sufficient repeatability and predictability. But even with those, any non-realtime system will introduce a range of scheduling introduced timing delays. Nonetheless, these numbers are perfectly fine to be used as a comparison of the upper bound. The actual raw delay might be lower as the scheduling and multitasking of the operation system is included in these values. In order average out single effects, a total of n=1000 tests were performed. Each of these tests followed the same pattern: At t=0 the LED was flashed and the delay between this event and the detection in the processed image was determined. The results are depicted in Table 9.1: The

| **Camera** | **Connection** | **Settings** | **Latency** | **SD** |
|---|---|---|---|---|
| PS3 Eye | USB | 640x480@50 Hz | 20.37 ms | 4.17 ms |
| Firefly MV | Firewire | 652x524@60 Hz | 25.92 ms | 8.54 ms |
| Dell webcam | USB | 640x480@30 Hz | 37.43 ms | 13.38 ms |

Table 9.1.: Results of the camera latency measurements.

lowest latency was found on the Playstation3 Eye camera with an average capturing delay of $20.37\,\mathrm{ms}$ ($\sigma$=$4.17\,\mathrm{ms}$). The Firefly MV camera was tested to have an average latency of $25.92\,\mathrm{ms}$ ($\sigma$=$8.54\,\mathrm{ms}$). And finally, the hidden scene camera on the robotic head showed an average latency of $37.43\,\mathrm{ms}$ ($\sigma$=$13.38\,\mathrm{ms}$). Even though the PS3 Eye camera possesses the lower latency, the final choice for the tracking camera was the Firefly MV camera due to size and weight constraints.

### 9.4.2. Projector Latency

With the delay of the input chain determined it is necessary to quantify the projector latency. Similar to the camera latency test, a light emitting diode was toggled on and off by the host PC. At the same time, a black or white image was displayed on the projector. The delay between the LED flash and the change in the projected image was measured by an external high speed camera capturing the scene at $333\,\mathrm{fps}$. Different output modalities (HDMI vs. VGA output) and software visualization frameworks (native vs. OpenGL) have been tested. The results, depicted in Table 9.2, show that

| Connection | Canvas | Latency | SD |
|---|---|---|---|
| HDMI | native | 48.83 ms | 5.31 ms |
| HDMI | openGL | 30.05 ms | 3.43 ms |
| VGA | native | 46.16 ms | 2.59 ms |
| VGA | openGL | 29.83 ms | 2.33 ms |

Table 9.2.: Results of the projector latency measurements.

there is only a minimal difference between the HDMI and VGA output. Because of the superior image quality and thin flexible HDMI cables the final design uses the HDMI port. Switching the drawing method from a native- to an openGL-canvas and disabling double- and triple-buffering reduced the latency on the HDMI port from $48.43\,\mathrm{ms}$ down to $30.05\,\mathrm{ms}$.

### 9.4.3. Camera-Projector Latency

In order to evaluate the whole pipeline delay and the resulting overall latency from a scene change to the image update on the projection screen it is necessary to execute an additional test. Similar to the previous tests, a light emitting diode was pulsed at non-periodic intervals. The time delay between this flash and its occurrence on the projection screen was measured with the high speed camera. The resulting graph for n=1000 tests is shown in Figure 9.9. Without any optimizations the average latency was measured as $144.3\,\mathrm{ms}$ ($\sigma$=$20.2\,\mathrm{ms}$). Outsourcing the image un-distortion functions to the graphics card resulted in computational time savings and a reduced average latency of $111.1\,\mathrm{ms}$ ($\sigma$=$10.6\,\mathrm{ms}$). In order to further reduce the total delay, additional techniques have been evaluated and deployed. The raw projection delay measurements from Section 9.4.2 showed a significant delay reduction by changing the drawing environment from a native- to an openGL-canvas and disabling double- and triple-buffering. The same effect was observed in this test. These measures successfully reduced the overall delay of the proposed system to $75.4\,\mathrm{ms}$ ($\sigma$=$10.3\,\mathrm{ms}$). The final result is a reduction by 50% of the initially measured delay. The measured order of magnitude sounds reasonable, a publication by Willis *et al.* describing a comparable processing a projection pipeline, not featuring the mentioned

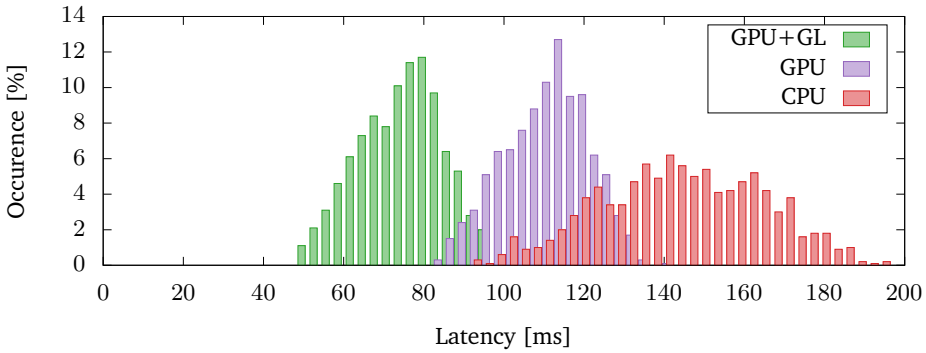



Figure 9.9.: Distribution of the different camera to projector latencies for n=1000 measurements using the one dimensional distortion correction.

optimizations, measured a delay of $121.45\,\mathrm{ms}$ ($\sigma$=$8.49\,\mathrm{ms}$) [175].

### 9.4.4. Robot Latency

The robotic head is driven by two separate data sources. The motion capture data drives a subset of facial joints whereas the head rotation data obtained from the IMU drives the head joints. This makes it necessary to consider both input sources and the associated delays separately. The first test evaluated the delay of a change in the facial expression to the activation and motion of the associated joint. The motion capture input was emulated by placing two light emitting diodes next to each other with only one activated at a given time. If the active- and inactive LED are switched, the tracking system is tricked into ”seeing” a moving marker. Additionally, the motor controller firmware was modified to flash an LED once the motor, controlling the given joint, physically starts to move. This test design allows to evaluate the whole processing chain consisting of frame capture, image processing, motion calculation, motor command generation, transmission to the control server, processing, delivery to the robot over the serial link, command decoding, and finally motion execution. In this test, the motor commands were sent out with a fixed rate of $60\,\mathrm{Hz}$ which is the same as

the camera framerate. The measured average latency for n=1000 frames was $35.58\,\mathrm{ms}$ with $\sigma=13.61\,\mathrm{ms}$. The IMU data transfer latency measurements were executed by a visual frame-wise inspection of a weight hitting the gyroscope and measuring the time when the motor movement count be observed. Due to this manual testing method this has been evaluated for a lower number of n=80 trials. The average latency amounts to $46.42\,\mathrm{ms}$ ($\sigma=10.58\,\mathrm{ms}$).

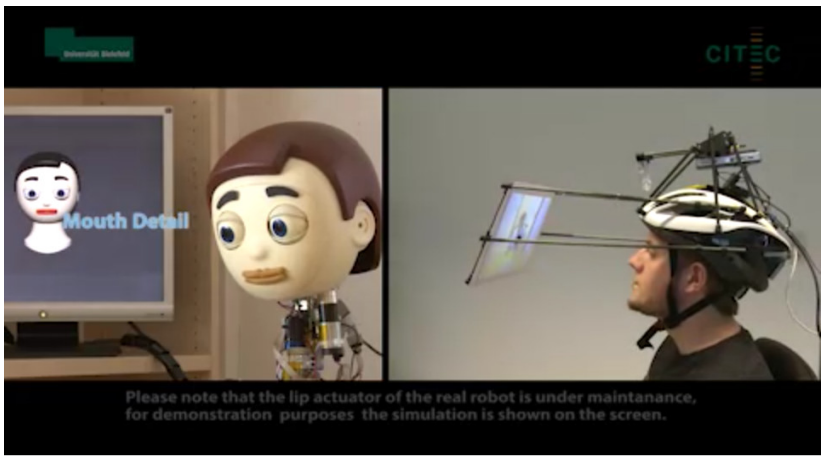
## 9.5. Discussion

The presented motion capture and control system allows an operator to take over the role of a humanoid robotic head. This system successfully combines readily available commercial off-the-shelf components and simple self-made structural assemblies to a portable, novel, and unique robot control interface that allows very intuitive tele-operation and -presence. Please refer to the accompanying video submission of the associated publication [166] for an impression of the overall system performance. This video shows an exemplary interaction installation where a person is playing a card game against the robot. Exemplary still frames extracted from this video are shown in Figure 9.10: The opening scene in Figure 9.10a shows the interaction scenario. The robotic head and a user are engaged in a card playing game. Subsequently, the setup of the operator wearing the proposed system is shown in Figure 9.10b. The direct motion and emotion transfer from the operator to the robot is shown side-by-side in the successive scene using split screen techniques (Figure 9.10c). The final scene, depicted in Figure 9.10d, additionally shows the facial marker tracking output side-by-side to the operator and a live rendering of the Flobi robot as the mouth construction of the physical robot was not operational during the video shooting.

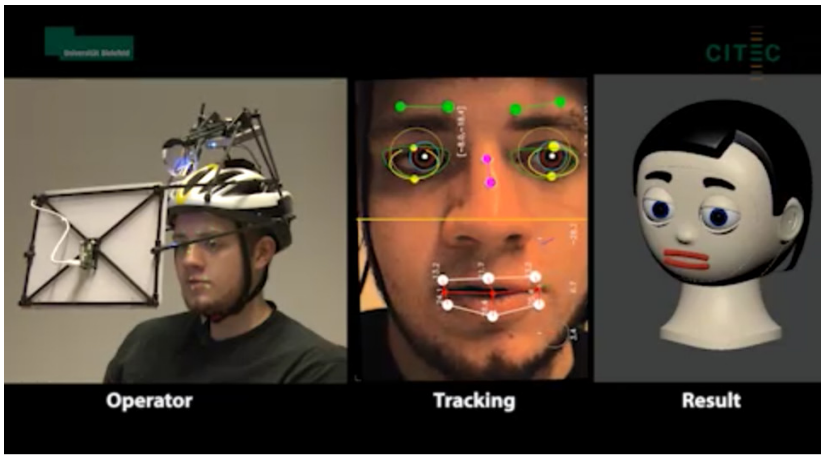
The complete design is made available at no charge under an opensource license. This facilitates the reproduction of this system at a low cost of less than 600 €. Building the structural parts out of 3D printed connectors and cheap off-the-shelf carbon fiber rods reduces the assembly time to an absolute minimum. The associated software stack for motion capture, distortion correction, and projection is available under an opensource license as well. The smart use of the readily available openCV based camera distortion cor-


(a)


(b)




(c)




(d)

Figure 9.10.: Still frames extracted from the demonstration video which was published in [166].

rection algorithms facilitate an accurate offset projection by removing distortion as well as chromatic aberration effects introduced by the optical path of the projection.

Thorough latency measurements and evaluation of different techniques in order to reduce hardware- and software-based latencies and delays have been carried out and the perceivable delay was reduced to a minimum. The measured latencies are close to numbers given as acceptable by So *et al.* in [176]. It is known that exceeding the tolerated amount of latency can degrade the performance, for example by causing tracking errors, and can cause simulator sickness [177]. In general, different numbers for an acceptable amount of latency while using a head mounted display can be found in literature. Studies have found the tolerable latency to be highly subject dependent [177] and to depend on the speed of the head motion as well [177], [178]. The limited scene field of view and the helmet weight might influence the user's behavior in terms of the frequency and amplitude of eye- and head- movements. There are some clues by Kollenberg *et al.* that even common, rather lightweight, head mounted display devices can affect eye and head movements by the altered field of view and additional weight [179]. It should be further investigated if this also applies to the presented system and what consequences arise from this e.g. in the proposed use case of robot tele-operation in social interaction studies. If proven necessary, hardware modifications are possible. The field of view could e.g. be enhanced by deploying multiple calibrated projectors. Recent advances in the field of head mounted displays could be another option. Head mounted displays with integrated eye tracking solutions as well as retrofit sets are now available. It remains open, however, whether it is possible to add the eye-lid and -brow tracking on top of this and how much of the mouth area concealed by these devices.

In addition to the tele-operation mode, this system has been used successfully to capture and record animation sequences for playback on the real robot. This allows to find first answers to the questions posed at the beginning of this section. The analysis of recordings provided important insights into the interplay, the expressiveness, and the importance of different motion patterns. Browsing through a set of recordings and selectively watching and disabling single actuators e.g. indicated the importance of the complex interplay of different motion patterns such as the eyeball following eyelid motion. This type of motion, introduced and described in

Section 2.8, turned out to be a very salient property of the robotic head animation. These observations influenced the modeling and implementation of the proposed human motion generation system introduced in Section 10.6 and especially the eyelid animation described in Section 10.6.1. It shows how apparently unimportant motion patterns can contribute a lot to the overall effect of a motion generation system. In order to answer the question of how a humanoid robot eye should be actuated it can be concluded that it is important to thoroughly study the human and the interplay of all single motion patterns even if they may appear unimportant. Please refer to Chapter 11 for a user survey that looks into this question from the users' point of view.

Chapter 10

# Motion Control

As introduced in Chapter 3, a convincing overall performance of a humanoid robotic head requires beside appropriate mechanics, proper and smooth motion control. Two prototypes, presenting a mechanical construction that can cope with the imposed requirements, have been introduced in Chapters 7 and 8. The generation and application of appropriate motion control will be the scope of this chapter.

As previously shown in Section 1.1, the replication of human motion patterns that are the result of years of evolution provide further benefits beyond the advantages of matching the users' expectation. For example the combination of active eye vision with the motion range of the neck allows to widen the perceivable field of view while, at the same time, replicating a motion pattern that humans are familiar with. Matching the users' expectations is one key for a pleasant perception of the robot and therefore an important factor for the control of a humanoid head. The human eye, combined eye-neck movements, and the underlying control patterns have been thoroughly studied by neuroscientists and psychologists over the last decades (see Chapter 2). However, the transfer and application on physical robotic heads is still a challenging issue due to differences with regard to joint limits, the number of degrees of freedom, or body structure (see Chapter 5).

The replication of human motion patterns on a robotic system requires to consider the motion generation chain as a whole, starting from the mechanics through the low-level motor control to high-level motion generation algorithms. This chain, consisting of single key components, builds on one another and the overall performance depends on the quality and interaction of each of these individual steps. These components will be subsequently

analyzed and described one by one in the following Sections 10.1 to 10.7.

## 10.1. Control Electronics

In order to facilitate smooth motion control of a robotic actuator, it is important to start optimizing at the lowest level possible. Smooth velocity control with a high update rate is mandatory. This, in combination with the high count of 19 single actuators and the limited space available inside the robotic head, required an in-house development of the motion control hardware: A typical example for an industrial motion control hardware that matches the used motors and allow position and velocity control is the Faulhaber MCDC3002. This solution is closed source and it is not possible to add pre-processing or run different control algorithms on this module. It can control a single brushed motor with up to 3A and measures $50\,\mathrm{mm} \times 25\,\mathrm{mm}$ with a height of $14\,\mathrm{mm}$ [180]. Placing the required amount of these modules (19) in the constrained space of the head is not feasible.

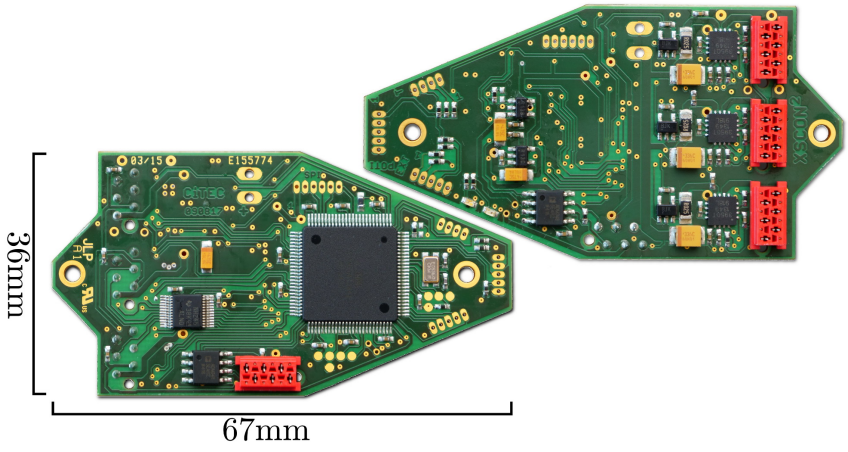



Figure 10.1.: The XSCON2 servo controller printed circuit board (front and back).

The custom, in-house developed motion control board XSCON2 solves this issue by controlling three motors per board. A picture of the devel-

oped motion control hardware is depicted in Figure 10.1. It measures $67\,\mathrm{mm} \times 36\,\mathrm{mm}$ with a height of $8\,\mathrm{mm}$. The core of this printed circuit board is an Atmel AVR XMega128 microcontroller operating at $32\,\mathrm{MHz}$. This controller was announced in 2008 and was meant to be an update for the Atmel AVR microcontroller series. It features a new DMA engine and a wide range of peripherals that were not available on previous AVR controllers. This controller is a perfect fit for the intended application as it features a large enough number of hardware PWM generators and a total of three hardware quadrature encoder inputs. At the time of construction this was the only available controller chipset that was small enough to be placed on the available board space and, at the same time, supported three quadrature encoder inputs allowing the simultaneous control of three motors at once. This motion control board runs a custom firmware written mostly in C with some minor parts written in assembler. The host interface is implemented by means of a high-speed, disturbance immune differential serial link according to the EIA-485 standard. The striking feature of this control module is that it can control a total of up to three brushed DC motors at once. Each motor is connected to the CPU via an Allegro Microsystems A3950 H-bridge [181]. This chip is fully protected against all fault conditions such as over current or over temperature that might happen during operation out of its specification. The current board is designed to operate motors from $9\,\mathrm{V}$ to $24\,\mathrm{V}$ that have a continuous current rating of up to $2.8\,\mathrm{A}$. The controller can measure the voltage, the temperature, and the current consumption of each individual motor. The special feature of the used microcontroller is that it features three hardware timer based quadrature decoders. These inputs can be used to read the high resolution relative encoders on the motor shafts at high update rates. In order to be versatile, this board can additionally be connected to a variety of different sensors through expansion ports. This is necessary to e.g. facilitate the use of a combination of high resolution relative and less accurate absolute encoders. The higher resolution of the relative encoder facilitates smooth motion control whereas the absolute encoder satisfies the technical requirement for a defined initial position on power up that was introduced in Section 3.4. As shown in Section 7.2, the typical gain in resolution is one order of magnitude for a typical sensor setup as used on the presented prototypes.

## 10.2. Low-Level Control Loop

With the hardware for in- and output in place, it is possible to run a custom motion control loop: The principle is straightforward, the control board will receive a target position and velocity from the host and the controller firmware will try to control the motor by comparing the current values to the target and setting the appropriate output. In our case the control loop uses a classical proportional–integral–derivative controller. This type of controller calculates the error or difference between a given setpoint and a measured variable. Based on this error, a proportional- (P), integral- (I), and derivate- (D) term is calculated and the sum is applied to the actuator in order to act as a correction. The control variable (position, velocity, force, ...) can be freely chosen. Often multiple PID controllers are cascaded in a chain structure. For the best performance in terms of smooth and natural looking motion, the low-level control loop is running as a velocity controller. For speed and throughput reasons, the control loop code has been optimized to use fixed point arithmetic. This allows the microcontroller to control three individual motors with each utilizing a separate control loop running at $500\,\mathrm{Hz}$. On top of that, a position control loop runs on the host machine and updates the velocity target at a rate of $100\,\mathrm{Hz}$. Furthermore, the motion control hardware is capable to capture temperature and current consumption values of each individual motor. Right now these values are only used to protect the motors from overheating and abuse. Future implementations could use the current consumption feedback for e.g. some kind of simulated force control.

## 10.3. Communication

There are multiple ways to communicate between a host machine and a set of motion control boards. The basic task of the communication system is to publish target values to the actuator nodes and to receive a variety of different measured response values. Starting at the physical layer, a variety of options exist. Well known representatives are e.g. different serial links, the CAN-Bus, or the Ethercat system. Both, CAN and Ethercat, require an

additional external chipset and thus more board space that was available. Moreover, in the case of the Ethercat system, the current consumption and the corresponding heat build-up came on top. At the time of construction of the motion control hardware, the best compromise in terms of necessary board space, data throughput, and current consumption was the implementation of a serial link. For better noise immunity the differential full-duplex EIA-485 standard was chosen. A total of 19 motors are controlled by a chain of seven motion control boards, each responsible for up to three motors. The motor control boards are arranged in a ring topology and implement a custom protocol that was heavily inspired by Ethercat link layer. Every $10\,\mathrm{ms}$ the host PC sends a long data packet that contains data for all nodes to the first node in the ring. This data packet is shifted through the whole chain of nodes and finally reaches the host PC again. This generates a steady stream of data and makes the protocol en-/decoder implementation very efficient. Based on different header fields each node detects which data is designated for itself and extracts the data. If data is to be requested from a node, the host prepares the datastream by inserting enough dummy bytes to contain the values to be transferred. The length of one data packet contains the same amount of bytes from start to end of the chain. Similar to ethercat, this data packed is neither queued nor cached on any node. All data is passed byte by byte with reading and writing values directly in place.

Depending on the latency and throughput of the communication system the amount of processing on the actuator node can be quite different. With hard-realtime systems and enough bandwidth it is possible to have dumb actuators that only forward the measured values to the host and output pre computed set points to the motor driver. In this case, the low-level control loop is running on the host system. The opposite would be a system where all calculation is done on the actuator node: The host sends a plain target value and all calculation and control is done entirely on the actuator node. In our case a mixture of this is applied: The inner fast realtime critical velocity control loop runs at the actuator nodes ($500\,\mathrm{Hz}$) and is fed by the external host application at $100\,\mathrm{Hz}$. This allows a convenient setup where the host PC can be a normal desktop machine without using any special realtime kernel or hardware modifications.

## 10.4. Joint-Level Motion Control

The low-level framework for the PC side control has been developed in C++ and is called the XSC3 server. A well known and widely used software package for hardware abstraction and message based control of robotic systems is the opensource ROS framework [182]. It provides a standard for robotics software development and a large set of tools necessary for robot control. All applications running on the host PC make extensive use of this toolchain. The XSC3 server opens a serial connection to multiple xscon2 controller boards. The bus communication is outsourced into a separate thread running at $100\,\mathrm{Hz}$. Once the control server is initialized, multiple PC side applications can access this server by using a well-defined ROS interface. It is possible to read and write all configuration values, read back all acquired data values, and send new target values. It is mandatory to pass a configuration file to the server binary on startup. This file defines a mapping and calibration of raw motor positions in encoder ticks to a unified degree representation. A very unique and handy feature is the possibility to use this configuration to run any subset or all motors in a virtual mode. In this mode the motors are emulated inside the PC host application. Please note that this is not a physical simulation, the motors are assumed to be ideal and the simulation does not use any physical model and does not take any friction into account. The pure virtual mode with all motors being emulated allows the user to develop and test applications without having the real robot connected to the development machine. When the user finally wants to run the software on the real robot, no code has to be changed. The only difference is that the robot is now connected to the PC and the server is restarted in the non-virtual mode. Special care has been taken that all interfaces, naming conventions, and configured min-/max-positions stay the same no matter if running in virtual or real mode.

## 10.5. Simulation

The virtual feature of the XSC3 server as such is not worth much without a proper visualization. Right now there are two possible visualization possibil-

ities. The first, very basic, visualization is the RViz 3D visualization tool for ROS. A custom robot model, which can be loaded in rviz, has been crated. It allows to visualize the state and appearance of the Flobi head in 3D. The XSC3 server will publish standardized joint state messages and the rviz visualization will always match the robot appearance. The second possibility is the currently preferable solution as its renderer produces a more aesthetic output. This simulation is based on the modular open robots simulation engine (MORSE) [183]. Morse itself is based on the game engine of the blender 3D modeling framework [184]. This origin makes the integration of new blender 3D models into Morse straightforward. A detailed Flobi and Floka 3D model was enhanced with a properly defined skeleton that replicates the internal joint structure. Some python glue code acts as a bridge between ROS and the Morse simulation. The future richness of blender and Morse makes it possible to spawn virtual cameras and to for example place virtual humans inside a simulated scene. This allows the user to evaluate and compare the behavior of a simulated robot to the physical robot [185]. It is important to say that the simulation of the robot motion itself is running inside the XSC3 server. Morse and rviz both use the data provided for visualization.

## 10.6. Human Motion Replication

At this stage, the actuators and sensors of the robot are connected to the host machine and basic motion control is possible. Custom trajectories can be sent to the robot's joints and smooth motion is the result. Actuating single joints in a non synchronized manner will not have the best possible effect. Perfectly timed and coupled motion of all involved joints is mandatory for a convincing overall performance. This orchestration is realized by means of a motion control framework. A literature overview of such gaze control frameworks for humanoid robotic heads can be found in Section 5.3.

This section covers the implementation of a library which is based on the physiological data and the requirements previously defined in Chapters 2 and 3, respectively. This highly customizable motion control library is called humotion and has been presented at the HAI conference in 2016 under the title "Humotion - A human inspired gaze control framework for anthropo-

morphic robot heads" [28]. This library can be configured in different ways by means of thresholds, delays, and repetition rates. One unique feature is the possibility to reduce the maximum accelerations and velocities used during the path planning producing jerky or incorrect motion patterns: All configuration variables are preset to the humans values found in literature and, most importantly, can be modified during runtime. This allows to fine tune and adapt the motion planning and generation algorithm to any given robotic platform even if it is less equipped in terms of reachable velocities and accelerations.

### 10.6.1. Control Model

The basic idea of a control model is straightforward: It will calculate synchronous trajectories for a set of joints based on a gaze target input. The proposed model includes different additional input cues such as a neck offset and an eyelid target for additional liveliness. The output will be the target values for the eye, neck, and eyelid joint. The implemented control model is depicted in Figure 10.2: In a first step, a threshold-based detection of eye- and neck- saccades (1) based on the current- and target-position triggers an update of the current angular neck target. The sum of dynamic neck offset values for roll, pitch, and yaw and a configurable anthropomorphic breathing pattern is then superimposed on top of the neck angles for additional liveliness (2). The breathing pattern is freely configurable and defaults to a rate of 12 breaths per minute [186] and a default deflection of 2 degrees. The default amount of deflection was chosen to be small while still being noticeable. This motion prevents the impression of a frozen, non-working, or switched off robot.

#### Neck Offsets

The dynamic offset values for all three neck joints can be used in different ways: It is possible to overlay animations such as shaking the head or nodding on top of the neck motion that has been calculated based on the gaze target. It is also possible to use those offsets for biasing or shifting the head orientation, e.g. for partial target alignment as a feedback for engagement during communication tasks without losing the visual focus. Last but not

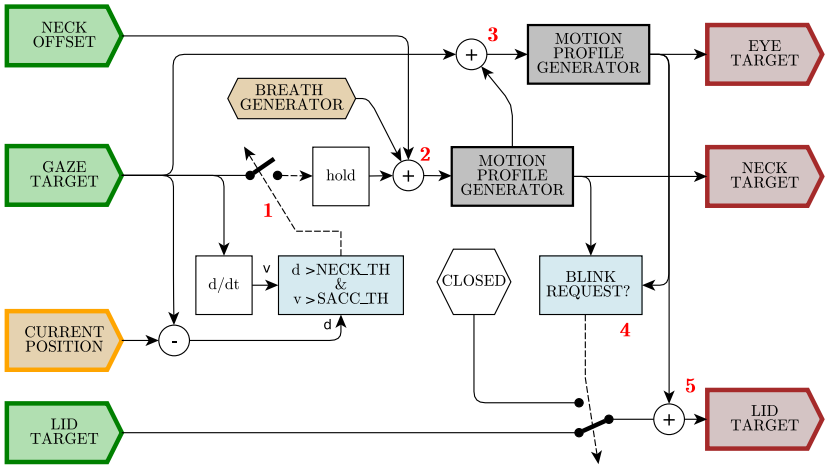



Figure 10.2.: The motion control model as implemented by humotion.

least it can be used to express internal robot states such as the robot's mood by offsetting single joints. The rest of the framework will compensate for those offsets: As long as it is mechanically possible, the robots gaze will always point to the correct location in space, no matter what the neck offset values are. The effect of this is shown in Figure 10.3: Both robots fixate the same target in all four pictures: One time with a neck yaw offset of $0°$ (Figures 10.3a and 10.3c) and in comparison with an offset of $-20°$ (Figures 10.3b and 10.3d). This example shows how these offsets can e.g. be used to express the level of engagement in communication tasks by keeping the robots head more aligned to its communication partner during the execution of gaze shifts. Much more importantly, this dynamic offset facilitates the overlay of head animations while retaining the overall gaze at the same visual focus. This unique feature allows a very natural human robot interaction by executing gestures such as shaking the head without interfering with gaze and vision processing tasks. Although being an important factor

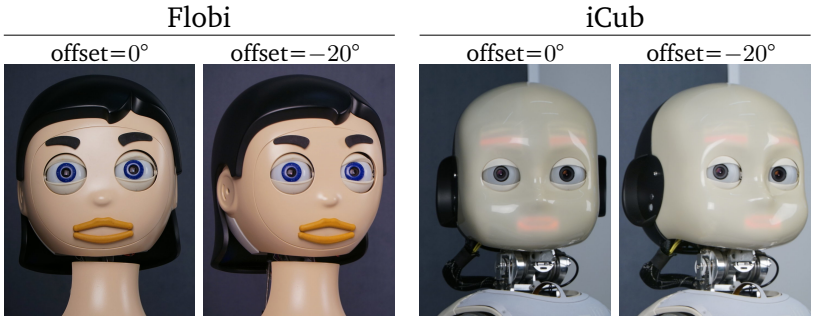



Figure 10.3.: Comparison of different neck yaw offsets and a visual comparison of the effect by running humotion on two exemplary robotic heads. Flobi is shown on the left and iCub on the right.

to facilitate authentic human-robot interaction, the direct support of this kind of angular neck offsets is often neglected in literature describing gaze control systems for robots.

