

The Role of Object Physical Properties in Human Handover Actions: Applications in Robotics

Nuno F. Duarte^{1,2} *Student, IEEE*, Aude Billard² *Fellow, IEEE*, and José Santos-Victor¹ *Senior Member, IEEE*

Abstract—Observing how humans interact with each other and how they manipulate objects, offer insight on interaction mechanisms and how these are influenced by the physical properties of the used objects. These insights can support a more informed design of robot controllers meant to manipulate objects in collaboration with humans.

In this work we study human-to-human handovers of cups filled with various amount of liquid and textures, and investigate to which extent the manipulation strategy depends on : (i) the individual preference, (ii) whether the cup is filled with water or not, and (iii) the cup physical properties. An analysis of the human giver's hand acceleration, velocity, and position during the handover of different cups under two liquid level conditions, allows to distinguish between careful and not-careful (normal) manipulation. We quantify to which extent the liquid level inside the cups influences the carefulness level of human manipulation. Lastly, our study reveals that the cups' physical properties, such as fragility, breakability, and deformability, play a role in shaping the carefulness of the manipulation.

We apply these findings to human-robot scenarios by developing a robot controller capable of detecting, in real-time, if the human is being more careful than normal, and adapting the robot's approach of interaction accordingly. Additionally, we show that the detection of a careful manipulation, depending on the experimental context, may provide the robot with information concerning the human partner's intention or need for (manipulation) assistance.

Index Terms—Human Motion Understanding, Human-Robot Interaction, Non-verbal Communication.

I. INTRODUCTION

IT is our common experience that transporting a container (such as cups, glasses, or mugs) filled with some liquid is much more delicate than when it is empty. Mayer et al. [1] put this common knowledge to the test by examining people walking with a mug filled with coffee. They have found that humans try to avoid spilling the content by either estimating the frequency of sloshing of the liquid (moving the hand so as to counteract the induced slosh), or by slowing down and adopting a more careful manipulation. The choice between these two strategies seems to be related to individual preference. When it comes to programming similar skills in

robots we argue it would be best to choose the latter option, as it will be the most effective to prevent accidental spills.

Intention can be communicated either verbally or non-verbally through our movements. However, the former is not only computationally expensive it is also cumbersome for anyone who has to dictate their intention back to the robot. As such the latter has been extensively researched as an intuitive and biologically-inspired alternative [2], [3], [4], [5]. A few works have analyzed aspects such as decoding the action intention from a robot's eye and arm movements [6], judging an objects' weight from robot lifting movements [7], [8], or understanding when to handover [9], [10]. Ortenzi et al. [11] presented recently a survey on handovers in robotics. The authors looked at human-to-human handovers studies and the current approaches on human-robot handovers either for robot giver (robot-to-human) and robot receiver (human-to-robot). They identify two important phases of the handover: the pre-handover phase, i.e. the approach, and the physical phase, i.e. grasping and releasing. In the pre-handover phase, which is the scope of this paper, there are several works that proposed strategies for human-robot handovers, inspired by human-to-human handovers. The existing approaches have focused mainly in reproducing human-like motions [12], [13], [14], estimating the handover location [15], [16], or user satisfaction [17], [18], [19]. The assumption in the state-of-the-art is that the human handover motion is a purely functional motion. Instead, we argue that it can be modulated (and therefore express) latent features related to the object or the human action intentions. The state-of-the-art does not study human-to-robot handovers where the robot is capable of distinguishing types of handovers: a normal handover, or a challenging handover where the human resorts to perform it with extra care, as presented by [1]. A recent work has addressed the human manipulation of full and empty cups [20]. The authors studied the kinematic motion during pick & place and were able to distinguish between careful and not careful motions by inspecting the complete trajectory of the motion. Instead, our work is on human-to-human handovers which adds an interactive variable of "informing" the other partner whether the cup requires extra care to manipulate. Additionally, our approach is not only capable of online classification, without needing the complete trajectory, but it has been applied to real-time human-robot interactions.

This paper aims at providing an in depth scope of the manipulation strategies for cups filled with water or completely empty. The objective of the analysis is to identify and extract features to recognize human careful and not careful motions during human-human trials and apply it to human-

Corresponding author: Nuno Ferreira Duarte.

This work was partially supported by the RBCog-Lab research infrastructure, Fundação para a Ciência e a Tecnologia (FCT) with reference UID/50009/2020, the CHIST-ERA program through the project CORSMAL, by the Lisbon Ellis Unit (LUMLS), and the PhD grant with reference PD/BD/135116/2017 of Nuno Ferreira Duarte.

¹Vislab, Institute for Systems and Robotics, Instituto Superior Técnico, Universidade de Lisboa, Portugal. {nferreiraduarte, jasv}@isr.tecnico.ulisboa.pt

²LASA, Swiss Federal Institute of Technology, Lausanne, Switzerland. {nuno.ferreiraduarte, aude.billard}@epfl.ch.

robot scenarios. The focus is not on the effects of the weight, which is a user-dependent variable (the stronger you are the lower the manifestation), but instead on the challenge of transporting liquids and the underlying effects of the object properties on the transportation. Potential applications can be a factory plant, where robots can infer inherent properties of objects from human manipulations, such as fragility or breakableness. Alternatively, a robot caretaker can study the senior residents various levels of musculoskeletal limitations and adapt its motor constraints when assisting them.

Our contribution is fourfold: (i) show how the kinematic data of human-to-human handover movements reveal two distinct giver's carefulness levels, that change with empty or full cups; (ii) learn "carefulness" hand-over models, both for careful and not-careful (careless) human-to-human handovers; (iii) apply the model to a robotic controller to recognize during human-to-robot interactions (handovers, pick & place, and box carrying) whether the human is being careful when transporting the object in free space; and (iv) adapt the robot's motor control approach to the type of carefulness motion.

We start in Section II by detailing the human-human interaction (HHI). The wrist's kinematic data from the HHI dataset during cup handover is employed in a feature extraction model. The model learns two manipulation strategies, careful and careless behaviour. The model, presented in Section III, is compared against a baseline model and evaluated in terms of distinguishing handovers of empty cups as natural manipulations (not careful), and water filled cups as careful manipulations. In Section IV the model is tested to understand the impact of cups properties and other datasets on the model's accuracy. The best model is incorporated in a robotic controller. Section V presents the real-time human-robot interaction (HRI) experiments. Section VI is reserved for discussion and presentation of future work.

II. HUMAN TO HUMAN HANDOVER

This section presents the human-to-human handovers from which human motions during handovers are extracted. The first scenario presented is used for analysing the human manipulation strategy when handling a cup under two liquid level conditions. Moreover, the proposed models are designed based on the first scenario findings. The other two scenarios, introduced later, are included to further evaluate the carefulness detection accuracy of the proposed models.

