From human action understanding to robot action execution: how the physical properties of handled objects modulate non-verbal cues

Nuno Ferreira Duarte[1,2], Konstantinos Chatzilygeroudis[1], José Santos-Victor[2] Aude Billard[1]

Abstract— Humans manage to communicate action intentions in a non-verbal way, through body posture and movement. We start from this observation to investigate how a robot can decode a human's non-verbal cues during the manipulation of an object, with specific physical properties, to learn the adequate level of "carefulness" to use when handling that object. We construct dynamical models of the human behaviour using a human-to-human handover dataset consisting of 3 different cups with different levels of fillings. We then included these models into the design of an online classifier that identifies the type of action, based on the human wrist movement. We close the loop from action understanding to robot action execution with an adaptive and robust controller based on the learned classifier, and evaluate the entire pipeline on a collaborative task with a 7-DOF manipulator. Our results show that it is possible to correctly understand the "carefulness" behaviour of humans during object manipulation, even in the pick and place scenario, that was not part of the training set.

I. INTRODUCTION

Humans are capable of expressing their actions and intentions, resorting to verbal and/or non-verbal communication. In verbal communication, humans use language to express, in structured linguistic terms, the desired action they wish to perform. Non-verbal communication refers to the expressiveness of the human body movements during the interaction with other humans, while manipulating objects, or simply navigating in the world. In a sense, all actions that require moving our musculoskeletal system contribute to expressing the intention concerning the completion of that action. Moreover, considering that all humans share a common motor-repertoire, i.e. the degrees of freedom and joint limits, excluding cultural or society-based influences, all humans express action intentions using a common non-verbal language. From walking along a corridor, to pointing to a painting on a wall, or handing over a cup to someone, communication is provided in the form of non-verbal "cues", that express action intentions [1].

Endowing robots with the ability to understand human action intentions from non-verbal cues will broaden the

robots' use-case scenarios. This ability is especially useful in scenarios where humans and robots need to collaboratively manipulate objects. For example, imagine that a robot and a human are performing hand-overs of different types of objects. A few questions arise: does the type of object (e.g., fragile vs non-fragile) change the motion of the handover? does the amount of filling (e.g., full vs empty) when handing over a container change the type of behaviour? We hypothesize that non-verbal cues extracted from the human body movement can reveal relevant information regarding the manipulation of the object. In other words, the object's intrinsic physical properties will influence the action execution and, therefore, the non-verbal cues. In human-robot interaction (HRI), the interpretation of non-verbal cues provides the robot with relevant information concerning the object to grasp, which can be adapted during the interaction, to comply with the object's physical properties. Mayer and Krechetnikov [2] studied the impact on human walking when preventing a cup of coffee from spilling, and noticed that in the scenario where participants were focused on not spilling, the overall average speed was reduced compared to the group of participants who walked without caring about spilling the coffee. This fits in line with our hypothesis that cups full of coffee or water will invoke a different arm-hand manipulation approach, i.e. carefulness level, into the manipulation.

Our hypothesis hence formulates three questions: How does the human motion profile change depending on the object's properties? Can the motion be modelled to reliably classify human "carefulness" during object manipulation? Can it be integrated in a HRI scenario for online robot adaptation to object properties decoded from human movement?

To test the hypothesis, we collected a dataset of humans handing over cups. To evaluate the effect of the perceived fragility of the object, different materials were used (hard plastic vs soft plastic). We used two conditions related to the object properties, the cups were either full of water, or empty. Our results show that the velocity of the movement is a distinctive factor for "carefulness" (*careful* vs *not-careful*) in handovers. Subjects were significantly slower when the object was fragile or the carrying full-cups. Using these observations, we built two models of the dynamics of motion: one characterizing the velocity of the *careful* motion, and another one for the *not-careful* motion. Then, we introduced a classifier to identify, in real-time, whether the human wrist motion corresponds to *careful* or *not-careful* dynamics. Finally, we implemented an adaptive controller for a robot to adjust the movement's "carefulness" during the handover and manipulation of the cup.

II. Related Work

Neuroscientists have investigated the hypothesis that some human brain circuits, the mirror-neurons, may be involved in both the execution and observation of goal-oriented, physical actions. It has been suggested that mirror-neurons could result from evolutionary pressure, and greatly facilitate action understanding [3]. Alaerts et al. [4] note that mirror-neurons in the primary motor cortex (M1) are activated when observing people lifting objects and heavier objects induced higher M1 excitability. This corroborates the ability for humans to extract knowledge of objects' weight, fragility, or contents, from human motion. Studies on human non-verbal cues found that joint kinematics and dynamics of hand manipulation are crucial features for object weight estimation [5], action duration [6], and absolute velocities [7].