### Rescale & Limiting

One of the big issues porting a control model to a given robot is that not all robotic heads facilitate human-like velocities and accelerations. Even if they are specially crafted to do so like the robots presented in Chapters 7 and 8, sometimes it might be a good idea to limit the maximum accelerations to a lower value in order not to stress the robot joints e.g. during development of a new tracking algorithm. In this case, some previous systems allowed the clipping and limiting of maximum acceleration and velocity values. This is far from perfect, if trajectories of single joints are modified because of clipping, unnatural combined motion trajectories can arise. The humotion library uses a different approach. Based on physiological measurements introduced in Section 2.6 the main configuration parameters are the maximum velocities and accelerations values per joint. Those values are merged and all limits are taken into account during the motion planning phase and do not cause unnatural trajectories as caused by the clipping of single joints.

A very special feature of humotion is the ability to rescale the resulting

overall neck motion by a given factor. This allows to either adept the resulting profile to slower mechanics or slow down the trajectories e.g. during development by the change of a simple slider. The different limits and scale parameters are applied before a suitable motion can be planned and executed. Simultaneously to the neck motion, a compensatory human-like motion for both eyes (3) is calculated based on the neck command, the gaze target, and a given vergence angle before being rescaled and limited in the same way as the neck. This optional rescale and limiting feature for the neck- and eye-motion is configurable at runtime and is one of the distinguishing features of the proposed model. It facilitates the use of the proposed library on a robotic head that is less equipped in terms of reachable velocities and maximum allowed accelerations. The naive approach of clipping single actuator velocities during execution would lead to incorrect and not human-like motion patterns. For example clipping the eye pan axis velocity during a combined pan and tilt motion will result in a strange motion where the single axes come to a rest independently. The limiting function of the proposed algorithm takes care of this situation by reducing the velocities and accelerations during the motion planning phase. This allows the proposed algorithm to generate proper and correct motion profiles for any given hardware limit. For the previous example the tilt axis velocity would be recalculated based on the limit based clipping of the pan axis velocity in a way that both axes come to a rest at the same time. On the other hand, the rescaling function allows to scale the overall motion to e.g. half of the human velocity and acceleration values found in literature. No manual tuning is required, setting this single parameter will take care of reducing all complex and interlinked motion patterns at the same time while maintaining the dependencies among themselves. The final motion pattern will still be human-like but will appear like human motion recorded with a slow-motion camera.

### Tracking Time

Humans need to compensate the latency of the visual system e.g. during smooth pursuit motion (Section 2.6.2). Technical systems are experiencing similar difficulties. A low latency loop from camera input to actuator motion is advantageous but not sufficient. In order to allow accurate tracking of visual stimuli the proposed control system keeps track of a timestamped

history of all joint values. Similar on how humans use an efference copy of a scene, this timestamped history can be used to infer the state of all robot joints at the time of image retrieval and to calculate the correct absolute gaze target based on the relative state error on image retrieval.

### Eyelid Motion

The first type of eyelid motion, namely eye blinks, can be triggered manually or by a variety of internally calculated sources (4). The internal sources are parameterized and uses the defaults introduced in Section 2.8 but, again, can be freely configured by the user during runtime. The programmatic sources include triggers activated by exceeding different thresholds such as high eye velocities, saccade frequency, or timers. In addition, the eyelids pursue the eyeball during vertical movements with a given offset (5). This offset can be dynamically altered by the user and can be used to give emotional or mood feedback.

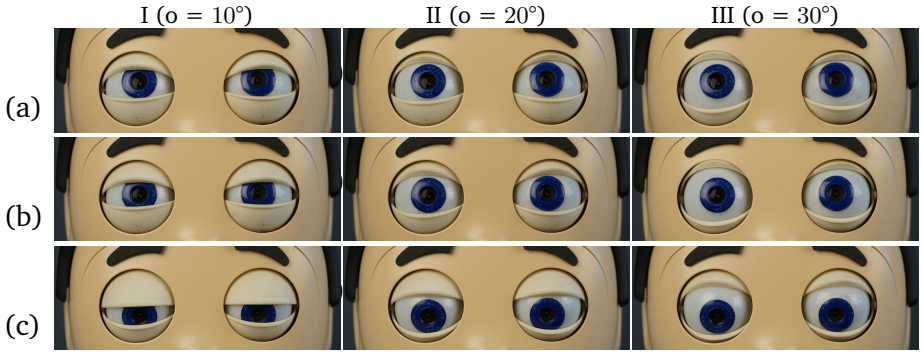



Figure 10.4.: The effect of different eyelid offsets applied on the Flobi robot.

Figure 10.4 gives a visual example how this eyeball pursuing eyelid motion looks like when executed on the Flobi robot: The eyelids seem to ”stick” to the moving eyeball during the vertical gaze shift from $20°$ (10.4a) to $-20°$ (10.4c). This is best seen in column I with an eyelid offset angle of $10°$. For reference, columns II and III show offset angles of $20°$ and $30°$, respectively. Despite not being shown here, it is also possible to set all four offsets independently. These offset changes can be used to express the robot's internal state: Smaller offsets, resulting in closer proximity to the pupil, can be used

to reflect sleepiness or sadness, whereas bigger offsets (eyes wide open) can support surprise or fear expressions. In addition to the impact on the robots expressiveness, there is also a pleasant side effect of the eye pursuing eyelid motion. Unlike eyelids in humans, the proposed eyelid prototype introduced in Chapter 8 is made of a solid shell and does not fold in front of the eye socket but rather moves behind the eyeball on opening up. In this case, the eye pursuing eyelid motion helps to avoid collisions of the wiring harness of the camera emerging from the back of the eyeball with the moved-back eyelids. This allows to extend the overall actuation range of the eyeball and the eyelids.

### 10.6.2. Trajectory Generation

The complex final task of synchronous online trajectory generation for multiple joints is handled by the online trajectory generator of the external motion generation library libreflexxes [187]. This library has been developed with efficiency as the main driving factor. It allows to calculate a large amount of synchronous trajectories for multiple joints under given constraints such as velocities, acceleration, or jerk in the range of microseconds. A typical output of the generated motion by the proposed humotion library for a step-like gaze change is depicted in Figure 10.5. This figure shows the

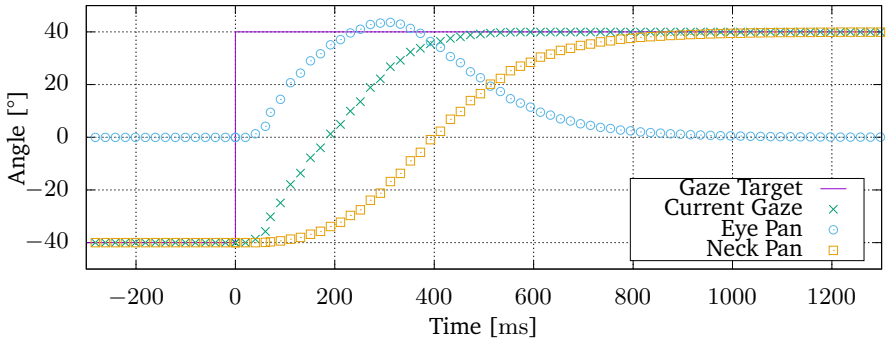



Figure 10.5.: An exemplary 80° gaze-shift as generated by humotion in response to a step-like change of the gaze target.

gaze target input (continuous line), the current gaze angle (x), the corresponding neck pitch angle (□), and the eye tilt angle (○) as generated by the proposed humotion library. At $t = 0ms$ the gaze target is changed by a total of $80°$ and the whole system starts to react with all its joints. The eye- and neck-angle start to move at the same time. Just like in the case of humans, the different model parameters allow the eye to accelerate much faster than the neck. The sum of eye- and neck-deflection and especially its steepness shows the effect of both joints moving at the same time. With the neck approaching its target position, the eye deflection gradually decreases and returns to a zero deflection at the end. A closer look at the overall gaze (x) shows the distinctive first and steep increasing acceleration phase ($t < 300ms$) with contribution of both accelerating joints. This phase is quite distinguishable from the slower deceleration phase ($t > 300ms$) before the eye and neck come to a rest. The calculated and executed motion resembles a human response to a similar stimulus as introduced in Section 2.7 and Figure 2.7. Please note that this graph depicts only the yaw angles of the neck- and eye-joint for better readability. It is important to note that the associated pitch and roll motion is executed synchronously so that all target angles are reached at the same point in time.

### 10.6.3. Configurability

One of the key features of this model is that all parameters are not only pre-configured to match the values found in literature on humans but can be changed during runtime as well. This feature is supplemented by the unique rescaling and limiting facilities of intermediate velocities and acceleration values during the motion planning phase. In particular this kind of high configurability facilitates the application of the proposed library on a variety of different equipped robot platforms. This allows e.g. to slow down the overall motion that is generated by the model through a single slider in order to be executed on a less capable robotic head. Unlike other approaches such as clipping at a given threshold, the complex interplay of different joints and the resulting human-like motion stays intact.

### 10.6.4. Portability

All interprocess communication and dynamic (re-)configuration of the model parameters are handled by a middleware abstraction layer in order to facilitate loose coupling and allow straightforward integration into an existing system. A general overview of the integration of the humotion library is depicted in Figure 10.6. Currently, interfaces for the ROS and the "Robotics Service Bus" (RSB) [188] framework have been implemented. Additionally a wrapper to integrate humotion into yarp [189] has been written.

As the humotion library was written with portability in mind, porting it to a new robot platform is straightforward and involves two steps: First, a humotion server object has to be created. This instance serves as the middleware connection for configuration- and control-inputs such as gaze targets, neck offsets, eye opening angle, and eyelid states. Second, a predefined joint interface has to be implemented for the robot target system. This interface will allow the humotion server instance to read and write joint angles and velocities. Once activated, the humotion library will take over control over all joints. Even though interaction with the proposed humotion library through plain middleware calls is possible, a convenience client API that wraps the actual middleware specific message calls is provided for as well. Currently, C++, Java, and Python are supported.

The portability and applicability of the proposed library was successfully shown by the integration on the Flobi, the Floka, the Meka S3 and the iCub robot platform. It was validated to be working as expected both in simulation and on the real robot hardware. Please refer to Chapter 11 for an in depth evaluation of the generated motion.

## 10.7. High Level Robot Control

Research with humanoid robots is often conducted by people from different disciplines with varying levels of experience and coding skills. A robot that needs deep understanding and requires manual control of single joints constitutes a serious obstacle in multidisciplinary research on human robot interaction. In order to foster the use of robots by non-expert users, a humanoid robot needs an intuitive control interface. It should provide an

easy-to-use way to control the robotic head and all of its features at once in a consistent way. The humotion library is one of the key components as it wraps and hides the synchronous animation of a number of single joints by using a biological inspired model and a straightforward interface that uses gaze targets instead. The developed high level robot control framework (hlrc) builds on top of the humotion framework and provides a set of basic primitives and features required for human robot interaction studies. In addition to the eye- and neck-control, this framework will also take multiple other robot features such as the mouth and eyebrows into account.

The architecture of the hlrc server and its client API is depicted in Figure 10.6. Just like the humotion library in Section 10.6.4, this client-server infrastructure also supports different middleware frameworks for the interprocess communication. In order to work, hlrc requires the target robot to support the humotion library. Currently, example code for the Flobi, Floka, iCub, and Meka robot exist. Please refer to Section 10.6.4 for porting instructions. When started, the hlrc server will open a middleware connection to a robotic platform of your choice and provide a set of different features to the user. A common requested feature from people doing experiments with our robot platforms is speech output. In order to be flexible in the actual program being used for synthesis the hlrc server uses a middleware connection to send a request to a configurable text to speech (TTS) service. Right now a wrapper TTS service that uses Mary TTS [190] to convert text to a list of phoneme-duration pairs for mouth animation and audio data is provided. This part is accessed through middleware connections and is interchangeable as long the TTS system can generate phoneme data. Once synthesized, the audio data is either played back locally or sent to an audio server infrastructure using the middleware for remote playback. The synchronous playback of lip animation and audio data is ensured by the synchronized start of the audio file playback and the strict compliance to the timestamp based playback of the individual lip animations associated to the single phonemes.

On top of that, an internal arbitration scheme takes care of the superimposition of the facial expression configured by the internal emotion state and the lip synchronization data. This facilitates the reproduction of speech and simultaneous emotion feedback without user intervention. The proposed implementation supports a configurable default and current emotion that can affect all face joints. The default emotion state is the facial ex-

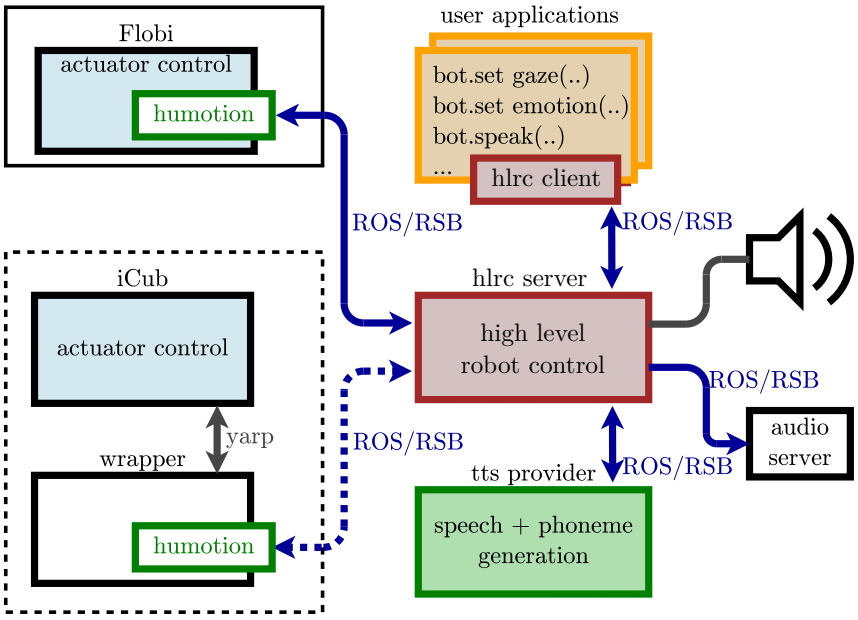



Figure 10.6.: The motion control model as implemented by hlrc.

pression to which the robot returns after the execution of a current emotion target ended. For the sake of simplicity these emotion expressions are implemented in a static way for now but future versions can additionally allow animations for further enhanced expressiveness.

All calls to the high level control server can be either asynchronous or blocking. In order to simplify it even further clients in C++, Python, and Java that hide the actual middleware communication from the user and provides a clean and consistent API for all supported features are provided:

`set_*_emotion(...)`
: The default and current emotion expression of the face can be set by a single call to the associated function. When the ”current emotion” is set, the user passes a timeout. Once timed out, the facial expression will return to the given default state.

`set_gaze_target(...)`
: The overall robot gaze target can be set by two means. This method will set the gaze towards a specific direction given in angles.

`set_lookat_target(...)`
: This is the second possible call to direct the robots' gaze to a given target. In this case, the target has to be provided in the metric coordinate system. This function will also take care of setting the proper eye vergence angle based on the objects' distance.

`set_mouth_target(...)`
: This call sets the mouth actuators to arbitrary values.

`set_head_animation(...)`
: A variety of predefined animations can be triggered by this function call. The list of predefined animations include examples such as nodding, shaking the head, or blinking with the eyes.

`set_speak(...)`
: This method triggers speech output that includes synchronized lip animations based on an input string.

The HLRC library is meant to be an example and proof of concept for the use of the humotion library and can be used for the scripting of simple interaction scenarios. It is important to note that this abstraction layer is not meant to be a behavior modeling language as this would go beyond the scope of this thesis.

## 10.8. Discussion

As stated before, the replication of human motion patterns requires to consider, to analyze, and optimize the whole motion generation chain starting from the mechanics through the low-level motor control to the high-level motion generation. This chapter demonstrates the efforts that are necessary to bring human-like motion to a humanoid robotic head. Given a finished mechanical design that is capable to replicate human-like velocities and acceleration, the appropriate motion control requires the interplay of a chain of necessary sub elements: Starting at the lowest level, a custom motion control circuit board has been developed. A control loop implementation takes care of the low-level motion control by driving brushed DC motors

based on a given target input and the data of different sensors. Based on this, a low-latency and high data throughput communication layer connects the host application and the motion control boards. The successive chain, from the communication server, over the human-inspired model used by the humotion control framework towards the high level robot control framework has been implemented with portability in mind. The humotion library is highly configurable and can also be used on less agile robotic heads by using its unique rescaling and limiting feature. This feature reduces the maximum accelerations and velocities used during the path planning phase and prevents the generation of incorrect or jerky motion patterns.

All the different human motion patterns, as listed in Section 2.6, are replicated by the humotion library and can be used by simple functions calls using the HLRC abstraction layer. Once the humotion library takes over the control of the robot, the vestibular ocular reflex is replicated based on measurements of the neck angles. human-like saccades and smooth motion phases are automatically triggered for gaze target inputs based on a human-inspired, threshold based control model. The eye vergence angle is automatically calculated based on the gaze targets distance, At the same time, human-like eyelid animation is replicated, executed, and enhanced by human-like eye blinks.

The high portability of the framework has been proven by the successful application of the humotion- and hlrc-framework on a variety of different robot platforms (Flobi, Floka, Meka, and iCub). In addition, please refer to the associated video submission of the HAI 2016 presentation [28][1] for a better impression of the overall system performance, the expressiveness, and a qualitative comparison of the proposed humotion library. In this video, two exemplary robotic heads, an iCub and a Flobi robot, actuated by the proposed humotion library, present the feature set of the proposed approach by having a conversation with each other using the humotion and hlrc framework.

In addition, please refer to Chapter 11 for a throughout evaluation of the expressiveness, liveliness, and quality of the motion generated by the application of the proposed control loops and motion control framework.

[1] http://youtu.be/AJm5IVMVKsI

# III

# Evaluation

This part of the thesis is dedicated to the evaluation of the proposed software framework and the different hardware prototypes that were introduced in Part II. The evaluation of the proposed software framework, that was presented in Chapter 10, in the course of a user study with a total of 132 people, is presented in Chapter 11. The participants, ranging from naive to expert users, were questioned in order to assess the quality of the motion generation of the proposed system and the general expectations of the different user groups on how a humanoid robotic head moves. Some parts of the results and the evaluation was presented at the International Conference on Human Agent Interaction (HAI) under the title "Humotion - A human inspired gaze control framework for anthropomorphic robot heads" [28]. Following up, the three different mechanical eye prototypes that have been previously presented will be thoroughly measured and evaluated. The dataset and the analysis of the Flobi robot as presented in Chapter 6 and the tendon-based prototype from Chapter 7 have been initially presented at the ICRA conference [26]. The second dataset and the final analysis of the Floka prototype (Chapter 8) was presented at the ICRA conference in 2019 [27]. Finally, a more elaborate evaluation, combining both datasets, is furthermore enriched with human data, allowing a direct comparison of all three systems to each other and additionally towards the human, is presented in Chapter 12.

Chapter 11

# Software Evaluation Study

In order to get an objective evaluation of the proposed motion generation system that was introduced in Chapter 10, a user study was conducted. This study was meant to evaluate what naive, non-expert users' expectations on a motion generation system for a humanoid robotic head are and how they rate the different aspects of the proposed system. Therefore, the main objectives to be answered by this study were identified as:

**Objective 1: Appreciation of Different Motion Properties**
*How does a user rate the generated motion if the combined motion generation is gradually enabled?*

**Objective 2: The Robot's Intention**
*Does the motion generation influence the users' ability to recognize the robot's intentions?*

**Objective 3: Popularity**
*Which level of activation is the most popular one among the users?*

**Objective 4: Importance of Properties**
*How much importance does a user attach to properties like lively-, fluid-, or human-like motion when interacting with a humanoid robotic head?*

The following Section 11.1 gives a detailed overview of the design and execution of the associated study. The analysis of the obtained results can

be found in Section 11.2. Finally, the gained results are interpreted and summarized in Section 11.3.

## 11.1. Study Design and Execution

The study was designed as an online survey consisting of multiple pages using the LimeSurvey toolkit [191]. All questions were asked and answered in German. In order to retain the first impression and prevent the participants to change their mind in the course of the survey, the survey was configured in a way that it was not possible to go back to previous pages. In addition to the following explanations, a full transcript of all survey pages can be found in Part V. The individual parts of the survey will be explained subsequently.

### 11.1.1. Introduction and Basic Data

The first page functions both as an introduction and as a manual for the study (see Figure 1 in the appendix). Subsequently, the participants were asked to indicate their age, gender, and give a short overview of their familiarity with robots (see Figure 2). In the course of this robot familiarity ranking the participants were asked to rank a given set of six robots from 0 (means nothing to me) to 4 (i have worked with it). An overview of all possible answers including translations can be found in the appendix Table 4.

### 11.1.2. Appreciation of Different Motion Properties

It was a design decision to start the questionnaire with a video and rating block in order to leave the participants completely unbiased at this point. The objective was to get a rating that did not rely on any direct comparison of what the participant's have seen before. Please refer to Figure 3 for the actual question layout.

This part of the survey was meant to gain an insight on the appreciation of different sub parts of the motion generation (Objective 1). In order to evaluate the contribution of the individual motion patterns, a set of four

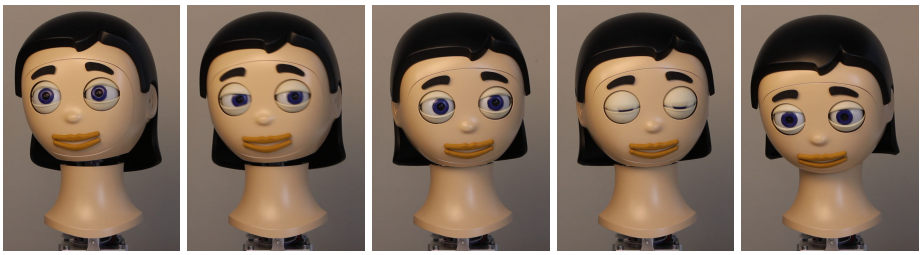


Figure 11.1.: Still frames from video #1 under condition $C_{FULL0.7}$.

different videos were recorded. Each video showed the Flobi robot switching focus between the same three target points while using the following different motion generation strategies:

$C_{NECK}$ : only neck motion. eyes, and eyelids fixed
$C_{NOLIDS}$: neck and eye motion, eyelids fixed
$C_{FULL0.7}$: full motion with reduced acceleration
$C_{FULL1.0}$: full motion

As the full accelerated motion condition $C_{FULL1.0}$ did put a lot of stress on the hardware and was visually a bit shaky, the maximum acceleration was reduced to 70% under condition $C_{FULL0.7}$. Please refer to Figure 11.1 for some exemplary still frame captures taken from the video $C_{FULL0.7}$.

Every participant got to see only one randomly selected video out of this set of four different videos and was asked to rate the following statements:
*The motion shown in this video...*
$S_{FLUID}$ : is fluid
$S_{LIVELY}$: makes a lively impression
$S_{HUMAN}$: appears human-like
$S_{MATCH}$: is well-matched to this robot

These statements were presented in a random order and each participant was asked to rate each on a scale from 1 (totally disagree) to 10 (totally agree). The ranking was designed to utilize the "forced choice" method by dividing the scale into an even number of choices, not letting the user choose a neutral element.

The next questionnaire page (Figure 4) featured an extended set of questions for video #2. This second video showed a short sequence of the robot executing a reading gesture under three different conditions. This time,

condition $C_{FULL1.0}$ was not tested in order to achieve a higher participant count per condition. The next form asked the participants to rate the same statements as for video #1.

### 11.1.3. The Robot's Intention

Beside evaluating the motion generation on a different gesture, the same questionnaire page (Figure 4), was used to investigate how the motion generated by the humotion library improves on how the user empathizes with the robot and understands its intention (Objective 2). Again, only one randomly selected video condition was presented to the participant before being asked to give a guess what the robot is actually doing. The participants were given no further hints or clues and were asked to fill out a free text form.

### 11.1.4. Popularity

The fifth page was explicitly designed to facilitate a *direct* comparison of the first three conditions as shown in video #1 ($C_{NOLIDS}$, $C_{FULL0.7}$, and $C_{NECK}$) in order to answer Objective 3. All conditions were played back synchronously and next to each other in a single video. The participants were asked to nominate the most lively robot by evaluating all possible pairs of two's. Furthermore, the participants were asked which condition they would prefer if they had to interact with the robot. Please refer to Figure 5 for the full question text.

### 11.1.5. Importance of Properties

In order to address our fourth research question (Objective 4), how much importance the user addresses to properties such as fluidness ($P_{FLUID}$), liveliness ($P_{LIVELY}$), and human-like motion ($P_{HUMAN}$) of a humanoid robotic head, the participants were asked to rate the importance of these properties on a scale from 1 (not important at all) to 10 (very important).

## 11.2. Evaluation of the Results

A total of 132 people (63 male and 69 female) participated in the survey over a period of 26 days. The participants' age was distributed between 18 and 60 years with an average age of 30.9 years and a standard deviation of $\sigma$=8.9 years. The random allocation of the different test conditions happened evenly distributed on beginning of the survey. Any differences between the number of participants per condition are due to the fact that only completed surveys were included in the evaluation.

### 11.2.1. Appreciation of Different Motion Properties

The different rating score results for the video #1 questionnaire as described in Section 11.1.2, are shown my means of box plots in Figure 11.2: The x-axes show the participants' rating and the y-axes show the single conditions. The subplot in Figure 11.2a, 11.2b, 11.2c, and 11.2d show the results for the different statements $S_{FLUID}$, $S_{LIVELY}$, $S_{HUMAN}$, and $S_{MATCH}$, respectively. The single boxes show the minimum, first quartile, median, third quartile, and maximum of this rating whereas any occurring outliers are represented by 'X' symbols. All four statements show very similar median values among the tested conditions for all four statements. Under condition $S_{FLUID}$ the median value for $C_{NECK}$ shows the biggest deviation from the other three conditions (5 vs. 7). A similar deviation can be found for $S_{LIVELY}$: Condition $C_{FULL1.0}$ shows a comparable lower median value than the other conditions (4.5 vs. 6). Statement $S_{HUMAN}$ shows median values 7, 6, 6, and 5 for the conditions $C_{NECK}$, $C_{NOLIDS}$, $C_{FULL0.7}$, and $C_{FULL1.0}$, respectively. The median values for $S_{MATCH}$ show the smallest variations, ranging from 7 to 7.5.