A. Handover Data

The first dataset used is from a collaboration between École polytechnique fédérale de Lausanne (EPFL) and Karlsruher Institut für Technologie (KIT) [21]. We refer to it as the EPFL-KIT dataset¹ from now. It involves picking up an object and passing it to another person which receives it and places it back on the table. This is repeated several times with each pair of participants. There is a total of 157 handovers, 81 of empty cups and 76 of full cups, respectively. Figure 1 shows the different cups for each of the participants. The

handover trajectories in the EPFL-KIT dataset are recorded at 120 Hz, taking on average 1-3 seconds, corresponding to 100-300 data points. A total of 4 participants, aged between 25-35 years old, with a graduate academic level, participated in the experiments. The recorded data includes motion tracking markers from the OptiTrack system on the participant's wrist and the cups, as well as data gloves from the CyberGlove system on the participant's hand.

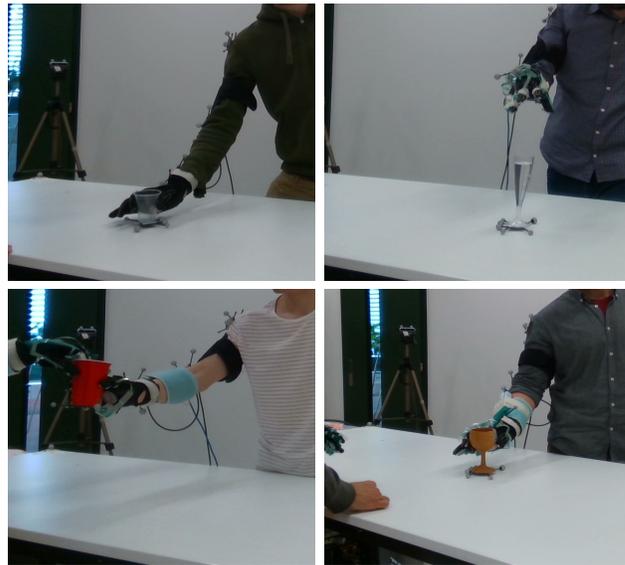


Fig. 1: Each of the four participants is holding one of the four cups. From left-to-right then top-to-bottom it is the transparent cup, champagne cup, red cup, and the wine glass. The first three are made of plastic and the last is glass.

The handovers of the cups happen under two distinct situations: (i) an empty cup, and (ii) a cup 90% filled with water. Each participant hands over the different cups to a second participant (also present in the dataset but not analysed during the handover) and each cup is manipulated in both conditions. The cups relevant for this work are, shown in Figure 1, the red plastic cup (bottom-left), the transparent plastic cup (top-left), the champagne plastic cup (top-right), and the opaque wine glass (bottom-right).

B. Handover Analysis

The handover motions are 3D Cartesian coordinates over time which begin at the moment of pickup (grasp) and finish when the cup is safely held by the other participant (handover). During the experiments, the participant could grasp the cup multiple ways (top of the cup, higher or lower from the side, etc) and the handover location could be in a 3D space bounded by the table as seen in Figure 1. This gave rise to several different grasp configurations and a disparity in the duration/length of the handover trajectories. Bear in mind that it was not possible to re-grasp the cup or change grasp configuration during the handover, and the cup would have to start upwards on the table for every interaction.

In order to analyse the kinematic of the wrist for all handovers for every participant, and every cup in both cup

¹<https://www.epfl.ch/labs/lasa/datasets/>

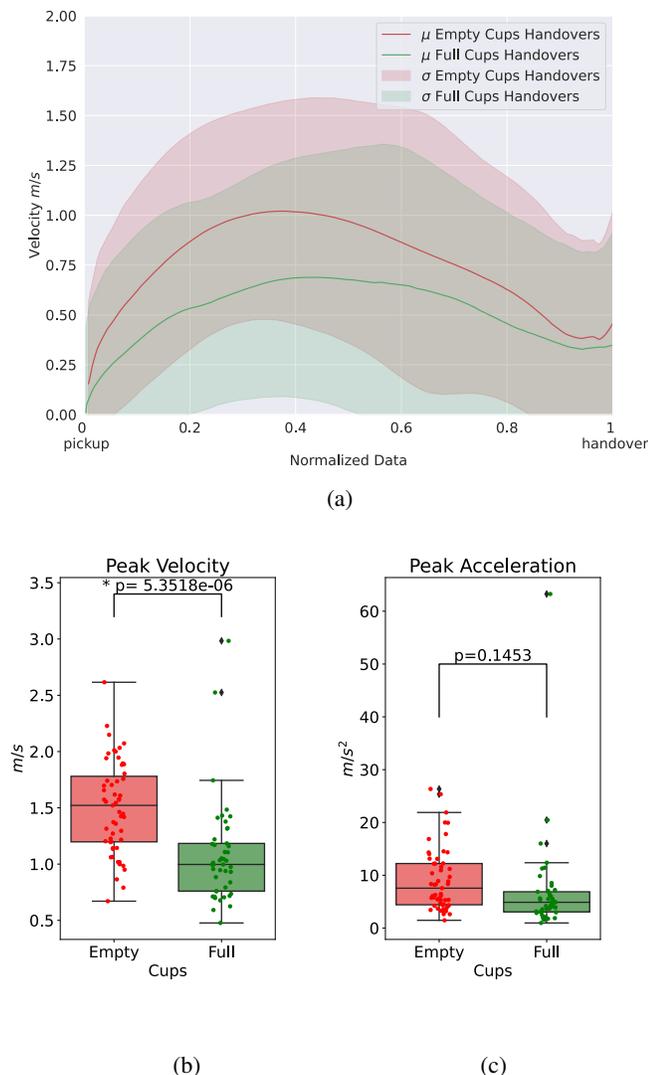


Fig. 2: (a) is the mean μ and standard deviation σ of the handover actions of the dataset separated in the two cups conditions (empty and full). The velocities of the human wrist are plotted over the trajectory of the handover. The trajectories are normalized. (b) is the peak velocity and (c) the peak acceleration, for each handover action and both conditions. p-values for peak velocities and acceleration of both cup levels are shown on the top each plot. Confirmation of significant difference is highlighted using a star.

conditions it was decided to normalize all the handovers to a standardized fixed length. Hence it was applied a min-max normalization before reducing the dimensionality to a vector by calculating the euclidean norm of the x, y , and z Cartesian coordinations. This re-scales the data to a $[0, 1]$ dimension where 1 is the final step, refer to as the handover, and 0 is the initial step, representing the moment of pickup. The plot in Figure 2 shows the velocities' mean and standard deviation for the human giver's hand throughout the handover in the entire EPFL-KIT dataset distinguishing both empty and full cup conditions.

Throughout our analysis, the common trait of all demonstrations is the typical bell-shape for the velocity profile as humans choose a minimum jerk approach for the hand trajectory. This has been identified in previous works in point-to-point human motion [22] and this behaviour manifests, likewise, in object handovers. In contrast, the most notable difference is on the bell-shape peak, i.e. the maximum velocity reached by the human. Analysing the peak velocity box plot in Figure 2, the difference is noticeable when distinguishing the cups by the level of water contents. The one-way ANOVA test revealed a statistically significant difference between the peak velocity of both cup conditions ($F(1,98) = 23.19, p < 0.001$). This is fairly straightforward as a cup filled with water presents an additional challenge during manipulation, in other words, transporting the contents inside without spilling or breaking. There is also the added weight of the liquid to the overall mass of the cup, however, we argue that the effect of a liquid oscillating inside a cup during human transportation is more impactful in deterring quick and jerky movements than a particularly heavy object. The peak acceleration, as seen in the box plot of Figure 2, is not as relevant to differentiate the two cup conditions. The one-way ANOVA test did not reveal a statistically significant difference between the peak acceleration of both cup conditions ($F(1,98) = 2.16, p = 0.1453$).

