Eye-Hand Behavior in Human-Robot Shared Manipulation

Reuben M. Aronson
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
rmaronson@cmu.edu

Thiago Santini
University of Tübingen
Tübingen, Germany
thiago.santini@uni-tuebingen.de

Thomas C. Kübler
University of Tübingen
Tübingen, Germany
thomas.kuebler@uni-tuebingen.de

Enkelejda Kasneci
University of Tübingen
Tübingen, Germany
enkelejda.kasneci@uni-tuebingen.de

Siddhartha Srinivasa
University of Washington
Seattle, Washington, USA
siddh@cs.uw.edu

Henny Admoni
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
henny@cmu.edu

ABSTRACT

Shared autonomy systems enhance people's abilities to perform activities of daily living using robotic manipulators. Recent systems succeed by first identifying their operators' intentions, typically by analyzing the user's joystick input. To enhance this recognition, it is useful to characterize people's behavior while performing such a task. Furthermore, eye gaze is a rich source of information for understanding operator intention. The goal of this paper is to provide novel insights into the dynamics of control behavior and eye gaze in human-robot shared manipulation tasks. To achieve this goal, we conduct a data collection study that uses an eye tracker to record eye gaze during a human-robot shared manipulation activity, both with and without shared autonomy assistance. We process the gaze signals from the study to extract gaze features like saccades, fixations, smooth pursuits, and scan paths. We analyze those features to identify novel patterns of gaze behaviors and highlight where these patterns are similar to and different from previous findings about eye gaze in human-only manipulation tasks. The work described in this paper lays a foundation for a model of natural human eye gaze in human-robot shared manipulation.

KEYWORDS

human-robot interaction, eye gaze, eye tracking, shared autonomy, nonverbal communication

ACM Reference Format:
Reuben M. Aronson, Thiago Santini, Thomas C. Kübler, Enkelejda Kasneci, Siddhartha Srinivasa, and Henny Admoni. 2018. Eye-Hand Behavior in Human-Robot Shared Manipulation. In *HRI '18: 2018 ACM/IEEE International Conference on Human-Robot Interaction, March 5–8, 2018, Chicago, IL, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3171221.3171287

Figure 1: Eye gaze detected in real time through a worn eye tracker can reveal human mental states and guide robot assistance in human-robot shared manipulation tasks.

1 INTRODUCTION

To be useful partners, robots that collaborate closely with people must predict their partners' goals or identify when they are having trouble with a task. *Human-robot shared manipulation* represents a special case of collaboration, in which people and robots jointly control the same robotic end effector. Because it is an especially close collaboration, human-robot shared manipulation particularly benefits from building this understanding.

One key application of human-robot shared manipulation is for physically assistive robots. Assistive robots, particularly wheelchair-mounted robot arms, provide a flexible, mobile, highly dexterous tool for performing activities of daily living without a caregiver's assistance. These robots are typically teleoperated by their user through a joystick or other input device, enabling users to grasp and move objects. For example, people can use the robot arm to eat by spearing bites of food on a fork in the robot's hand (Fig. 1).

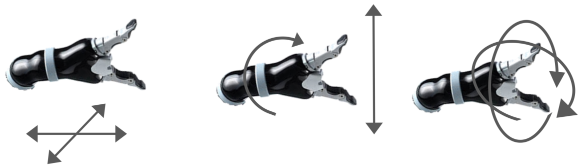
Assistive robot control has been improved with *shared autonomy* algorithms, which reduce the amount of control people need to exert by 1) predicting people's goals and 2) taking assistive actions toward those goals [5, 27, 28, 33]. In order to accurately predict people's goals, the system must use a prior understanding of how humans behave while controlling a robot. State-of-the-art shared autonomy methods (e.g., [5, 33]) rely exclusively on monitoring direct user input to the robot system. That is, the system's predictions of user goals are based on input signals from a synthetic interface such as a joystick. In this paper, we show that humans generally follow some predictable patterns while providing joystick control.

In addition, we explore another source of information for goal recognition: eye gaze. People constantly and automatically communicate their goals, future actions, and mental states through their eye gaze behavior [9, 26, 31, 35, 46, 65]. If we can leverage these unconscious, natural, revealing signals of intent, we can improve the ability of shared autonomy systems to understand how to assist.

Using eye gaze for shared autonomy algorithms requires understanding gaze patterns during the interaction. The psychological literature has characterized gaze behavior of people performing

HRI '18, March 5–8, 2018, Chicago, IL, USA

ACM ISBN 978-1-4503-4953-6/18/03…$15.00
https://doi.org/10.1145/3171221.3171287



(a) x-y mode **(b) z-yaw mode** **(c) pitch-roll mode**

Figure 2: In this study, the Kinova MICO was operated in six DOFs with a two-DOF joystick. Fewer input DOFs means mode switching is required to control the robot.

certain manipulation tasks with their own hands, like moving objects around obstacles [35] or making tea [47]. These studies show that gaze follows the objects involved in the task [26] and that eye gaze precedes hand motion [35, 46]. However, there is little information about eye gaze while operating a robot arm. In this paper, we provide novel insights into the dynamics of eye gaze in human-robot shared manipulation tasks. As the results show, human-robot shared manipulation elicits distinct gaze patterns.

This paper's contribution follows these three steps:

(1) We conduct a data collection study to record eye gaze during a human-robot shared manipulation task, both with and without shared autonomy assistance (Section 3),
(2) We apply signal processing algorithms to the data to extract eye gaze and task features (Section 4), and
(3) We analyze those features to identify control and gaze behavior during human-robot shared manipulation (Section 5).

The resulting data supports several conclusions about control and gaze behavior in this interaction. People's joystick use is heavily axis-aligned and depends on the assistance mode. People's pupil sizes also increase when they operate the joystick, suggesting that pupil response is linked to task activity. In addition, we find a distinct pattern of visual monitoring of the robot end-effector during translation and not rotation. Finally, our results show that gaze patterns when people are involved in the shared manipulation task differ from when people watch a robot perform the task without input. This work provides the basis for new models of human behavior, particularly using eye gaze, during human-robot shared manipulation. These insights can help us build gaze-responsive shared autonomy to improve robots' assistance capabilities.

2 RELATED WORK

Shared autonomy for assistive robots. The current work is motivated by improving shared autonomy for physically assistive robots. Activities of daily living, such as pouring a glass of water or picking up a bite of food, present significant challenges to people with upper motor disability. Robot manipulators such as the Kinova JACO and MICO [39], and the Exact Dynamics iArm [17] provide a platform for accomplishing these tasks using simple input devices.

Assistive robots can be challenging to control, however, because positioning the end effector in 3D space requires managing six degrees of freedom (DOFs). Controlling this many DOFs is a challenge, especially for people with motor disabilities. Input devices may be limited to a two-DOF joystick or even a single DOF through a sip-n-puff device or a head array switch. To translate low-dimensional input into high-dimensional robot control, users must toggle through control modes to access their desired degree of freedom (Fig. 2), increasing the time and cognitive load required to complete a task.

Shared autonomy algorithms address this control problem by trying to predict the operator's goal, then providing robotic assistance—combined with existing user control—to complete that goal [5, 27, 33]. To do so, shared autonomy systems use models of human behavior to try to predict the user's goal from their control input. Shared autonomy algorithms can reduce the amount of time required to grasp an object [33] and can automatically put the robot in the correct control mode for the user's current goal [27]. Some research has proposed methods for gaze-responsive shared autonomy [3], but this work is in its earliest stages.