Additionally, a descriptive statistical analysis of the gathered data for all four statements ($S_{...}$) under all four conditions ($C_{...}$) is presented in Table 11.1. Of particular note is that the gathered ratings for video #1 show a high variability: Considering the results shown in Table 11.1, all ratings span at least 80% of the possible range (min to max) and show a high standard deviation (SD). Even though most of the rating medians shown in Table 11.1 and Figure 11.2 are above the imaginary neutral point, a first

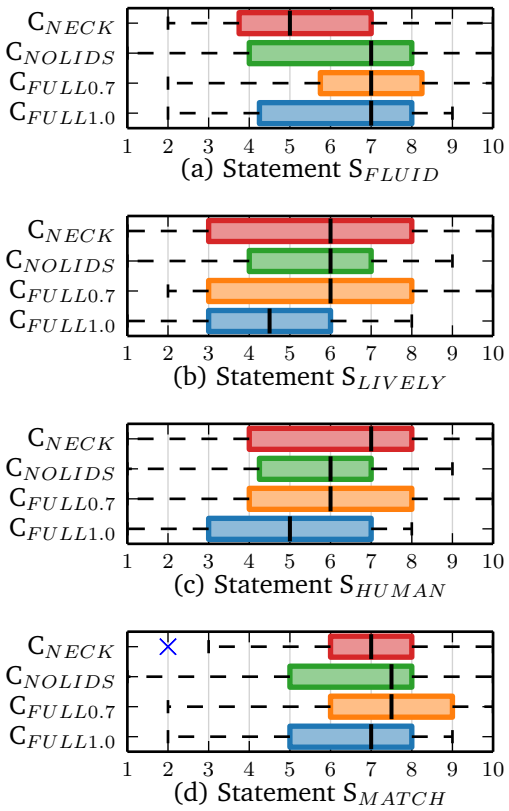



Figure 11.2.: Results of the user evaluation of video #1.

glimpse onto this data reveals no apparent favorite that stands out among the tested conditions. In order to statistically evaluate this first impression, an one-way between-groups analysis of variance was conducted to explore the impact of the four different test conditions C$_{NECK}$ to C$_{FULL1.0}$ for every statement S$_{FLUID}$, S$_{LIVELY}$, S$_{HUMAN}$, and S$_{MATCH}$. The participants were divided into four groups according to the video condition C$_{...}$ that was presented to them. Levene's test for homogeneity of variances show that the dataset does not violate the homogeneity of variance assumption at the $p < 0.05$ level (see Table 11.2). The results of the between-groups analysis are shown in Table 11.2. As suspected initially, this test showed no statisti-

Table 11.1.: Descriptive data analysis for Video #1

| | | N | Min | Max | Mean | Median | SD | SE | 95% conf. Lower | 95% conf. Upper |
|---|---|---|---|---|---|---|---|---|---|---|
| $S_{FLUID}$ | $C_{NECK}$ | 32 | 2 | 10 | 5.56 | 5.0 | 2.35 | 0.42 | 4.71 | 6.41 |
| | $C_{NOLIDS}$ | 34 | 1 | 10 | 6.15 | 7.0 | 2.41 | 0.41 | 5.30 | 6.99 |
| | $C_{FULL0.7}$ | 36 | 2 | 10 | 6.69 | 7.0 | 2.28 | 0.38 | 5.92 | 7.47 |
| | $C_{FULL1.0}$ | 30 | 2 | 9 | 6.13 | 7.0 | 2.18 | 0.40 | 5.32 | 6.95 |
| | Total | 132 | 1 | 10 | 6.15 | 7.0 | 2.32 | 0.20 | 5.75 | 6.55 |
| $S_{LIVELY}$ | $C_{NECK}$ | 32 | 1 | 10 | 5.62 | 6.0 | 2.71 | 0.48 | 4.65 | 6.60 |
| | $C_{NOLIDS}$ | 34 | 1 | 9 | 5.62 | 6.0 | 2.40 | 0.41 | 4.78 | 6.45 |
| | $C_{FULL0.7}$ | 36 | 2 | 10 | 5.94 | 6.0 | 2.76 | 0.46 | 5.01 | 6.88 |
| | $C_{FULL1.0}$ | 30 | 1 | 8 | 4.67 | 4.5 | 1.86 | 0.34 | 3.97 | 5.36 |
| | Total | 132 | 1 | 10 | 5.49 | 6.0 | 2.49 | 0.22 | 5.06 | 5.92 |
| $S_{HUMAN}$ | $C_{NECK}$ | 32 | 1 | 10 | 6.22 | 7.0 | 2.74 | 0.49 | 5.23 | 7.21 |
| | $C_{NOLIDS}$ | 34 | 1 | 9 | 5.44 | 6.0 | 2.23 | 0.38 | 4.66 | 6.22 |
| | $C_{FULL0.7}$ | 36 | 1 | 10 | 5.81 | 6.0 | 2.36 | 0.39 | 5.01 | 6.61 |
| | $C_{FULL1.0}$ | 30 | 1 | 8 | 5.03 | 5.0 | 2.19 | 0.40 | 4.22 | 5.85 |
| | Total | 132 | 1 | 10 | 5.64 | 6.0 | 2.40 | 0.21 | 5.22 | 6.05 |
| $S_{MATCH}$ | $C_{NECK}$ | 32 | 2 | 10 | 6.59 | 7.0 | 2.24 | 0.40 | 5.79 | 7.40 |
| | $C_{NOLIDS}$ | 34 | 1 | 10 | 6.71 | 7.5 | 2.41 | 0.41 | 5.87 | 7.55 |
| | $C_{FULL0.7}$ | 36 | 2 | 10 | 7.08 | 7.5 | 2.16 | 0.36 | 6.35 | 7.81 |
| | $C_{FULL1.0}$ | 30 | 2 | 9 | 6.20 | 7.0 | 2.04 | 0.37 | 5.44 | 6.96 |
| | Total | 132 | 1 | 10 | 6.67 | 7.0 | 2.22 | 0.19 | 6.29 | 7.05 |

Table 11.2.: One-way ANOVA for Video #1

| | | Levene | ANOVA |
|---|---|---|---|
| Video 1 | $S_{FLUID}$ | 0.9485 | 0.2591 |
| | $S_{HUMAN}$ | 0.5148 | 0.2449 |
| | $S_{LIVELY}$ | 0.1216 | 0.1991 |
| | $S_{MATCH}$ | 0.7664 | 0.4532 |

cally significant difference at the $p < 0.05$ level in scores for the four tested groups within each of the four statements.

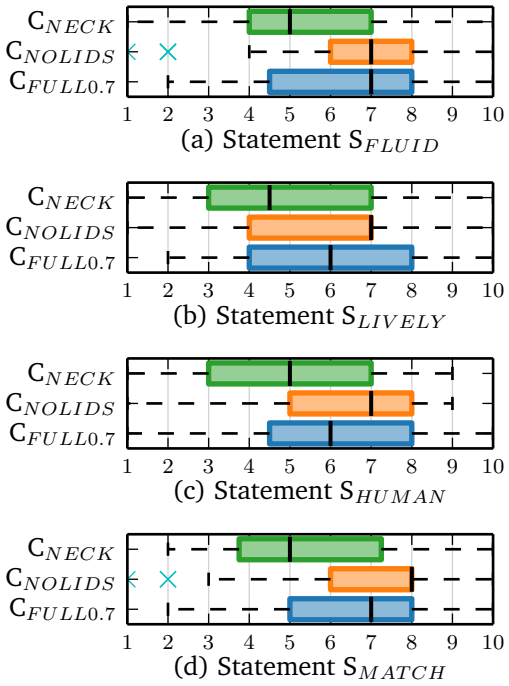



Figure 11.3.: Results of the user evaluation of video #2

The dataset for video #2 and the associated four statements $S_{FLUID}$ to $S_{MATCH}$ were analyzed by the same means: Again, the rating scores for all tested permutations are presented as box plots (see Figure 11.3) in combination with a detailed table that contains statistical descriptives as well (see Table 11.3). The first thing to mention is the higher participant count per condition due to the fact that this time, only three instead of four conditions were tested. A comparison of the minimum and maximum ratings show similar, wide spread use of all available scores and a comparable standard deviation as observed for video #1. This time, the ratings' median values for the test condition $C_{NECK}$ are 1.5 to 3 points lower than for the other conditions under all tested statements. The differences between $C_{NOLIDS}$ and $C_{FULL0.7}$ are rather low, reaching a one point difference at best.

Table 11.3.: Descriptive data analysis for Video #2

| | | N | Min | Max | Mean | Median | SD | SE | 95% conf. Lower | 95% conf. Upper |
|---|---|---|---|---|---|---|---|---|---|---|
| $S_{FLUID}$ | $C_{NECK}$ | 44 | 1 | 10 | 5.23 | 5.0 | 2.19 | 0.33 | 4.56 | 5.89 |
| | $C_{NOLIDS}$ | 49 | 1 | 10 | 6.37 | 7.0 | 2.11 | 0.30 | 5.76 | 6.97 |
| | $C_{FULL0.7}$ | 39 | 2 | 10 | 6.41 | 7.0 | 2.50 | 0.40 | 5.60 | 7.22 |
| | Total | 132 | 1 | 10 | 6.00 | 6.0 | 2.30 | 0.20 | 5.60 | 6.40 |
| $S_{LIVELY}$ | $C_{NECK}$ | 44 | 1 | 10 | 4.68 | 4.5 | 2.36 | 0.36 | 3.96 | 5.40 |
| | $C_{NOLIDS}$ | 49 | 1 | 10 | 5.78 | 7.0 | 2.33 | 0.33 | 5.11 | 6.44 |
| | $C_{FULL0.7}$ | 39 | 2 | 10 | 6.08 | 6.0 | 2.29 | 0.37 | 5.34 | 6.82 |
| | Total | 132 | 1 | 10 | 5.50 | 6.0 | 2.38 | 0.21 | 5.09 | 5.91 |
| $S_{HUMAN}$ | $C_{NECK}$ | 44 | 1 | 9 | 4.93 | 5.0 | 2.30 | 0.35 | 4.23 | 5.63 |
| | $C_{NOLIDS}$ | 49 | 1 | 9 | 6.16 | 7.0 | 2.32 | 0.33 | 5.50 | 6.83 |
| | $C_{FULL0.7}$ | 39 | 1 | 10 | 6.21 | 6.0 | 2.48 | 0.40 | 5.40 | 7.01 |
| | Total | 132 | 1 | 10 | 5.77 | 6.0 | 2.42 | 0.21 | 5.35 | 6.18 |
| $S_{MATCH}$ | $C_{NECK}$ | 44 | 2 | 10 | 5.45 | 5.0 | 2.24 | 0.34 | 4.77 | 6.13 |
| | $C_{NOLIDS}$ | 49 | 1 | 10 | 7.10 | 8.0 | 2.14 | 0.31 | 6.49 | 7.72 |
| | $C_{FULL0.7}$ | 39 | 2 | 10 | 6.79 | 7.0 | 2.05 | 0.33 | 6.13 | 7.46 |
| | Total | 132 | 1 | 10 | 6.46 | 7.0 | 2.25 | 0.20 | 6.07 | 6.85 |

Table 11.4.: One-way ANOVA for Video #2

| | | Levene | ANOVA | $\eta^2$ |
|---|---|---|---|---|
| Video 2 | $S_{FLUID}$ | 0.4455 | 0.0232 | 0.0567 |
| | $S_{HUMAN}$ | 0.8299 | 0.0186 | 0.0599 |
| | $S_{LIVELY}$ | 0.9225 | 0.0162 | 0.0620 |
| | $S_{MATCH}$ | 0.3929 | 0.0008 | 0.1039 |

The results of further statistic tests are shown in Table 11.4: Again, the column marked "Levene", depicting the result of the precondition test, showed no violation of the homogeneity of variance assumption so that a one-way between-groups analysis of variance could be applied. This time, a statistically significant difference was found for $S_{FLUID}$, $S_{LIVELY}$, $S_{HUMAN}$, and $S_{MATCH}$ at the $p < 0.05$ level with a $p$ value of 0.023, 0.016, 0.019, and 0.001, respectively. The effect size, calculated using $\eta^2$, was calculated as 0.057, 0.062, 0.060, and 0.104, respectively. Despite reaching statistical significance, the actual difference in mean scores between the groups $C_{NECK}$, $C_{NOLIDS}$, and $C_{FULL0.7}$ was small to just medium for all statements.

In order to gain a detailed insight for the pair-wise differences additional post-hoc comparisons using the Tukey HSD test indicate that the mean score for the following pairs was significantly different: For Statements $S_{FLUID}$, $S_{HUMAN}$, and $S_{MATCH}$ the conditions $C_{NECK}$ was found to be significantly different from $C_{NOLIDS}$ and $C_{FULL0.7}$. No significant difference was found between $C_{NOLIDS}$ and $C_{FULL0.7}$. The data for statement $S_{LIVELY}$ indicated only a significant difference between $C_{NECK}$ and $C_{FULL0.7}$. Further statistical analysis showed that the effect size was rather low, reaching only just medium at best.

### 11.2.2. The Robot's Intention

In order to evaluate how, or if, the motion generation algorithm can support and guide the understanding of the robot's intentions, the participants were asked to give a guess about the robot's current activity based on the shown video sequence. A full overview of all given answers and the associated categorization is depicted in the appendix (see Table 3). For further analysis, a manually categorized set of the six most named answers is shown in Table 11.5: Roughly one third of the participants that got to see the least actuated condition $C_{NECK}$ gave the correct guess ("reading") for the sequence of movements shown. This rate increases to 40.8% for the more human-like animated test condition $C_{NOLIDS}$ and reaches 48.7% for the full actuated condition $C_{FULL0.7}$. The very similar looking action "looking at something" was cited second most frequently with 25.0%, 32.7%, and 33.3%, respectively. Altogether it can be said that those two highest rated answers are practically correct descriptions for the presented activity. The

Table 11.5.: Guessed robot action for Video #2

| Category | Tested condition | | | | | |
|---|---|---|---|---|---|---|
| | $C_{NECK}$ | | $C_{NOLIDS}$ | | $C_{FULL0.7}$ | |
| reading something | 15 | 34.1% | 20 | 40.8% | 19 | 48.7% |
| looking at something | 11 | 25.0% | 16 | 32.7% | 13 | 33.3% |
| searching something | 5 | 11.4% | 0 | 0.0% | 1 | 2.6% |
| moving | 4 | 9.1% | 2 | 4.1% | 2 | 5.1% |
| head-shaking/disapproval | 3 | 6.8% | 0 | 0.0% | 0 | 0.0% |
| being ashamed | 2 | 4.5% | 1 | 2.0% | 3 | 7.7% |
| no answer | 4 | 9.1% | 10 | 20.4% | 1 | 2.6% |
| Sum | 44 | 100.0% | 49 | 100.0% | 39 | 100.0% |

difference between reading a book and looking at it might be to subtle to notice. If both answers are accepted as correct and the results are combined the correct identification ratio amounts to 59.1%, 73.5%, and 82%, respectively. Considering the incorrect identification ratio, the following picture emerges: In the least actuated condition $C_{NECK}$ 40.9% of the answers were incorrect. For $C_{NOLIDS}$ a total amount of 26.5% of the answers were incorrect. This trend continues, for the most actuated condition $C_{FULL0.7}$, only 18% of the given answers were wrong. Summarizing it can be said that the more actuated the robot shown in the test condition was, the higher the correct identification ratio. The amount of incorrect answers can be halved by enabling the full motion generation and thus switching from $C_{NECK}$ to $C_{FULL0.7}$.

### 11.2.3. Popularity

The evaluation of the ratings for the first two videos revealed that only a slight user preference towards the more actuated conditions could be identified among the conditions tested. It is important to note that these results were obtained in a scenario where the participant had no direct object of comparison. In contrast, the next form asked the participants to rank the different conditions in direct comparison to each other: The comparison of $C_{NOLIDS}$ and $C_{NECK}$ (52.3% vs. 47.7%), resulted in a slight advantage for

$C_{NOLIDS}$. A clear majority of 90.2% ranked $C_{FULL0.7}$ to be more lively than $C_{NECK}$ and 96.2% of the participants ranked $C_{FULL0.7}$ to be more lively as $C_{NOLIDS}$. When being asked with which of the differently actuated robots the participants preferred to interact the most, a total of 92.4% choose condition $C_{FULL0.7}$. Taken as a whole, these results show that, in direct comparison, the participants clearly prefer our motion generation algorithm as used in condition $C_{FULL0.7}$.

### 11.2.4. Importance of Properties

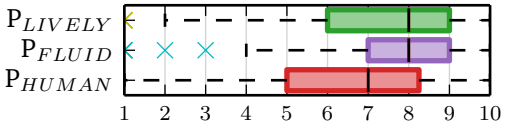



Figure 11.4.: Results of the importance voting.

Finally, the participants were asked to vote the importance of the properties $P_{FLUID}$, $P_{LIVELY}$, and $P_{HUMAN}$ concerning robot motion during an interaction with this robot. These results are shown as box plots in Figure 11.4: All three properties were ranked as quite important, the properties fluidness and liveliness got the highest median ranking ($M = 8$). The property $P_{HUMAN}$ ranked lower ($M = 7$) and showed a tendency of more rankings towards lower scores with a lower quantile of 5. Further statistical analysis by means of an one-way between-groups analysis of variance indicates a statistical difference at the $p < 0.05$ level among the three statements. However, a post-hoc comparisons using the Tukey HSD test showed no differences in the mean scores for all possible pairs of two's.

In order to analyze if different levels of importance are attributed to the properties among different user groups the participants were allocated to different groups based on the robot familiarity survey: The unfamiliar group consists of people that never heard the name of any of the named robots (21.2%). People that only heard the name of at least one of the robots (22.7%) are categorized into the novice group. The seen-live group consists of the participants that have seen at least one of the robots, but never worked with any of the robots (38.6%). The remainder, people categorized

into the experts group (17.4%), have actually worked with at least one of the robots. In order to see if members of these groups have different wishes on how a humanoid robot should move, again an one-way between groups analysis was conducted. The results show no statistically relevant difference among the importance voting of these four groups.

In addition to that, neither a gender- nor an age-specific significant difference was found using statistical tests when the whole participant group was split into groups based on gender and by age using $<$30 years (N=69), 30-40 years (N=48), and $>$40 years (N=15) as thresholds. In closing it can be stated that all three properties fluidness, liveliness, and human-likeness are of equally high importance to all participants, regardless of age, gender, or robot experience.

## 11.3. Discussion

The first part of the study was meant to provide insights and answers to Objective 1. The evaluation of the associated data showed that the compilation of a ranking for the motion shown by the robot without any comparison produced very widespread results. Despite the large differences in actuation for the four tested conditions, the very first assessment for video #1 showed no significant differences at all in terms of fluidness, liveliness, human-likeness, and well matched motion. Even though, due to the reduced number of conditions tested, the number of participants for each of the three tested conditions of video #2 was slightly higher than for video #1, the obtained results were still surprising to some extend. The comparatively lower ranking of $C_{NECK}$ (in comparison to $C_{NOLIDS}$ and $C_{FULL0.7}$) matched the assumptions made beforehand but overall a clearer outcome, especially for $C_{FULL0.7}$, was expected.

With regard of Objective 2, it can be said that evidence suggests that a more human-like actuation can boost the users ability to identify the activity carried out by the robot and the user is able to better sense the robot's intentions. That the more actuated the robot shown in the test condition was, the higher the correct identification ratio. The transition from $C_{NECK}$ to $C_{FULL0.7}$ effectively halves the amount of incorrect answers.

The unclear outcome of the first part evaluating Objective 1 was surpris-

ing. However, if multiple conditions are shown in a direct comparison, the vast majority of participants clearly voted for the proposed human-like motion generation algorithm in terms of liveliness as well as their most preferable interaction partner (Objective 3). These results could suggest that it is difficult for participants to give a ranking without a direct comparison of another condition in juxtaposition. This is not surprising considering that every participant has a different background in robot knowledge and different expectations on anthropomorphic robots.

With regard to Objective 4, it has been shown that all three properties were ranked as very important among all user groups, regardless of gender, age, or robot familiarity. Particularly the fact, that this was not just rated important among experts but true among all user groups, shows that the way the motion is generated plays an important role for the end-user acceptance of an anthropomorphic robotic head and should not be underrated or neglected.

In summary, it can be said that the analysis of the rankings given by the participants confirmed the position that proper motion generation plays an important role in human robot interaction. As hypothesized, the participants indeed preferred the proposed motion generation algorithm, which combines human-like neck-eye coordination with eyelid animation, to the other conditions tested.

Chapter 12

# Hardware Analysis

In order to evaluate that all requirements introduced in Chapter 3 have been fulfilled, a throughout evaluation of all three prototypes has been executed. A variety of different tests were performed and the collected data was analyzed. The underlying dataset and the analysis of the Flobi and the Tendon prototype have been initially presented at the ICRA conference in 2012 [26] and the dataset and analysis of the Floka prototype was presented at the ICRA conference in 2019 [27]. The combined results will be presented in more detail in the following sections.

## 12.1. Range of Motion

In order to evaluate the different systems in terms of the available range of motion, the internal cameras were used to track the gaze direction by utilizing an external marker. The dataset for the Flobi and Tendon prototype was collected by using a custom marker tracking software based on openCV. For those runs, a single $10\,\mathrm{cm} \times 10\,\mathrm{cm}$ marker was placed in a straight line in front of the eyeball at a distance of $0.2\,\mathrm{m}$. The Floka prototype was evaluated by using a marker map in combination with the Aruco tracking library [164]. This time, the eye prototype was placed $0.3\,\mathrm{m}$ away from a wall equipped with a marker map which was $1\,\mathrm{m}$ high and $2\,\mathrm{m}$ wide with 200 equidistant placed $50\,\mathrm{mm} \times 50\,\mathrm{mm}$ markers. An example of the visual scene including an overlay of the results of the marker detection algorithm is depicted in Figure 8.5. In order to assess the maximum range of motion all software joint safety limits were disabled and the eyes' pan

and tilt motors were moved to all extremes in a ray pattern originating from looking straight ahead. In order to compare all three prototypes and the human capabilities, all datasets were merged with human data extracted from [31]. The resulting polar plot is depicted in Figure 12.1. The Flobi

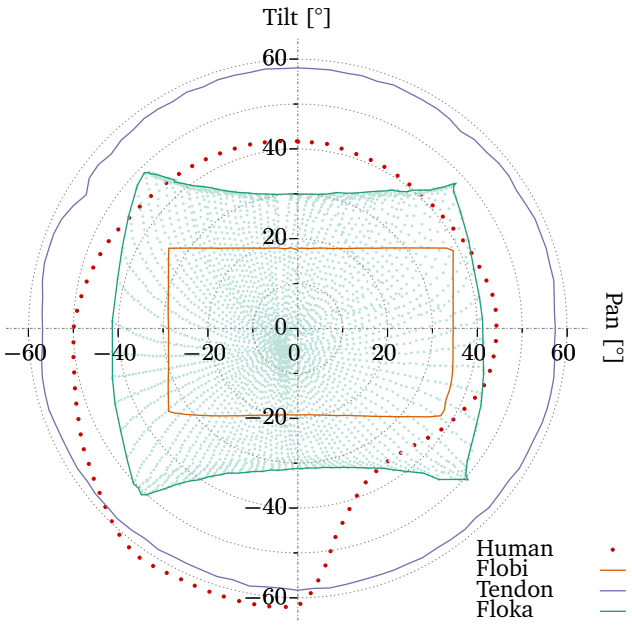



Figure 12.1.: Comparison of the range of motion between the Human- (red), Flobi- (orange), Tendon- (purple), and the new Floka-eye (green).

robot possesses the smallest range of motion with roughly $\pm 30°$ horizontally and $\pm 20°$ vertically. This is considerable different from the human capabilities as introduced in Section 2.4 and the requirements specified in Section 3.3 as it covers only about 33.6% of the human range. Worse still, this only covers 75.5% of the most used practical field of view. This restriction became apparent during the use of the first Flobi robot and lead to the development of the next two prototypes: The first prototype, the tendon-based eye, covers a region of circularly $55°$, which overlaps with 97.3% of

the human range and can reach additional regions that are not accessible for the human. Although this version would be the perfect match in terms of reach, tests presented in Section 12.5 showed that it suffers uncontrollable eye torsion during these extreme movements. In addition, it turned out to be mechanically more complex to integrate as initially thought. The final prototype, the eye mechanics integrated into our new Floka robot, was evaluated to be able to reach around $\pm 30°$ vertically and about $\pm 40°$ horizontally, which amounts to 69.2% of the human oculomotor range. This is twice as much as the initial Flobi design could reach, and around 70% of what the tendon prototype achieved.

## 12.2. Mutual Axis Dependence

The mechanical construction and the associated kinematics of the tendon-based prototype ensure that the axis of rotation for one actuator is always in line with the attachment points of the other actuator. For that reason both actuators can be moved independently by the associated single motor. This is different for the Flobi and Floka prototype Due to the location of the lever attachment points and the fulcra, the pan- and tilt- actuation systems are in theory not independent of each other. In particular at the actuator extreme positions, a change in the position of one motor influences both end-effector positions at the same time. However, for the Flobi prototype, the combination of a small range of motion and the choice of the lever attachment points ensure that, in practice, this effect is neglectable. The increased range of motion of the Floka prototype and the associated change of the lever attachment points made this effect clearly visible. Although less pronounced, this also holds for the opposite where the tilt motor influences the pan end-effector. This dependency can be illustrated with the help of Figures 12.2 and 12.3: These two plots show the mapping of the tilt- (x) and pan-motor (y) coordinates in encoder ticks to the angular deflection of the eye tilt and pan end-effector in each case. For an ideal system these transfer functions would be flat planes. However, the plots show a clear influence of the both motor position on the individual single gaze angles. In order to compensate for that at controller level, the fourth order polynomial transfer function given by Equation (12.1) has been fitted to the datasets for

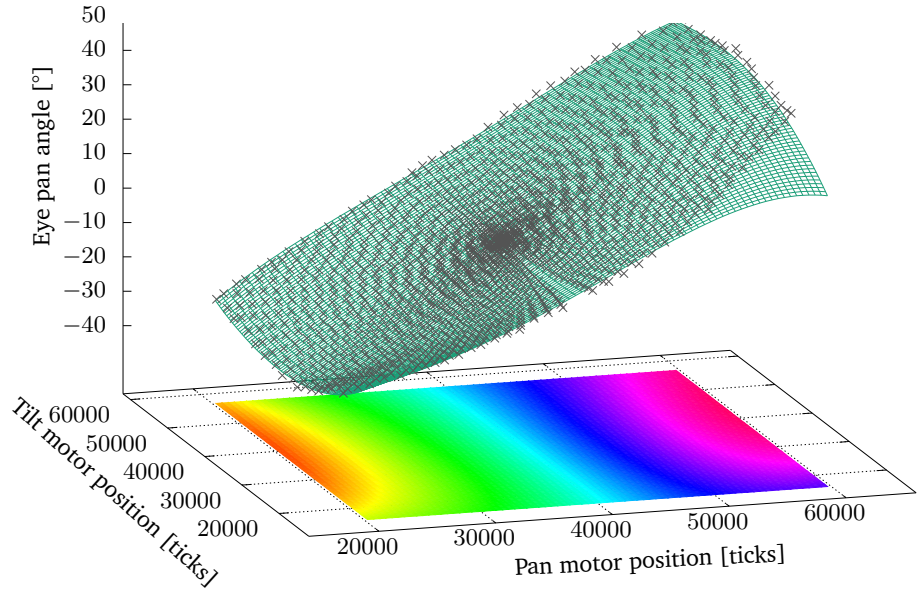



Figure 12.2.: Mapping from motor coordinates x,y to gaze pan angle for the Floka prototype.

both axes.

$$f_*(x,y) = c_1y^4 + c_2xy^3 + c_3x^2y^2 + c_4x^3y + c_5x^4 + c_6y^3 + c_7xy^2 \\ c_8x^2y + c_9x^3 + c_{10}y^2 + c_{11}xy + c_{12}x^2 + c_{13}y + c_{14}x + c_{15} \tag{12.1}$$

The fit result for the parameters $c_*$ given in Table 12.1 is depicted as the green grid in Figures 12.2 and 12.3. After fitting, the total root mean square error of residuals amounts to $RMS_{pan} = 0.54°$ and $RMS_{tilt} = 0.64°$ for the pan- and tilt- actuator, respectively. The variance of residuals amounts to $\sigma_{pan} = 0.29$ for the pan- and $\sigma_{pan} = 0.41$ for tilt- actuator. A closer look at Table 12.1 with respect to $f_{pan}$ reveals that the higher order terms are quite small and do not contribute much in this case. This matches the observation that the plotted dataset for the pan actuator, depicted in Figure 12.2, looks almost planar in the center region and that the effect is visually less pronounced when moving the pan actuator of the physical robot. The overall fitting error for the gaze pan- and tilt-angle is depicted in Figure 12.4a and Figure 12.4b, respectively. Except for six single outliers, it can be quantified

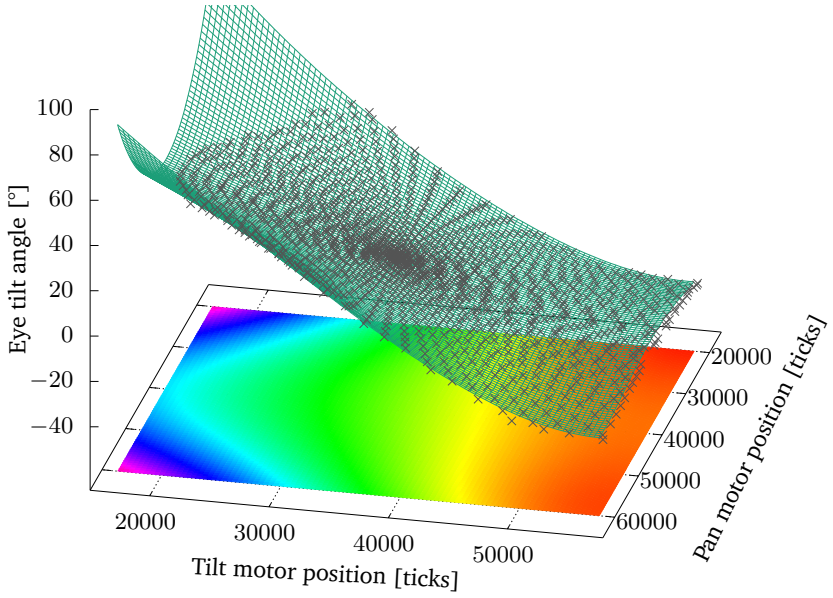


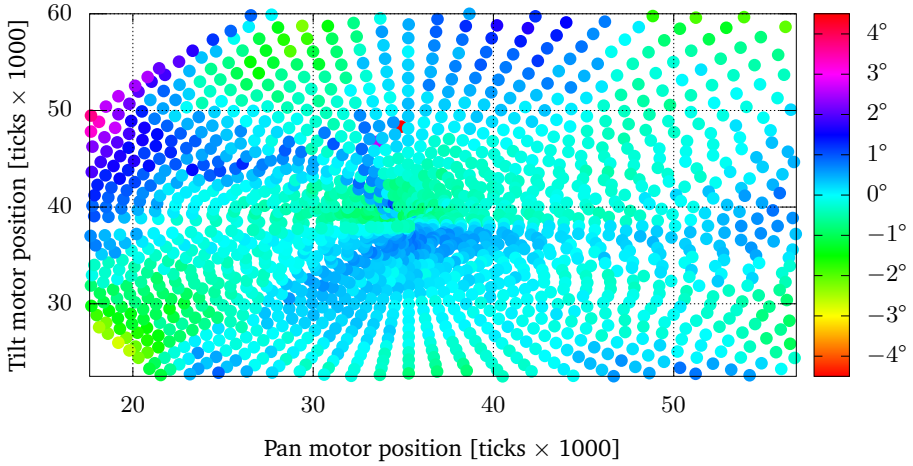

Figure 12.3.: Mapping from motor coordinates x,y to gaze tilt angle for the Floka prototype.

as less than $\pm 3.0°$ for $f_{pan}$ globally with $\pm 1.0°$ for the circular center region with $r = 30°$. Similar values can be derived for $f_{tilt}$: The overall fitting error amounts to $\pm 4.0°$ globally and $\pm 1.5°$ in the center region if 10 single outliers are omitted as measuring errors.