Sciutti et al. [8], [9] devise social interactions to analyse human reaction to diverse robot behaviours. Work from [10] points that human non-verbal cues from eyes, head, and arm movements decode action intention, and when incorporating onto a robot, it provides similar information to read the robot's intention. In the case of object manipulation, object affordances was popularized in robotics [11], [12], relating to (i) the action associated with the object, (ii) a physical property, or (iii) the type of behaviour required to manipulate the object. Kjellström et al. [13] inferred object affordances from human demonstrations to classify cups and glasses as drinkable and pourable. Work on affordance reasoning [14], [15], [16], which reasons with the object's properties, e.g. inferring the water level in cups [17], although water levels are only possible in transparent cups and glasses. Our approach explores the human kinematic motion to infer the impact of cups water content on human motion behaviour, irrespective of cup's transparency. Additionally, depending on the type of cup, these behaviours may be more predominant or less, which may indicate that the difficulty of manipulation impacts the human motion [18]. Therefore, we complement previous works and focus on the arm motion when varying object type and its content.

In robotics our work falls within the domain of learning from human demonstrations, with a particular focus on manipulation. Several works on grasping and handover motion by studying the visuomotor coordination in humans and robots during interactions [10], [19], taking inspiration on human grasping for adaptive robot grasps [20], [21], or learning how to grasp and release during handovers from humans [22], [23]. Learning robot dynamics model from human motion improves human-robot collaboration making it more predictable. As such, we intend to explore human-human interactions (HHI) to extract non-verbal cues that express "carefulness" during manipulation and handover of cups. This human non-verbal cues applied in robotics provide information for adaptive motion dynamics in accordance with the object's properties. To our knowledge there is no previous research on human "carefulness" identification from non-verbal cues during manipulation of cups or other objects, with distinct physical properties, and its applications to HRI.

III. Human Motion Study

A. Dataset Description

A total of 6 participants (male, 25-35 years old, academic employees) took part in the experiments. The experimental task involves grasping a cup from a table and hand it over to a subject on the opposite side that places it back in the table (Figure 1). The experiment is performed under two
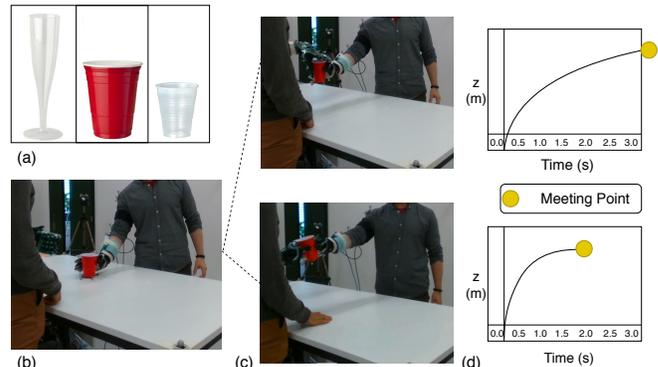

Fig. 1: Representation of handover actions. (a) the 3 types of cups: champagne cup, red cup, and plastic cup, in that order; (b) t_0 frame of hand-over action; (c) t_f frame is the final frame of a *not careful* motion (bottom) and a *careful* motion (top); (d) the duration of each type of motion.

conditions: (i) empty cup, and (ii) cup 90% filled with water. For each condition the experiment is repeated a minimum of 4 times per participant for each cup manipulated. There are 3 cups: red cup, plastic cup, and champagne cup. Each participant had to grasp the cups with their preferred hand (right hand for all) and there were no restriction on the type of grasp. Motion capture system (OptiTrack) recorded right-hand wrist's location for each participants as well as the cup's location. A total of 30 right-hand wrist trajectories were collected for either of the two conditions. This HHI dataset was gathered in collaboration with the High Performance Humanoid Technologies Lab (H2T) of the Karlsruher Institut für Technologie (KIT) [1] [24].