Natural gaze in psychology and HRI. Psychologists have established that eye gaze is an important part of human-human interactions, revealing people's mental states and facilitating cooperation [6, 7, 13, 38, 41]. Eye gaze is especially useful for communicating about objects and locations in the environment. People look at objects one second or less before referencing them [22, 24, 69], and listeners can use this gaze communication to predict their partner's reference [9, 61] and disambiguate unclear language [1, 24]. When people act to manipulate objects, their gaze typically reaches the target before they even begin moving their hands [46], and shifts to the next target before their hands reach the current target [35]. Objects that are not task-relevant are rarely gazed at [26]. This work aims to identify whether similar gaze patterns exist during human-robot shared manipulation.

There is a robust body of research investigating eye gaze in human-robot interactions [2]. In the domain of manipulation, researchers have developed robot behavior models that use eye gaze to predict human intentions during collaborative object manipulations [56] and handovers [23, 62]. Robot motion planners can optimize trajectories so that objects remain visible to people during handovers [60] and teleoperation [28], improving the effectiveness of the robot's actions. Nevertheless, gaze behavior during comanipulation remains unexplored.

Eye tracking as input signal. By combining eye tracking technology with machine learning, multiple works have identified users' activities and actions based on eye movements, both in post-experimental analysis (e.g. [11, 18, 19]) and online (e.g., [10]). HRI researchers have used head-mounted eye trackers to guide robot behaviors that are sensitive to real-time human gaze, such as cooperative and responsive gaze behaviors during conversation [67, 68], natural gaze during game-playing [51], and anticipatory assistive actions from a manipulator robot [30]. In driver activity recognition during conditionally autonomous driving, eye tracking is used in real time to trigger assistive driving systems [10].

Eye tracker data can also reveal mental states like cognitive load. There is a well-studied correlation between pupillary dilation and cognitive factors such as workload [12], fatigue [34], surprise [42], attention [16], and emotional arousal [21]. This effect is called the task-evoked pupillary response (TEPR), which denotes the change in pupil diameter that is caused by a specific task [8]. In this work we analyze pupil sizes during various phases of shared manipulation. The main challenge when using pupillary information is the noise in the pupillary signal, which arises due to changes in the illumination or other persistent conditions; various methods have been presented to compensate for these variations [37, 45, 49, 52].

3 DATA COLLECTION STUDY

We begin by collecting gaze and input data during a human-robot shared manipulation task. We conducted a data collection study to investigate eye gaze during human-robot shared manipulation with and without assistance. Details about this data collection can be found in prior work [32], but we summarize them here.

3.1 Design

We designed a within-subjects study to simulate eating, an important activity of daily living. Participants sat in front of a plate containing three bite-sized morsels of food (marshmallows and cake, chosen for their spearability) and were asked to spear one of the three morsels using a fork held in the end-effector of a Kinova MICO robot (Fig. 1). To do so, participants maneuvered the robot above their desired piece of food using a 2-axis joystick. They then pressed a button that prompted the robot to autonomously lower the fork, spear the food, and serve it to the participant.

Because many degrees of freedom (DOFs) are required to complete this task, the robot was operated in modal control, where each mode corresponded to two DOFs operated by the two joystick axes (Fig. 2). One control mode moved the robot in the x and y directions along a single plane parallel to the table; a second control mode moved the robot in the z direction and also controlled the yaw rotation of the end effector; a third mode controlled the pitch and roll orientations of the end effector.

Participants completed the task multiple times under four levels of **robot assistance**:

(1) *Teleoperation*: participants fully controlled the robot using the joystick with no assistance
(2) *Autonomy*: the robot autonomously selected one morsel at random and speared it without participant intervention
(3) *Shared autonomy*: the robot attempted to predict the participant's target morsel and assist toward retrieving that morsel using a state-of-the-art shared autonomy framework [33]
(4) *Blend*: the robot and human provided separate control inputs which were combined through an arbitration function based on the robot's confidence [15].

Condition (1) mimics how current users of the Kinova MICO robot primarily interact with their device [48]. Other than some preprogrammed motions, people generally control the MICO directly using the same interface they use to drive their powered wheelchair.

Condition (2) represents the opposite end of the assistance spectrum: the robot acts completely autonomously. In this mode, the robot used a wrist-mounted depth camera to identify morsel positions automatically, then plans and executes a path to the target.

In the two assistance conditions (3 and 4), the robot combines the user's joystick control with some autonomous control based on predictions of the user's intent. The shared autonomy method (condition 3) models the user as a partially observable Markov decision process (POMDP), where user goals are the latent states and joystick inputs are the observations [33]. The robot then assists toward goals by solving the POMDP for the optimal action using hindsight optimization. The blend method (condition 4) calculates an autonomous robot policy and blends it with the human's joystick input based on the robot's confidence. Until the robot is within a confidence threshold of a morsel, the robot provides no assistance, so control effectively replicates teleoperation.

We collect gaze using the Pupil Labs Pupil [53], a head-mounted eye tracker, which consists of two cameras worn on a glasses-like frame. An IR camera records the eye, and a forward-facing camera records the world. Software provided by the eye tracker locates the pupil position in the frame of the eye camera and matches it to a pixel location in the world camera through a degree-2 polynomial mapping, calibrated by having the user look at specific points.

3.2 Procedure

We recruited 24 able-bodied participants from the local community (11 male, 13 female, ages 19 to 59). Participants were compensated \$10 for their participation. One participant was excluded from the final analysis for failure to follow directions.

First, participants were instructed on how to control the robot and given about 5 minutes to practice, in order to reduce the effect of novelty. Then, participants completed five trials under each level of robot assistance, for a total of twenty trials. All five trials of one assistance condition were completed sequentially, and the order of trials was fully counterbalanced across the participant pool. Each trial lasted between 30 seconds and 6 minutes, depending on user success at positioning the fork. The eye tracker was individually calibrated at the beginning of the study and recorded participant eye gaze during each trial. Between each trial, the robot was reset to a constant starting position (about 30 cm above the plate).

4 ANALYSIS

We process the joystick control signals for analysis. We also filter and process the raw eye tracker data to extract meaningful features.

4.1 Processing Joystick Input Data

To characterize joystick use, we performed the following featurization. First, we isolated periods of active joystick use by finding stretches of time during which the magnitude of the joystick input remained above $\epsilon = 0.0001$ and fusing stretches 0.1 seconds or less apart. Since joystick motion is highly axis-aligned (see 5.1), we labeled each isolated **joystick operation** with its primary direction of motion (forward/back or left/right) and its corresponding robot twist direction (x, y, z, pitch, yaw, or roll).

4.2 Filtering Gaze Data

Gaze data quality depends heavily on the initial calibration, the position of the eye tracker over time (i.e., slippage), and individual user characteristics such as eye lashes, eyelid shape, and makeup. We collected data from a wide variety of participants, some of whom yielded high quality gaze data and others who did not. In order to analyze only high quality gaze data, we established filtering criteria.