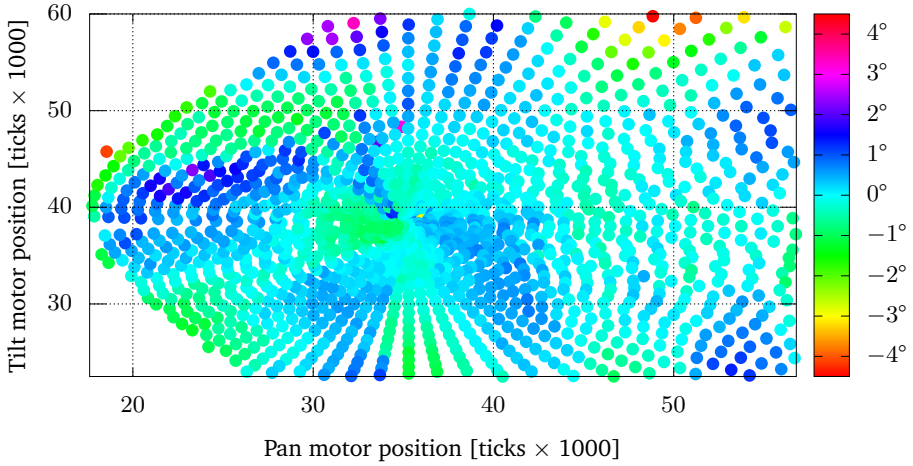
| | $f_{Tilt}$ | $f_{Pan}$ |
|---|---|---|
| $c_1$ | 2.61e-05 | 4.15e-18 |
| $c_2$ | 8.76e-06 | 1.02e-17 |
| $c_3$ | 9.11e-05 | -4.13e-18 |
| $c_4$ | 3.28e-05 | 3.64e-17 |
| $c_5$ | 2.59e-05 | -2.95e-17 |
| $c_6$ | -0.003567 | -9.83e-13 |
| $c_7$ | -0.008642 | -1.98e-12 |
| $c_8$ | -0.014081 | -4.03e-12 |

| | $f_{Tilt}$ | $f_{Pan}$ |
|---|---|---|
| $c_9$ | -0.005265 | 2.77e-12 |
| $c_{10}$ | 0.350315 | 1.15e-07 |
| $c_{11}$ | 1.02003 | 2.72e-07 |
| $c_{12}$ | 0.645692 | -4.86e-08 |
| $c_{13}$ | -25.389 | -0.0073842 |
| $c_{14}$ | -33.2974 | -0.0016023 |
| $c_{15}$ | 656.14 | 48.2059 |

Table 12.1.: Parameters of fitting Equation (12.1) to the acquired dataset for the Floka prototype.




(a) Measured for the pan actuator / $f_{pan}$




(b) Measured for the tilt actuator / $f_{tilt}$

Figure 12.4.: Floka actuator fitting error measured in degrees between the approximation formula $f_{...}$ and the dataset points. The raw motor positions are shown on the x- and y-axis, the fitting error is encoded in the point color.

## 12.3. Velocity and Acceleration

In order to quantify and compare the performance of all three systems to each other and to the human, a total of 1000 different length eye saccades were executed. Just like the maximum velocities of the human oculomotor system, the reachable eye end velocity starting from standstill correlates to the total traveled distance. The graphs in Figures 12.5 and 12.6 show this effect by plotting the length of the different saccades on the x-axis and the reached end velocity on the y-axis. The plots are based on datasets acquired using the Flobi-, the Tendon-, and the Floka- prototype. In addition, typical human values extracted from [57] (see Figure 2.5) are included in the plots for reference. With the continuous yellow line representing the average- and the dashed line representing the standard deviation of the reached saccade velocity of the tested subjects during the study by Leigh *et al.* [57]. The gray sector marks the typical maximum velocities humans can reach as introduced in Section 2.6.3. Both plots show that all proposes systems reach velocities of up to $800\,°/s$ to $900\,°/s$ that are well above the human counterpart. Solely the tilt actuator of the Floka prototype stagnates at around

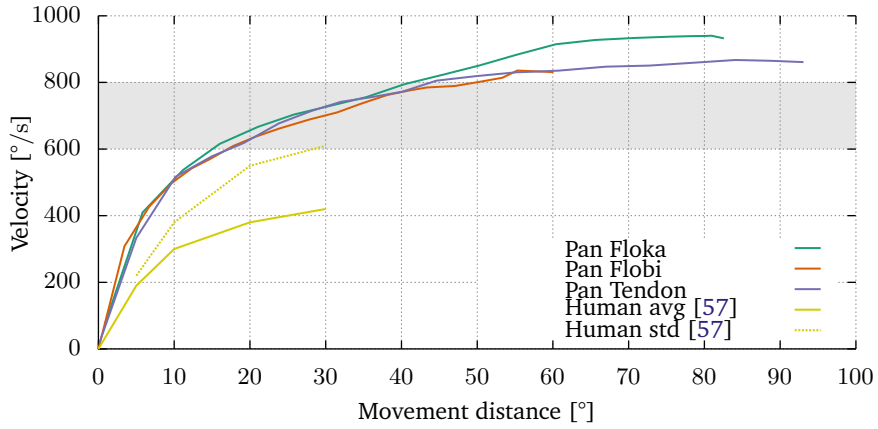



Figure 12.5.: Maximum reachable velocity from standstill for the pan actuator of the different prototypes and the human.

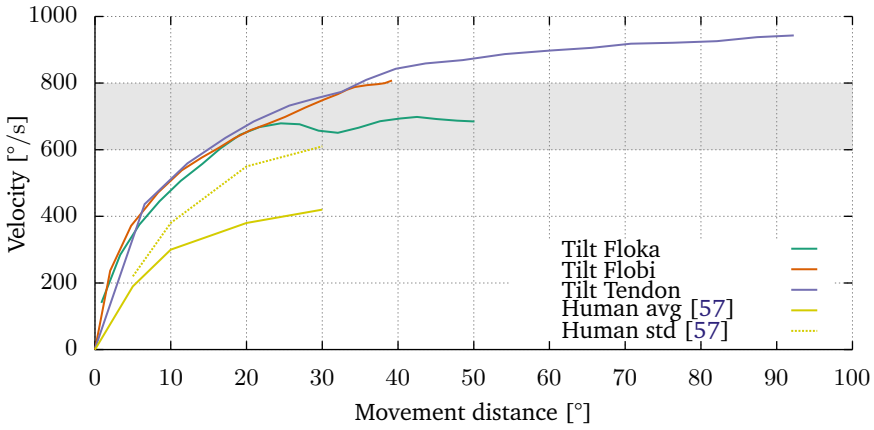



Figure 12.6.: Maximum reachable velocity from standstill for the tilt actuator of the different prototypes and the human.

$700\,°/s$. A potential fix could be the extension of the actuating lever and thus optimizing the translation factor between motor rotation and eye tilt motion if deemed necessary. But even without this fix, the reached maximum velocity of this actuator is well within the human range.

Beside logging the maximum velocity during these trials additional acceleration measurements were conducted. Again, the results for all three systems tested and the human are shown in Figures 12.7 and 12.8: The total distance traveled during a saccade is plotted on the x-axis and the maximum acceleration that was reached, measured in $°/s^2$, is depicted on the y-axis. In addition, human values, manually extracted from Bahill *et al.* [59], are given for reference. The Flobi prototype reached between $70\,000\,°/s^2$ to $80\,000\,°/s^2$ for the tilt and pan actuator, respectively. The biggest peak acceleration was measured for the tendon-based prototype with $90\,000\,°/s^2$ to $100\,000\,°/s^2$. The Floka prototype scored lower, with around $60\,000\,°/s^2$ for the pan- and $40\,000\,°/s^2$ to $75\,000\,°/s^2$ for the tilt-actuator. All three prototypes are well within the defined Requirement $R_{3.3}$ and some even exceed the human performance given in Section 2.6.3.

Similar tests to these saccade measurements have been executed for the evaluation of the Floka prototype's eyelids and the maximum accelerations and velocities were recorded. The eyelids of the Floka prototype reached

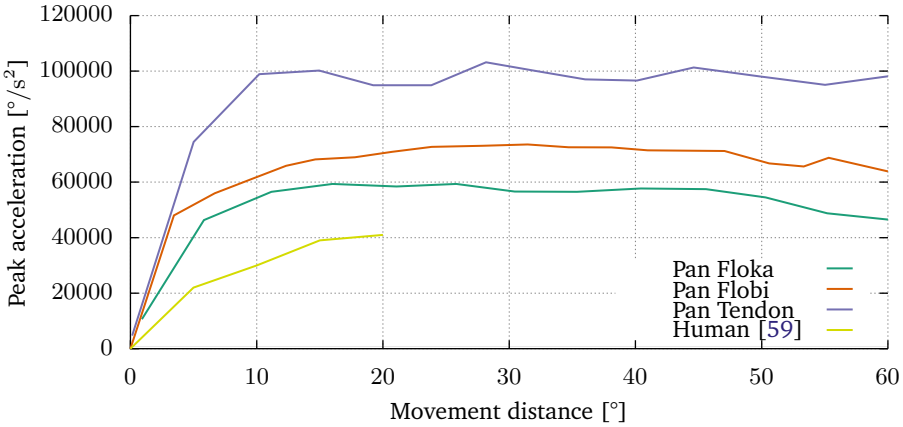



Figure 12.7.: Maximum acceleration during a saccade of a given length for the pan actuator of the different prototypes and the human.

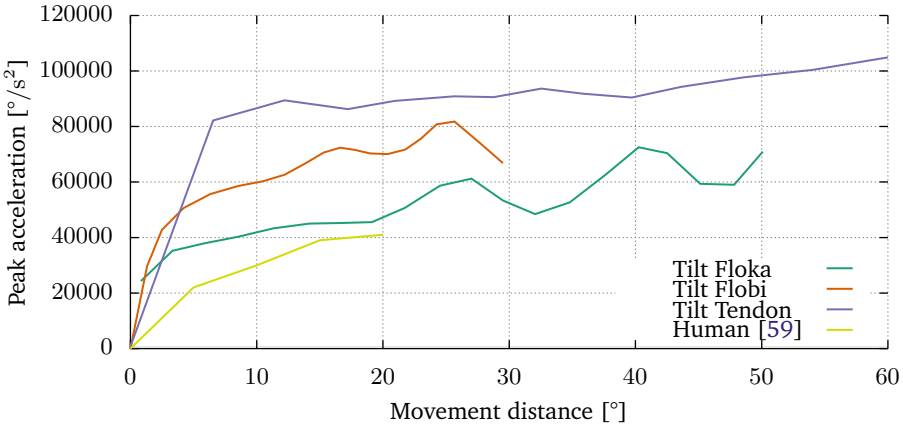



Figure 12.8.: Maximum acceleration during a saccade of a given length for the tilt actuator of the different prototypes and the human.

peak-velocities of up to $1100\,°/s$ and -accelerations of up to $130\,000\,°/s^2$. This is allows Flokas eyelid actuation system to undercut the human counterpart and execute a full eye blink motion (close and open up again) in less than $90\,ms$, which results in Requirement $R_{3.5}$ being more than fulfilled.

The tendon-based prototype does not feature eyelids and has thus not been evaluated. The previous Flobi prototype reached similar values with eye blinks lasting less than $120\,ms$ but had problems in terms of mechanical stability. The attachment levers were prone to break on e.g. eyelid collisions. The mechanical construction of the Floka eyelids with a passively closing spring assembly works around this issue.

## 12.4. Gearbox Backlash

One key factor in accuracy optimization is the minimization of gearbox backlash. The starting point, the initial Flobi design, was facing a large positioning error caused by the gearbox backlash. Two different mitigation strategies have been deployed and evaluated on the tendon- (Chapter 7) and the Floka prototype (Chapter 8). The tendon-based prototype removed the influence of the gearbox backlash by placing a high resolution positioning sensor after the gearbox. In this case, the elimination of the backlash comes at the cost of a less controllable system as any change in the motor rotation is measured with a delay at the positioning sensor. The second mitigation strategy was deployed on the Floka prototype. The key ingredient of this system is the special zero backlash gearbox. According to the gearbox datasheet [156], there should be no measurable backlash at the gearbox output under no load. However, in our application the total backlash, measured at the eyeball gaze angle, additionally depends on any potential slack in the lever actuation mechanism. In order to qualify the total system backlash from motor input to end-effector position, a series of measurements have been acquired for all three systems. Again, the eyeball was facing and tracking a marker as in previous tests and the difference between two target approaches from different directions have been recorded. Our starting point, the Flobi prototype, possesses an average backlash of $1.62°$ and $1.41°$ for the pan- and tilt- axis, respectively. The performance of this prototype is not surprising as it uses a gearbox that is specified to have a backlash

of up to $1.5°$ [192]. The tendon prototype was evaluated to possess an average backlash of $0.42°$ and $0.5°$. Even though this system uses the same motor gearbox combination as the Flobi prototype, the placement of the position sensor on the gearbox output shaft successfully reduced the backlash to roughly one third. The total backlash measured for the Floka prototype amounts to $0.53°$ and $0.29°$ for the pan- and tilt- axis, respectively. Using a similar linked actuation system as the Flobi prototype, the backlash on the Floka robot could be significantly reduced by using the proposed zero backlash gearbox.

## 12.5. Unconstrained Eye Torsion

The tendon prototype possesses the largest range of motion and features very high velocity and acceleration values combined with a very high angular accuracy. However, these benefits come at the cost of unconstrained eye torsion: Given the fact that the eye torsion is not actively being controlled by a third actuator, the eyeball can, to some degree, execute unconstrained rotations. The eyeball is usually attracted to the untwisted zero degree torsion during a move as the posture which applies the minimum additional force on the tendons will be the one occupied most likely. In order to verify this assumption several random saccades were executed and the camera was used to track a visual marker. This marker allowed to measure the variation in the angular eye torsion after each move. Figure 12.9 shows a histogram of the values acquired by executing 800 test runs. The absolute angular deviation is depicted on the x- and the frequency of occurrence on the y-axis. Over 50% of the single saccades showed less than $1.6°$ deviation. There were no saccades detected that were twisted more than $9°$. One advantage of the tendon approach is that it is possible to optimize the tendon attachment- and the contact-points in a way to replicate the human eye torsion caused by Listings' law as introduced in Section 2.5 on a mechanical basis [193]. The current tendon prototype was not optimized in this regard. The mechanical construction of the Flobi and the Floka prototype do not allow eye torsion by design due to the missing degree of freedom and thus do not suffer unconstrained eye torsion. In order to mimic the human eye torsion described by Listings' law a third degree of freedom would be

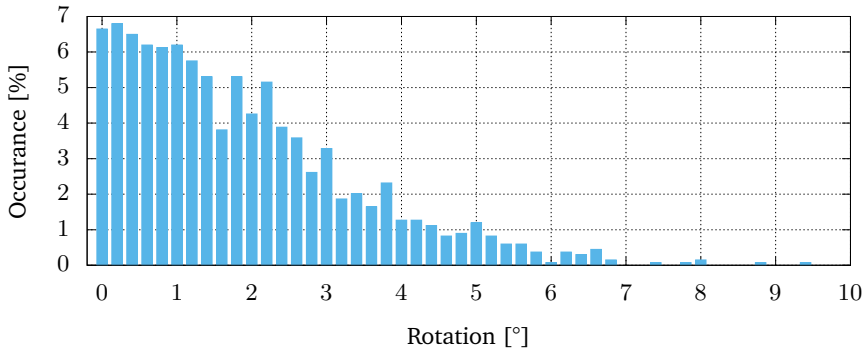



Figure 12.9.: Histogram of angular eye torsion of the tendon prototype.

necessary. Due to missing features on the iris the rotation itself would not visible by the user observing the robot eye from the outside. If necessary for human-like processing of image features, this torsion has to be replicated by rotating the captured image in software.

## 12.6. Discussion

In this chapter, the two different mechanical robotic eye prototypes as introduced in Part II have been analyzed and evaluated. The first prototype is inspired by the human muscles and tendon system: It uses strings that attach to the eyeball in a similar fashion as the tendons do on the human eyeball. There are some design ideas that worked out quite well on this prototype whereas some aspects will need some more work and improvements. First of all, crafting the main structure and the redirecting pulleys using 3D printed parts has been a success: Arbitrary shapes, short lead time combined with uncomplicated assembly, and good overall stiffness can be seen as a positive outcome. In addition, the reduced cost and manufacturing time speeds up the development process and results in a faster evolution of robotic heads. This approach features the largest range of motion of all three presented systems and its maximum velocities and accelerations met all requirements. There is some unconstrained eye torsion, which can be

coped with in certain applications during human robot interaction, however, it is to high for calculating e.g. depth images using disparity information from the two eyeball cameras. In order to have better control over the unconstrained eye torsion, future designs could use some additional bearings in order to guide the tendon at a position closer to the eyeball. In total, the backlash error of the whole system has been reduced by two-third by using magnetic linear encoders that read out a magnetic strip glued onto the actuating pulley. An aspect that needs work is the tendon tensioning system which definitely needs some minor adjustments: The contact pressure of the screw based tensioning system is not high enough, after long periods of use or end position collisions the tendon might get loose, which requires a re-fixation and recalibration.

The second prototype has been successfully integrated into the humanoid robot research platform Floka. It does ot only feature human-like velocities and accelerations, but also integrates fast responding eyelids into an overall system. Its enclosed and non-coupled nature facilitates integration into any arbitrary humanoid robotic head design. The proposed system does not rely on any coupled joints and therefore each individual eyeball with its eyelids and actuators can be placed and mounted on a freely chosen position without restrictions. For a standard binocular setup two mirrored copies of this system are installed next to each other. All mechanical parts are hidden from the users' view by using matching exterior shells. In addition, the aged and low-resolution firewire camera has been replaced by a state-of-the-art high resolution camera with a modern USB3 host interface. A throughout evaluation verified the human-like performance in terms of required velocities, accelerations, range of motion, and accuracy. The presented eyelid actuation scheme even exceeds the human performance. The main advantages of this prototype, are the reduced backlash, the increased range of motion, high velocities, high accelerations, and the integrated high resolution camera. Even though the measured range of motion of this prototype is less than found on the tendon-based approach, it covers 69.2% of the human oculomotor range, which is as much as the initial Flobi design could reach.

In sum, this prototype exceeds the previous Flobi based system and most other robot eyes in all respects, and is the only one in the literature to fulfill the appearance, the motion, and the sensory requirements outlined all in one. By using state-of-the-art and off-the-shelf components such as motors,

| | | Flobi | Tendon | Floka |
|---|---|---|---|---|
| Design | $R_{1.1}$: Closed Shell | ✓ | ✗ possible addon | ✓ |
| | $R_{1.2}$: Eyeball Design | ✓ | ✓ | ✓ |
| | $R_{1.3}$: Eyelids | ✓ | ✗ possible addon | ✓ |
| Camera | $R_{2.1}$: Included Camera | ✓ | ✓ | ✓ |
| | $R_{2.2}$: Single Lens | ✓ | ✓ | ✓ |
| | $R_{2.3}$: Camera FOV | ✗ | ✗ | ✓ |
| | $R_{2.4}$: Camera Acuity | ✗ | ✗ | ✓ |
| Motion | $R_{3.1}$: Motion Range (practical) | (✓) | ✓ | ✓ |
| | $R_{3.2}$: Motion Range (full) | ✗ | ✓ | (✓) |
| | $R_{3.3}$: Velocity & Acceleration | ✓ | ✓ | ✓ |
| | $R_{3.4}$: Smooth Motion Control | ✓ | ✓ | ✓ |
| | $R_{3.5}$: Agile eyelids | ✓ | ✗ | ✓ |

Table 12.2.: Compliance analysis of the three prototypes towards the requirements introduced in Chapter 3

encoders, and the camera subsystem this will foster the further evolution of robotic heads. The CAD model of this prototypes is available free of charge on request for non-commercial applications.

In order to objectively evaluate the presented systems, it is mandatory to objectively assess the individual compliance of all three prototypes in regards of the initial requirements compiled in Chapter 3. The compiled results, depicted in Table 12.2, clearly demonstrates the major drawbacks of the initial Flobi design. As already shown in Section 4.5, it meets all design requirements but fails to meet the camera Requirements $R_{2.3}$ and $R_{2.4}$ as the used camera and lens does neither allow the required FOV, nor deliver the necessary acuity. The biggest issue with this system is the limited range of motion: The mechanical construction of the eyeball actuation system massively limited the available range of motion, covering only 33.6% of the human oculomotor range and yet worse, only 74.5% of the most used practical field of view. This is where the tendon prototype best displays its strengths. It is the most capable prototype in terms of range and speed of motion as it covers 100% of the human oculomotor range and outperforms all other presented prototypes (Requirements $R_{3.1}$, $R_{3.2}$, $R_{3.3}$, $R_{3.4}$). As it was meant for technology evaluation, there was no planned eyelid support and it used the same camera as the Flobi prototype. This is why it can not

fulfill the Requirements $R_{1.3}$, $R_{1.1}$, $R_{2.3}$, $R_{2.4}$, and $R_{3.5}$.

The Floka prototype fulfills all design and all camera based requirements. This comes at the cost of a less pronounced range of motion as the tendon prototype but clearly outperforms the initial Flobi design in all aspects. The maximum velocities and accelerations of the tilt actuator are a bit lower but can be potentially optimized with a different lever length and thus translation ratio. The only debatable point of this prototype is the extend of the full range of motion coverage of 69.2% of the human oculomotor range. Visual inspection of the coverage graph in Figure 12.1 gives the impression that the covered range should be sufficient. Most of the human saccades happen in the so-called practical field of view, which is covered to 100% by this prototype. The restrictions when it comes to the full range of motion might be acceptable considering the drastic improvement in terms of camera resolution and its much wider field of view.

# IV

# Discussion

The final part of this thesis will recapitulate the contributions of the individual chapters, analyze the scientific impact, and will give prospective points for future work.

Chapter 13

# Conclusion

With this thesis, I have presented a set of interlocking building blocks that are necessary for the design, construction, control, and the evaluation of a human-inspired robotic eye for humanoid robotic heads.

Based on insights gained by studying the human eye, including its fundamental functionality and its performance numbers, it became quite apparent that none of the eye mechanics found in literature were able to compete with all the requirements, that were derived from the human model, at once. Similar to a chain, whose strength is only as good as its weakest link, all components of this technical system have to work hand in hand and it is mandatory to focus on and to comply to a defined set of requirements based on the human model right from the start. Therefore, in contrast to previous systems, my approach for the replication of human eye motion considered this problem as a whole, from the mechanical through electronic to software design and construction in order to achieve the best possible overall performance.

In a first step, a list of requirements was narrowed down to a set of technical and human-inspired specifications with regard to the outer design, the camera capabilities, and the motion performance. The analysis of related humanoid robotic heads revealed that their associated publications rarely publish exact numbers in terms of performance and that none of those robotic heads complied to the full set of the defined requirements. In addition, further literature research demonstrated the lack of comparative studies and the lack of publications of robot independent gaze control frameworks for humanoid robots. In order to change that, this work used the previously defined set of requirements to document the design of a human-inspired robotic eye and an associated, robot independent, toolchain for

the human-inspired model-driven control of the eye region of a humanoid robot.

The feasibility of a mechanical replica of the human eye region was shown by the design, construction, and evaluation of two prototypes under the premise of the previously set goals. The first system, mimicking the human eye actuation principle by means of artificial tendons, possesses human-like motion in terms of range, velocity, and accelerations. The second one, using a more classical four bar linkage actuation principle, adds agile expressive eyelids and a novel camera system with a human-like field of view and acuity. As it turned out, the specifications resulting from the key requirements can be contradictory and it required careful design decisions and trade-offs for the simultaneous fulfillment of all key requirements. For example the enhancement in terms of camera capabilities of the second prototype came at the cost of a somewhat reduced oculomotor range.

With hardware that is capable of human-like motion in place, first ideas for the motion control framework were verified by means of a novel bi-directional motion capture system that allows an operator to control a humanoid robotic head in an intuitive way. This system made it possible to observe and to analyze the uninfluenced interplay of different eye motion patterns that are described in literature, by placing the operator and the robot in a direct feedback loop. The insights gained during these experiments paved the way for the subsequent design and implementation of a human-inspired motion control framework.

In addition, the approach of viewing the system as a whole and subsequently optimizing all chain elements, also led to the development of custom motion control electronics. This hardware can not only drive three motors with a similar space requirement as commercial single device controllers, but does allow custom communication and motion control routines to be implemented. The developed low-level motion generation firmware allows state-of-the-art, velocity based, smooth actuation of the individual joints.