From the kinematic motion, it is clear that in any handover motion there are two distinct stages, an acceleration stage and a deceleration stage. This is in line with a minimum-jerk motion which starts and ends with the wrist in rest positions. It is also evident the disparity of the two water conditions for those two stages. The empty cup condition is showing a much steeper acceleration and, consequently deceleration, to reach the rest position. This feature, which has been addressed above, can be utilized to differentiate the two types of manipulation strategies: careful and not careful. Another point is related to familiarization with the task and object. As humans repeat an exercise multiple times they tend to gather prior knowledge from past events, and in this particular scenario, estimate the object's mass and the required force to manipulate. As a result, a novel object with an unexpected heavy mass might invoke a slower manipulation in the first trial but after some attempts, there is a pre-activation of muscles and joints to anticipate the requirements which may result in a more natural manipulation. This familiarization with the object properties can occur when there is liquid inside but the risk of spilling is constantly present. Hence the manipulation strategy will not change significantly because it is bound to the risk assessment. For this reason, we consider the level of water to be the most important factor.

In this section, we presented the human-human experiments studies and the datasets collected. From the first dataset, the handover trajectories were post-processed to extract the acceleration and deceleration phase of the wrist movement for any handover of empty or full water cups. The analysis has shown that manipulating cups full of water would usually originate in a slower, less abrupt motion, compared to empty cups. In the next section, the handover segmentations will be used for detecting different manipulation strategies.

III. MODELLING OF HANDOVER MANIPULATIONS

In this section, we present the models for human manipulation of cups in the two conditions: (i) empty cup, or (ii) water level at around 90 %. From the discussion in Section II we can argue that there are two possibilities for modelling the human manipulation in the handover context: (i) the acceleration phase, and (ii) the deceleration phase. This has been shown in pick-and-place actions [23] where goal-oriented biological motions are typically a minimum-jerk control problem with an acceleration and deceleration phase [22]. The acceleration phase begins with the object at rest position, the human grasps the object, and then the object increases in velocity as it is lifted up for transportation. The deceleration phase indicates the approach stage to handover the cup to another person, indicated by a gradual decrease in velocity until a stationary state is reached for completion of the handover.

Figure 3 is a diagram of the control system for both the acceleration and deceleration phase models. The overall structure is identical for both models, the main differences are the information used from the handover (Training Data) and the Modelling technique itself. The Classifier and the Human-in-the-loop control is identical. The following subsections describe the two modelling techniques, the classifier applied to the control loop and the advantages and disadvantages of both models.

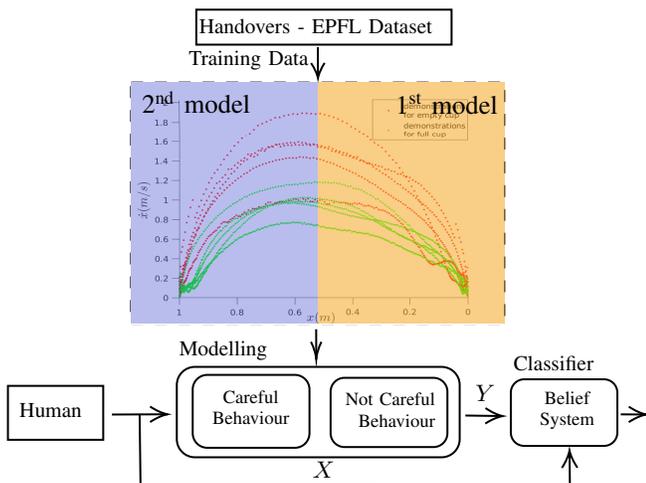


Fig. 3: Carefulness detection controller loop for both models. The 1st model learns from the deceleration phase of human handovers (right-side of the trajectory - yellow region). The 2nd model learns from the acceleration phase (left-side of the trajectory - blue region).

A. 1st model - Deceleration Phase

The deceleration phase was first used in the previous paper [24] and it provided a direct comparison of velocities for the two types of behavior. This was possible as the extraction of the deceleration phase during the cup manipulation revealed the maximum velocities and its evolution towards the resting stage, i.e. handover meeting point. From this approach, we model the velocity as a function of the distance towards the

handover for both situations. In this approach, the input $x \in D \subset \mathbb{R}^+$ denotes the distance of the human wrist towards the meeting point. Considering two time-independent dynamical systems (Ds),

$$\dot{x} = \mathbf{f}(x) \quad (1)$$

one for the *not careful* manipulation (empty cups), and another for the *careful* manipulation (full cups), where $\mathbf{f} : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is a continuous differentiable function, with one equilibrium point set at the origin with guaranteed Lyapunov conditions for global asymptotic stability. Each DS is encoded using Gaussian Mixture Models (GMM) which defines a joint distribution function between the velocity and distance of the wrist, and Gaussian Mixture Regression (GMR) generates the desired velocity at each location during the handover for each of the two models.

B. 2nd model - Acceleration Phase

In order to learn the latent features in the acceleration phase, a new approach is selected. The reason for opting for a new approach instead of applying the previous model technique is due to not being capable of finding distinct features for the two DS in the acceleration phase. This happens because all handovers, regardless of the empty or full condition, start with zero velocity and in a stationary position. On account of this, the DS output would render the generated desired velocities of both Careful and Not Careful behaviours indistinguishable. As an alternative to GMR to model the acceleration phase of the handovers we use the covariance matrix of the GMMs. The covariance matrix Σ of the GMMs that encode $\dot{x} = \mathbf{f}(x)$ expresses the correlation between the velocity and position in the handover space. The 1st Gaussian represents the steepest phase in the acceleration and from $\Sigma \vec{v} = \lambda \vec{v}$ the 1st eigenvector \vec{v} for both cup conditions is indicative of the direction of largest data variance. Given that we analyse the handover data as velocity and position, the eigenvector components $\vec{v} = [\vec{v}_x, \vec{v}_x]^T$ are the velocity and distance component respectively, and the

$$\frac{\vec{v}_x}{\vec{v}_x}$$

gives the inverse of time (velocity divided by position), i.e. the frequency of change of the wrist. As discussed in Section II, the velocity profiles are usually distinct when manipulating empty and filled with water cups, therefore the acceleration model learned the “frequency” of the wrist ($\frac{\vec{v}_x}{\vec{v}_x}$) for either condition which in the Modelling block represents the Careful and Not Careful Behaviour.