First, we excluded any gaze point that the Pupil Labs eye tracker detected with less than 60% confidence, as recommended by the vendor. Next, we defined an *extended calibration rectangle* by taking the smallest bounding rectangle enclosing all calibration points and increasing its dimensions by 25% in each direction. Gaze points near the calibration points (especially those within the convex hull of calibration points) are likely to be the most accurate; outside the extended rectangle, the extrapolation is less reliable. In our analysis, we included only trials with at least 80% of gaze points within the extended calibration rectangle. This filtering process left us with 36% of the original trials. While our reasonably stringent requirements for quality led to a significant reduction in data, this still represents 95 minutes of data from 155 trials with 16 participants.

4.3 Extracting Gaze Features

Using the eye tracker, we collected the following data at 30Hz: (1) raw world camera images, (2) the pixel location of the gaze position in the world camera image (with detection confidence) and (3) pupil position and shape ellipse in pixels (with detection confidence). These data can be processed to extract spatio-temporal features of gaze such as fixations, saccades, and smooth pursuits [37].

Visual **fixations** maintain the focus of gaze on a single location. Fixation duration varies based on the task, but one fixation is typically $100 - 500$ms, although they can be as short as 30ms [29]. **Saccades** are rapid, ballistic eye movements (usually between $20 - 200$ms) that abruptly change the point of fixation. They range in amplitude from small movements made while reading to much larger movements made while gazing around a room. Saccades can be elicited voluntarily, but they occur reflexively whenever the eyes are open, even when fixating on a target. **Smooth pursuit** movements are slower tracking movements of the eyes that keep a moving stimulus on the fovea. Such movements are voluntary in that the observer can choose to track a moving stimulus, but only highly trained people can make smooth pursuit movements without a target to follow. **Vestibulo-ocular movements** stabilize the eyes relative to the external world to compensate for head movements. These reflex responses prevent visual images from slipping on the surface of the retina as head position changes.

4.3.1 Stabilizing vestibulo-ocular movements. We stabilized the videos to compensate for vestibulo-ocular movements. Since participants' heads were not stationary during the trials, these movements (head movements with the eyes fixed) appeared identical to smooth pursuits (eye movements with head fixed), since each contain smooth motion of the focal point relative to the head frame.

To counteract this effect, we performed *ego-motion compensation* [40]. Unlike [40], we used feature-based video stabilization, which is more reliable when moving objects (i.e., the robot) are present. Throughout the task, a scale-rotation-translation transformation between adjacent frames was detected, using FAST feature points [54] and HOG features [14] to determine correspondences, and implemented using MATLAB [50] built-in routines. The transformation was detected using MSac [63], and transforms with more than 10 inliers were accepted. Gaze points were then transformed to a common reference frame, and these stabilized gaze points were used for subsequent analysis. Over all trials, 99.96% of frames had stable transformations. The presence of stabilization reduced the rate of pursuit detection from 15.1% to 14.0% of all clusters, indicating that it likely compensated for some vestibulo-ocular motion.

4.3.2 Extracting fixations, saccades, and smooth pursuits. Though eye movements can be extracted online [36, 59], here we used an offline multistage detection approach, which allowed us to account for only high-quality gaze signals. Given that all subjects were positioned roughly at the same distance to the robot arm, no subject-specific parameter optimization was performed.

The approach starts by identifying fixation candidates based on a Dispersion Threshold Identification (I-DT) [57] filter with a minimum *duration* (dur_{min}) of 80ms and maximum *dispersion* (dis_{max}) of 25 pixels as implemented by Eyetrace [44].

Although fixations are identified reliably, smooth pursuits are clustered in multiple adjacent fixations as the dispersion threshold is continually exceeded during motion. Thus, adjacent fixation candidate clusters are merged if the dispersion between their adjacent gaze positions does not exceed dis_{max}. Resulting clusters are then classified as

$$\begin{cases} \text{smooth pursuit} & \text{if } ED > 2 * dis_{max} \text{ ;} \\ \text{fixation} & \text{otherwise,} \end{cases}$$

where *ED* is the Euclidean distance between the first and last point in the cluster; this approach favors a more robust detection of longer pursuits to the detriment of shorter ones.

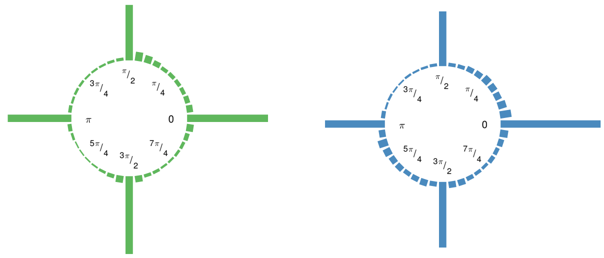
This process might merge fixations at the beginning or end of smooth pursuits into the pursuit cluster. Therefore, smooth pursuit clusters are analyzed with a second I-DT pass ($dur_{min} = 300$ ms, $dis_{max} = 5$ pixels) to separate and reclassify such fixations correctly. Remaining non-classified gaze points are set to

$$\begin{cases} \text{saccade} & \text{if } \textit{pupil confidence} > 0.6 \text{ ;} \\ \text{noise} & \text{otherwise.} \end{cases}$$

4.3.3 Identifying gaze points in the world. In general we did not attempt to identify the target of gaze points in the world, for two reasons. First, we want to focus our analysis on the *dynamics* of gaze, using features that can be extracted online from gaze movements. There is evidence that gaze dynamics alone can reveal much about an interaction, and our results support this observation. Second, our study was not set up to reliably recognize the real-world targets of gaze. To robustly identify the relevant objects in the scene, either a detector would need to be trained or the videos would need extensive manual coding, which was beyond the scope of this project. In the future, fiducial markers can simplify this detection problem in a constrained environment; robust object detection in unconstrained environments is an active topic of computer vision research.

However, we did manually classify glances to the goal morsel. A fixation was labeled a morsel glance if it had a clear preceding saccade towards the morsel, and a clear following saccade away from the morsel. Distinguishing morsel glances was difficult when the robot operated very close to the plate, as eye motion was small and hard to distinguish from noise. In these cases, morsel fixations were labeled only when the robot was not also moving in the direction of the gaze motion and when independent fixations were detected. Thus, morsel glances were identified conservatively; it is possible that the coding scheme underestimated the actual number of morsel fixations that occurred. Morsel glances were only coded for conditions when robot assistance was always on (shared autonomy) or never on (teleoperation).

Morsel glances were first performed with an automated heuristic and then confirmed by manually checking all of the heuristic's



(a) Teleoperation **(b) Shared autonomy**

Figure 3: Histogram of joystick control signal direction across all trials, separated by robot assistance condition. Joystick control was highly axis-aligned.

classification against videos of the gaze points. A second coder manually labeled 10% of the data (randomly selected); the resulting Cohen's $\kappa = 0.9415, p < 0.001$ indicates high inter-rater reliability.

5 RESULTS

Understanding how people operate robots can provide vital insights that enable better goal recognition for shared autonomy. Using the data we collected (Section 3), we first establish some facts about how people use the joystick (Sections 5.1 and 5.2). Then, we analyze pupil size (Section 5.3) and scanpaths (Sections 5.4 and 5.5) to draw insights about how gaze patterns reveal aspects of the interaction.