B. Motion Analysis

Fig. 2 shows that the duration and maximum velocity are affected by change in the object properties as mentioned in [6], [7]. The difference is detected between the two conditions: empty, referred as *not careful* motion, and full, as a *careful* motion. Following our hypothesis that for the same action the human wrist motion of the object manipulation changes depending on the human intention: whether the human wants to be *careful* or *not careful* in their motion. As stated in Section II, the object weight impacts the maximum speed of the lifting motion [6], in our human study it is observed that the human motion during handovers of cups full of water causes similar effects on the velocity and consequently the action's duration. It is then our decision

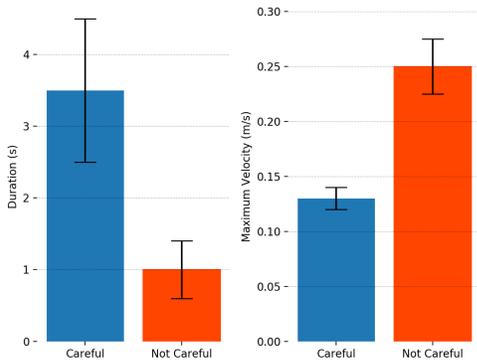to pick as input of our pipeline the human wrist position with respect to the handover meeting.

Fig. 2: Average and standard deviation for duration and maximum velocities of *careful* and *not careful* handovers.

IV. METHOD

A. "Carefulness" Detection Pipeline

The control loop of our pipeline shown in Fig. 3 is composed of three (3) distinct parts: (i) the model, (ii) the classifier, and (iii) the controller. The model refers to the two
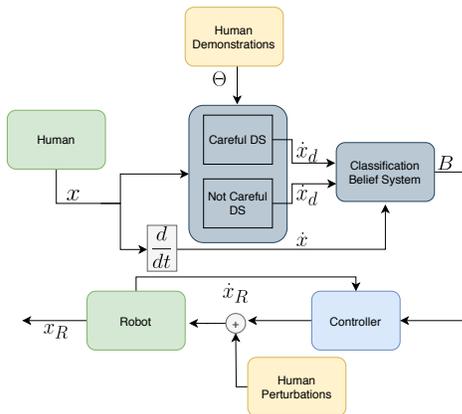

Fig. 3: Pipeline for "carefulness" detection in a HRI.

learned dynamical systems (DS) from human demonstrations of the HHI dataset: the first model represents the *not careful* behaviour of humans for handover of cups, and the second model represents the *careful* behaviours. From the human demonstrations we extract the important cues Θ (velocity profile, etc) which are used to build the DS. The model is explained in greater detail in Section IV-B, the classifier in Section IV-C, and the controller in Section IV-D.

The input of the pipeline is the human wrist position and during the manipulation of the cup the velocities are computed. The position of the human wrist is used by the two learned models in order to compute the desired velocity for the two "carefulness" conditions, *careful* and *not careful*. The real velocity of the human is then evaluated against the generated velocities of the models to provide a belief of the "carefulness" behaviour of the human. The pipeline

runs at every time step and the belief system is updated with new information from the human motion to classify whether the human is manipulating the cup in a *careful* or *not careful* motion. The output of the classifier is an instruction command to the controller that, depending on whether the human is performing a *careful* or *not careful* behaviour, the robot acts accordingly. This means that it will adapt the robot's behaviour to either grasp and move the cup in a *careful* or *not careful* manner. This distinction refers to the stiffness and smoothness of the robot's controller.

B. Model Learning from Human Demonstrations

Let $x \in D \subset \mathbb{R}^+$ denote the distance of the human wrist towards the handover meeting point. Consider a behavior encoded as a state-dependent dynamical system (DS)

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

where $\mathbf{f} : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is a continuous and continuously differentiable function, with a single equilibrium point $\dot{x}_d^* = \mathbf{f}(x^*)$. x^* is set at the origin and it is globally asymptotic stable such that $\dot{x}^* = \mathrm{f}(x^*) = 0$ which is guaranteed under a Lyapunov function $V(x) : \mathbb{R}^+ \rightarrow \mathbb{R}^+$.

Our approach defines each "carefulness" condition, *careful* and *not careful*, as two distinct DS. Each DS is encoded using Gaussian Mixture Models (GMM) which defines a joint distribution function $\mathcal{P}(x^t{}_n, \dot{x}^t{}_n | \Theta) = \sum_{k=1}^K \pi^k \mathcal{N}(x^t{}_n, \dot{x}^t{}_n, \mu^k, \Sigma^k)$ over the data as mixture of K Gaussian distributions [25], where π^k, μ^k, and Σ^k are, respectively, the prior component, mean, and covariance matrix of the kth Gaussian. $x^t{}_n$ is nth trajectory of x at time t, and $\dot{x}^t{}_n$ is its derivative. Fig. 4 illustrates the position (x) and velocity ($\dot{x}$) relations for *careful* and *not careful* motions. To compute the DS from Eq. (1) the posterior mean of $\mathcal{P}(\dot{x}^t{}_n | x^t{}_n)$ is estimated which approximates it to:

$$\hat{x} = \sum_{n=1}^K h^k(x)(\Sigma^k_{\dot{x}x}(\Sigma^k_{xx})^{-1}(x - \mu^k_x) + \mu^k_{\dot{x}}) \tag{2}$$

where $h^k(x) = \frac{\pi^k \mathcal{N}(x^t, \dot{x}^t, \mu^k, \Sigma^k)}{\sum_{i=1}^K \pi^k \mathcal{N}(x^t{}_n, \dot{x}^t{}_n, \mu^i, \Sigma^i)}$, $h^k(x) > 0$, and $\sum_{n=1}^K h^k(x) = 1$. The GMMs are computed using the stable estimator of dynamical systems (SEDS) approach [25].
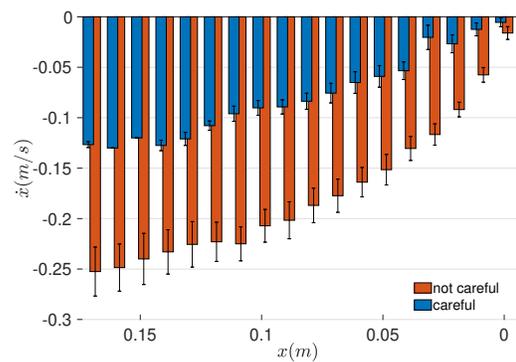

Fig. 4: Human handover trajectories from *careful* and *not careful* behaviours.

Fig. 4 shows that the distinction between *careful* and *not careful* is more noticeable in the beginning of the handover than at the end. Based on this observation, we propose a classifier to predict the type of behaviour *during* the movement instead of recognizing the action only after its completion. By doing so, we intrinsically embed an anticipatory capability to the overall pipeline.

C. Classification

For the purpose of classifying human motions, when interacting with either humans or robots, a belief system was implemented. The objective is to compare the human wrist motion of the handover against the learned *careful* vs *not careful* motions. The classification method uses as input the human wrist data (position, velocity) and, at each time step, it computes the desired velocities for the two DS models, the error metric to compare the velocities, and then outputs a belief system of the "carefulness" level. The expression for the classification follows [26] and is:

$$\dot{\hat{b}}_i^s = \epsilon(\dot{e}^T\mathbf{f}_i(x) + (b_i^s - 0.5)|\mathbf{f}_i(x)|^2) \qquad (3)$$

where $\dot{e} = \dot{x} - \dot{x}_d$, $\epsilon \in \mathbb{R}^+$ is the adaptation rate, $\dot{x}_d = \sum_{i=1}^2 b_i^s\mathbf{f}_i(x)$, $\mathbf{f}_i(x)$ is the desired velocity for each DS given the current x. b_i^s is the belief, at step s, for each DS in $i = \{1, 2\}$, and $\sum_{i=1}^2 b_i^s = 1$.

Eq. 3 provides a vector of belief-updates $\dot{\hat{B}}^s = [\dot{\hat{b}}_1^s, \dot{\hat{b}}_2^s]$ which are updated following a winner-take-all process. The winner-take-all aims at favoring the DS model which is considered most similar to the real human motion. The final step is reserved to update the belief $B = [b_1, b_2]$, where $b_i^{s+1} \leftarrow b_i^s + \dot{b}_i^s\Delta t$, for $i = 1, 2$. The belief system B converges to $b_1 = 1$ or $b_2 = 1$ depending on whether the human motion resembles a *not careful* behavior, or a *careful* behavior, respectively. Fig. 5 shows the output of B at each time step s for various human trajectories.

D. Robot Control

The robot is represented as a fixed rigid-body chain with n degrees of freedom. Since the desired end-effector's accelerations are known, we use an inverse dynamics formulation in order to compute the required joint-level torques needed to achieve the target accelerations. We formulate the problem as a quadratic programming problem (QP):

$$\min_{\mathcal{X}} -0.5\mathcal{X}^T\boldsymbol{G}\mathcal{X} + \boldsymbol{g}^T\mathcal{X}$$
$$\text{s.t. } \boldsymbol{A}_E\mathcal{X} = \boldsymbol{b}_E$$
$$\boldsymbol{A}_I\mathcal{X} \geq \boldsymbol{b}_I \qquad (4)$$