C. Classifier - Belief System

Inspired in the adaptation mechanism of [25], the Classifier is a Belief System that compares the Human-in-the-loop input X with the generated Y for each behaviour (*careful* and *not careful* motions). For the deceleration model, the real-time velocity is the current velocity

$$X = \dot{x}^t$$

and the generated $Y = \mathbf{f}$ are the DS velocities, as for the acceleration model, the real-time “frequency” of the wrist is computed as

$$X = \frac{\dot{x}^t - \dot{x}^{t-1}}{x^t - x^{t-1}}$$

and the generated $Y = \frac{\dot{x}}{\ddot{x}}$, as the “frequency” of the wrist for Careful and Not Careful. The Belief System is as follows:

$$\begin{aligned} e &= X - \sum_{i=1}^2 b_i^t Y_i \\ b_i^t &= \epsilon(e^T \mathbf{f}_i + (b_i^t - 0.5)|\mathbf{f}_i|^2) \quad \epsilon \in \mathbb{R}^+ \\ b_i^{t+1} &\leftarrow b_i^t + \dot{b}_i^t \Delta t \\ B &= [b_1, b_2] \quad \sum_{i=1}^2 b_i = 1 \end{aligned} \quad (2)$$

where B provides the belief that the new handover is one of the carefulness motions. This is calculated as the error function comparing the real input data and the output of the trained models. The adaptation rate ϵ is the hyperparameter common in both models. It weighs the effect of past information on the current step, i.e. memory from the beginning. \mathbf{f}_i is the model output for each of the motion behaviour, b_i^t is the classification output (belief), at time t , for each model $i = 1, 2 := (\text{not careful}, \text{careful})$. For real-time classification, the B vector is read at each time step, and when one of the beliefs (b_1^t or b_2^t) reaches 1 (100%) the information is sent to the robot to update its state depending on the HRI scenario in Section V.

The previous modelling approach (the deceleration phase model) has some limitations. Foremost, it is focused on the latter stage (the deceleration phase), resulting in a later classification. The Belief System classifies the motion at the beginning of the deceleration trajectory where the two DS diverge. However, for handover data outside the trained region, i.e. regions where the velocities are far greater than in the dataset, the data can not be accurately compared with the two DS. This new handover trajectory can occur outside the joint distribution of both DS which would generate unpredictable GMR outputs. One drawback of the second modelling approach (the acceleration phase model) relates to segmentation. It is more challenging to extract the acceleration phase since it involves identifying the precise moment of the pickup which, due to sensor noise and occlusions during grasping, is prone to errors. This problem does not occur in the deceleration phase, making it simpler to extract from the dataset. As a workaround, it was decided to add a low pass filter during training and testing. This low-pass filter ignores the small velocities, which mainly occur right after pickup and in the final stage of the handover (which is not part of the second model). This solution improves the classification accuracy without influencing the real-world performance since the first samples that are ignored by the low-pass filter are not informative enough to distinguish the motion.

The models discussed in this section, the acceleration and deceleration models, provide two possibilities of understanding human cup manipulations in the presence of varying liquid levels. The next section is reserved for analysing these two models in great detail. It starts by comparing both models on the dataset of Section II, it then evaluates the effect of the ϵ parameter in the classification step, proceeding with

Type of Cup	Train	Test	Carefulness Detection				
			Predicted	Acceleration Model		Deceleration Model	
				Empty	Full	Empty	Full
Red Cup	Red Cup	Not Careful	0.77	0.17	1	0.2	
		Careful	0.23	0.83	0	0.8	
Champagne	Champagne	Not Careful	0.4	0	0.82	0.27	
		Careful	0.6	1	0.18	0.73	
Transparent Cup	Transparent Cup	Not Careful	0.43	0.33	0.65	0.39	
		Careful	0.57	0.73	0.35	0.61	
Wine Glass	Wine Glass	Not Careful	0.57	0.23	0.5	0.43	
		Careful	0.43	0.77	0.5	0.57	

TABLE I: Train set: 50% One cup; Test set: 50% Same cup. Higher accuracy in the prediction is marked in **bold**

an in depth exploration of the novel model, the acceleration phase. This involves studying the impact of different cup materials and properties while testing for the other two datasets (QMUL and IST datasets). These datasets will present unseen challenges such as new cups, participants, and new data acquisition techniques.

IV. EXPERIMENTAL RESULTS

A. Evaluation of both models

From the results in Table I the acceleration model is better than the previous deceleration model and those conclusions are present below. The evaluation metric was chosen as the classification accuracy of each model when splitting the dataset into types of cups. This means that careful accuracy is how many transportations of full cups are considered careful manipulations, and not careful accuracy is how many empty cups are considered not careful manipulations. Table I shows that the acceleration model is better at detecting handovers of full cups as careful manipulations. This is desirable given the challenging nature of transporting water in cups. Furthermore, it can be concluded, from the model, that an empty cup does not imply a not careful (careless) manipulation. Although the deceleration model is better at detecting handovers of empty cups as not careful manipulation it comes at a cost of not detecting most full cups as careful. Since the HHI experiment allowed participants to choose a preferable handover strategy, an empty cup restricts the movement less than the same cup filled with water (restriction on the orientation, oscillations, velocity, etc). As a result, the empty cup handover should reflect the user preference, either handover normally (not careful), or restricted (careful), and assuming that the dataset is a fair representation of both types of people, the empty cup should not reflect any preferred carefulness motion.

B. Results for adaptation rate (ϵ) values

The ϵ is the hyperparameter present in the classifier. It is a weighted parameter on the knowledge of past iterations. Figure 4 shows that as ϵ increases the careful accuracy drops, while the not careful accuracy increases. When increasing the adaptation rate the system is sensitive to initial noise and spurious data, it reaches a classification quicker (quicker response time) which results in more incorrect decisions. The not careful accuracy increasing as the ϵ rises are the result of being influenced by initial spurious points in the trajectory.

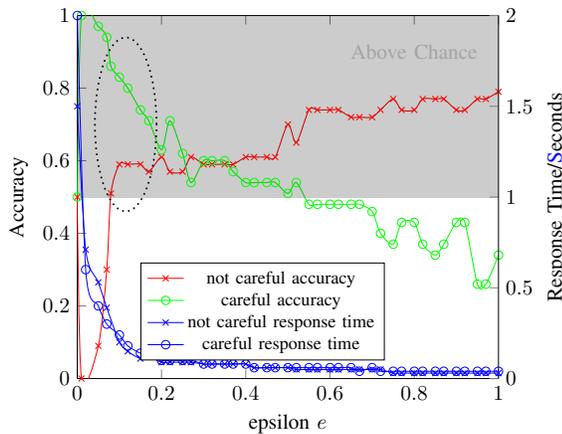


Fig. 4: The evolution of models accuracy and respective response time of the prediction for each value of epsilon.

The region for the best accuracy models is the ϵ interval between $[0.05 - 0.2]$ (indicated in Figure 4 by the dotted ellipse). This is a trade-off between a faster prediction time and good accuracy. Inside this region, 75% or more of full cups are classified as careful manipulations, and around 50% of empty cups are not careful manipulation (which is ideal given the personal preference choice when handling empty cups).

C. Results on types of cups

Table II shows the results of the model's accuracy for each of the train and test splits. From Table II it can be concluded that unknown full cups are predominantly classified as careful manipulations irrespective of the type of cup the model is trained on. While empty cups, regardless of the type of cups trained on, give rise to a non-preferential manipulation.