5.1 Joystick Control is Axis Aligned

The joystick in this study provides two-axis control in three modes: x-y mode, z-yaw mode, and pitch-roll mode (Fig. 2). People's joystick control was strongly aligned to the cardinal directions of movement. Even though participants could use the joystick to control two degrees of freedom simultaneously by pushing along a diagonal (thus, for example, moving the robot in x and y at the same time), people rarely moved the joystick in anything but a cardinal direction. Fig. 3 shows the direction and magnitude of joystick control in teleoperation and shared autonomy conditions. There is slightly greater variance in the joystick directions in shared autonomy than in teleoperation. (Standard deviations for shared autonomy: +x: 0.27, +y: 0.30, -x: 0.28, -y: 0.29; SD for teleoperation: +x: 0.25, +y: 0.24, -x: 0.22, -y: 0.19; all in radians.)

5.2 Robot Assistance Affects Joystick Use

There are six possible joystick control directions (x, y, z, roll, pitch, and yaw). During shared autonomy, most of people's input to the joystick was in x and y directions, whereas during teleoperation, people's joystick control inputs were more uniformly distributed (Fig. 4). To identify whether the distribution of direction inputs is different between robot assistance conditions, we conducted a χ^2 test of homogeneity. We found that there is a significant difference in the frequencies of each control direction between shared autonomy and teleoperation conditions ($\chi^2(5) = 23.376, p = 0.032$).

This is not terribly surprising: in shared autonomy, the robot's assistance took care of much of the z and rotation movements that people had to handle themselves in teleoperation mode. Selecting a particular morsel then became a matter of positioning the robot along a 2D plane by moving it in the x and y directions. This result, while unsurprising, underscores the fact that human behaviors are different during shared autonomy and teleoperation modes.

5.3 Pupil Size Increases During Joystick Use

To understand how people are behaving while they operate a robot, it can be valuable to monitor their real-time cognitive load. One available metric is pupil size, which several studies [8, 12, 49] demonstrate is correlated with a person's cognitive effort. We analyzed participants' pupil sizes while operating the robot.

First, we found that people's pupil sizes varied by what assistance the robot was providing. A one-way repeated measures ANOVA tested the effect of assistance mode on pupil size, independent of joystick control, and found a significant effect ($F(3, 18) = 7.774, p = 0.002$). Post-hoc analysis with Bonferroni correction revealed no significant pairwise differences, though pupils were smaller in the autonomous mode than in blend mode at a marginal level ($p = 0.061$). This discrepancy between significance of main effects and pairwise analysis is likely due to the low numbers of unfiltered trials in this analysis; the relatively few acceptable trials and statistically insignificant effects require confirmation in future experiments. Nevertheless, examination of pupil size averages (Fig. 5) suggests that pupils were generally smaller in the autonomous condition than any other condition, which could reflect the increase cognitive load required when controlling the robot at all.

We also found that people's pupils were larger while they were controlling the joystick than while they were not (Fig. 5). This effect held across teleoperation, shared autonomy and blend assistance conditions. (The autonomous condition had no joystick control, so it was omitted from this analysis.) A two-way repeated measures ANOVA tested the effect of assistance mode (teleoperation, shared autonomy, or blend) and joystick actuation (on or off) on average pupil size. There was a significant effect of joystick actuation ($F(1, 8) = 9.231, p = 0.016$), but not assistance mode ($F(2, 16) = 0.434, p = 0.655$). The interaction effect was not significant, though it was marginal ($F(2, 16) = 3.180, p = 0.069$). Thus, we propose that the user's increased pupil size while providing actual control input reflects a higher cognitive load, though we look forward to confirmation with more data.

The pupil size effect, while significant, involves very small differences. In prior work, pupil sizes changed about 0.1mm for a 4mm pupil, a 2.5% increase during cognitive load [8]. We see a similar 2.5% increase in pupil sizes when the joystick was engaged relative to pupil size during autonomous trials that had no joystick input.

Pupil sizes can also be affected by other factors such as ambient lighting or personal eye characteristics. This study involved a naturalistic interaction, so we did not rigidly control visual stimuli during each trial. However, participants experienced each condition sequentially, so the ambient lighting varied very little between trials. Since the order of conditions was counterbalanced, we do not believe there was a systematic effect of ambient lighting on pupil size. Furthermore, because we perform the analysis as a repeated measures test, we account for any systematic personal characteristics like emotional state or natural pupil size variability.

In addition, task-based factors other than cognitive effort may have influenced pupil size. For example, because the eye tracker reports pupil sizes as visible pixels in the eye image, it may be that

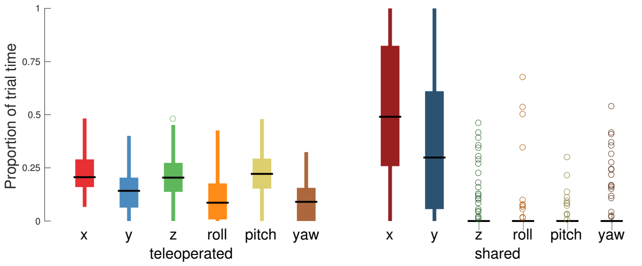


Figure 4: Proportion of active control time spent actuating the joystick in each control direction by robot assistance type (teleoperation versus shared autonomy).

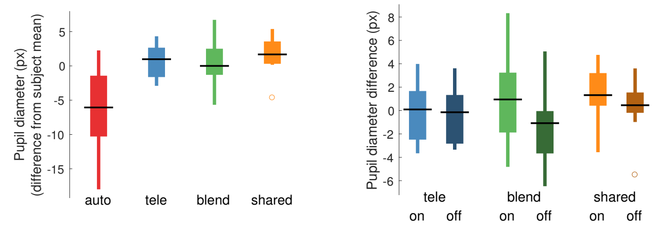


Figure 5: Pupil sizes by (a) robot assistance mode and (b) the presence of joystick actuation. Pupil sizes are systematically larger during joystick actuation. The *autonomous* assistance mode is omitted from (b) since there is no joystick control.

pupil size differences reflect peripheral versus central gaze. Another possibility is that blink behaviors occurred more frequently during non-joystick control periods, and that the eyelid partially occluded the pupil during blinks which led to smaller sizes detected. To support the cognitive load hypothesis over the other possibilities, future data collections should strictly control for lighting, blinks, and other confounds. Future data collection should also supplement pupil data with well-established subjective measures of pupil size, such as the NASA TLX [25], to further establish their validity.

5.4 People Use Visual Feedback For Alignment

By analyzing a participant's eye behavior during a single trial, we can identify persistent patterns of gaze behavior. Fig. 6 shows the vertical position of the participant's gaze during a teleoperation trial after stabilization. Since the robot remains above the plate for the duration of the trial, vertical gaze position is suggestive of what object the participant is looking at: the participant moves their gaze up to look at the robot and down to look at the plate. Morsel glances (see Section 4.3.3) are circled.

The participant begins by glancing at the morsel (at 1s) before moving the robot at all. Then, the participant performs rough positioning in x and y (1s-6s). At 7s, the participant begins moving the robot down (in z) to just above the plate, and we see a clear pattern wherein the participant alternates looking at the morsel and at the end-effector, likely to monitor their distance visually. Then, from 12 s to 24 s, the participant rotates the end-effector and looks only at different parts of the robot (presumably to check for internal collisions). Next (24s-33s), the participant performs a fine positioning step in x and y, with repeated glances between the end-effector and the morsel to ensure alignment. Finally (34s-36s), the participant does some last minor adjustments (with gaze patterns too small to be distinguished by the fixation classifier) and triggers the autonomous spearing action (41s).