This hardware built the basis for a host control framework, which in turn uses the model-based actuation control framework that drives the whole eye region based on a human-inspired model. In contrast to other frameworks, this work does not rely on a specific robotic platform, but was developed with portability in mind. Portability achieved by the use of loose coupling based on a standardized communication layer (ROS) and an in-

tegrated scaling and limiting feature that allows intuitive adjustments of the internal control model. In particular, this configurability allows the deployment even on less agile humanoid robotic heads where it will produce slower motion, but the human characteristics and the overall composition of the interlocked motion patterns will be preserved. This flexibility was proven by the exemplary control of the two commercially available robotic heads that are used the iCub and the Meka robot. In particular this flexibility will foster comparative studies of different humanoid robotic heads by others in the future.

The quality of the generated motion has been analyzed by means of a user study. It showed that the motion generated by the fully enabled human motion model does, as postulated, boost the users' ability to identify the activity executed by the robot. Unfortunately the results did not show a clear favorite between the tested conditions in terms of appreciation of the different motion patterns when tested between subjects. However, when tested in direct comparison, the robot that was using all features of the proposed model, was chosen to be the preferred interaction partner.

In addition to the user study, the performance of the mechanical prototypes was verified by means of an extensive evaluation based on detailed measurements of the final hardware. As stated before, the literature on humanoid robotic heads often only give design targets of the performance to be expected, if at all. In contrast to that, this work provided a throughout evaluation including the measurements of all relevant performance values based on the manufactured prototype. This allowed a direct and objective comparison between measured values of a previous design, the two prototypes, and the human. These results showed that the two prototypes are indeed an improvement to the previous design and that both can compete with the human standards. The mostly 3D printed tendon-based prototype was meant as a technology evaluation. In the sake of simplicity, this first prototype did not implement eyelids and used the same low-resolution camera as previous designs. It showed the highest performance numbers in terms of velocities and accelerations but fell short on the design and camera requirements. Whereas the lever-based prototype fulfilled all requirements in terms of the design and the camera capabilities, but was not able to replicate the full range of human motion down to the last extend. Based on the practical experience gained during the operation of the previous prototypes and given the fact that the most used practical field of view is completely

covered, this limitation is neglectable. Considering its unique camera capabilities and that it outperforms other eye mechanisms found in literature, this prototype has been chosen and was successfully included in the new humanoid robotic head Floka.

With the prototypes and tools developed in this thesis I have shown that it is feasible to design and control a human-inspired robotic eye that features human-like performance and can replicate human movement patterns. Viewing the system as a whole and tackling the optimization right from the start has proved effective in the practice and resulted in a very capable overall system. As pointed out in this thesis, all robotic heads that feature movable eyes will benefit from the combined, model-driven actuation of all actuators. To the present day, I am not aware of any studies comparing different humanoid robotic head designs that are all actuated by the same human-inspired control model. In order to change that, the model-based, human-like actuation toolchain has been made publicly available under an opensource license. I hope that this work will contribute to the way future humanoid robotic systems will be designed, controlled and finally evaluated.

Chapter 14

# Outlook

As for all technical solutions, there is always room for improvement and the work presented in this thesis is no exception. The first prototype, actuated by artificial tendons, showcased the benefits of 3D printing and provided a convincing performance in terms of motor capabilities. Its actuation scheme showed a huge potential and it is definitely conceivable that this system could be revised and a more recent camera module and actuated eyelids can be added. Even though this direction has a huge potential, it would require further research in order to deal with the introduced unconstrained rotation issue. This could be accomplished in a number of ways, e.g. through a reduction by design or algorithmically using a defined or trained model. The second mechanical prototype has proven its matureness but could be still optimized further to cover the last remainder of the full human range of motion if proven necessary. In addition, possible future camera developments in the fast-growing smartphone sector could allow the integration of smaller sensors that feature a similar quality in terms light sensitivity, resolution, and framerate.

The quality of the motion, generated by the humotion toolchain, and the results of the user study should be further assessed. Even though the condition that used the fully enabled software stack turned out to be the preferred interaction partner, a more clear outcome was expected. Therefore, future research should be conducted in more on a broader scale or using a more realistic settings, featuring a bi-directional interaction part. In addition, future studies should aim to replicate the results on different robotic heads in order e.g. assess the differences in the perception of those robots.

Even though the presented bi-directional motion capture system is still an useful tool, it should be upgraded to use state-of-the-art technologies.

The recent advances in the sensor technology and the marker-less motion tracking systems such as OpenFace [134] should be integrated in order to upgrade the system. This way, the tedious and time-consuming process of attaching the markers to the actors face would become unnecessary and would make wearing the helmet more pleasant.

Concluding it can be said the methodical approach of optimizing each part of the full chain from the very beginning worked out quite well. Both presented prototypes outperform other systems found in literature and the model control model allows an intuitive and realistic animation of all related joints of a humanoid robotic head at once. On a broader scale, it will be interesting to see how the presented prototypes and control framework will find its application in the field of humanoid robotic heads.

# V

# Appendices

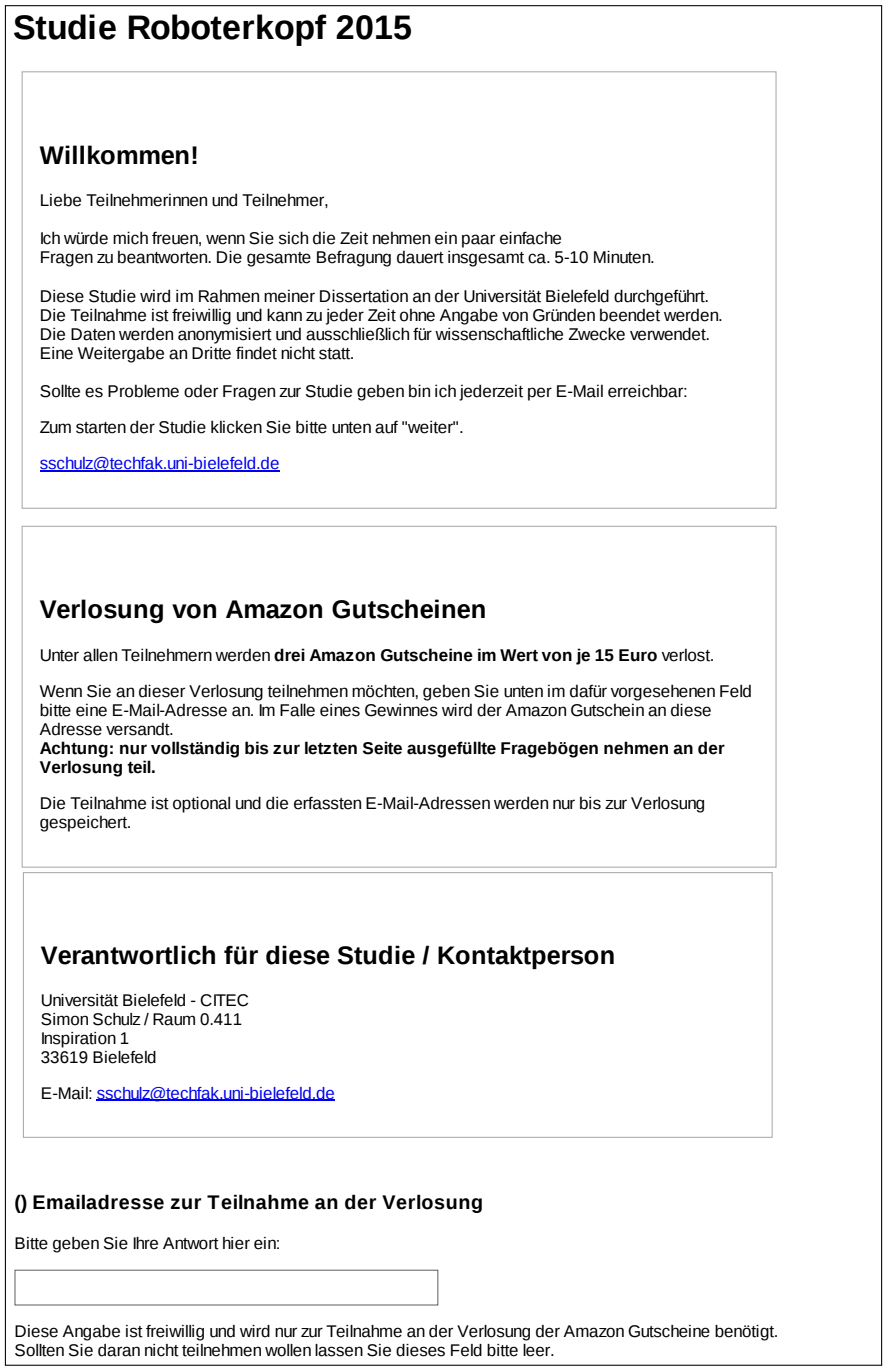
This part contains supplementary materials and the bibliography.

# Supplementary Userstudy Material

The first section of this appendix chapter shows the original Questionnaire used by the Evaluation study that was conducted as described in Chapter 11.

## Questionnaire

The following screenshots show the single pages of the questionnaire as they were presented to the participants.



# Studie Roboterkopf 2015

## Willkommen!

Liebe Teilnehmerinnen und Teilnehmer,

Ich würde mich freuen, wenn Sie sich die Zeit nehmen ein paar einfache
Fragen zu beantworten. Die gesamte Befragung dauert insgesamt ca. 5-10 Minuten.

Diese Studie wird im Rahmen meiner Dissertation an der Universität Bielefeld durchgeführt.
Die Teilnahme ist freiwillig und kann zu jeder Zeit ohne Angabe von Gründen beendet werden.
Die Daten werden anonymisiert und ausschließlich für wissenschaftliche Zwecke verwendet.
Eine Weitergabe an Dritte findet nicht statt.

Sollte es Probleme oder Fragen zur Studie geben bin ich jederzeit per E-Mail erreichbar:

Zum starten der Studie klicken Sie bitte unten auf "weiter".

sschulz@techfak.uni-bielefeld.de

## Verlosung von Amazon Gutscheinen

Unter allen Teilnehmern werden **drei Amazon Gutscheine im Wert von je 15 Euro** verlost.

Wenn Sie an dieser Verlosung teilnehmen möchten, geben Sie unten im dafür vorgesehenen Feld bitte eine E-Mail-Adresse an. Im Falle eines Gewinnes wird der Amazon Gutschein an diese Adresse versandt.
**Achtung: nur vollständig bis zur letzten Seite ausgefüllte Fragebögen nehmen an der Verlosung teil.**

Die Teilnahme ist optional und die erfassten E-Mail-Adressen werden nur bis zur Verlosung gespeichert.

## Verantwortlich für diese Studie / Kontaktperson

Universität Bielefeld - CITEC
Simon Schulz / Raum 0.411
Inspiration 1
33619 Bielefeld

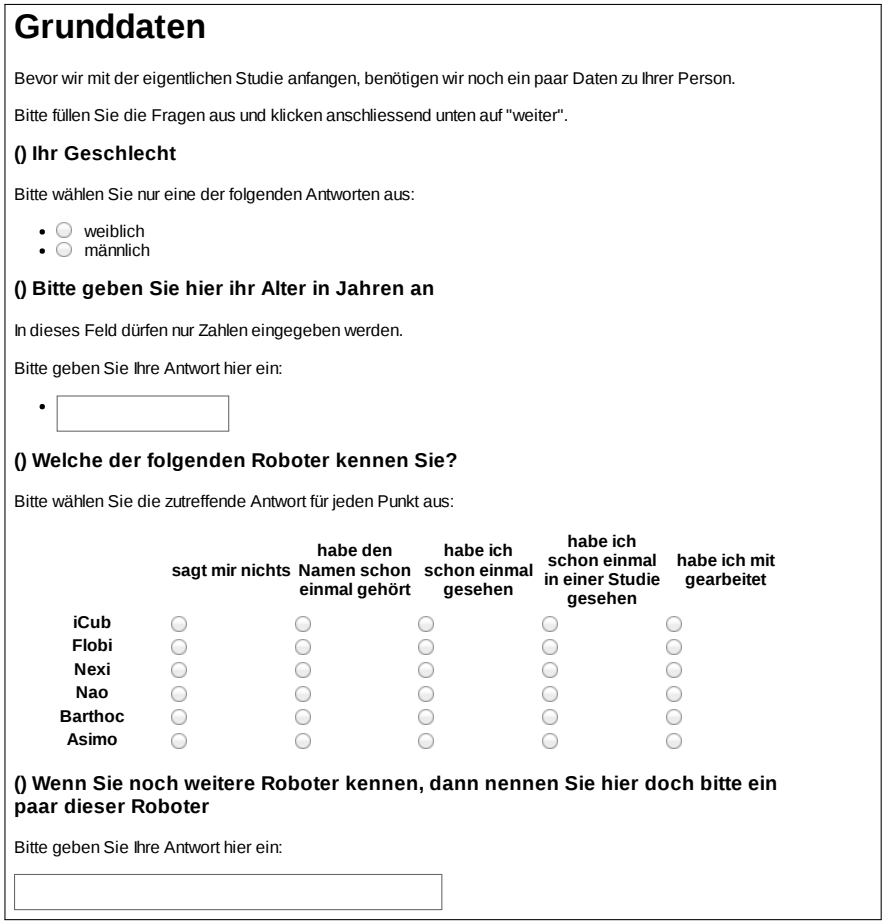
E-Mail: sschulz@techfak.uni-bielefeld.de

**() Emailadresse zur Teilnahme an der Verlosung**

Bitte geben Sie Ihre Antwort hier ein:

Diese Angabe ist freiwillig und wird nur zur Teilnahme an der Verlosung der Amazon Gutscheine benötigt.
Sollten Sie daran nicht teilnehmen wollen lassen Sie dieses Feld bitte leer.

Figure 1.: Question Block #1



## Grunddaten

Bevor wir mit der eigentlichen Studie anfangen, benötigen wir noch ein paar Daten zu Ihrer Person.

Bitte füllen Sie die Fragen aus und klicken anschliessend unten auf "weiter".

**() Ihr Geschlecht**

Bitte wählen Sie nur eine der folgenden Antworten aus:

- ○ weiblich
- ○ männlich

**() Bitte geben Sie hier ihr Alter in Jahren an**

In dieses Feld dürfen nur Zahlen eingegeben werden.

Bitte geben Sie Ihre Antwort hier ein:

-

**() Welche der folgenden Roboter kennen Sie?**

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

| | sagt mir nichts | habe den Namen schon einmal gehört | habe ich schon einmal gesehen | habe ich schon einmal in einer Studie gesehen | habe ich mit gearbeitet |
|---|---|---|---|---|---|
| **iCub** | ○ | ○ | ○ | ○ | ○ |
| **Flobi** | ○ | ○ | ○ | ○ | ○ |
| **Nexi** | ○ | ○ | ○ | ○ | ○ |
| **Nao** | ○ | ○ | ○ | ○ | ○ |
| **Barthoc** | ○ | ○ | ○ | ○ | ○ |
| **Asimo** | ○ | ○ | ○ | ○ | ○ |

**() Wenn Sie noch weitere Roboter kennen, dann nennen Sie hier doch bitte ein paar dieser Roboter**

Bitte geben Sie Ihre Antwort hier ein:

Figure 2.: Question Block #2

## Video1

### Anleitung

Auf den nachfolgenden Seiten werden nun mehrere Videos gezeigt und einzelne Fragen dazu gestellt.

Drücken Sie bitte auf den abspielen Button (in der Mitte des Videos) und schauen Sie sich das Video an. Das Video kann mehrfach angesehen werden, dazu einfach nach dem Ende des Videos erneut auf den abspielen Button klicken. Das gezeigte **Video ist ohne Tonspur**. Nachdem Sie das Video gesehen haben beantworten Sie bitte die unten stehenden Fragen.

Bei der Beantwortung der Fragen gibt es kein richtig oder falsch, wichtig ist vielmehr den eigenen Eindruck zu beschreiben.

Bitte klicken Sie nach dem ausfüllen der Fragen unten auf "weiter".

Sollten die Videos nicht angezeigt werden oder sonst irgendwelche Fehler auftreten wäre ich sehr dankbar für einen Hinweis per E-Mail.


Bitte bewerten Sie nun die nachfolgenden Aussagen auf einer Skala von 1 bis 10 (1=trifft überhaupt nicht zu, ..., 10=trifft voll und ganz zu)

*** () Die im Video zu sehenden Bewegungen...**

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **sind sehr flüssig** | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| **passen gut zu diesem Roboterkopf** | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| **wirken auf mich lebendig** | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| **sind menschenähnlich** | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

1=trifft überhaupt nicht zu, ... , 10=trifft voll und ganz zu

Figure 3.: Question Block #3

## Video2

### Anleitung

Schauen Sie sich bitte zuerst das Video an. Dazu drücken Sie bitte auf den abspielen Button (in der Mitte des Videos) und schauen Sie sich das Video an. Das Video kann mehrfach angesehen werden, dazu einfach nach dem Ende des Videos erneut auf den abspielen Button klicken. Das gezeigte **Video ist ohne Tonspur**.
Nachdem Sie das Video gesehen haben beantworten Sie bitte die unten stehenden Fragen und klicken anschliessend auf "weiter".


**() Bitte beschreiben Sie kurz und knapp was der Roboter in dem Video gerade tut.**

Bitte geben Sie Ihre Antwort hier ein:

Wenn Sie nicht sicher sind raten Sie bitte

*** () Die im Video zu sehenden Bewegungen...**

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **sind sehr flüssig** | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| **passen gut zu diesem Roboterkopf** | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| **wirken auf mich lebendig** | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| **sind menschenähnlich** | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

Bitte bewerten Sie nun die Aussagen auf einer Skala von 1 bis 10 (1=trifft überhaupt nicht zu, ..., 10=trifft voll und ganz zu)

Figure 4.: Question Block #4

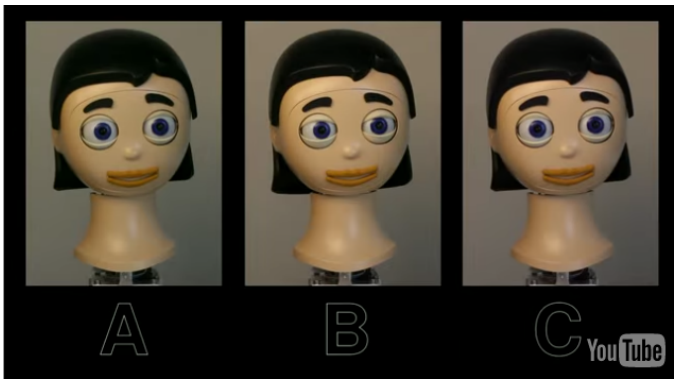
## Video3

### Anleitung

Sehen Sie sich bitte zuerst das Video an. Dazu drücken Sie bitte auf den abspielen Button (in der Mitte des Videos) und schauen Sie sich das Video an. Dieses Video läuft in einer Dauerschleife, d. h. Sie werden dieselbe Sequenz immer und immer wieder sehen. Das gezeigte **Video ist ohne Tonspur**.

Schauen Sie sich bitte ein paar Sequenzen an und beantworten dann die unten stehenden Fragen. Sie können während der Beantwortung der Fragen zwischendurch immer wieder das Video ansehen.

Nachdem Sie alle Fragen beantwortet haben klicken Sie bitte unten auf "weiter".




*** () Im direkten Vergleich...**

Bitte wählen Sie nur eine der folgenden Antworten aus:

- ○ wirkt Roboter A (Links) lebendiger als Roboter B (Mitte)
- ○ wirkt Roboter B (Mitte) lebendiger als Roboter A (Links)

*** () Im direkten Vergleich...**

Bitte wählen Sie nur eine der folgenden Antworten aus:

- ○ wirkt Roboter C (Rechts) lebendiger als Roboter B (Mitte)
- ○ wirkt Roboter B (Mitte) lebendiger als Roboter C (Rechts)

*** () Im direkten Vergleich...**

Bitte wählen Sie nur eine der folgenden Antworten aus:

- ○ wirkt Roboter A (Links) lebendiger als Roboter C (Rechts)
- ○ wirkt Roboter C (Rechts) lebendiger als Roboter A (Links)

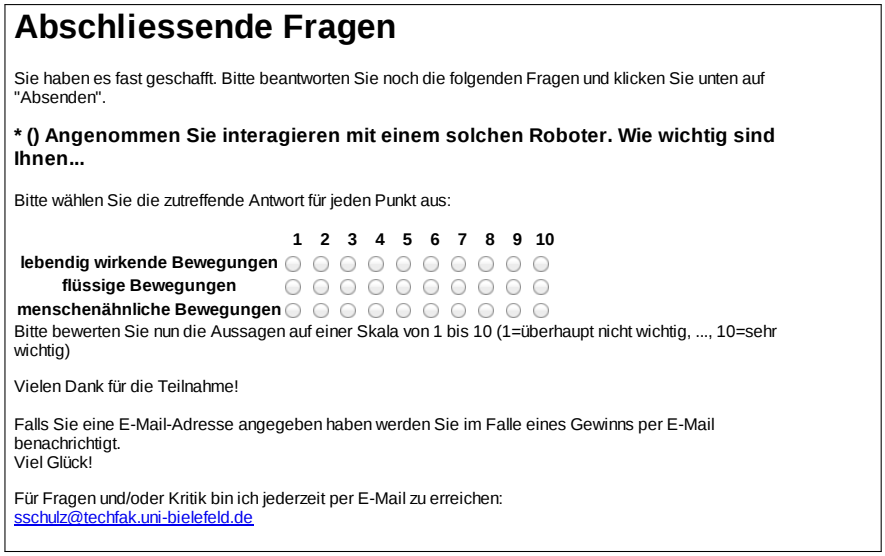
*** () Mit welchem der drei gezeigten Roboter würden Sie am liebsten interagieren?**

Bitte wählen Sie nur eine der folgenden Antworten aus:

- ○ Roboter A (links)
- ○ Roboter B (mitte)
- ○ Roboter C (rechts)



Figure 5.: Question Block #5



## Abschliessende Fragen

Sie haben es fast geschafft. Bitte beantworten Sie noch die folgenden Fragen und klicken Sie unten auf "Absenden".

*** () Angenommen Sie interagieren mit einem solchen Roboter. Wie wichtig sind Ihnen...**

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **lebendig wirkende Bewegungen** | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| **flüssige Bewegungen** | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| **menschenähnliche Bewegungen** | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

Bitte bewerten Sie nun die Aussagen auf einer Skala von 1 bis 10 (1=überhaupt nicht wichtig, ..., 10=sehr wichtig)

Vielen Dank für die Teilnahme!

Falls Sie eine E-Mail-Adresse angegeben haben werden Sie im Falle eines Gewinns per E-Mail benachrichtigt.
Viel Glück!

Für Fragen und/oder Kritik bin ich jederzeit per E-Mail zu erreichen:
sschulz@techfak.uni-bielefeld.de

Figure 6.: Question Block #6

## Result Dataset

The following tables show the unprocessed answers as given by all participants during the user study. Please refer to Section 11.2 for a detailed evaluation and interpretation of the results. Please note that the ID numbers are not continuous due to technical restrictions of the survey software. Every request to the survey website generated a new ID but only completed surveys were stored in the server database.

Table 1.: Unprocessed survey results

| ID | Gender | Age | Familarity iCub | Familarity Flobi | Familarity Nexi | Familarity Nao | Familarity Barthoc | Familarity Asimo | Video 1 | Video1 $S_{FLUID}$ | Video1 $S_{MATCH}$ | Video1 $S_{LIVELY}$ | Video1 $S_{HUMAN}$ | Video 2 | Video2 $S_{FLUID}$ | Video2 $S_{MATCH}$ | Video2 $S_{LIVELY}$ | Video2 $S_{HUMAN}$ | Compare A/B | Compare B/C | Compare C/A | Favourite | $P_{LIVELY}$ | $P_{FLUID}$ | $P_{HUMAN}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | M | 33 | 3 | 4 | 2 | 4 | 2 | 2 | A | 1 | 1 | 1 | 1 | A | 1 | 1 | 1 | 1 | B | B | C | B | 1 | 1 | 1 |
| 6 | F | 27 | 2 | 3 | 0 | 2 | 0 | 2 | C | 10 | 8 | 10 | 9 | A | 10 | 8 | 10 | 9 | B | B | C | B | 9 | 10 | 9 |
| 12 | M | 26 | 2 | 3 | 0 | 2 | 2 | 2 | C | 7 | 7 | 8 | 8 | A | 6 | 7 | 7 | 7 | B | B | A | B | 6 | 8 | 7 |
| 27 | F | 36 | 1 | 0 | 0 | 0 | 0 | 0 | C | 4 | 6 | 2 | 1 | A | 4 | 6 | 2 | 1 | B | B | C | B | 5 | 8 | 1 |
| 36 | F | 29 | 3 | 4 | 1 | 4 | 2 | 3 | A | 8 | 8 | 7 | 7 | C | 8 | 8 | 7 | 7 | B | B | A | B | 5 | 10 | 4 |
| 37 | M | 54 | 2 | 3 | 0 | 3 | 3 | 3 | A | 8 | 10 | 9 | 8 | A | 8 | 10 | 9 | 9 | B | B | C | B | 10 | 8 | 6 |
| 41 | F | 28 | 2 | 4 | 0 | 3 | 2 | 2 | B | 10 | 10 | 10 | 9 | C | 10 | 10 | 10 | 9 | B | B | A | B | 9 | 10 | 7 |
| 48 | M | 27 | 1 | 2 | 0 | 2 | 2 | 3 | B | 8 | 6 | 7 | 6 | A | 8 | 7 | 8 | 7 | B | B | A | B | 6 | 8 | 4 |
| 52 | M | 30 | 3 | 3 | 0 | 3 | 2 | 3 | C | 10 | 10 | 10 | 10 | B | 10 | 10 | 10 | 10 | B | B | C | B | 10 | 10 | 6 |
| 57 | M | 23 | 2 | 0 | 0 | 0 | 0 | 0 | D | 7 | 7 | 3 | 3 | A | 7 | 7 | 4 | 3 | B | B | A | B | 8 | 8 | 7 |
| 59 | M | 28 | 2 | 4 | 0 | 3 | 2 | 2 | B | 10 | 10 | 10 | 10 | B | 10 | 10 | 10 | 10 | B | B | A | B | 1 | 1 | 1 |
| 61 | F | 27 | 1 | 1 | 0 | 0 | 0 | 0 | B | 8 | 7 | 7 | 7 | B | 8 | 7 | 6 | 6 | B | B | A | B | 6 | 7 | 6 |
| 64 | F | 52 | 0 | 2 | 0 | 0 | 0 | 1 | C | 6 | 8 | 8 | 9 | B | 6 | 7 | 7 | 9 | B | B | A | B | 9 | 9 | 9 |
| 65 | F | 56 | 0 | 1 | 0 | 0 | 0 | 0 | A | 7 | 7 | 7 | 7 | A | 7 | 7 | 7 | 7 | B | B | A | B | 10 | 10 | 10 |
| 75 | F | 26 | 0 | 0 | 0 | 0 | 0 | 0 | B | 6 | 8 | 5 | 5 | A | 7 | 8 | 5 | 5 | B | B | C | B | 4 | 4 | 4 |
| 76 | F | 26 | 0 | 1 | 0 | 0 | 0 | 0 | C | 2 | 2 | 2 | 4 | C | 2 | 2 | 2 | 3 | B | B | C | B | 4 | 4 | 4 |
| 83 | M | 35 | 0 | 0 | 0 | 0 | 0 | 0 | A | 8 | 8 | 6 | 5 | A | 6 | 2 | 2 | 2 | B | B | A | B | 10 | 10 | 10 |
| 85 | F | 22 | 0 | 1 | 0 | 0 | 0 | 1 | D | 4 | 6 | 3 | 3 | A | 4 | 6 | 3 | 2 | B | B | C | B | 10 | 8 | 9 |
| 86 | F | 30 | 2 | 3 | 0 | 2 | 3 | 2 | D | 4 | 3 | 5 | 4 | C | 4 | 3 | 5 | 5 | B | B | C | B | 3 | 5 | 4 |

To be continued on the following page...