Training on one plastic type and testing solely on the other plastics give rise to the conclusion that training and testing on the same cup material (such as plastics) achieves the best results since it induces similar characteristics, e.g. risk of breaking, friction, weight, etc. Training the model on glass and testing on plastics it can be concluded that training a model on glass and testing on plastics worsens the likelihood of detecting full cups as careful manipulations. Although the dataset only has one glass cup we hypothesize that this could be induced by the risk of breaking, and the fact that glass is heavier than plastic.

Further discussion allows to infer differences when comparing types of plastic cups. Soft plastics are deformable due to their physical structure and material composition, thereby are prone to deforming. This is exacerbated when are filled to the top with a liquid. Rigid plastics are non-deformable as they do not present the same structural flaws as soft plastic, and are non-breakable, contrarily to glass which can break and shatter easily. The model trained on soft plastics (red and transparent cup) and tested on rigid plastics (champagne cup) produce worse careful accuracy. We argue this is the cause of deformability which makes it difficult to handle soft plastics when filled with water compared to rigid plastics. The model

Type of Cup	Train	Test	Acceleration Model			
			Predicted		Real	
			Empty	Full	Empty	Full
Transparent Cup	Red Cup	Champagne	0.688	0.15	0.46	0.1
		Wine Glass	0.312	0.85	0.54	0.90
Champagne	Transparent Cup	Red Cup	0.5	0.1	0.53	0.15
		Wine Glass	0.5	0.9	0.47	0.85
Red Cup	Transparent Cup	Champagne	0.47	0	0.5	0.16
		Wine Glass	0.53	1	0.5	0.84
Wine Glass	Transparent Cup	Red Cup	0.5	0.25	0.59	0.17
		Champagne	0.5	0.75	0.41	0.83
Champagne	Transparent Cup	Red Cup	0.5	0.1	0.55	0.08
		Champagne	0.5	0.9	0.45	0.92
Red Cup	Transparent Cup	Champagne	0.47	0	0.5	0.09
		Champagne	0.53	1	0.5	0.91
Transparent Cup	Red Cup	Champagne	0.625	0.15	0.52	0.04
		Champagne	0.375	0.85	0.48	0.96

TABLE II: One vs Rest Classification. Training set: 100% One cup type; Testing set: 100% Other cup types. **Plastic cups;** **Glass cups**

learned (intrinsically) the deformability feature in the full cup case and since the testing set does not have cups with that property, it became difficult to detect the full cup cases in the test split. On the other hand, training a model on rigid plastics and testing on soft plastics provide the highest level of careful accuracy (with a bias to plastics only). The rigid plastic is not affected by any of the latent features (deformability or breakability), hence the model did not learn to be extra sensitive. In the testing set, the deformability feature was present (soft plastics only) and since the effect is mainly present in full conditions, the model was capable of easily distinguishing the two carefulness levels.

Our conclusion is that a model trained on soft (deformable) plastics learns that full cup actions are extremely difficult, hence these actions for Rigid plastics have a higher likelihood to be classified as not careful since the rigid plastic is considered to be a cup with no inherent challenging properties. A model trained on rigid plastics learns the opposite, considering full cup actions for soft plastics as mostly classified as careful manipulations.

D. Results of other datasets

In this subsection we extend the analysis of our model's performance to other datasets with new participants, cups, sensor's data, and overall new experimental scenarios to the first dataset.

The second dataset used is the CORSMAL Containers Manipulation dataset² from the Queen Mary University of London (QMUL). The participants performed a series of tasks on a set of containers. The tasks involve pouring water, rice, or pasta into containers such as cups or boxes, and initiating a handover towards a robot. The dataset, referred to as the QMUL dataset, includes four cameras, one attached to the human, one attached to the robot, and two looking from each

²https://corsmal.eecs.qmul.ac.uk/containers_manip.html

side (top images in Figure 5 shows two frames of each external cameras), and one microphone. The cup location is estimated using a multi-view projective geometry which provides a 2D centroid of the cup from the two side cameras at 30 Hz sampling frequency [26].

The third dataset is the Human Manipulation of Cups with Water³ from Instituto Superior Técnico (IST). It involves participants in pairs interacting with cups where they both perform pick & place and handover actions. The dataset, referred as the IST dataset, includes two head-mounted eye trackers (bottom images in Figure 5), one on each participant, and OptiTrack markers on the head and wrist of the participants to record the motion (recorded at 120 Hz). All three datasets are publicly available in their corresponding academia websites. The first and third datasets were gathered by one or more authors but not for the purpose of this work. None of the authors were involved in the elaboration of the second dataset.



Fig. 5: Extracted frames of handover actions from the QMUL dataset (top row) and the IST dataset (bottom row).

Table III shows the results for the three datasets: (i) the EPFL-KIT, (ii) the QMUL, and (iii) the IST. The experiments were accomplished by training the models using varying percentages of dataset (i). The accuracy results of each model indicate that it can identify most of the full cups handovers as careful manipulations and corroborate the idea that handovers of empty cups are dependent on human preference, and not conditioned by cup properties. An interesting find is seeing the not careful accuracy in the training set decreases when trained on large datasets while the testing set accuracy for both classifications remains fairly similar. This, we argue, is another proof that for large datasets, the model that best generalizes is the one that assumes that empty cups do not necessarily invoke a not careful (natural) manipulation. The conclusion are three fold: (i) the model generalizes well for unknown people and cups, (ii) all 3 datasets show no preference for the empty condition, and (iii) the Carefulness detection controller can

³https://vislab.isr.tecnico.ulisboa.pt/datasets_and_resources/#hcups_water

Train	Test	Real Predicted	Acceleration Model			
			Training Set		Testing Set	
			Empty	Full	Empty	Full
EPFL 10%	EPFL 90%	Not Careful	0.83	0	0.53	0.19
		Careful	0.17	1	0.47	0.81
EPFL 20%	EPFL 80%		0.89	0.12	0.5	0.16
			0.11	0.88	0.5	0.84
EPFL 30%	EPFL 70%		0.69	0.09	0.51	0.15
			0.31	0.91	0.15	0.85
EPFL 40%	EPFL 60%		0.6	0.13	0.55	0.1
			0.4	0.87	0.45	0.9
EPFL 50%	EPFL 50%		0.61	0.16	0.5	0.1
			0.39	0.84	0.5	0.9
EPFL 60%	EPFL 40%		0.55	0.08	0.5	0.15
			0.45	0.92	0.5	0.85
EPFL 10%	QMUL	Not Careful	1	0	0.5625	0.2
		Careful	0	1	0.4375	0.8
EPFL 20%	QMUL		0.92	0.14	0.5	0.2
			0.08	0.86	0.5	0.8
EPFL 40%	QMUL		0.94	0.18	0.5	0.13
			0.06	0.82	0.5	0.87
EPFL 10%	IST	Not Careful	0.45	0.02	0.65	0.18
		Careful	0.55	0.97	0.35	0.82
EPFL 20%	IST		0.58	0.11	0.51	0.22
			0.42	0.88	0.49	0.78
EPFL 40%	IST		0.55	0.10	0.52	0.23
			0.45	0.89	0.47	0.77

TABLE III: Top: Train set - sample of EPFL; Test set - rest of EPFL dataset. Middle: Train set - sample of EPFL; Test set - 100% QMUL dataset (new people and new cups). Bottom: Train set - sample of EPFL; Test set - 100% IST dataset (new people and transparent cup).

achieve good accuracy for either precise data points (MoCap markers) and 3D point estimation (from stereo vision).