We can identify several patterns during this process. First, we distinguish between two types of morsel glances: planning and monitoring. **Monitoring glances** are plate glances that occur *during* joystick control operations, while the joystick is engaged. **Planning glances** are plate glances that occur *between* joystick control operations, when the joystick is not engaged.

We found that people perform *planning* glances to the morsel before initiating motion, as at the beginning in Fig. 6, in 76% of trials. This gaze behavior accords with observations during human manipulation: people saccade to a manipulation target before their hand begins moving [35, 46]. However, we failed to find regular *planning* glances in a one-second window before each manipulator movement. For both teleoperation and shared autonomy conditions, only a small proportion of the one-second windows preceding joystick control contained any planning glances to the plate (14% for teleoperation and 13% for shared autonomy). Thus, we see overall planning behavior, but individual motions do not exhibit the same effect.

A second pattern of gaze behavior suggested by Fig. 6 is that people use visual feedback (*monitoring* glances) mainly during translation. This pattern holds across all teleoperation trials: a one-way ANOVA showed a significant effect of joystick control direction (x, y, z, pitch, yaw, roll) on frequency of monitoring glances ($F(5, 144) = 4.5, p < 0.001$). A post-hoc analysis revealed that people displayed significantly more plate-monitoring glances while operating the joystick in the y direction than roll ($p = 0.035$), pitch ($p = 0.018$), or yaw ($p = 0.002$) directions. Similarly, people displayed significantly more plate-monitoring glances while operating the joystick in the x direction than the yaw ($p = 0.029$) direction. The inferential analysis was performed only for the teleoperation condition; in shared autonomy, there were almost no instances of joystick control other than in x-y mode (Section 5.2), so the analysis was not performed. The frequency of morsel glances by robot motion direction are shown in Fig. 7.

This pattern of people using visual feedback is further established by examining the frequency of *repeated monitoring* glances, instances of operation during a single mode that contain two or more *monitoring* glances. Fig. 6 shows two examples of these repeated glances (shaded in the figure): from 7s to 14s, the participant repeatedly looks at the morsel to monitor the distance between it and the end-effector, and from 24s to 33s, the participant checks the x-y position of the morsel while aligning the end-effector in that plane. Across the data, we find this pattern occurs more often during translation than rotation. Specifically, a one-way ANOVA for the effect of joystick control mode (x-y, z-yaw, pitch-roll) on frequency of repeated monitoring glance sequences (length ≥ 2) finds a significant difference ($F(2, 16) = 7.810, p = 0.004$). A post-hoc test with Bonferroni correction revealed that repeated monitoring glances occurred significantly more often in x-y mode than in pitch-roll mode (mean difference = 0.292, $p = 0.015$). However, the absolute frequency of modes with repeated glances is less than half in any control mode. We note several possible reasons for this effect. First, during coarse motion far from the morsel, repeated visual feedback may not be necessary since perfect alignment is

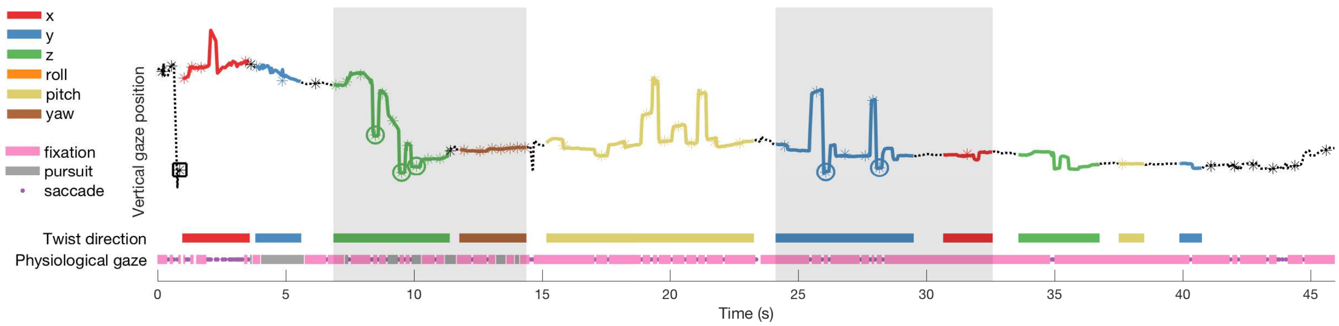


Figure 6: Vertical position of gaze points in the world image over time from a representative trial. *Twist direction* colors indicate which DOF is being controlled by the participant through the joystick; *physiological gaze* colors and dots indicate detected fixations, smooth pursuits, and saccades. Plate glances are outlined with either a black square (planning glance) or colored circle (monitoring glance). Shaded sections highlight two examples of repeated monitoring glances.

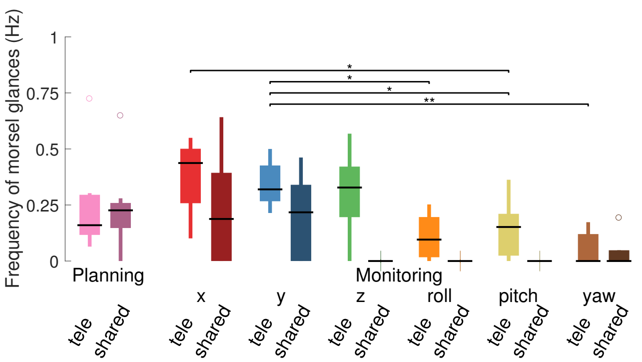


Figure 7: Mean frequency of planning and monitoring glances to the plate during each robot assistance mode. Monitoring glances are subdivided by joystick control direction. * indicates significance at the $\alpha = 0.05$ level; ** at $\alpha = 0.01$.

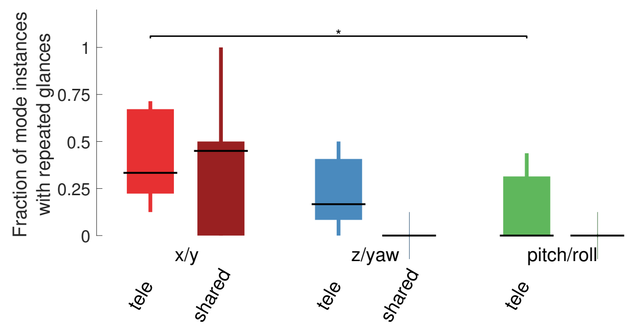


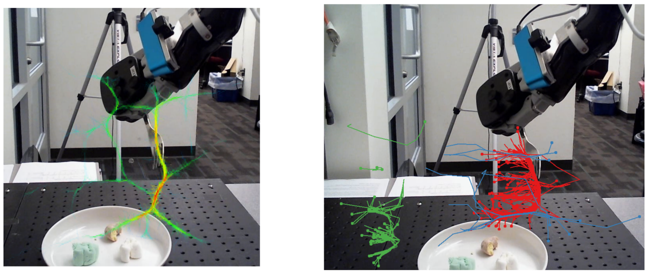
Figure 8: Proportion of joystick control sequences of the same mode that contained multiple ($\geq$ 2) monitoring glances, subdivided by their control mode. * indicates significance at the $\alpha = 0.05$ level.

not required (in Fig. 6, the participant uses a single *planning* glance before the first x-y motion but does not monitor during the process). Second, when the robot is operating very close to the morsel, two effects may occur: the end-effector tip and morsel may be too close to distinguish separate fixations, as in Fig. 6 near 35 s, and the participant may be using peripheral vision for feedback and not using separate glances at all. Thus, while we see that this pattern occurs more often in translation than in rotation, users need not deploy it consistently during robot operation. In addition, when we analyzed repeated monitoring glances by *individual* joystick directions, rather than 2-axis mode, we failed to find a significant effect. A one-way ANOVA testing the effect of joystick direction (x, y, z, roll, pitch, yaw) on frequency of repeated monitoring glances was not significant ($F(5, 14.110) = 1.115, p = 0.348$, with Greenhouse-Geisser correction because sphericity assumption is violated). Qualitatively, it appears that there are many examples of people switching between x and y directions during sequences of repeated monitoring glances.