. . . continued

| ID | Gender | Age | Familarity iCub | Familarity Flobi | Familarity Nexi | Familarity Nao | Familarity Barthoc | Familarity Asimo | Video 1 | Video1 $S_{FLUID}$ | Video1 $S_{MATCH}$ | Video1 $S_{LIVELY}$ | Video1 $S_{HUMAN}$ | Video 2 | Video2 $S_{FLUID}$ | Video2 $S_{MATCH}$ | Video2 $S_{LIVELY}$ | Video2 $S_{HUMAN}$ | Compare A/B | Compare B/C | Compare C/A | Favourite | $P_{LIVELY}$ | $P_{FLUID}$ | $P_{HUMAN}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 87 | M | 32 | 4 | 2 | 0 | 4 | 2 | 3 | C | 6 | 6 | 4 | 7 | A | 7 | 8 | 7 | 8 | B | B | C | B | 9 | 7 | 7 |
| 88 | F | 22 | 0 | 0 | 0 | 0 | 0 | 0 | D | 8 | 7 | 6 | 7 | C | 7 | 6 | 5 | 6 | B | B | A | B | 7 | 6 | 4 |
| 89 | M | 60 | 0 | 2 | 0 | 2 | 0 | 0 | C | 8 | 8 | 7 | 7 | B | 8 | 8 | 7 | 6 | B | B | C | B | 7 | 8 | 6 |
| 92 | M | 23 | 0 | 0 | 0 | 0 | 0 | 2 | B | 4 | 3 | 2 | 2 | C | 4 | 4 | 2 | 2 | B | B | C | B | 6 | 9 | 6 |
| 93 | F | 30 | 1 | 2 | 0 | 2 | 0 | 2 | C | 3 | 6 | 4 | 4 | A | 4 | 3 | 3 | 3 | A | B | A | A | 4 | 8 | 4 |
| 94 | F | 20 | 2 | 2 | 0 | 4 | 0 | 0 | D | 9 | 9 | 7 | 6 | B | 9 | 9 | 8 | 8 | B | B | C | B | 7 | 9 | 7 |
| 96 | M | 22 | 0 | 0 | 0 | 0 | 0 | 0 | B | 9 | 10 | 9 | 9 | B | 9 | 10 | 9 | 9 | B | B | A | B | 10 | 9 | 9 |
| 98 | F | 28 | 1 | 1 | 0 | 0 | 1 | 2 | C | 7 | 8 | 7 | 9 | A | 7 | 8 | 7 | 9 | B | B | A | B | 10 | 9 | 8 |
| 106 | F | 22 | 0 | 0 | 0 | 0 | 0 | 0 | B | 9 | 7 | 3 | 3 | C | 6 | 3 | 3 | 3 | B | B | A | B | 8 | 10 | 8 |
| 107 | F | 27 | 2 | 4 | 2 | 3 | 2 | 2 | A | 8 | 9 | 9 | 9 | A | 8 | 8 | 8 | 8 | B | B | C | B | 9 | 9 | 7 |
| 110 | M | 19 | 0 | 1 | 0 | 0 | 0 | 2 | C | 3 | 9 | 6 | 8 | C | 5 | 7 | 7 | 7 | B | B | A | B | 7 | 9 | 6 |
| 111 | M | 22 | 1 | 1 | 0 | 0 | 0 | 0 | D | 7 | 5 | 4 | 4 | C | 7 | 4 | 4 | 4 | B | B | C | B | 5 | 8 | 4 |
| 113 | F | 19 | 0 | 0 | 0 | 0 | 0 | 1 | D | 8 | 8 | 7 | 8 | A | 6 | 7 | 7 | 9 | B | B | C | B | 8 | 6 | 7 |
| 114 | F | 19 | 3 | 0 | 2 | 2 | 3 | 0 | D | 5 | 8 | 3 | 3 | C | 6 | 6 | 4 | 4 | B | B | A | B | 7 | 8 | 6 |
| 122 | F | 21 | 0 | 0 | 0 | 0 | 0 | 0 | D | 9 | 9 | 7 | 8 | B | 9 | 9 | 7 | 8 | B | B | C | B | 9 | 10 | 9 |
| 125 | F | 27 | 0 | 0 | 0 | 3 | 0 | 0 | A | 2 | 2 | 2 | 3 | C | 2 | 2 | 2 | 3 | B | B | C | B | 8 | 8 | 6 |
| 127 | F | 52 | 0 | 3 | 0 | 3 | 0 | 2 | B | 2 | 2 | 2 | 3 | A | 2 | 2 | 2 | 3 | B | C | C | B | 8 | 8 | 8 |
| 128 | M | 27 | 3 | 2 | 1 | 4 | 2 | 3 | C | 7 | 9 | 7 | 8 | A | 7 | 10 | 7 | 8 | B | C | C | B | 7 | 10 | 6 |
| 133 | F | 34 | 1 | 0 | 0 | 0 | 1 | 0 | B | 8 | 8 | 7 | 7 | B | 8 | 8 | 8 | 8 | B | B | A | B | 8 | 9 | 9 |
| 134 | F | 28 | 1 | 1 | 0 | 0 | 0 | 0 | D | 2 | 2 | 2 | 3 | C | 2 | 2 | 2 | 3 | B | C | C | A | 10 | 10 | 10 |

To be continued on the following page...

. . . continued

| ID | Gender | Age | Familarity iCub | Familarity Flobi | Familarity Nexi | Familarity Nao | Familarity Barthoc | Familarity Asimo | Video 1 | Video1 $S_{FLUID}$ | Video1 $S_{MATCH}$ | Video1 $S_{LIVELY}$ | Video1 $S_{HUMAN}$ | Video 2 | Video2 $S_{FLUID}$ | Video2 $S_{MATCH}$ | Video2 $S_{LIVELY}$ | Video2 $S_{HUMAN}$ | Compare A/B | Compare B/C | Compare C/A | Favourite | $P_{LIVELY}$ | $P_{FLUID}$ | $P_{HUMAN}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 136 | F | 18 | 0 | 0 | 1 | 0 | 1 | 0 | A | 6 | 9 | 5 | 5 | A | 7 | 9 | 6 | 7 | B | B | A | B | 5 | 9 | 6 |
| 138 | M | 31 | 0 | 3 | 2 | 0 | 0 | 0 | D | 6 | 7 | 4 | 3 | B | 6 | 5 | 4 | 4 | B | B | A | B | 7 | 6 | 8 |
| 139 | M | 26 | 2 | 2 | 0 | 2 | 0 | 2 | A | 3 | 5 | 5 | 6 | C | 4 | 5 | 5 | 5 | B | B | C | B | 6 | 8 | 3 |
| 140 | M | 27 | 2 | 2 | 0 | 2 | 0 | 2 | A | 9 | 8 | 9 | 9 | A | 6 | 8 | 7 | 8 | B | B | A | B | 10 | 9 | 10 |
| 164 | M | 55 | 0 | 3 | 0 | 0 | 0 | 2 | B | 7 | 7 | 5 | 5 | C | 7 | 7 | 5 | 5 | B | B | A | B | 8 | 8 | 8 |
| 179 | M | 32 | 2 | 2 | 0 | 2 | 2 | 2 | C | 5 | 6 | 4 | 5 | C | 4 | 5 | 4 | 5 | B | B | A | B | 8 | 7 | 9 |
| 182 | M | 32 | 4 | 4 | 0 | 2 | 2 | 2 | B | 7 | 8 | 8 | 6 | C | 7 | 8 | 7 | 7 | B | B | C | B | 9 | 7 | 7 |
| 183 | F | 23 | 0 | 0 | 0 | 0 | 0 | 1 | B | 9 | 8 | 9 | 8 | B | 8 | 9 | 8 | 8 | B | B | C | B | 8 | 8 | 7 |
| 184 | M | 27 | 0 | 0 | 0 | 2 | 0 | 2 | A | 3 | 4 | 2 | 1 | C | 2 | 4 | 1 | 1 | B | B | A | B | 1 | 10 | 3 |
| 192 | F | 19 | 1 | 0 | 1 | 3 | 0 | 1 | C | 5 | 7 | 6 | 8 | B | 5 | 7 | 7 | 7 | B | B | A | A | 9 | 6 | 5 |
| 194 | F | 25 | 0 | 0 | 0 | 0 | 0 | 0 | A | 3 | 3 | 2 | 2 | C | 3 | 3 | 3 | 3 | B | B | A | B | 8 | 8 | 7 |
| 197 | F | 22 | 0 | 1 | 0 | 0 | 0 | 1 | D | 9 | 5 | 5 | 8 | C | 9 | 5 | 5 | 7 | B | C | C | B | 7 | 9 | 10 |
| 203 | M | 38 | 0 | 2 | 0 | 0 | 0 | 2 | C | 5 | 7 | 3 | 6 | C | 6 | 5 | 3 | 6 | B | B | C | B | 7 | 5 | 8 |
| 206 | M | 36 | 0 | 0 | 0 | 2 | 0 | 2 | C | 3 | 4 | 1 | 2 | C | 4 | 4 | 3 | 3 | B | B | C | B | 6 | 8 | 8 |
| 209 | M | 32 | 1 | 3 | 0 | 2 | 0 | 2 | B | 6 | 4 | 2 | 6 | B | 9 | 6 | 5 | 9 | B | B | A | A | 8 | 10 | 2 |
| 212 | F | 23 | 0 | 0 | 0 | 0 | 0 | 0 | B | 7 | 8 | 10 | 9 | B | 7 | 8 | 10 | 9 | B | B | A | B | 7 | 6 | 8 |
| 225 | F | 23 | 0 | 2 | 0 | 4 | 0 | 1 | D | 5 | 9 | 4 | 7 | A | 5 | 9 | 4 | 6 | B | B | A | B | 10 | 8 | 9 |
| 236 | M | 32 | 2 | 4 | 0 | 0 | 2 | 3 | B | 7 | 7 | 6 | 7 | A | 7 | 7 | 6 | 7 | B | B | C | B | 6 | 8 | 10 |
| 237 | M | 25 | 2 | 2 | 0 | 2 | 0 | 2 | A | 3 | 5 | 7 | 6 | A | 6 | 6 | 8 | 8 | B | B | A | B | 3 | 5 | 2 |
| 243 | M | 36 | 3 | 3 | 0 | 2 | 2 | 2 | D | 6 | 7 | 5 | 5 | A | 8 | 8 | 7 | 6 | B | B | A | B | 9 | 9 | 10 |

To be continued on the following page...

. . . continued

| ID | Gender | Age | Familarity iCub | Familarity Flobi | Familarity Nexi | Familarity Nao | Familarity Barthoc | Familarity Asimo | Video 1 | Video1 $S_{FLUID}$ | Video1 $S_{MATCH}$ | Video1 $S_{LIVELY}$ | Video1 $S_{HUMAN}$ | Video 2 | Video2 $S_{FLUID}$ | Video2 $S_{MATCH}$ | Video2 $S_{LIVELY}$ | Video2 $S_{HUMAN}$ | Compare A/B | Compare B/C | Compare C/A | Favourite | $P_{LIVELY}$ | $P_{FLUID}$ | $P_{HUMAN}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 251 | F | 31 | 1 | 2 | 1 | 0 | 0 | 0 | C | 5 | 7 | 7 | 9 | C | 4 | 7 | 8 | 8 | B | B | C | B | 10 | 8 | 10 |
| 252 | M | 33 | 2 | 1 | 1 | 0 | 0 | 2 | D | 9 | 8 | 7 | 8 | A | 9 | 8 | 7 | 8 | B | B | C | B | 8 | 7 | 6 |
| 287 | F | 43 | 0 | 0 | 0 | 0 | 0 | 0 | C | 3 | 10 | 3 | 4 | B | 3 | 9 | 3 | 4 | B | B | C | B | 8 | 7 | 9 |
| 292 | M | 34 | 2 | 4 | 0 | 2 | 1 | 2 | D | 7 | 7 | 4 | 7 | B | 7 | 7 | 4 | 6 | B | B | A | B | 8 | 9 | 6 |
| 298 | F | 33 | 0 | 1 | 0 | 0 | 0 | 0 | C | 7 | 7 | 8 | 8 | B | 8 | 8 | 8 | 8 | B | C | C | B | 6 | 8 | 5 |
| 299 | M | 53 | 0 | 0 | 0 | 0 | 0 | 1 | A | 3 | 3 | 2 | 2 | C | 3 | 3 | 2 | 2 | B | B | A | B | 4 | 4 | 5 |
| 300 | M | 34 | 0 | 0 | 0 | 0 | 0 | 0 | D | 9 | 5 | 6 | 5 | B | 8 | 5 | 6 | 6 | B | B | A | B | 10 | 10 | 10 |
| 303 | F | 22 | 0 | 0 | 0 | 0 | 0 | 0 | D | 7 | 8 | 6 | 8 | C | 5 | 7 | 6 | 8 | B | B | A | B | 8 | 8 | 8 |
| 313 | F | 36 | 0 | 0 | 1 | 0 | 0 | 0 | B | 3 | 3 | 2 | 2 | C | 3 | 3 | 3 | 3 | B | B | A | B | 7 | 8 | 7 |
| 318 | F | 33 | 0 | 0 | 0 | 0 | 0 | 0 | A | 7 | 8 | 7 | 7 | C | 7 | 8 | 7 | 8 | B | B | C | B | 9 | 9 | 9 |
| 320 | F | 31 | 2 | 4 | 0 | 2 | 1 | 2 | A | 9 | 10 | 6 | 8 | A | 9 | 10 | 8 | 8 | B | B | C | B | 8 | 9 | 5 |
| 322 | M | 40 | 0 | 0 | 0 | 0 | 0 | 0 | B | 4 | 4 | 3 | 4 | A | 4 | 4 | 3 | 4 | B | B | C | B | 4 | 6 | 5 |
| 328 | F | 23 | 0 | 0 | 0 | 0 | 0 | 0 | C | 4 | 3 | 3 | 2 | B | 2 | 3 | 3 | 2 | B | B | A | B | 4 | 5 | 5 |
| 336 | F | 45 | 0 | 0 | 0 | 0 | 0 | 1 | C | 7 | 6 | 8 | 8 | C | 4 | 5 | 4 | 4 | B | B | A | B | 7 | 8 | 8 |
| 340 | F | 34 | 0 | 0 | 0 | 0 | 0 | 0 | C | 2 | 2 | 2 | 2 | B | 2 | 2 | 2 | 2 | B | B | A | B | 1 | 1 | 1 |
| 343 | M | 33 | 1 | 1 | 0 | 0 | 0 | 2 | D | 8 | 8 | 7 | 7 | C | 8 | 8 | 7 | 7 | B | B | C | B | 9 | 7 | 9 |
| 344 | M | 34 | 0 | 0 | 0 | 0 | 0 | 2 | A | 7 | 6 | 5 | 6 | C | 6 | 6 | 5 | 5 | B | B | C | B | 8 | 7 | 6 |
| 350 | F | 33 | 0 | 0 | 0 | 0 | 0 | 0 | B | 9 | 9 | 9 | 9 | A | 7 | 9 | 7 | 7 | B | B | A | B | 9 | 9 | 9 |
| 354 | F | 32 | 0 | 0 | 0 | 0 | 0 | 0 | A | 7 | 8 | 7 | 5 | B | 7 | 7 | 7 | 7 | B | B | A | B | 6 | 9 | 6 |
| 355 | M | 29 | 1 | 2 | 0 | 2 | 1 | 2 | A | 9 | 8 | 9 | 6 | C | 8 | 8 | 8 | 6 | B | B | C | B | 9 | 10 | 8 |

To be continued on the following page...

...continued

| ID | Gender | Age | Familarity iCub | Familarity Flobi | Familarity Nexi | Familarity Nao | Familarity Barthoc | Familarity Asimo | Video 1 | Video1 $S_{FLUID}$ | Video1 $S_{MATCH}$ | Video1 $S_{LIVELY}$ | Video1 $S_{HUMAN}$ | Video 2 | Video2 $S_{FLUID}$ | Video2 $S_{MATCH}$ | Video2 $S_{LIVELY}$ | Video2 $S_{HUMAN}$ | Compare A/B | Compare B/C | Compare C/A | Favourite | $P_{LIVELY}$ | $P_{FLUID}$ | $P_{HUMAN}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 356 | F | 37 | 0 | 0 | 0 | 0 | 0 | 0 | B | 7 | 6 | 6 | 6 | A | 6 | 7 | 6 | 6 | B | B | A | B | 6 | 4 | 3 |
| 359 | M | 37 | 0 | 0 | 0 | 0 | 0 | 0 | B | 8 | 9 | 7 | 6 | A | 8 | 8 | 7 | 8 | B | B | C | B | 7 | 7 | 7 |
| 363 | M | 39 | 2 | 2 | 0 | 2 | 2 | 2 | A | 10 | 10 | 9 | 8 | A | 10 | 10 | 9 | 8 | B | B | A | B | 10 | 9 | 8 |
| 365 | F | 56 | 0 | 0 | 0 | 0 | 0 | 0 | A | 6 | 7 | 7 | 7 | A | 6 | 7 | 7 | 7 | B | B | C | B | 10 | 10 | 10 |
| 367 | F | 25 | 1 | 0 | 0 | 0 | 0 | 0 | A | 9 | 4 | 6 | 6 | B | 9 | 5 | 6 | 6 | B | B | C | B | 9 | 8 | 9 |
| 370 | M | 30 | 1 | 4 | 0 | 3 | 2 | 2 | D | 8 | 7 | 6 | 6 | B | 8 | 7 | 6 | 6 | B | B | A | B | 4 | 8 | 3 |
| 373 | M | 29 | 0 | 3 | 0 | 3 | 0 | 2 | A | 8 | 7 | 7 | 5 | B | 8 | 6 | 8 | 5 | B | B | A | B | 9 | 8 | 7 |
| 375 | F | 37 | 0 | 1 | 0 | 0 | 0 | 0 | B | 5 | 5 | 3 | 1 | B | 4 | 5 | 3 | 1 | B | B | A | B | 8 | 8 | 3 |
| 376 | M | 33 | 1 | 0 | 0 | 0 | 1 | 0 | D | 7 | 6 | 1 | 1 | C | 6 | 5 | 1 | 1 | B | B | A | B | 2 | 10 | 2 |
| 378 | M | 30 | 1 | 1 | 0 | 0 | 0 | 0 | D | 7 | 7 | 6 | 3 | B | 7 | 7 | 4 | 3 | B | B | A | B | 7 | 7 | 3 |
| 379 | F | 25 | 3 | 3 | 0 | 3 | 0 | 0 | D | 7 | 5 | 4 | 3 | B | 7 | 5 | 4 | 4 | B | B | A | B | 8 | 8 | 8 |
| 384 | F | 42 | 4 | 2 | 0 | 4 | 2 | 4 | B | 9 | 10 | 10 | 8 | A | 9 | 10 | 9 | 8 | B | B | C | B | 10 | 10 | 10 |
| 385 | M | 40 | 1 | 1 | 0 | 1 | 0 | 1 | C | 8 | 7 | 6 | 7 | A | 8 | 7 | 6 | 7 | A | B | A | B | 8 | 8 | 8 |
| 392 | M | 39 | 0 | 0 | 0 | 0 | 0 | 2 | A | 4 | 6 | 3 | 7 | C | 4 | 6 | 3 | 7 | B | B | A | B | 6 | 8 | 6 |
| 393 | F | 41 | 1 | 1 | 0 | 0 | 0 | 0 | C | 3 | 3 | 2 | 2 | C | 3 | 3 | 2 | 2 | B | B | A | B | 3 | 4 | 1 |
| 407 | F | 31 | 0 | 0 | 0 | 0 | 0 | 0 | A | 7 | 9 | 5 | 3 | A | 6 | 8 | 2 | 2 | B | B | A | B | 5 | 8 | 4 |
| 412 | M | 28 | 0 | 0 | 0 | 0 | 0 | 2 | A | 8 | 7 | 4 | 6 | A | 7 | 6 | 4 | 6 | B | B | A | B | 10 | 10 | 10 |
| 413 | F | 35 | 0 | 1 | 0 | 0 | 0 | 2 | B | 6 | 8 | 2 | 2 | C | 6 | 8 | 2 | 2 | B | B | C | B | 4 | 9 | 2 |
| 414 | F | 28 | 0 | 0 | 0 | 0 | 0 | 0 | B | 6 | 8 | 5 | 5 | A | 7 | 7 | 6 | 6 | A | B | A | A | 6 | 8 | 7 |
| 416 | M | 43 | 0 | 2 | 0 | 0 | 0 | 2 | C | 6 | 7 | 6 | 5 | A | 7 | 8 | 8 | 8 | B | B | C | B | 10 | 10 | 10 |

To be continued on the following page...

. . . continued

| ID | Gender | Age | Familarity iCub | Familarity Flobi | Familarity Nexi | Familarity Nao | Familarity Barthoc | Familarity Asimo | Video 1 | Video1 $S_{FLUID}$ | Video1 $S_{MATCH}$ | Video1 $S_{LIVELY}$ | Video1 $S_{HUMAN}$ | Video 2 | Video2 $S_{FLUID}$ | Video2 $S_{MATCH}$ | Video2 $S_{LIVELY}$ | Video2 $S_{HUMAN}$ | Compare A/B | Compare B/C | Compare C/A | Favourite | $P_{LIVELY}$ | $P_{FLUID}$ | $P_{HUMAN}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 424 | F | 21 | 1 | 4 | 0 | 3 | 0 | 1 | B | 6 | 8 | 7 | 4 | C | 6 | 8 | 7 | 5 | B | B | C | B | 8 | 8 | 7 |
| 425 | M | 25 | 1 | 0 | 0 | 0 | 0 | 1 | A | 8 | 9 | 6 | 7 | A | 9 | 8 | 7 | 8 | B | B | A | B | 10 | 8 | 10 |
| 439 | F | 24 | 0 | 0 | 0 | 0 | 0 | 0 | C | 4 | 7 | 5 | 6 | B | 4 | 7 | 5 | 6 | B | B | A | B | 7 | 8 | 4 |
| 451 | F | 59 | 2 | 2 | 0 | 0 | 2 | 0 | D | 5 | 4 | 4 | 4 | C | 5 | 4 | 4 | 4 | B | B | C | B | 10 | 10 | 10 |
| 478 | M | 34 | 3 | 3 | 0 | 3 | 4 | 2 | C | 9 | 9 | 9 | 8 | A | 9 | 9 | 9 | 8 | B | B | A | B | 9 | 9 | 7 |
| 481 | F | 23 | 0 | 0 | 0 | 0 | 0 | 0 | D | 4 | 3 | 6 | 6 | B | 3 | 4 | 5 | 6 | B | C | C | B | 5 | 9 | 7 |
| 490 | M | 27 | 3 | 4 | 0 | 2 | 2 | 2 | C | 9 | 6 | 8 | 6 | C | 9 | 7 | 7 | 7 | B | B | A | B | 10 | 6 | 8 |
| 495 | M | 27 | 0 | 0 | 0 | 0 | 0 | 1 | D | 4 | 7 | 3 | 5 | B | 3 | 7 | 4 | 5 | B | C | C | B | 7 | 4 | 4 |
| 497 | F | 28 | 2 | 2 | 0 | 4 | 0 | 2 | A | 8 | 9 | 8 | 6 | B | 8 | 9 | 8 | 6 | B | C | C | B | 8 | 9 | 7 |
| 499 | F | 27 | 0 | 2 | 0 | 0 | 1 | 2 | A | 4 | 8 | 2 | 1 | A | 2 | 8 | 3 | 3 | B | B | A | B | 3 | 9 | 2 |
| 503 | F | 46 | 0 | 0 | 0 | 0 | 0 | 0 | D | 3 | 5 | 2 | 2 | A | 2 | 5 | 2 | 2 | B | B | A | B | 7 | 6 | 8 |
| 516 | F | 24 | 0 | 2 | 0 | 0 | 0 | 1 | B | 8 | 7 | 8 | 8 | B | 7 | 8 | 8 | 8 | B | C | C | B | 7 | 8 | 8 |
| 535 | F | 26 | 1 | 0 | 0 | 2 | 2 | 2 | C | 5 | 7 | 6 | 8 | A | 6 | 8 | 7 | 8 | B | C | C | C | 8 | 9 | 6 |
| 537 | M | 27 | 3 | 3 | 1 | 3 | 2 | 2 | D | 3 | 2 | 2 | 4 | C | 3 | 2 | 3 | 3 | B | B | C | B | 2 | 4 | 4 |
| 541 | F | 18 | 0 | 0 | 0 | 0 | 0 | 0 | A | 5 | 8 | 6 | 4 | C | 5 | 8 | 8 | 8 | B | B | A | B | 10 | 10 | 10 |
| 542 | F | 28 | 0 | 2 | 0 | 0 | 0 | 0 | B | 2 | 4 | 2 | 3 | B | 2 | 4 | 3 | 4 | A | C | C | A | 4 | 3 | 6 |
| 548 | F | 28 | 0 | 0 | 0 | 4 | 0 | 0 | B | 9 | 9 | 6 | 8 | A | 8 | 9 | 5 | 8 | B | C | C | C | 8 | 9 | 5 |
| 553 | F | 22 | 2 | 1 | 0 | 0 | 0 | 0 | B | 9 | 9 | 9 | 9 | C | 9 | 9 | 9 | 9 | B | B | A | B | 10 | 9 | 10 |
| 559 | M | 32 | 0 | 0 | 0 | 0 | 0 | 0 | D | 2 | 4 | 3 | 2 | B | 2 | 4 | 3 | 2 | B | B | A | B | 2 | 2 | 4 |
| 563 | M | 25 | 2 | 2 | 2 | 3 | 2 | 2 | C | 5 | 9 | 9 | 8 | B | 5 | 8 | 7 | 7 | B | B | C | B | 9 | 8 | 9 |

To be continued on the following page...

...continued

| ID | Gender | Age | Familarity iCub | Familarity Flobi | Familarity Nexi | Familarity Nao | Familarity Barthoc | Familarity Asimo | Video 1 | Video1 $S_{FLUID}$ | Video1 $S_{MATCH}$ | Video1 $S_{LIVELY}$ | Video1 $S_{HUMAN}$ | Video 2 | Video2 $S_{FLUID}$ | Video2 $S_{MATCH}$ | Video2 $S_{LIVELY}$ | Video2 $S_{HUMAN}$ | Compare A/B | Compare B/C | Compare C/A | Favourite | $P_{LIVELY}$ | $P_{FLUID}$ | $P_{HUMAN}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 568 | M | 30 | 3 | 3 | 0 | 3 | 2 | 2 | B | 8 | 9 | 8 | 6 | A | 7 | 8 | 7 | 6 | B | B | C | B | 9 | 9 | 4 |
| 577 | M | 24 | 1 | 0 | 0 | 0 | 0 | 0 | B | 8 | 7 | 5 | 5 | B | 6 | 6 | 6 | 6 | B | B | C | B | 6 | 7 | 8 |
| 581 | M | 37 | 0 | 0 | 0 | 0 | 0 | 0 | C | 8 | 8 | 8 | 10 | B | 8 | 8 | 8 | 10 | A | B | A | A | 6 | 10 | 6 |
| 593 | M | 26 | 4 | 3 | 0 | 2 | 2 | 3 | A | 4 | 4 | 4 | 5 | C | 5 | 5 | 5 | 5 | B | B | A | B | 9 | 8 | 8 |
| 598 | M | 27 | 3 | 3 | 0 | 4 | 2 | 2 | B | 6 | 9 | 6 | 5 | A | 6 | 8 | 6 | 5 | B | B | C | B | 5 | 8 | 2 |
| 622 | M | 22 | 1 | 2 | 0 | 4 | 1 | 2 | B | 2 | 5 | 3 | 4 | B | 2 | 3 | 2 | 2 | B | B | A | B | 7 | 10 | 3 |
| 629 | M | 27 | 1 | 2 | 0 | 0 | 0 | 2 | D | 5 | 8 | 8 | 8 | C | 6 | 8 | 6 | 7 | B | B | C | B | 6 | 5 | 6 |
| 654 | M | 25 | 0 | 0 | 0 | 0 | 0 | 1 | A | 6 | 5 | 6 | 4 | A | 4 | 5 | 4 | 6 | B | C | C | C | 9 | 9 | 8 |
| 657 | F | 30 | 0 | 0 | 1 | 0 | 0 | 0 | B | 4 | 6 | 3 | 6 | A | 4 | 6 | 2 | 6 | B | B | C | B | 8 | 9 | 10 |
| 665 | F | 21 | 0 | 0 | 0 | 0 | 0 | 0 | A | 4 | 5 | 3 | 6 | A | 4 | 3 | 5 | 6 | B | B | A | B | 5 | 7 | 6 |
| 678 | F | 38 | 2 | 2 | 0 | 2 | 2 | 2 | B | 5 | 7 | 8 | 6 | C | 5 | 7 | 8 | 7 | B | B | C | B | 8 | 7 | 6 |
| 679 | M | 24 | 0 | 0 | 0 | 0 | 0 | 0 | A | 7 | 8 | 8 | 7 | B | 8 | 8 | 8 | 9 | B | B | C | B | 3 | 8 | 8 |
| 691 | M | 38 | 0 | 0 | 0 | 0 | 0 | 0 | C | 2 | 2 | 1 | 1 | C | 1 | 2 | 2 | 1 | B | B | A | B | 9 | 10 | 9 |

Table 2.: Other known robots (skipped empty replies)

| ID | Other Robots i heared about |
|---|---|
| 3 | iCat, Biron/Tobi |
| 12 | Biron, Keepon, icat |
| 37 | Mekabot |
| 52 | Geminoid, Geminoid F, Paro, Kismet |
| 59 | T-101 |
| 87 | Coman, Pleo, BigDog, Petman, Oncilla |
| 92 | Außer R2D2 und co keine weiteren |
| 164 | R2D2 |
| 182 | Geminoid, Scitos G5, Peoplebot, P3-AT, Baxter |
| 206 | Atlas |
| 237 | KUKA-Industriearme, ATLAS, Big Dog, ARMAR, HRP-4C, Kismet |
| 251 | Robbenrobotor, Fussbalrobotor...keine gezielt mit Namen |
| 292 | Shadow Hand, barret hand, mekabot, Gifu hand, Allegro Hand |
| 299 | Tobi |
| 320 | ToBi |
| 344 | r2d2 |
| 359 | Terminator |
| 363 | Kismet |
| 373 | Data |
| 478 | Biron |
| 490 | Biron |
| 497 | Big Dog, Aibo, Chimp, PR2, ... |
| 499 | Big Dog, Lola |
| 537 | Biron, Meka |
| 559 | aibo |
| 568 | biron, hitchBOT, pleo, iuro, T-1000 |
| 593 | AMiRo, HECTOR |
| 598 | BIRON |
| 622 | Tobi |

To be continued on the following page...