E. Conclusion

In terms of cup properties, the analysis can be summed by Figure 6. It is known that glass is breakable while plastic usually is not. Soft and hard plastic cups although are sharing most of the properties, when filled with water the soft plastic may deform due to the weight of the liquid inside. As a result, soft plastics are characterized as deformable but not-breakable, while hard plastics are not-deformable and not-breakable.

From the conclusions on Section IV-C, hard plastic cups are the easiest to manipulate as they do not present a challenge of breakability or deformability. The glass is not-deformable but breakable hence it is more challenging to manipulate than the hard plastic cups. Soft plastic cups are deformable but not-breakable so are also more difficult than hard plastics. It is hard to quantify which of the two challenges has a higher priority when it comes to manipulation strategy. We could argue that breaking a glass cup is worse and irreversible compared to deforming a cup. However, in this situation, the deformability only manifests when the cup is filled with water, hence it is fair to conclude that a glass cup is the hardest to manipulate and would influence the manipulation strategy the most.

V. HUMAN-ROBOT APPLICATIONS

In this section are detailed the HRI experiments performed to evaluate the acceleration model. The Carefulness detection

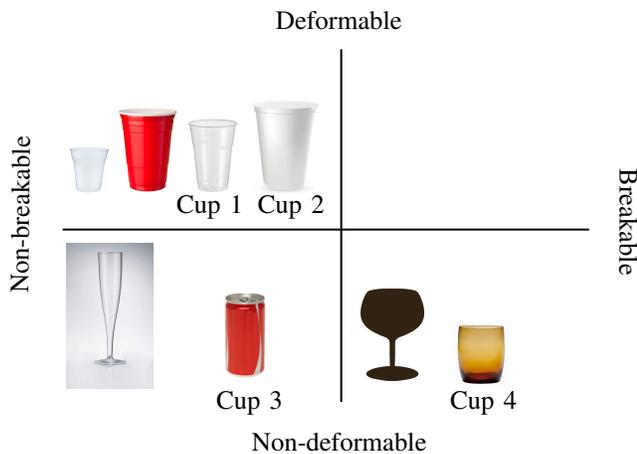


Fig. 6: An illustration scheme of the important features of cups during manipulation: deformability, and breakability. Deformability is present solely when filled with water, while breakability is an inherent property of the cup.

controller is applied to recognize whether the human is being careful with the object or careless (not careful). The robot proceeds to pick the cups and, according to the observed human manipulation, manipulate them differently and separate the careful from the not careful cups. These HRI experiments are three-fold: (i) pick-and-place, (ii) handover task, and (iii) robot assistance. The robot platform used is the Kinova gen3 with a Robotiq gripper attached to the end-effector as seen in Figure 7. The Kinova robot was controlled using the `kortex_ros`⁴ package for ROS and velocity commands were used to control the end-effector in Cartesian coordinates at 40 Hz for linear (m/s) and angular (rad/s) velocities. For each of the three HRI experiments, the objects are tracked by the OptiTrack MoCap system which is streaming the data to ROS at 120 Hz. The same velocity PI controller is used in all experiments for the careful motion we lower the gain to have the robot move slower (and cap the maximum velocity). The model from Section III-B picked has the accuracy of Table III in the EPFL 40%/60% train/test split case. The ϵ value is picked inside the region mentioned in Section IV-B.

The control loop system presented in Figure 3 is a human-in-the-loop controller running at 120 Hz during the human-robot scenario. For simplification, the human hand position and velocity are calculated as the cup's or box's which always starts stationary either on the table or the floor, respectively. The human-in-the-loop system begins predicting the human manipulation the moment the cup has non-zero velocity. The Belief System classification B then outputs to the robot controller either not-careful (1) manipulation or careful (2) manipulation. This result can then be applied to several HRI scenarios. In the following subsections, we will address the experiments performed and the overall evaluation.

⁴The official repository to interact with the Kinova robot https://github.com/Kinovarobotics/ros_kortex

TABLE IV: Results of Pick & Place and Handover experiments.

		Not Careful	Careful
Cup 1	Empty	0.65	0.35
	Full	0.10	0.9
Cup 2	Empty	0.6	0.4
	Full	0.057	0.95
Cup 3	Empty	0.68	0.32
	Full	0.22	0.78
Cup 4	Empty	0.55	0.45
	Full	0.15	0.85

A. Human and Robot Pick and Place

The deceleration phase model used in the previous paper [24] has this HRI experiment setup. The experiments begins by the participant picking a cups from the table and placing it in another location. The robot then proceeds to grasp the same cup and move it to a third location depending on the model's output. In the not-careful manipulation option, the robot transports the cup to a bottom shelf, without worrying about any danger of spilling (tilting the cup). As for the careful option, the robot transports the cup, keeping the orientation fixed while slowing its velocity, and placing it on the top shelf. Table IV shows the successfulness of the human-in-the-loop system of adapting correctly to the present cup conditions. A total of 4 subjects participated in the experiment and manipulated 4 different cups (identified in Figure 6) with the two conditions (empty and full of water). Two cups are from the category of non-breakable and deformable properties (soft plastics), and the other two are a rigid plastic cup and a glass cup, respectively. Each participant manipulated the cup 10 times per cup and per condition. As a comparable variable, the cup present in the previous paper [24] is also included in the experiments as Cup 1.

B. Human-Robot Handover

The second scenario involves a human-robot handover of cups where the robot tries to infer, from human motion, whether the cup requires a careful manipulation or not. The main difference to the previous scenario of Section V-A has to do with the proactiveness of the robot that, instead of waiting for the cup to be placed, it meets the human in order to perform the handover. This, as mentioned before, is one of the advantages of analysing the acceleration phase of the motion with regards to the previous method in [24]. The results are present in Table IV since no major changes to the controller were implemented. The same subjects participated in both experiments and manipulated the same cups for both conditions. The classification results are calculated during the acceleration phase hence there is no difference in waiting for the object to be picked up or handed over.