5.5 Scanpath Predicts Assistance Condition

Repetitive patterns, such as the above-mentioned plate glances or pursuits of the robot trajectory, represent behavioral strategies. In addition to manually identifying patterns, we performed automatic classification of one-second-long sequences of eye movement patterns to discern their discriminative capability [43]. Showing that these subsequences can be distinguished by assistance mode further demonstrates that eye gaze provides a rich source of information about the operator's intentions.

We first establish a metric of similarity between the sequences. We used an edge bundling approach [64] that allows us to visually adjust the clustering strength (Fig. 9a). This is an extension of the mean-shift clustering approach for trajectory data and shifts all trajectories towards the neighboring area of highest trajectory density. Data of all participants and trials was clustered jointly, and the number of iterations and advection speed was adjusted manually as shown in Figure 9a. Next, we performed a k-means assignment of trajectories to pattern clusters. Occurrence frequencies of cluster representatives were used as features in a k-nearest neighbor classifier (with $k = 3$) after feature selection using a *fast correlation-based filter* [70]. Three out of 200 such patterns were found to be most discriminative between autonomous and teleoperation conditions (Fig. 9b) and allowed for a classification accuracy of 83% in a leave-one-out cross-validation. Interestingly, the discriminative patterns seem to identify transitions between the table and an upper location, probably the robot arm, during the critical aiming phase. Because this analysis was conducted on non-stabilized data, the clusters



(a) Edge bundling

(b) Discriminative patterns

Figure 9: Certain clusters of one-second-long gaze trajectories could discriminate between autonomous and teleoperation conditions with 83% accuracy. Dots mark the beginning of a trajectory as direction might be relevant.

may correspond to the same gaze transitions between end-effector and table as discussed above (red and green in Fig. 9b), though the clusters are translated in the image frame by head rotation. Because of this head rotation, the position of the clusters in Fig. 9b does not match the table position in the chosen representative image, but does so in the video at the time that they occurred.

Since it only used trajectory frequencies, not trial duration or robot movement, the . method of extracting gaze trajectories employed here could be used in continuous real-time scenarios. Future work could train a classifier to infer user intentions from these pure motion trajectories. The classification here further demonstrates the richness of gaze information within the domain of human-robot shared manipulation.

6 DISCUSSION

The biggest challenge of the gaze portion of our analysis was poor *data quality*. As described in Section 4.2, we established filtering criteria based on the relative position of gaze points to known calibration points, which aimed to yield reliable gaze information. Data quality is affected by many factors, including number and breadth of calibration points, individual differences in eye shape, and positioning of the eye camera relative to the eye. Participants in this study were only calibrated once at the beginning of the trial, which affected the consistency of gaze data over time. Furthermore, the low-cost eye tracker we used for this study was monocular and had a fixed eye camera, so if the eye tracker was positioned poorly for the user, the data quality was inherently poor. Our experimentation with an adjustable, binocular eye tracker (acquired after this study) indicate more reliable tracking behavior, suggesting that one main issue was that the fixed camera position worked well for some users and poorly for others. By retaining only the users for which the positioning and calibration was reliable, we are confident that data used for analysis is valid. Additional strategies for ensuring higher data quality in the future include: 1) more sophisticated calibration techniques (e.g., [58]), 2) a wider-angle world camera to capture the complete robot arm at all times, and 3) higher sampling rates, which improves the detection and characterization of the eye movements.

Eye trackers only report central gaze points, so our data doesn't reveal if and when people use their *peripheral vision* to complete a task. However, people do use peripheral vision and memory to accomplish manipulation tasks—like when we eat bites of food while reading on our phones—particularly as expertise with the task increases [4, 35, 55]. Thus, our participants may have used memory or peripheral vision to monitor the plate's position, even without making a saccade to it. Current eye tracking techniques are insufficient to detect such peripheral gaze, but future studies can employ task manipulations (e.g., gaze locking [35]) to reveal the effect of peripheral gaze during shared autonomy manipulation.

One factor we did not control for is *expertise*. We expect, for example, that people's joystick control and mode switching behavior may change as they become more accustomed to the robot. In addition, studies show that gaze patterns change as people gain expertise in a task [20, 55, 66]. For example, novices' gaze tends to lag behind cursor position in a screen-based task, whereas experts' gaze leads cursor positions [55]. The practice period given to each subject before data collection reduced the effect of novelty within this study, but we expect that as a user learns to more fluently control the robot, their gaze patterns will change. We look forward to evaluating the effect of expertise in this scenario by bringing participants back for repeated sessions with the robot.

7 CONCLUSION

The goal of this paper was to learn about human control and gaze behavior during human-robot shared manipulation. To accomplish this task, we first performed a study in which participants used an assistive robot manipulator to perform an activity of daily living at several different assistance levels. We recorded people's gaze with a monocular eye tracker as they used the robot in all modes.

We processed gaze data from this study using state-of-the-art signal processing algorithms to extract features like fixations, saccades, and smooth pursuits. We stabilized the world-facing videos to account for vestibulo-ocular movements and manually coded plate-directed glances.

Our analysis of the data revealed that people behave in distinct patterns. People's commands to the robot were heavily axis-aligned. People's pupil sizes were bigger when people were controlling the robot than when they were not. In addition, we identify specific gaze patterns we termed *monitoring* and *planning* glances and discovered repeated sequences of monitoring glances during x-y control of the robot's end effector. Gaze patterns between robot assistance conditions are so distinct that a classifier trained on one-second sequences of gaze could successfully discriminate between teleoperation and autonomous trials 83% of the time purely using the dynamics of the gaze scanpaths.

There are three key insights from this work: (1) eye gaze is a meaningful signal that can reveal important aspects of a human-robot shared manipulation interaction, as evidenced by our finding that pupil sizes increased during robot control; (2) both control signals and gaze behavior during this interaction contain distinct patterns of behavior; and (3) gaze patterns change when people's behavior in the interaction changes (e.g., because of robot assistance), which suggests that models of eye gaze for human-robot shared manipulation must also consider the robot's behavior. These insights inform the design of strategies for task recognition and assistance during human-robot shared manipulation.

REFERENCES

[1] Henny Admoni and Brian Scassellati. 2014. Data-Driven Model of Nonverbal Behavior for Socially Assistive Human-Robot Interactions. In *ACM International Conference on Multimodal Interaction (ICMI)*. 196–199.

[2] Henny Admoni and Brian Scassellati. 2017. Social Eye Gaze in Human-Robot Interaction: A Review. *Journal of Human-Robot Interaction* 6, 1 (2017), 25–63.