. . . continued

| ID | Other Robots i heared about |
|---|---|
| 678 | Mekabot |

Table 3.: Guessed Robot Action

| ID | Guessed action | Category |
|---|---|---|
| 3 | Was unten lesen. | reading something |
| 6 | Flobi schaut sich ein/mehrere vor ihm liegende Objekt/e an und schaut anschließend mich an | looking at something |
| 12 | lesen | reading something |
| 27 | Der Roboter neigt den Kopf nach rechts unten, fixiert etwas mit den Augen ,es folgt eine Augenbewegung (folgt einer Bewegung mit den Augen?) und richtet den Kopf wieder geradeaus. | moving |
| 36 | Der Roboter sucht etwas. | searching something |
| 37 | Floppy betrachtet ein Objekt | looking at something |
| 41 | nach unten rechts gucken, als würde er etwas suchen | searching something |
| 48 | Er dreht sich nach rechts und bewegt die Augen, alsob er etwas lesen würde. | reading something |
| 52 | Aufgrund der mikrosakkadischen Augenbewegungen würde ich vermuten, dass Flobi ein Wort liest. | reading something |
| 57 | er bewegt die augen und schaut nach unten links in die ecke. kehrt dann wieder zurueck und bewegt erneut die augen. | looking at something |

To be continued on the following page...

. . . continued

| ID | Guessed action | Category |
|---|---|---|
| 59 | Guckt nach (von sich aus) unten rechts und liest ein paar Zeilen Text | reading something |
| 61 | er bewegt seinen Kopf um nach unten sehen zu können, er öffnet und schließt seine Augen | looking at something |
| 64 | er ist traurig, schaut zu Boden und schämt sich | being ashamed |
| 65 | etwas verfolgen | looking at something |
| 75 | Er sieht in die Kamera, sieht dann schräg nach unten und bewegt die Augen, als ob er etwas lesen würde. | reading something |
| 76 | er schaut auf den Boden | looking at something |
| 83 | | no answer |
| 85 | Auf den Boden gucken | looking at something |
| 86 | Der Roboter guckt auf den Boden, suchend. | searching something |
| 87 | | no answer |
| 88 | der Roboter neigt den Kopf und bewegt sich nach rechts und zurück | moving |
| 89 | | no answer |
| 92 | Einen Text lesen | reading something |
| 93 | | no answer |
| 94 | Er liest. | reading something |
| 96 | Der Roboter guckt kurz schräg nach unten und bewegt dabei die Augen nachdenklich von links nach rechts | looking at something |
| 98 | Guckt sich etwas an und denkt drüber nach | looking at something |
| 106 | Schaut auf den Boden | looking at something |

To be continued on the following page...

. . . continued

| ID | Guessed action | Category |
|---|---|---|
| 107 | | no answer |
| 110 | Niedergeschlagen den Kopf schütteln | head-shaking / disapproval |
| 111 | Ausgehend aus der Mittelposition senkt sich sein Kopf nach unten links bzw. rechts. Unten angekommen schüttelt er leicht den Kopf | moving |
| 113 | lesen | reading something |
| 114 | Er schaut nach unten und wirkt 'nachdenklich' | looking at something |
| 122 | Der Roboter senkt seinen Kopf nach links unten, bewegt seine Augen ein wenig hin und her, und hebt den Kopf wieder und schaut in die Kamera. Dabei blinzelt er in regelmäßigen Abständen. | looking at something |
| 125 | Schaut sich etwas unten an, scheint etwas bewegliches zu sein | looking at something |
| 127 | er neigt den Kopf nach unten | moving |
| 128 | lesen | reading something |
| 133 | etwas lesen | reading something |
| 134 | nach rechts unten gucken | looking at something |
| 136 | Der Roboter schaut nach unten und liest etwas. | reading something |
| 138 | schaut nach unten, bewegt die augen | looking at something |
| 139 | Er ist verlegen. | being ashamed |
| 140 | | no answer |
| 164 | lesen | reading something |
| 179 | Schaut sich etwas an. | looking at something |

To be continued on the following page...

. . . continued

| ID | Guessed action | Category |
|---|---|---|
| 182 | Liest etwas das vor Ihm liegt, sucht nach einem Gegenstand von dem er weiss wo er sich ungefaehr befindet. | reading something |
| 183 | Für mich scheint der Roboter zu lesen, da er nach unten schaut und sich die Augen nach rechts und wieder zurück bewegen. | reading something |
| 184 | lesen | reading something |
| 192 | es sieht aus, als würde er kurz etwas lesen bevor er wieder sein gegenüber anschaut | reading something |
| 194 | Kopfschütteln | head-shaking / disapproval |
| 197 | Der Roboter schaut nach links unten (aus meiner Sicht) bzw. nach rechts unten (aus Roboters Sicht) | looking at something |
| 203 | antwortet mit nein | head-shaking / disapproval |
| 206 | as if it was reading although obviously cannot read | reading something |
| 209 | schaut in die rechte untere Ecke und es scheint, als würde der Roboter zwei Zeilen lesen, schaut dann wieder nach vorn | reading something |
| 212 | ”Überlegen”, oder etwas tiefer liegendes betrachten | looking at something |
| 225 | es sieht aus als würde er lesen (einen kurzen Satz oder ein bis zwei Wörter) | reading something |
| 236 | | no answer |

To be continued on the following page...

. . . continued

| ID | Guessed action | Category |
|---|---|---|
| 237 | Der Roboter wird anscheinend von einem visuellen Reiz zu ihrer/seiner Rechten abgelenkt und schaut kurz dorthin, ehe der Blick wieder in Normalposition zurückkehrt (etwa: Die Betrachterin/den Betrachter fokussiert) | looking at something |
| 243 | Er bewegt seinen Kopf nach unten und bewegt seine Augen. Es scheint, als würde er etwas lesen. | reading something |
| 251 | er schaut tnach unten, bewegt den Kopf nach rechts und links und sieht dann wieder hoch, es wirklt traurig und als ob er etwas ”verneint” | looking at something |
| 252 | Lesen eines Textes. | reading something |
| 287 | sieht etwas am unteren linken rand udn blickt mir dann in die augen | looking at something |
| 292 | kuckt oder sucht mit die Augen etwas nach unten und schaut wieder nach Vorne. Blinkt mit die Augen | searching something |
| 298 | er sieht nach unten rechts und blickt dann wieder auf- er blinzelt mehrfach | looking at something |
| 299 | Lesen | reading something |
| 300 | lesen | reading something |

To be continued on the following page...

. . . continued

| ID | Guessed action | Category |
|---|---|---|
| 303 | Bewegt/neigt den Kopf schräg nach unten, dann nach links, dann wieder zurück und hebt ihn wieder an | moving |
| 313 | kopf senken, seitlich bewegen, sieht aus wie lesen | reading something |
| 318 | er ist traurig, guckt nach unten und schüttelt den Kopf | looking at something |
| 320 | schaut sich etwas an und ist etwas ”traurig” | looking at something |
| 322 | sich schämen | being ashamed |
| 328 | Der Roboter richtet seine Gesicht und die Augen nach unten rechts. Es sieht so aus, als würde mit seinen Augen etwas betrachten, da die Pupillen von hin und her wandern. Auch ein Zwinkern ist dabei. | looking at something |
| 336 | | no answer |
| 340 | Er liest sich eine kurze Notiz durch, dafür schaut er zunächst runter und bewegt dann seine Augen. | reading something |
| 343 | etwas lesen oder verfolgen | reading something |
| 344 | Er schaut verlegen nach unten | being ashamed |
| 350 | guckt nach unten rechts | looking at something |
| 354 | schaut nach unten und zwinkert mit den Augen, sieht aus als wenn er mit den Augen etwas verfolgt... | looking at something |
| 355 | lesen | reading something |
| 356 | sich etwas auf dem Boden anschauen | looking at something |

To be continued on the following page...

. . . continued

| ID | Guessed action | Category |
|---|---|---|
| 359 | blickt nach unten, scheint etwas zu lesen | reading something |
| 363 | Der Roboter schaut nach rechts unten und liest. | reading something |
| 365 | Die Augen gucken nach links unten und wieder hoch. | looking at something |
| 367 | Roboter neigt den Kopf nach unten, richtet nach einem Moment wieder den Kopf auf und zwinkert währendessen zweimal | moving |
| 370 | Er schaut sich etwas an das rechts unten von ihm liegt | looking at something |
| 373 | Er schaut auf einen kurzen Text, liest ihn und schaut wieder auf | reading something |
| 375 | Er schaut sich etwas an und liest ggf. einen Text | reading something |
| 376 | Den Kopf nach rechts unten bewegen. | moving |
| 378 | Blinzelt, schaut nach unten und blickt dann wieder auf | looking at something |
| 379 | senkt den blick, schaut von links nach rechts, blinzelt | looking at something |
| 384 | Der Roboter verfolgt eine Aktion mit den Augen. | looking at something |
| 385 | scheint (auf dem Handy o.ä.) zu lesen | reading something |
| 392 | schaut nach rechts unten und bewegt den Kopf leicht hin und her | looking at something |
| 393 | lesen | reading something |
| 407 | | no answer |
| 412 | Etwas lesen | reading something |

To be continued on the following page...

. . . continued

| ID | Guessed action | Category |
|---|---|---|
| 413 | Der Roboter guckt nach unten und sucht da irgendwas (vielleicht auf dem Boden). | searching something |
| 414 | gucken, lesen, dann gucken | reading something |
| 416 | | no answer |
| 424 | lesen | reading something |
| 425 | Er neigt seinen Kopf nach rechts unten, blickt mit den Augen in die gleiche Richtung und schaut danach nach links. | looking at something |
| 439 | Lesen | reading something |
| 451 | Er schaut auf etwas herunter, folgt dem mit seinem Blick, fast so als würde er etwas lesen. | reading something |
| 478 | Lesen | reading something |
| 481 | den Kopf nach unten rechts neigen | moving |
| 490 | schaut nach unten | looking at something |
| 495 | Sich schämen... Peinlich berührt sein | being ashamed |
| 497 | Den Kopf zu einer bestimmten Stelle bewegen und dort versucht er etwas zu erkennen. Durch die sich bewegenden Pupillen sieht es aus, als würde er lesen. | reading something |
| 499 | | no answer |
| 503 | Der Roboter schaut nach unten, anscheinend um etwas nachzuschauen. | looking at something |
| 516 | lesen | reading something |
| 535 | Lesen | reading something |
| 537 | Der Roboter liest etwas. | reading something |
| 541 | Der Roboter liest etwas. | reading something |

To be continued on the following page...

. . . continued

| ID | Guessed action | Category |
|---|---|---|
| 542 | Schämen | being ashamed |
| 548 | | no answer |
| 553 | Der Roboter guckt nach unten (auf den Boden), so als würde er dort etwas suchen oder dort etwas mit dem Blick verfolgen | searching something |
| 559 | Der Roboter neigt den Kopf nach schräg unten links und blinzelt. Er bewegt die Pupillen und auch leicht den Kopf in dieser geneigten Position. Dann blickt er wieder hoch zum Bildmittelpunkt und neigt den Kopf dabei nach Oben | looking at something |
| 563 | Er schaut nach unten und liest | reading something |
| 568 | lesen oder etwas vor sich suchen | reading something |
| 577 | Der Roboter schaut in die untere rechte Ecke. ablesen? | reading something |
| 581 | Liest 2 Zeilen Text oder folgt einer Bewegung in seinem unteren Blickfeld | reading something |
| 593 | | no answer |
| 598 | lesen | reading something |
| 622 | Er liest einen text links unten | reading something |
| 629 | | no answer |
| 654 | nach unten links schauen | looking at something |
| 657 | der Roboter dreht den Kopf, wendet den Kopf nach unten und mcht Augenbewegungen von links nach Rechts und schaut dann wieer hoch | looking at something |
| 665 | gucken, lesen | reading something |
| 678 | lesen | reading something |

To be continued on the following page...

. . . continued

| ID | Guessed action | Category |
|---|---|---|
| 679 | Text lesen | reading something |
| 691 | | no answer |

| Answer Code | Answer |
|---|---|
| 0 | sagt mir nichts<br>(means nothing to me) |
| 1 | habe den Namen schon einmal gehört<br>(Heard the name before) |
| 2 | habe ich schon einmal gesehen<br>(I have seen one like that before) |
| 3 | habe ich schon einmal in einer Studie gesehen<br>(I have seen it in a study) |
| 4 | habe ich mit gearbeitet<br>(I have worked with it) |

Table 4.: Robot familiarity "Welche der folgenden Roboter kennen Sie?"

| Code | Condition | Description |
|---|---|---|
| A | $C_{NOLIDS}$ | neck and eye motion, eyelids fixed |
| B | $C_{FULL0.7}$ | full motion with reduced acceleration |
| C | $C_{NECK}$ | only neck motion, eyes, and eyelids fixed |
| D | $C_{FULL1.0}$ | full motion |

Table 5.: Mapping of keys to video condition.

# Bibliography


[1] W. L. Stone, **The History of Robots**, in *Robotics and Automation Handbook*, T. R. Kurfess, Ed. 2005.

[2] H. Abdul-muhsin and V. Patel, **History of Robotic Surgery**, in *Robotics in General Surgery*, Kim, Chae, 2014, pp. 3–8.

[3] T. Standage, **The Turk: The Life and Times of the Famous Eighteenth-Century Chess-Playing Machine**, Berkley Trade, 2002, vol. 127.

[4] R. Clarke, **Asimov's Laws of Robotics: Implications for Information Technology**, *Computer*, vol. 27, no. 1, pp. 57–66, 1993,

[5] J. F. Engelberger, **Historical Perspective and Role in Automation**, in *Handbook of Industrial Robotics*, S. Y. Nof, Ed., 2nd. 1999.

[6] B. Siciliano and O. Khatib, **Introduction**, in *Springer Handbook of Robotics*, B. Siciliano and O. Khatib, Eds. Springer, 2008, pp. 1–9.

[7] iRobot. (2013). **iRobot Launches Braava Floor Mopping Robot in North America**, [Online]. Available: `http://media.irobot.com/press-releases?item=122399` (visited on 01/20/2017).

[8] ——, (2017). **History: iRobot**, [Online]. Available: `http://www.irobot.com/About-iRobot/Company-Information/History.aspx` (visited on 02/01/2017).

[9] International Federation of Robotics, **Executive Summary of World Robotics 2019 Service Robots**, 2019.

[10] S. Thrun, **Toward a Framework for Human-Robot Interaction**, *Human–Computer Interaction*, vol. 19, pp. 9–24, 2004,

[11] B. Reeves and C. Nass, **How People Treat Computers, Television, and New Media Like Real People and Places**, in *The Media Equation*, 3. 1998, vol. 34, pp. 19–36.

[12] C. Breazeal, **Toward sociable robots**, *Robotics and Autonomous Systems*, vol. 42, no. 3-4, pp. 167–175, 2003,

[13] C. C. Kemp, P. Fitzpatrick, H. Hirukawa, K. Yokoi, K. Harada, and Y. Matusmoto, **Human-centered and Life-Like Robotics**, in *Springer Handbook of Robotics*, B. Siciliano and O. Khatib, Eds., Springer, 2008, ch. 56, pp. 1308–1327.

[14] C. E. Looser and T. Wheatley, **The Tipping Point of Animacy. How, When, and Where We Perceive Life in a Face**, *Psychological science*, vol. 21, no. 12, pp. 1854–1862, Dec. 2010.

[15] A. D. Dragan, S. Bauman, J. Forlizzi, and S. S. Srinivasa, **Effects of Robot Motion on Human-Robot Collaboration**, in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '15, New York, NY, USA: ACM, 2015, pp. 51–58.

[16] M. L. Knapp and J. A. Hall, **An Introduction to the Study of Nonverbal Communication**, in *Nonverbal Communication in Human Interaction*, 7th ed. Bosten, USA: Wadsworth: Cengage Learning, 2010, ch. 1.

[17] M. Imai, T. Kanda, T. Ono, H. Ishiguro, and K. Mase, **Robot Mediated Round Table: Analysis of the Effect of Robot's Gaze**, in *Proceedings. 11th IEEE International Workshop on Robot and Human Interactive Communication*, 2002.

[18] C. C. Kemp, P. Fitzpatrick, H. Hirukawa, K. Yokoi, K. Harada, and Y. Matusmoto, **Interfacing with people**, in *Springer Handbook of Robotics*, B. Siciliano and O. Khatib, Eds. Springer, 2008.

[19] K. Dautenhahn, **Socially Intelligent Robots: Dimensions of Human-Robot Interaction**, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 362, no. 1480, pp. 679–704, 2007.

[20] M. Mori, K. F. MacDorman, and N. Kageki, **The Uncanny Valley**, in *Robotics & Automation Magazine, IEEE*, 33-35, 2. Energy, 2012, vol. 19, pp. 98–100.

[21] M. Mori, K. MacDorman, and N. Kageki, **The Uncanny Valley [From the Field]**, *IEEE Robotics & Automation Magazine*, vol. 19, pp. 98–100, Jun. 2012.

[22] C. Bartneck, T. Kanda, H. Ishiguro, and N. Hagita, **Is the Uncanny Valley an Uncanny Cliff?** In *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 2007.

[23] B. R. Duffy and G. Joue, **I, Robot Being**, in *Proceedings of the Intelligent Autonomous Systems Conference, IAS*, vol. 2005, 2004.

[24] D. Moore, H. Tennent, N. Martelaro, and W. Ju, **Making Noise Intentional: A Study of Servo Sound Perception**, in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17, Vienna, Austria: ACM, 2017, pp. 12–21, ISBN: 978-1-4503-4336-7.

[25] F. Hegel, **Gestalterisch konstruktiver Entwurf eines sozialen Roboters**, Ph.D. dissertation, 2010.

[26] S. Schulz, I. Lütkebohle, and S. Wachsmuth, **An Affordable, 3D-printable Camera Eye with Two Active Degrees of Freedom for an Anthropomorphic Robot**, in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct. 2012, pp. 764–771.

[27] S. Schulz, S. Meyer zu Borgsen, and S. Wachsmuth, **See and Be Seen – Rapid and Likeable High-Definition Camera-Eye for Anthropomorphic Robots**, in *Proceedings - IEEE International Conference on Robotics and Automation, ICRA 2019*, May 2019.

[28] S. Schulz, F. Lier, A. Kipp, and S. Wachsmuth, **Humotion - A human inspired gaze control framework for anthropomorphic robot heads**, in *Proceedings of the 4th International Conference on Human Agent Interaction, HAI*, 2016.

[29] R. K. Sharma and B. Ehinger, **Development and Structure of the Retina**, in *Adler's Physiology of the Eye: Clinical Application*. Mosby, 2003, pp. 319–374.

[30] R. Jogi, **Basic Sciences**, in *Textbook of Ophthalmology*, S. Agarwal, A. Agarwal, and D. J. Apple, Eds. Jaypee Brothers Medical Publishers (P) Ltd, 2002, ch. 4, pp. 22–27.

[31] H. Kaufmann and H. Steffen, **Anatomie und Physiologie der Orbita und des Bewegungsapparates**, in *Strabismus*, H. Kaufmann and H. Steffen, Eds., 4th ed. Thieme, 2003, ch. 1.2, pp. 39–72.

[32] R. J. Leigh and D. S. Zee, **The Neurology of Eye Movements (Contemporary Neurology Series)**, F.A. Davis Company, 1991.

[33] P. Riordan-Eva and J. P. Whitcher, **Anatomy and Embryology of the Eye**, in *Vaughan & Asbury's General Ophthalmology*, 2008, ch. 1.

[34] A. H. Schultz, **The size of the orbit and of the eye in primates**, *American Journal of Physical Anthropology*, vol. 26, no. 1, pp. 389–408, 1940,

[35] M. Sachsenweger, **Bulbus oculi (Augapfel)**, in *Augenheilkunde*, M. Sachsenweger, Ed. Stuttgart: Georg Thieme Verlag, 2003, ch. 2.

[36] F. Rüfer, A. Schröder, and C. Erb, **White-to-White Corneal Diameter: Normal Values in Healthy Humans Obtained With the Orbscan II Topography System**, *Cornea*, vol. 24, pp. 259–261, May 2005.

[37] J. D. Reynolds and S. E. Olitsky, **Anatomy and Physiology of the Retina**, *Pediatric Retina*, 2011,

[38] H. Kolb, **Simple Anatomy of the Retina**, in *Webvision: The Organization of the Retina and Visual System*. 1995.

[39] N. Dodgson, **Variation and Extrema of Human Interpupillary Distance**, vol. 5291, Jan. 2004, pp. 36–46.

[40] R. H. Spector, **Visual Fields**, *Clinical Methods: The History, Physical, and Laboratory Examinations*, pp. 565–572, 1990.

[41] L. Sawides, A. de Castro, and S. A. Burns, **The Organization of the Cone Photoreceptor Mosaic measured in the Living Human Retina**, *Vision Research*, vol. 132, pp. 34–44, 2017,

[42] F. N. Low, **Peripheral Visual Acuity**, *JAMA Ophthalmology*, vol. 45, no. 1, pp. 80–99, Jan. 1951,

[43] D. Miller, P. Schor, and P. Magnante, **Optics of the Normal Eye**, in *Ophthalmology*, M. Yanoff and J. Duker, Eds. Mosby, Incorporated, 1998, ch. 2.6.

[44] D. Guitton and M. Volle, **Gaze Control in Humans: Eye-Head Coordination during Orienting Movements to Targets Within and Beyond the Oculomotor Range**, *Journal of Neurophysiology*, vol. 58, no. 3, pp. 427–459, Sep. 1987.

[45] A. T. Bahill, D. Adler, and L. Stark, **Most Naturally Occurring Human Saccades Have Magnitudes of 15 Degrees or Less**, *Investigative Ophthalmology*, vol. 14, pp. 468–9, Jul. 1975.

[46] F. C. Donders, **Beitrag zur Lehre von den Bewegungen des menschlichen Auges**, in *Holländische Beiträge zu den anatomischen und physiologischen Wissenschaften*, Bd. 1, Böttcher, 1984.

[47] H. von Helmholtz and A. König, **Handbuch der physiologischen Optik**, ser. Handbuch der physiologischen Optik Bd. 1. L. Voss, 1896, ch. 27.

[48] M. Hansard and R. Horaud, **Cyclorotation Models for Eyes and Cameras**, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 1, pp. 151–161, Feb. 2010,

[49] R. J. Leigh and D. S. Zee, **Diagnosis of Peripheral Ocular Motor Palsies and Strabismus**, in *The Neurology of Eye Movements (Contemporary Neurology Series)*. F.A. Davis Company, 1991, ch. 9, pp. 293–377.

[50] ——, **A Survey of Eye Movements**, in *The Neurology of Eye Movements (Contemporary Neurology Series)*. F.A. Davis Company, 1991, ch. 1, pp. 3–14.

[51] R. Gellman, J. Carl, and F. Miles, **Short Latency Ocular-Following Responses in Man**, *Visual Neuroscience*, vol. 5, no. 02, pp. 107–122, 1990,

[52] H. Collewijn and J. B. Smeets, **Early Components of the Human Vestibulo-Ocular Response to Head Rotation: Latency and Gain**, *Journal of Neurophysiology*, vol. 84, no. 1, pp. 376–389, Sep. 2000.

[53] A. Roncone, U. Pattacini, G. Metta, and L. Natale, **Gaze Stabilization for Humanoid Robots: a Comprehensive Framework**, *IEEE-RAS International Conference on Humanoid Robots*, vol. abs/1411.3, 2014.

[54] C. Rashbass, **The Relationship between Saccadic and Smooth Tracking Eye Movements**, *The Journal of Physiology*, vol. 159, no. 2, pp. 326–338, Dec. 1961,

[55] M. Steinbach, **Eye Tracking of Self-Moved Targets: The Role of Efference**, *Journal of Experimental Psychology*, vol. 82, no. 2, p. 366, 1969,

[56] D. A. Robinson, **Control of Eye Movements**, in *Comprehensive Physiology*, Hoboken, NJ, USA: John Wiley & Sons, Inc., Jan. 1981.

[57] R. J. Leigh and D. S. Zee, **The Saccadic System**, in *The Neurology of Eye Movements (Contemporary Neurology Series)*. F.A. Davis Company1, 1991, ch. 3, pp. 79–138.

[58] D. Purves, G. J. Augustine, D. Fitzpatrick, L. C. Katz, A.-S. LaMantia, J. O. McNamara, and S. M. Williams, **Types of Eye Movements and Their Functions**, in *Neuroscience*, D. Purves, G. J. Augustine, D. Fitzpatrick, L. C. Katz, A.-S. LaMantia, J. O. McNamara, and S. M. Williams, Eds. Sinauer Associates, 2001, ch. 19.

[59] A. T. Bahill, A. Brockenbrough, and B. T. Troost, **Variability and Development of a Normative Data Base for Saccadic Eye Movements**, in *Investigative Ophthalmology & Visual Science*, 1, vol. 21, 1981, pp. 116–125.

[60] A. Huber, **Eye Movements Recording**, *Bulletin de la Societe belge d'Ophtalmologie*, vol. 302, pp. 337–346, 1983.

[61] R. A. Abrams, D. E. Meyer, and S. Kornblum, **Speed and accuracy of saccadic eye movements: Characteristics of impulse variability in the oculomotor system**, *Journal of Experimental Psychology: Human Perception and Performance*, vol. 15, no. 3, pp. 529–543, Aug. 1989.

[62] G. Westheimer, **Eye Movement Responses to a Horizontally Moving Visual Stimulus**, *A.M.A. Archives of Ophthalmology*, vol. 52, no. 6, pp. 932–941, 1954,

[63] L. Young and L. Stark, **Variable Feedback Experiments Testing a Sampled Data Model for Eye Tracking Movements**, *IEEE Transactions on Human Factors in Electronics*, vol. HFE-4, no. 1, pp. 38–51, Sep. 1963,

[64] R. J. Leigh and D. S. Zee, **Vergence Eye Movements**, in *The Neurology of Eye Movements (Contemporary Neurology Series)*, F.A. Davis Company, 1991, ch. 8, pp. 264–290.

[65] A. L. Yarbus, **Eye Movements During Change of Stationary Points of Fixation in Space**, in *Eye Movements and Vision*, Boston, MA: Springer US, 1967, pp. 147–158.

[66] J. Müller, **Vom Gesichtssinn**, in *Handbuch der Physiologie des Menschen - Bd. 2*, Hölscher, 1840, ch. I, p. 336.

[67] R. J. Leigh and D. S. Zee, **Eye-Head Movements**, in *The Neurology of Eye Movements (Contemporary Neurology Series)*, F.A. Davis Company, 1991, ch. 7, pp. 232–263.