It can be concluded from the pick & place scenario that we reach good if not better results in detecting cups filled with water as careful manipulations. When comparing the results from the previous deceleration model, the Cup 1 results match the ones observed in the previous work validating both

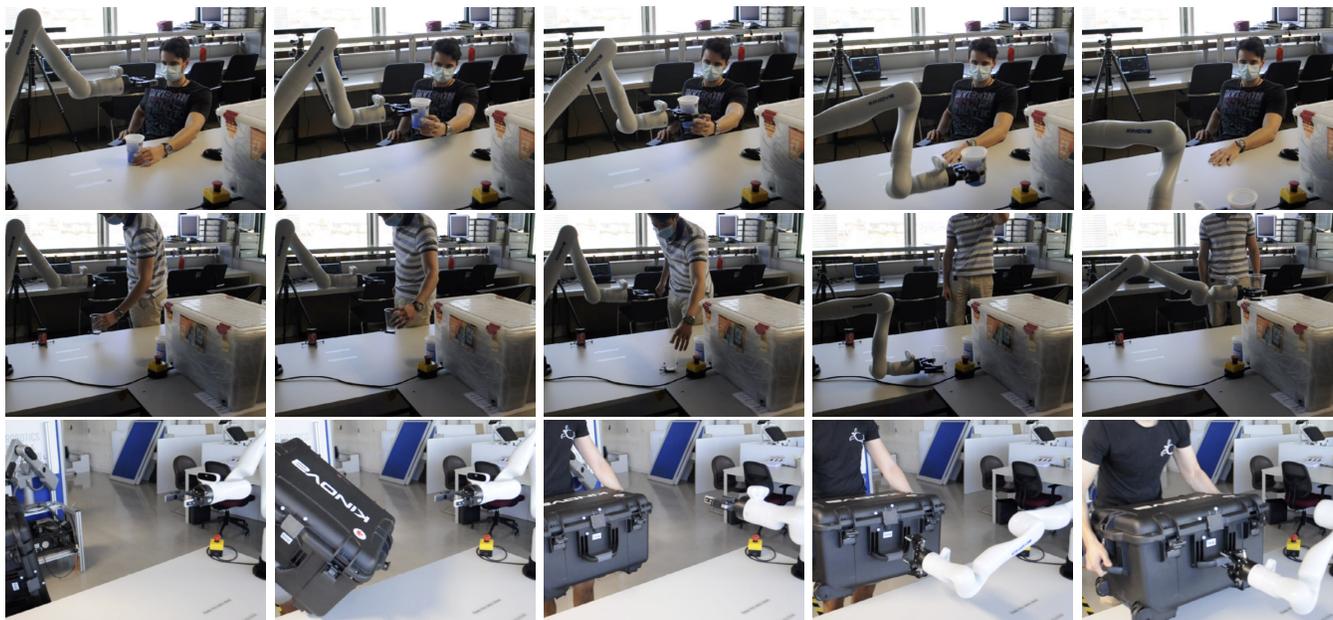


Fig. 7: Each row of images illustrate the three HRI scenarios where the Carefulness detection controller using the acceleration model is applied. The first row is the human-robot handover, the second row is the human-robot pick & place of cups, and the third row is the robot assistance to a human carrying a heavy box. In the third application setup the human was lifting the box at the right side of the robot to prevent occlusions of the box’s markers.

TABLE V: Results for Robot Assistance experiments.

		Easy	Hard
Box	Light	0.78	0.22
	Heavy	0.12	0.88

models as good detection mechanisms for human manipulation during pick & place tasks. However, these results also extend to other HRI applications such as handovers and given the architecture of the model the results achieve the same accuracy for both applications. To note, that due to the risk of spilling water, the participants’ trials ran in real-time but without robot participation. The Carefulness detection controller did output the commands to the robot and Figure 7 illustrates how the robot interacts in each scenario in a careful and not careful situation. The supplementary video shows clearly the different robot responses for each scenario.

C. Robot Assistance

In the last scenario, the context is changed. On a different note, we decided to move away from the realm of cups and carefulness manipulation detection and aim for another potential use case. The recognition of different manipulation strategies can also be applied to household activities such as lifting boxes, furniture, appliances, etc. The human is carrying a large box and the robot has to infer if the human requires assistance in lifting the box due to being too heavy. The robot if it detects that the human is struggling to transport the box it would grasp onto the side handle of the box and pull, in order to assist the person in placing the box on the table. If, on the contrary, the human does not exhibit any challenge

in lifting and transporting the box, it would not interact and leave the human unassisted. Table V shows the results for detecting whether the human was having trouble lifting the box given the human’s motion behaviour. This final experiment shows that it is possible to detect, fairly accurately, when a human is having difficulty in moving a large or heavy object by simply observing the motion pattern of the item being moved. Although light objects, similarly to empty cups, do not reflect one particular strategy, which once again we argue is indicative of human preference.

VI. CONCLUSION

We have studied the human hand motion during handovers of cups in two conditions: carrying an empty cup, or a cup filled with water. These experiments explored several datasets with data acquired from different sensors during handovers between two humans, or handovers simulated by a single person. Each dataset had different types of cups with different materials, and several participants manipulated each cup multiple times, in the two conditions mentioned above. We believe our results provide a broad and overall general analysis of the human motion behaviour during handovers of cups in two relevant conditions (an empty vs a full water cup). From these two conditions, we were able to detect a distinguishable motion strategy from humans when manipulating cups filled with water. This is a more secure, risk-free, option of moving objects when there is an apparent risk of spilling or danger compared to a normal handover between humans. Based on these findings, we developed a computational model describing careful/careless handovers, learned from human-to-human handover motion data. This computational model provides the robot with anticipatory knowledge of the type

of manipulation, careful or not careful, thus facilitating the robot's motor preparation and the adaptation, prior to the interaction, allowing for a better understanding of the object inherent properties. We provide a link to a video that shows examples of the HRI scenarios working in real-time, the operation of the controller and the online classification of the action's carefulness - video.PropertiesCupsHRI.ieee-2022.

The overall conclusions from the HRI experiments are that the acceleration model clearly shows its advantages over the previous model with its multi-use in different robot applications. While the previous one had been only applied to pick & place due to its limitation of having to set the final meeting point, this new model gives information the moment the object (cup or box) is picked by the human, making it versatile. As it was mentioned in the introduction, this model can be useful for many robot situations where humans play a vital role. As future work we intend to study the impact of the carefulness on the robot execution, if participants appreciate more this approach or not and whether the robot spills more or less water. Robots can learn a lot from humans and should take advantage of how humans tackle problems to better understand the world surrounding them. As a result, this robot controller aims to enhance the robot capabilities in understanding object properties from human manipulations.