[3] Henny Admoni and Siddhartha S. Srinivasa. 2016. Predicting User Intent Through Eye Gaze for Shared Autonomy. In *Proceedings of the AAAI Fall Symposium: Shared Autonomy in Research and Practice*. 298–303.

[4] Soichi Ando, Noriyuki Kida, and Shingo Oda. 2001. Central and Peripheral Visual Reaction Time of Soccer Players and Nonathletes. *Perceptual and Motor Skills* 92, 3 (2001), 786–794. https://doi.org/10.2466/pms.2001.92.3.786

[5] Brenna D Argall. 2015. Turning assistive machines into assistive robots. In *Proc. SPIE 9370, Quantum Sensing and Nanophotonic Devices XII*. International Society for Optics and Photonics.

[6] Michael Argyle. 1972. Non-verbal communication in human social interaction. In *Non-verbal communication*, R. A. Hinde (Ed.). Cambirdge University Press, Oxford, England.

[7] Michael Argyle and Mark Cook. 1976. *Gaze and Mutual Gaze*. Cambridge University Press, Oxford, England.

[8] Jackson Beatty. 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin* 91, 2 (1982), 276.

[9] Jean-David Boucher, Ugo Pattacini, Amelie Lelong, Gerrard Bailly, Frederic Elisei, Sascha Fagel, Peter Ford Dominey, and Jocelyne Ventre-Dominey. 2012. I Reach Faster When I See You Look: Gaze Effects in Human-Human and Human-Robot Face-to-Face Cooperation. *Frontiers in neurorobotics* 6, May (Jan. 2012), 1–11. https://doi.org/10.3389/fnbot.2012.00003

[10] C. Braunagel, E. Kasneci, W. Stolzmann, and W. Rosenstiel. 2015. Driver-activity recognition in the context of conditionally autonomous driving. In *IEEE 18th International Conference on Intelligent Transportation Systems*. IEEE, 1652–1657.

[11] Andreas Bulling, Jamie A Ward, Hans Gellersen, and Gerhard Tröster. 2009. Eye movement analysis for activity recognition. In *Ubicomp*. ACM.

[12] Tom Carlson and Yiannis Demiris. 2012. Collaborative control for a robotic wheelchair: evaluation of performance, attention, and workload. *IEEE Transactions on Systems, Man, and Cybernetics* (2012).

[13] Mark Cook. 1977. Gaze and Mutual Gaze in Social Encounters: How long—and when—we look others "in the eye" is one of the main signals in nonverbal communication. *American Scientist* 65, 3 (1977), 328–333.

[14] Navneet Dalal and Bill Triggs. 2005. Histograms of Oriented Gradients for Human Detection. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*. IEEE Computer Society, Washington, DC, USA, 886–893. https://doi.org/10.1109/CVPR.2005.177

[15] Anca Dragan and Siddhartha Srinivasa. 2013. A Policy Blending Formalism for Shared Control. *The International Journal of Robotics Research* (May 2013).

[16] Wolfgang Einhäuser, James Stout, Christof Koch, and Olivia Carter. 2008. Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry. *Proceedings of the National Academy of Sciences* (2008).

[17] Exact Dynamics 2016. Exact Dynamics. http://www.exactdynamics.nl. (2016).

[18] Alireza Fathi, Yin Li, and James M Rehg. 2012. Learning to recognize daily actions using gaze. In *ECCV*. Springer.

[19] J Randall Flanagan, Gerben Rotman, Andreas F Reichelt, and Roland S Johansson. 2013. The role of observers' gaze behaviour when watching object manipulation tasks: predicting and evaluating the consequences of action. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 368, 1628 (2013), 20130063.

[20] Andreas Gegenfurtner, Erno Lehtinen, and Roger Säljö. 2011. Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review* 23, 4 (2011), 523–552.

[21] Eric Granholm and Stuart R Steinhauer. 2004. Pupillometric measures of cognitive and emotional processes. *International Journal of Psychophysiology* 52, 1 (2004), 1–6.

[22] Zenzi M. Griffin and Kathryn Bock. 2000. What the Eyes Say About Speaking. *Psychological Science* 11, 4 (July 2000), 274–279. https://doi.org/10.1111/1467-9280.00255

[23] Elena Corina Grigore, Kerstin Eder, Anthony G. Pipe, Chris Melhuish, and Ute Leonards. 2013. Joint Action Understanding improves Robot-to-Human Object Handover. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Tokyo, Japan.

[24] Joy E. Hanna and Susan E. Brennan. 2007. Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language* 57 (2007), 596–615.

[25] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. https://doi.org/10.1177/154193120605000909

[26] Mary Hayhoe and Dana Ballard. 2005. Eye movements in natural behavior. *Trends in Cognitive Sciences* 9, 4 (2005), 188–194. https://doi.org/10.1016/j.tics.2005.02.009

[27] Laura V Herlant, Rachel M Holladay, and Siddhartha S Srinivasa. 2016. Assistive Teleoperation of Robot Arms via Automatic Time-Optimal Mode Switching. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 35–42.

[28] Rachel Holladay, Laura Herlant, Henny Admoni, and Siddhartha S. Srinivasa. 2016. Visibility Optimization in Manipulation Tasks for a Wheelchair-Mounted Robot Arm. In *RO-MAN Workshop on Human-Oriented Approaches for Assistive and Rehabilitation Robotics (HUMORARR)*.

[29] Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. 2011. *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

[30] Chien-Ming Huang and Bilge Mutlu. 2016. Anticipatory robot control for efficient human-robot collaboration. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 83–90.

[31] Laurent Itti and Christof Koch. 2001. Computational modelling of visual attention. *Nature Reviews Neuroscience* (March 2001), 194–203. Issue 2.

[32] Shervin Javdani, Henny Admoni, Stefania Pellegrinelli, Siddhartha S. Srinivasa, and J. Andrew Bagnell. 2017. Shared Autonomy via Hindsight Optimization for Teleoperation and Teaming. *arXiv preprint arXiv:1706.00155* (2017).

[33] Shervin Javdani, Siddhartha S. Srinivasa, and J. Andrew Bagnell. 2015. Shared autonomy via hindsight optimization. In *Robotics: Science and Systems (RSS)*.

[34] Qiang Ji, Zhiwei Zhu, and Peilin Lan. 2004. Real-time nonintrusive monitoring and prediction of driver fatigue. *IEEE transactions on vehicular technology* 53, 4 (2004), 1052–1068.

[35] R S Johansson, G Westling, A Bäckström, and J R Flanagan. 2001. Eye-hand coordination in object manipulation. *The Journal of Neuroscience* 21, 17 (Sept. 2001), 6917–6932.

[36] Enkelejda Kasneci, Gjergji Kasneci, Thomas C Kübler, and Wolfgang Rosenstiel. 2015. Online recognition of fixations, saccades, and smooth pursuits for automated analysis of traffic hazard perception. In *Artificial neural networks*. Springer, 411–434.

[37] Enkelejda Kasneci, Thomas Kübler, Klaus Broelemann, and Gjergji Kasneci. 2017. Aggregating physiological and eye tracking signals to predict perception in the absence of ground truth. *Computers in Human Behavior* (2017).

[38] Adam Kendon. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica* 26, 1 (1967), 22–63.