[68] W. H. Zangemeister and L. Stark, **Gaze Latency: Variable Interactions of Head and Eye Latency**, *Experimental Neurology*, vol. 75, no. 2, pp. 389–406, 1982,

[69] C. Evinger, K. a. Manning, and P. a. Sibony, **Eyelid Movements: Mechanisms and Normal Data**, *Investigative Ophthalmology and Visual Science*, vol. 32, no. 2, pp. 387–400, 1991.

[70] N. J. Gandhi and H. A. Katnani, **Interactions of Eye and Eyelid Movements**, in *The Oxford Handbook of Eye Movements*, March 2013, S. P. Liversedge, I. Gilchrist, and S. Everling, Eds., Oxford University Press, 2012, ch. 17, pp. 323–336.

[71] K. Schmidtke and J. A. Büttner-Ennever, **Nervous Control of Eyelid Function**, *Brain*, vol. 115, no. 1, pp. 227–247, 1992.

[72] H. Schiffman, **The Visual System**, in *Sensation and Perception: An Integrated Approach*, H. Schiffman, Ed. Wiley, 2000, ch. 3.

[73] D. Cramon and J. Zihl, **Die Häufigkeit von schnellen Augenbewegungen und Blinks als Aktivationsindikator**, *Journal of Neurology*, vol. 215, no. 2, pp. 115–125, 1977,

[74] C. Evinger, K. a. Manning, J. J. Pellegrini, M. a. Basso, a. S. Powers, and P. a. Sibony, **Not Looking While Leaping: The Linkage of Blinking and Saccadic Gaze Shifts**, *Experimental Brain Research*, vol. 100, no. 2, pp. 337–344, 1994,

[75] K. Lorenz, **Ganzheit und Teil in der tierischen und menschlichen Gemeinschaft**, in *Über tierisches und menschliches Verhalten: aus dem Werdegang der Verhaltenslehre*. R. Piper & Co Verlag München, 1965.

[76] S. Gould, **A Biological Homage to Mickey Mouse**, *Ecotone*, vol. 4, Jan. 2008.

[77] World Health Organisation. (2015). **Change the Definition of Blindness**, [Online]. Available: http://www.who.int/blindness/Change%20the%20Definition%20of%20Blindness.pdf (visited on 07/14/2015).

[78] Maxon Motor AG. (2019). **RE-max 13 Motor**. Datasheet, [Online]. Available: https://www.maxonmotor.de/medias/sys_master/root/8807108608030/13-373-EN.pdf (visited on 09/08/2019).

[79] K. Oh, D. Hanson, W. Kim, Y. Han, J. Kim, and I. Park, **Design of Android type Humanoid Robot Albert HUBO**, in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct. 2006, pp. 1428–1433.

[80] F. Hegel, T. Spexard, B. Wrede, G. Horstmann, and T. Vogt, **Playing a Different Imitation Game: Interaction with an Empathic Android Robot**, in *2006 6th IEEE-RAS International Conference on Humanoid Robots*, Dec. 2006, pp. 56–61.

[81] D. Sakamoto, T. Kanda, T. Ono, H. Ishiguro, and N. Hagita, **Android as a Telecommunication Medium with a Human-like Presence**, in *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction*, ser. HRI '07, Arlington, Virginia, USA: ACM, 2007, pp. 193–200.

[82] **Geminoid DK Mechanical Test**, https://www.youtube.com/watch?v=eZlLNVmaPbM, (visited on 25/09/2019).

[83] Hanson Robotics. (2019). **Zeno, Image**, [Online]. Available: https://www.hansonrobotics.com/wp-content/uploads/2018/11/zeno.jpg (visited on 03/24/2019).

[84] AIST. (2016). **HRP4C, Image**, [Online]. Available: https://unit.aist.go.jp/is/humanoid/index.html (visited on 06/09/2016).

[85] D. Glas, T. Minato, C. Ishi, T. Kawahara, and H. Ishiguro, **ERICA: The ERATO Intelligent Conversational Android**, Aug. 2016.

[86] Hanson Robotics. (2019). **Sophia**, [Online]. Available: https://www.hansonrobotics.com/sophia (visited on 08/08/2019).

[87] M. Hackel, S. Schwope, J. Fritsch, B. Wrede, and G. Sagerer, **Humanoid Robot Platform suitable for studying Embodied Interaction**, in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Aug. 2005, pp. 2443–2448.

[88] S. Nishio, H. Ishiguro, and N. Hagita, **Geminoid: Teleoperated Android of an Existing Person**, in *New Developments In Humanoid Robots*, 2007, pp. 343–352.

[89] D. Hanson, S. Baurmann, T. Riccio, R. Stuart Margolin, T. Dockins, M. Tavares, and K. T. Carpenter, **Zeno: A Cognitive Character**, in *AAAI*, 2008.

[90] I. Ranatunga, J. Rajruangrabin, D. Popa, and F. Makedon, **Enhanced Therapeutic Interactivity using Social Robot Zeno**, May 2011, p. 57.

[91] **DYNAMIXEL All-in-one Smart Actuator**, `http://www.robotis.us/dynamixel/`, visited on 2019/12/15.

[92] K. Kaneko, F. Kanehiro, M. Morisawa, T. Tsuji, K. Miura, S. Nakaoka, S. Kajita, and K. Yokoi, **Hardware improvement of Cybernetic Human HRP-4C for Entertainment use**, in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sep. 2011, pp. 4392–4399.

[93] A. Edsinger and U.-M. O'Reilly, **Designing a Humanoid Robot Face to Fulfill a Social Contract**, *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 2000.

[94] C. Breazeal, A. Edsinger, P. Fitzpatrick, B. Scassellati, and P. Varchavskaia, **Social Constraints on Animate Vision**, *Intelligent Systems and their Applications, IEEE*, vol. 15, pp. 32–37, Aug. 2000.

[95] **Domo, Images**, `http://people.csail.mit.edu/edsinger/domo_images.htm`, (visited on 2016/09/13).

[96] D. Biamino, G. Cannata, M. Maggiali, and A. Piazza, **MAC-EYE: A Tendon Driven Fully Embedded Robot Eye**, in *Proceedings of 2005 5th IEEE-RAS International Conference on Humanoid Robots*, 2005, pp. 62–67.

[97] T. Asfour, K. Welke, P. Azad, A. Ude, and R. Dillmann, **The Karlsruhe Humanoid Head**, in *8th IEEE-RAS International Conference on Humanoid Robots, Humanoids*, Feb. 2008, pp. 447–453.

[98] X.-y. Wang, Y. Zhang, X.-j. Fu, and G.-s. Xiang, **Design and Kinematic Analysis of a Novel Humanoid Robot Eye Using Pneumatic Artificial Muscles**, *Journal of Bionic Engineering*, vol. 5, no. 3, pp. 264–270, 2008,

[99] T. Villgrattner and H. Ulbrich, **Design and Control of a Compact High-Dynamic Camera-Orientation System**, *IEEE/ASME Transactions on Mechatronics*, vol. 16, no. 2, pp. 221–231, Apr. 2011,

[100] T. Asfour, J. Schill, H. Peters, C. Klas, J. Bücker, C. Sander, S. Schulz, A. Kargov, T. Werner, and V. Bartenbach, **ARMAR-4: A 63 DOF Torque Controlled Humanoid Robot**, in *13th IEEE-RAS International Conference on Humanoid Robots, Humanoids*, Oct. 2013, pp. 390–396.

[101] T. Makabe, K. Kawaharazuka, K. Tsuzuki, K. Wada, S. Makino, M. Kawamura, A. Fujii, M. Onitsuka, Y. Asano, K. Okada, K. Kawasaki, and M. Inaba, **Development of Movable Binocular High-Resolution Eye-Camera Unit for Humanoid and the Evaluation of Looking Around Fixation Control and Object Recognition**, Nov. 2018, pp. 840–845.

[102] R. A. Brooks, C. Breazeal, M. Marjanović, B. Scassellati, and M. M. Williamson, **The Cog Project: Building a Humanoid Robot**, in *Computation for Metaphors, Analogy, and Agents*, C. L. Nehaniv, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 52–87.

[103] B. Scassellati, **A Binocular, Foveated Active Vision System**, in *MIT AI Memo 1628*. Mar. 1999.

[104] C. Breazeal, A. Edsinger, P. Fitzpatrick, and B. Scassellati, **Active Vision for Sociable Robots**, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. XX, no. 5, pp. 1–12, Sep. 2000,

[105] A. Edsinger, **Robot Manipulation in Human Environments**, Ph.D. dissertation, Aug. 2007.

[106] A. Edsinger and J. Weber, **Domo: A force sensing Humanoid Robot for Manipulation Research**, in *4th IEEE/RAS International Conference on Humanoid Robots*, vol. 1, Nov. 2004, pp. 273–291.

[107] L. Aryananda and J. Weber, **MERTZ: A Quest for a Robust and Scalable Active Vision Humanoid Head Robot**, in *4th IEEE/RAS International Conference on Humanoid Robots*, vol. 2, Dec. 2004, pp. 513–532.

[108] Shadow Robot Company, **Shadow Dexterous Hand C5 Technical Specification**, 2008.

[109] T. Villgrattner and H. Ulbrich, **Optimization and Dynamic Simulation of a Parallel Three Degree-of-Freedom Camera Orientation System**, in *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, Oct. 2010, pp. 2829–2836.

[110] Y. Asano, T. Kozuki, S. Ookubo, M. Kawamura, S. Nakashima, T. Katayama, I. Yanokura, T. Hirose, K. Kawaharazuka, S. Makino, Y. Kakiuchi, K. Okada, and M. Inaba, **Human mimetic Musculoskeletal Humanoid Kengoro toward Real World Physically Interactive Actions**, in *IEEE-RAS 16th International Conference on Humanoid Robots, Humanoids*, Nov. 2016, pp. 876–883.

[111] **Robothespian**, `https://www.engineeredarts.co.uk/robothespian/`, (visited on 2019/06/30).

[112] **Robothespian SociBot**, `https://www.engineeredarts.co.uk/product/socibot-head-instead-of-robothespian-head/`, (visited on 2019/04/01).

[113] R. Loureiro, A. Lopes, C. Carona, D. Almeida, F. Faria, L. Garrote, C. Premebida, and U. J. Nunes, **ISR-RobotHead: Robotic Head with LCD-based Emotional Expressiveness**, in *IEEE 5th Portuguese Meeting on Bioengineering, ENBENG*, Feb. 2017, pp. 1–4.

[114] A. Parmiggiani, L. Fiorio, A. Scalzo, A. V. Sureshbabu, M. Randazzo, M. Maggiali, U. Pattacini, H. Lehmann, V. Tikhanoff, D. Domenichelli, A. Cardellino, P. Congiu, A. Pagnin, R. Cingolani, L. Natale, and G. Metta, **The Design and Validation of the R1 Personal Humanoid**, in *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, Sep. 2017, pp. 674–680.

[115] S. Al Moubayed, J. Beskow, G. Skantze, and B. Granström, **Furhat: A Back-Projected Human-Like Robot Head for Multiparty Human-Machine Interaction**, *Multiparty Human-Machine Interaction in Cognitive Behavioural Systems, COST*, Jan. 2012.

[116] **Emotion Expression Humanoid Robot WE-4RII**, `http://www.takanishi.mech.waseda.ac.jp/top/research/we/we-4rii/index.htm`, (visited on 2016/03/17).

[117] M. Poel, D. Heylen, A. Nijholt, M. Meulemans, and A. van Breemen, **Gaze Behaviour, Believability, Likability and the iCat**, *AI & Society*, vol. 24, no. 1, pp. 61–73, 2009,

[118] **iCub**, `https://www.iit.it/research/lines/icub`, (visited on 2018/07/21).

[119] I. Lütkebohle, F. Hegel, S. Schulz, M. Hackel, B. Wrede, S. Wachsmuth, and G. Sagerer, **The Bielefeld Anthropomorphic Robot Head "Flobi"**, in *Proceedings - IEEE International Conference on Robotics and Automation*, 2010.

[120] **Official MDS Robot Video - First Test of Expressive Ability**, `https://www.youtube.com/watch?v=aQS2zxmrrrA`, (visited on 25/09/2019).

[121] **S2 Humanoid Head**, `http://mekabot.com/products/humanoid-head`, (visited on 2012/10/11).

[122] Robopec. (2019). **Reeti: Photo & Video Gallery**, [Online]. Available: `http://wiki.reeti.fr/display/MEDIA/Expressions?preview=/1638679/1638750/IMG_2341.JPG` (visited on 06/01/2019).

[123] **Kuri**, `https://blog.heykuri.com/`, (visited on 2017/01/30).

[124] T. Asfour, L. Kaul, M. Wächter, S. Ottenhaus, P. Weiner, S. Rader, R. Grimm, Y. Zhou, M. Grotz, F. Paus, D. Shingarey, and H. Haubert, **ARMAR-6: A Collaborative Humanoid Robot for Industrial Environments**, Nov. 2018, pp. 447–454.

[125] A. K. Pandey and R. Gelin, **A Mass-Produced Sociable Humanoid Robot: Pepper: The First Machine of Its Kind**, *IEEE Robotics and Automation Magazine*, Jul. 2018.

[126] H. Miwa, T. Okuchi, H. Takanobu, and A. Takanishi, **Development of a New Human-like Head Robot WE-4**, in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 3, Sep. 2002, pp. 2443–2448.

[127] A. van Breemen, X. Yan, and B. Meerbeek, **iCat: An Animated User-Interface Robot with Personality**, in *AAMAS '05 Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, New York, NY, USA: ACM, 2005, pp. 143–144.

[128] R. Beira, M. Lopes, M. Praça, J. J. Santos-Victor, A. Bernardino, G. Metta, F. Becchi, R. Saltarén, M. Praga, J. J. Santos-Victor, A. Bernardino, G. Metta, F. Becchi, and R. Saltaren, **Design of the Robot-Cub (iCub) Head**, in *Proceedings 2006 IEEE International Conference on Robotics and Automation, ICRA*, vol. 2006, IEEE, Jun. 2006, pp. 94–100.

[129] A. Parmiggiani, M. Randazzo, M. Maggiali, F. Elisei, G. Bailly, and G. Metta, **An Articulated Talking Face for the iCub**, in *14th IEEE-RAS International Conference on Humanoid Robots, Humanoids*, Nov. 2014, pp. 1–6.

[130] Robopec. (2019). **Reeti: Technical Details**, [Online]. Available: `http://www.reeti.fr/index.php/en/detailen` (visited on 06/01/2019).

[131] **Kuri**, `https://www.youtube.com/watch?v=4pWUPk0OrF8`, (visited on 2017/08/10).

[132] **Vicon Motion Capture Systems**, `https://www.vicon.com`, (visited on 2017/02/01).

[133] T. Weise, S. Bouaziz, H. Li, and M. Pauly, **Realtime Performance-based Facial Animation**, in *ACM SIGGRAPH 2011*, Vancouver, British Columbia, Canada: ACM, 2011, 77:1–77:10.

[134] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, **OpenFace: A general-purpose face recognition library with mobile applications**, 2016, CMU-CS-16-118, CMU School of Computer Science.

[135] W. Clark, J. Golinski, and S. Schaffer, **Enlightened Automata**, in *The Sciences in Enlightened Europe*. The University of Chicago Press, 1999, ch. 5, pp. 126–165.

[136] B. Bickel, M. Lang, M. Botsch, M. A. Otaduy, and M. Gross, **Pose-space Animation and Transfer of Facial Details**, in *Proceedings of the 2008 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, Dublin, Ireland: Eurographics Association, 2008, pp. 57–66.

[137] Y.-H. Li, Y.-T. Hu, J. Shen, M. Bogdan Preda, A. Drexler, C. Sosoiu, D. F. Stanculescu, P. Liu, and J. Ye, **Ultrafast Facial Tracker Using Generic Cameras with Applications in Intelligent Lifestyle**, *International Conference on Virtual, Augmented and Mixed Reality*, 2016.

[138] R. B. i. Ribera, E. Zell, J. Lewis, J. Noh, and M. Botsch, **Facial Retargeting with Automatic Range of Motion Alignment**, *ACM Transactions on Graphics*, vol. 36, Jul. 2017.

[139] P. Ekman and W. V. Friesen, **Facial Action Coding System: A Technique for the Measurement of Facial Movement**, Palo Alto: Consulting Psychologists Press, 1978, vol. 12, pp. 271–302.

[140] M. Sagar, **Facial Performance Capture and Expressive Translation for King Kong**, in *Sketches on ACM SIGGRAPH 2006*, New York, NY, USA: ACM, 2006, p. 26.

[141] B. Lance and S. C. Marsella, **The Expressive Gaze Model: Using Gaze to Express Emotion**, *Computer Graphics and Applications, IEEE*, vol. 30, no. 4, pp. 62–73, Aug. 2010.

[142] S. P. Lee, J. B. Badler, and N. I. Badler, **Eyes Alive**, *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 637–644, Jul. 2002.

[143] Z. Deng, J. P. Lewis, and U. Neumann, **Automated Eye Motion using Texture Synthesis**, *IEEE Computer Graphics and Applications*, vol. 25, no. 2, pp. 24–30, Mar. 2005,

[144] T. Pejsa, B. Mutlu, and M. Gleicher, **Stylized and Performative Gaze for Character Animation**, *Computer Graphics Forum*, vol. 32, no. 2, pp. 143–152, 2013.

[145] U. Pattacini, **Modular Cartesian Controllers for Humanoid Robots: Design and Implementation on the iCub**, Ph.D. dissertation, University of Genoa, Italy, IIT, 2011.

[146] H. Lehmann, U. Pattacini, and G. Metta, **Blink-Sync: Mediating Human-Robot Social Dynamics with Naturalistic Blinking Behavior**, HRI Workshop on "Behavior Coordination between Animals, Humans and Robots", 2015.

[147] Dr. Fritz Faulhaber GmbH & Co KG. (2019). **DC Micromotors 2224 Series, 2224U012SR**. Datasheet, [Online]. Available: https://www.faulhaber.com/fileadmin/Import/Media/EN_2224_SR_DFF.pdf (visited on 09/08/2019).

[148] ——, (2019). **Spur Gearheads, 20/1R Series**. Datasheet, [Online]. Available: https://www.faulhaber.com/fileadmin/Import/Media/EN_20-1R_FMM.pdf (visited on 09/08/2019).

[149] ——, (2019). **Spur Gearheads, 22/5 Series**. Datasheet, [Online]. Available: https://www.faulhaber.com/fileadmin/Import/Media/EN_22-5_FMM.pdf (visited on 09/08/2019).

[150] ——, (2019). **IE2-128: IE2 Encoder Series**. Datasheet, [Online]. Available: https://www.faulhaber.com/fileadmin/Import/Media/EN_IE2-1024_DFF.pdf (visited on 09/08/2019).

[151] Austria Microsystems AG. (2019). **AS5045: 12-Bit Programmable Magnetic Rotary Position Sensor**. Datasheet, [Online]. Available: https://ams.com/documents/20143/36005/AS5045_DS000101_2-00.pdf/3561fcce-e9f0-22a2-880f-bdfff26b93f2 (visited on 09/08/2019).

[152] ——, (2019). **AS5311: High Resolution Magnetic Linear Encoder**. Datasheet, [Online]. Available: https://ams.com/documents/20143/36005/AS5311_DS000200_2-00.pdf/575d252f-096b-e5ab-5d8f-d504fd005062 (visited on 09/08/2019).

[153] Point Grey Research Inc. (2010). **Dragonfly 2 Camera**. Datasheet, [Online]. Available: https://www.ptgrey.com/support/downloads/10258 (visited on 04/21/2010).

[154] Sony Corporation, **ICX204: Diagonal 6mm (1/3") Progressive Scan CCD Image Sensor with Square Pixel**, Datasheet, 2005.

[155] Dr. Fritz Faulhaber GmbH & Co KG. (2019). **DC Micromotors 1524 Series, 1524E009SR**. Datasheet, [Online]. Available: https://www.faulhaber.com/fileadmin/Import/Media/EN_1524_SR_DFF.pdf (visited on 09/08/2019).

[156] ——, (2019). **Spur Gearheads, 15/8 Series**. Datasheet, [Online]. Available: https://www.faulhaber.com/fileadmin/Import/Media/EN_15-8_FMM.pdf (visited on 09/08/2019).

[157] ——, (2019). **IEH2-4096: IEH2 Encoder Series**. Datasheet, [Online]. Available: https://www.faulhaber.com/fileadmin/Import/Media/EN_IEH2-4096_DFF.pdf (visited on 09/08/2019).

[158] ——, (2019). **DC Micromotors 1331 Series, 1331T006SR**. Datasheet, [Online]. Available: https://www.faulhaber.com/fileadmin/Import/Media/EN_1331_SR_DFF.pdf (visited on 09/08/2019).

[159] ——, (2019). **Spur Gearheads, 15/5 Series**. Datasheet, [Online]. Available: https://www.faulhaber.com/fileadmin/Import/Media/EN_15-5_FMM.pdf (visited on 09/08/2019).

[160] ——, (2019). **IE2-400: IE2 Encoder Series**. Datasheet, [Online]. Available: https://www.faulhaber.com/fileadmin/Import/Media/EN_IE2-1024_DFF.pdf (visited on 09/08/2019).

[161] Austria Microsystems AG. (2019). **AS5048: Magnetic Rotary Encoder (14-Bit Angular Position Sensor)**. Datasheet, [Online]. Available: https://ams.com/documents/20143/36005/AS5048_DS000298_4-00.pdf/910aef1f-6cd3-cbda-9d09-41f152104832 (visited on 09/08/2019).

[162] Ximea GmbH. (2017). **MQ042CG-CM-BRD: xiQ Camera Series**. Datasheet, [Online]. Available: https://www.ximea.com/de/produkte/anwendungsspezifische-oem-kameras/board-level-kameras/mq042cg-cm-brd (visited on 10/12/2017).

[163] Austria Microsystems AG, formerly CMOSIS. (2015). **CMV4000: 4.2 Megapixel Machine Vision CMOS Image Sensor**. Datasheet, [Online]. Available: https://ams.com/documents/20143/36005/CMV4000_DS000436_1-00.pdf/f03c3681-6a4c-e63b-b61a-7faa07088580 (visited on 02/03/2015).

[164] F. J. Romero-Ramirez, R. Muñoz-Salinas, and R. Medina-Carnicer, **Speeded Up Detection of Squared Fiducial Markers**, *Image and Vision Computing*, vol. 76, pp. 38–47, 2018.

[165] ShenZhen JSD optoelectronics co., Ltd. (2019). **JSD3428: 5MP HD Non Distotion Lens 2G2P**. Datasheet, [Online]. Available: http://www.jsdoptical.com/m/en/product/51.html (visited on 09/08/2019).

[166] S. Schulz, F. Lier, I. Lutkebohle, and S. Wachsmuth, **Robot Reality - A Motion Capture System that makes Robots Become Human and Vice Versa**, in *Proceedings - IEEE International Conference on Robotics and Automation, ICRA*, 2013.

[167] Z. Zhang, **Microsoft Kinect Sensor and Its Effect**, *IEEE Multimedia*, vol. 19, pp. 4–10, Feb. 2012.

[168] S. Schulz. (2013). **RobotReality - Wiki**, [Online]. Available: https://opensource.cit-ec.de/projects/robotreality/wiki.

[169] Point Grey Research Inc. (2012). **Firefly MV**. Datasheet, [Online]. Available: http://www.ptgrey.com/products/fireflymv/fireflymv.pdf (visited on 04/28/2012).

[170] x-IO Technologies. (2012). **x-IMU Sensor**. Datasheet, [Online]. Available: http://www.x-io.co.uk/res/doc/ximu_user_manual_4_3.pdf (visited on 04/28/2012).

[171] R. H. Y. So and M. J. Griffin, **Compensating Lags in Head-coupled Displays using Head Position Prediction and Image Deflection**, *Journal of Aircraft*, vol. 29, pp. 1064–1068, 6 1992.

[172] J. Munkres, **Algorithms for the Assignment and Transportation Problems**, *Journal of the Society for Industrial and Applied Mathematics*, vol. 5, no. 1, pp. 32–38, 1957.

[173] C. Breazeal, A. Takanishi, and T. Kobayashi, **Social Robot Embodiment**, in *Springer Handbook of Robotics*, B. Siciliano and O. Khatib, Eds. Springer, 2008.

[174] S. Schulz. (2013). **libximu - An Opensoure C++ Library Implementing Access to x-IMU Devices**, [Online]. Available: https://opensource.cit-ec.de/projects/libximu (visited on 04/09/2017).

[175] K. D. Willis, I. Poupyrev, S. E. Hudson, and M. Mahler, **SideBySide: Ad-hoc Multi-user Interaction with Handheld Projectors**, in *Proceedings of the 24th annual ACM symposium on User interface software and technology*, Santa Barbara, California, USA: ACM, 2011, pp. 431–440.

[176] R. H. Y. So and M. J. Griffin, **Head-coupled Virtual Environment with Display Lag**, in *Simulated and Virtual Realities*, K. Carr and R. England, Eds. Bristol, PA, USA: Taylor & Francis, Inc., 1995, pp. 103–111.

[177] J. Jerald, **Scene-motion- and Latency-perception Thresholds for Head-mounted Displays**, Ph.D. dissertation, 2009.

[178] R. S. Allison, L. Harris, M. Jenkin, U. Jasiobedzka, and J. E. Zacher, **Tolerance of Temporal Delay in Virtual Environments**, in *Proceedings IEEE Virtual Reality*, Mar. 2001, pp. 247–254.

[179] T. Kollenberg, A. Neumann, D. Schneider, T.-K. Tews, T. Hermann, H. Ritter, A. Dierker, and H. Koesling, **Visual Search in the (Un)Real World: How Head-mounted Displays affect Eye Movements, Head Movements and Target Detection**, in *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, Austin, Texas: ACM, 2010, pp. 121–124.

[180] Dr. Fritz Faulhaber GmbH & Co KG. (2019). **MCDC3002S: Motion Controller**. Datasheet, [Online]. Available: `https://www.faulhaber.com/fileadmin/Import/Media/DE_MCDC3002S_V2-5_DFF.pdf` (visited on 09/08/2019).

[181] ALLEGRO Microsystems. (2014). **A3950: DMOS Full-Bridge Motor Driver**. Datasheet, [Online]. Available: `https://www.allegromicro.com/~/media/Files/Datasheets/A3950-Datasheet.ashx` (visited on 07/02/2014).

[182] M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, **ROS: An Open-source Robot Operating System**, in *ICRA Workshop on Open Source Software*, 2009.

[183] G. Echeverria, N. Lassabe, A. Degroote, and S. Lemaignan, **Modular Open Robots Simulation Engine: MORSE**, in *IEEE International Conference on Robotics and Automation*, Jun. 2011, pp. 46–51.

[184] Blender Foundation. (2013). **Blender - A 3D Modelling and Rendering Package**, [Online]. Available: http://www.blender.org (visited on 08/09/2015).

[185] F. Lier, S. Schulz, and S. Wachsmuth, **Reality Check! – A Physical Robot Versus its Simulation**, in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction, HRI*, Mar. 2014, p. 331.

[186] K. E. Barrett, S. M. Barman, S. Boitano, and B. Heddwen, **Ganong's Review of Medical Physiology**, McGraw-Hill; 23rd edition, 2010.

[187] T. Kröger, **Opening the Door to New Sensor-Based Robot Applications – The Reflexxes Motion Libraries**, *Proceedings 2011 IEEE International Conference on Robotics and Automation, ICRA*, 2011.

[188] J. Wienke and S. Wrede, **A Middleware for Collaborative Eesearch in Experimental Robotics**, in *IEEE/SICE International Symposium on System Integration, SSI*, Dec. 2011, pp. 1183–1190.

[189] G. Metta, P. Fitzpatrick, and L. Natale, **YARP: Yet Another Robot Platform**, *International Journal on Advanced Robotics Systems*, p. 2006,

[190] M. Schröder and J. Trouvain, **The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching**, *International Journal of Speech Technology*, vol. 6, no. 4, pp. 365–377, Oct. 2003.

[191] C. Schmitz, **Limesurvey: An Open Source Survey Tool**, Hamburg, Germany, 2012.

[192] Maxon Motor AG. (2019). **Planetary Gear GP 13 A**. Datasheet, [Online]. Available: https://www.maxongroup.com/medias/sys_master/root/8831026757662/2018EN-323.pdf (visited on 09/08/2019).

[193] G. Cannata and M. Maggiali, **Implementation of Listing's Law for a Tendon Driven Robot Eye**, in *IEEE International Conference on Intelligent Robots and Systems, IROS*, Nov. 2006, pp. 3940–3945.