REFERENCES

- [1] H. C. Mayer and R. Krechetnikov, "Walking with coffee: Why does it spill?," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 85, no. 4, pp. 1–7, 2012.
- [2] N. Mavridis, "A review of verbal and non-verbal human–robot interactive communication," *Robotics and Autonomous Systems*, vol. 63, no. 1, pp. 22–35, 2015.
- [3] A. Sciutti, M. Mara, V. Tagliasco, and G. Sandini, "Humanizing human-robot interaction: On the importance of mutual understanding," *IEEE Technology and Society Magazine*, vol. 37, no. 1, pp. 22–29, 2018.
- [4] H. Admoni, T. Weng, B. Hayes, and B. Scassellati, "Robot nonverbal behavior improves task performance in difficult collaborations," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, (Christchurch, New Zealand), pp. 51–58, IEEE, Mar. 2016.
- [5] H. Ravichandar, A. S. Polydoros, S. Chernova, and A. Billard, "Recent Advances in Robot Learning from Demonstration," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 3, pp. 297–330, May 2020.
- [6] N. F. Duarte, M. Rakovic, J. Tasevski, M. I. Coco, A. Billard, and J. Santos-Victor, "Action Anticipation: Reading the Intentions of Humans and Robots," *IEEE Robotics and Automation Letters*, vol. 3, pp. 4132–4139, Oct. 2018.
- [7] O. Palinko, A. Sciutti, L. Patané, F. Rea, F. Nori, and G. Sandini, "Communicative lifting actions in human-humanoid interaction," *IEEE-RAS International Conference on Humanoid Robots*, vol. 2015-Febru, pp. 1116–1121, 2015. ISBN: 9781479971749.
- [8] A. F. C. Hamilton, D. W. Joyce, J. R. Flanagan, C. D. Frith, and D. M. Wolpert, "Kinematic cues in perceptual weight judgement and their origins in box lifting," *Psychological Research*, vol. 71, no. 1, pp. 13–21, 2007.
- [9] J. R. Medina, F. Duvallet, M. Karnam, and A. Billard, "A human-inspired controller for fluid human-robot handovers," *IEEE-RAS International Conference on Humanoid Robots*, pp. 324–331, 2016. ISBN: 9781509047185.
- [10] S. Kato, N. Yamanobe, G. Venture, E. Yoshida, and G. Ganesh, "The where of handovers by humans: Effect of partner characteristics, distance and visual feedback," *PLOS ONE*, vol. 14, p. e0217129, June 2019.
- [11] V. Ortenzi, A. Cosgun, T. Pardi, W. P. Chan, E. Croft, and D. Kulic, "Object Handovers: A Review for Robotics," *IEEE Transactions on Robotics*, pp. 1–19, 2021.
- [12] K. Yamane, M. Revfi, and T. Asfour, "Synthesizing object receiving motions of humanoid robots with human motion database," in *2013 IEEE International Conference on Robotics and Automation*, (Karlsruhe, Germany), pp. 1629–1636, IEEE, May 2013.
- [13] A. Sidiropoulos, E. Psomopoulou, and Z. Doulgeri, "A human inspired handover policy using Gaussian Mixture Models and haptic cues," *Autonomous Robots*, pp. 1–16, 2018. Publisher: Springer US.
- [14] G. J. Maeda, G. Neumann, M. Ewerton, R. Lioutikov, O. Kroemer, and J. Peters, "Probabilistic movement primitives for coordination of multiple human–robot collaborative tasks," *Autonomous Robots*, vol. 41, pp. 593–612, Mar. 2017.
- [15] H. Nemlekar, D. Dutia, and Z. Li, "Object Transfer Point Estimation for Fluent Human-Robot Handovers," in *2019 International Conference on Robotics and Automation (ICRA)*, (Montreal, QC, Canada), pp. 2627–2633, IEEE, May 2019.
- [16] D. Widmann and Y. Karayiannidis, "Human Motion Prediction in Human-Robot Handovers based on Dynamic Movement Primitives," in *2018 European Control Conference (ECC)*, (Limassol), pp. 2781–2787, IEEE, June 2018.
- [17] D. Vogt, S. Stepputtis, B. Jung, and H. B. Amor, "One-shot learning of human–robot handovers with triadic interaction meshes," *Autonomous Robots*, vol. 42, no. 5, pp. 1053–1065, 2018. Publisher: Springer US.
- [18] M. K. Pan, E. A. Croft, and G. Niemeyer, "Evaluating Social Perception of Human-to-Robot Handovers Using the Robot Social Attributes Scale (RoSAS)," *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18*, pp. 443–451, 2018. ISBN: 9781450349536.
- [19] S. Parastegari, B. Abbasi, E. Noohi, and M. Zefran, "Modeling human reaching phase in human-human object handover with application in robot-human handover," *IEEE International Conference on Intelligent Robots and Systems*, vol. 2017-Septe, pp. 3597–3602, 2017. ISBN: 9781538626825.
- [20] L. Lastrico, A. Carfi, A. Vignolo, A. Sciutti, F. Mastrogiovanni, and F. Rea, "Careful with That! Observation of Human Movements to Estimate Objects Properties," in *Human-Friendly Robotics 2020* (M. Savariano, E. Renaudo, A. Rodríguez-Sánchez, and J. Piater, eds.), (Cham), pp. 127–141, Springer International Publishing, 2021.
- [21] J. Starke, K. Chatzilygeroudis, A. Billard, and T. Asfour, "On Force Synergies in Human Grasping Behavior," in *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*, (Toronto, ON, Canada), pp. 72–78, IEEE, Oct. 2019.
- [22] N. Fliège, J. McIntyre, and P. van der Smagt, "Minimum jerk for human catching movements in 3D," in *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechanics (BioRob)*, (Rome, Italy), pp. 581–586, IEEE, June 2012.
- [23] T. Flash, Y. Meirovitch, and A. Barliya, "Models of human movement: Trajectory planning and inverse kinematics studies," *Robotics and Autonomous Systems*, vol. 61, no. 4, pp. 330–339, 2013.
- [24] N. F. Duarte, K. Chatzilygeroudis, J. Santos-Victor, and A. Billard, "From human action understanding to robot action execution: how the physical properties of handled objects modulate non-verbal cues," in *2020 Joint IEEE 10th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, (Valparaiso, Chile), pp. 1–6, IEEE, Oct. 2020.
- [25] M. Khoramshahi and A. Billard, "A dynamical system approach to task-adaptation in physical human–robot interaction," *Autonomous Robots*, vol. 43, pp. 927–946, Apr. 2019.
- [26] A. Xompero, R. Sanchez-Matilla, Andrea Cavallaro, and R. Mazzon, "Corsmal containers manipulation," *CIS, School of Electronic Engineering and Computer Science, Queen Mary University of London*, 2020.



Nuno Ferreira Duarte is a Ph.D. student in a Joint-Degree between Instituto Superior Técnico, in Lisbon, Portugal and the Swiss Institute of Technology Lausanne (EPFL), in Switzerland, advised by Professor José Santos-Victor and Professor Aude Billard. He received his BSc/MSc in Electrical and Computer Engineering with specialization in Systems, Decision and Control from Instituto Superior Técnico in September 2016. He previously interned at California Institute of Technology under the supervision of professor Richard Murray. His research

interests include HRI, Machine Learning, and Robotics.



Aude Billard is a full professor and head of the LASA laboratory at the School of Engineering at the Swiss Institute of Technology Lausanne (EPFL). A.B. holds a B.Sc and M.Sc. in Physics from EPFL (1995) and a Ph.D. in Artificial Intelligence (1998) from the University of Edinburgh. A. B.'s research spans the fields of machine learning and robotics with a particular emphasis on learning from sparse data and performing fast and robust retrieval. This research received best paper awards from IEEE T-RO, RSS, ICRA, IROS, etc. Aude Billard is IEEE

Fellow, president-elect of the IEEE Robotics and Automation Society (RAS) and former IEEE RAS vice-president of publication activities, and was the recipient of the IEEE-RAS Best Reviewer Award and IEEE-RAS Distinguished Service Award.



José Santos-Victor is a full professor at the Department of Electrical and Computer Engineering, Instituto Superior Técnico (IST), Universidade de Lisboa, Portugal. He is the President of the Institute for Systems and Robotics in Lisbon. He conducts multidisciplinary research, with neuroscientists and psychologists, in the areas of cognitive and bio-inspired computer vision and robotics, including the design of advanced humanoid robotic platforms like the iCub. He is a member of the Academy of Sciences of Lisbon.