[39] Kinova Robotics, Inc. 2017. Kinova Robotics, Inc. http://kinovarobotics.com. (2017).

[40] Thomas Kinsman, Karen Evans, Glenn Sweeney, Tommy Keane, and Jeff Pelz. 2012. Ego-motion Compensation Improves Fixation Detection in Wearable Eye Tracking. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12)*. ACM, New York, NY, USA, 221–224. https://doi.org/10.1145/2168556.2168599

[41] Chris L. Kleinke. 1986. Gaze and eye contact: A research review. *Psychological Bulletin* 100, 1 (July 1986), 78–100.

[42] Niels A Kloosterman, Thomas Meindertsma, Anouk M Loon, Victor AF Lamme, Yoram S Bonneh, and Tobias H Donner. 2015. Pupil size tracks perceptual content and surprise. *European Journal of Neuroscience* (2015).

[43] Thomas C Kübler, Colleen Rothe, Ulrich Schiefer, Wolfgang Rosenstiel, and Enkelejda Kasneci. 2017. SubsMatch 2.0: Scanpath comparison and classification based on subsequence frequencies. *Behavior research methods* 49, 3 (2017), 1048–1064.

[44] Thomas C Kübler, Katrin Sippel, Wolfgang Fuhl, Guilherme Schievelbein, Johanna Aufreiter, Raphael Rosenberg, Wolfgang Rosenstiel, and Enkelejda Kasneci. 2015. Analysis of eye movements with Eyetrace. In *International Joint Conference on Biomedical Engineering Systems and Technologies*. Springer, 458–471.

[45] Andrew L Kun, Oskar Palinko, and Ivan Razumenić. 2012. Exploring the effects of size and luminance of visual targets on the pupillary light reflex. In *Proceedings of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 183–186.

[46] M F Land and M Hayhoe. 2001. In what ways do eye movements contribute to everyday activities? *Vision Research* 41, 25-26 (Jan. 2001), 3559–65.

[47] Michael F Land, N Mennie, and J Rusted. 1999. Eye movements and the roles of vision in activities of daily living: making a cup of tea. *Perception* 28, 4 (1999), 1311–1328.

[48] Veronique Maheu, Julie Frappier, Philippe S. Archambault, and Francois Routhier. 2011. Evaluation of the JACO robotic arm: Clinico-economic study for powered wheelchair users with upper-extremity disabilities. In *IEEE/RAS-EMBS International Conference on Rehabilitation Robotics (ICORR)*. 1–5. https://doi.org/10.1109/ICORR.2011.5975397

[49] Sandra P Marshall. 2002. The index of cognitive activity: Measuring cognitive workload. In *Proceedings of the IEEE Conference on Human Factors and Power Plants*. IEEE.

[50] MATLAB. 2016. *version 9.1.0 (R2016b)*. The MathWorks, Inc., Natick, MA.

[51] Gregor Mehlmann, Markus Häring, Kathrin Janowski, Tobias Baur, Patrick Gebhard, and Elisabeth André. 2014. Exploring a Model of Gaze for Grounding in Multimodal HRI. In *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI)*. ACM, New York, NY, USA, 247–254.

[52] Oskar Palinko, Andrew L Kun, Alexander Shyrokov, and Peter Heeman. 2010. Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 symposium on eye-tracking research & applications*. ACM, 141–144.

[53] Inc. Pupil Labs. 2016. Pupil Headset Website. (2016).

[54] Edward Rosten and Tom Drummond. 2005. Fusing points and lines for high performance tracking.. In *IEEE International Conference on Computer Vision*, Vol. 2. 1508–1511. https://doi.org/10.1109/ICCV.2005.104

[55] Uta Sailer, J Randall Flanagan, and Roland S Johansson. 2005. Eye-hand coordination during learning of a novel visuomotor task. *Journal of Neuroscience* 25, 39 (2005), 8833–8842.

[56] K. Sakita, K. Ogawara, S. Murakami, K. Kawamura, and K. Ikeuchi. 2004. Flexible cooperation between human and robot by interpreting human intention from gaze information. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)*, Vol. 1. https://doi.org/10.1109/IROS.2004.1389458

[57] Dario D Salvucci and Joseph H Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*. ACM, 71–78.

[58] Thiago Santini, Wolfgang Fuhl, and Enkelejda Kasneci. 2017. CalibMe: Fast and Unsupervised Eye Tracker Calibration for Gaze-Based Pervasive Human-Computer Interaction. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2594–2605.

[59] Thiago Santini, Wolfgang Fuhl, Thomas Kübler, and Enkelejda Kasneci. 2016. Bayesian identification of fixations, saccades, and smooth pursuits. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*. ACM, 163–170.

[60] Emrah Akin Sisbot and Rachid Alami. 2012. A Human-Aware Manipulation Planner. *IEEE Transactions on Robotics* 28, 5 (Oct. 2012), 1045–1057.

[61] Maria Staudte and Matthew W Crocker. 2011. Investigating joint attention mechanisms through spoken human-robot interaction. *Cognition* 120 (Aug. 2011), 268–291. https://doi.org/10.1016/j.cognition.2011.05.005

[62] Kyle Strabala, Min Kyung Lee, Anca Dragan, Jodi Forlizzi, Siddhartha S. Srinivasa, Maya Cakmak, and Vincenzo Micelli. 2013. Toward Seamless Human-Robot Handovers. *Journal of Human-Robot Interaction* 2, 1 (2013), 112–132.

[63] P.H.S. Torr and A. Zisserman. 2000. MLESAC. *Comput. Vis. Image Underst.* 78, 1 (April 2000), 138–156. https://doi.org/10.1006/cviu.1999.0832

[64] Matthew van der Zwan, Valeriu Codreanu, and Alexandru Telea. 2016. CUBu: universal real-time bundling for large graphs. *IEEE transactions on visualization and computer graphics* 22, 12 (2016), 2550–2563.

[65] Roel Vertegaal, Robert Slagter, Gerrit van der Veer, and Anton Nijholt. 2001. Eye Gaze Patterns in Conversations: There is More to Conversational Agents Than Meets the Eyes. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2001) (CHI '01)*, Vol. 3. ACM, Acm, 301–308. Issue 1.

[66] Joan N Vickers. 2009. Advances in coupling perception and action: the quiet eye as a bidirectional link between gaze, attention, and action. *Progress in brain research* 174 (2009), 279–288.

[67] Tian Xu, Hui Zhang, and Chen Yu. 2013. Cooperative gazing behaviors in human multi-robot interaction. *Interaction Studies* 14, 3 (2013), 390–418.

[68] Yuichiro Yoshikawa, Kazuhiko Shinozawa, Hiroshi Ishiguro, Norihiro Hagita, and Takanori Miyamoto. 2006. The effects of responsive eye movement and blinking behavior in a communication robot. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '06)*. 4564–4569. https://doi.org/10.1109/IROS.2006.282160

[69] Chen Yu, Paul Schermerhorn, and Matthias Scheutz. 2012. Adaptive Eye Gaze Patterns In Interactions with Human and Artificial Agents. *ACM Transactions on Interactive Intelligent Systems* 1, 2 (January 2012).

[70] Lei Yu and Huan Liu. 2003. Feature selection for high-dimensional data: A fast correlation-based filter solution. In *Proceedings of the 20th international conference on machine learning (ICML-03)*. 856–863.