We will define the optimization variables, $\mathcal{X}$, and all the matrices and vectors (for the objective and the constraints) shortly. The equations of motion and constraint equations for a robot with rigid bodies can be described as[2]:

$$\boldsymbol{M}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \boldsymbol{C}_g(\boldsymbol{q},\dot{\boldsymbol{q}}) = \boldsymbol{\tau}$$
$$\boldsymbol{J}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \dot{\boldsymbol{J}}(\boldsymbol{q},\dot{\boldsymbol{q}})\dot{\boldsymbol{q}} = \ddot{\boldsymbol{x}} \qquad (5)$$

---

[2]We ignore the possible contact points for brevity/clarity as we are not using them in this work.

where $\boldsymbol{q}$ is the full state of the system , $\boldsymbol{M}(\boldsymbol{q})$ is the inertia matrix, $\boldsymbol{C}_g(\boldsymbol{q},\dot{\boldsymbol{q}})$ is the sum of the gravitational, centrifugal and Coriolis forces, $\boldsymbol{J}$ is the concatenation of the Jacobians of all the contact points, and $\boldsymbol{x}$ is the concatenation of the poses (containing position and orientation) in Cartesian space of all the contacts. x_R is the end effector pose of the robot. We re-write the equations of motion as:

$$\begin{bmatrix} \boldsymbol{M}(\boldsymbol{q}) & -\boldsymbol{S} \end{bmatrix} \begin{bmatrix} \ddot{\boldsymbol{q}} \\ \boldsymbol{\tau} \end{bmatrix} + \boldsymbol{C}_g(\boldsymbol{q},\dot{\boldsymbol{q}}) = 0 \qquad (6)$$

where $\boldsymbol{S}$ is a selection matrix where the first 6 rows are all zeros and the rest is the identity matrix. Given this formulation, the state $(\boldsymbol{q},\dot{\boldsymbol{q}})$ give rise to linear equations for the motion with respect to $\begin{bmatrix} \ddot{\boldsymbol{q}} & \boldsymbol{\tau} \end{bmatrix}^T$. By defining $\mathcal{X} = \begin{bmatrix} \ddot{\boldsymbol{q}} & \boldsymbol{\tau} \end{bmatrix}^T$, it is now possible to formulate the inverse dynamics as a QP problem. In particular, we turn the equations of motion to equality constraints ($\boldsymbol{A}_E$ and $\boldsymbol{b}_E$), and we turn joint limits and other constraints into inequality constraints ($\boldsymbol{A}_I$ and $\boldsymbol{b}_I$). Finally, we define desired accelerations of some end-effector by filling $\boldsymbol{G}$ and $\boldsymbol{g}$ appropriately. In this paper, we compute the desired end-effector accelerations by:

$$\ddot{\boldsymbol{x}}^* = K_p(\boldsymbol{x}_d^* - \boldsymbol{x}) + K_d(\dot{\boldsymbol{x}}_d^* - \dot{\boldsymbol{x}}_d) + \ddot{\boldsymbol{x}}_d^* \qquad (7)$$

where $\boldsymbol{x}_d^*, \dot{\boldsymbol{x}}_d^*, \ddot{\boldsymbol{x}}_d^*$ are specified by a higher-level controller, can change over time, and define the current task. Depending on the output of the classifier, different gains (K_p, K_d) are chosen to perform the high-level tasks.

V. RESULTS

In this section we first test the learned DS models with unknown subjects and cups present in the HHI dataset, and then we evaluate the "carefulness" pipeline in a HRI scenario.

A. HHI dataset

The DS models were learned with a training set composed of 3 subjects (identified as #1, #2, and #3). Fig. 5 illustrates the belief system B at each s step and Fig. 6 (a) represents the predicted output of B as a confusion matrix. The DS models distinguished correctly the *careful* and *not careful* motions for the training set. The testing set has 3 new subjects (#4, #5, and #6) and 2 new cups (champagne and plastic cup). Fig. 5 shows the classifier output of the training and testing set of human motion trajectories over time, while Fig. 6 (b) represents the final predicted believed/understood action for each subject and each cup.

The results presented from the HHI dataset indicate that our method can generalize and distinguish most situations of *careful* and *not careful* manipulations of cups, where the *not careful* situation is considered a normal human handover motion. As Fig. 5 demonstrates, it is only required around 20-40 time steps in the human handover motion to accurately predict the "carefulness" behaviour. We expect good generalization of the classification to new subjects and cups (not seen during training), as humans follow similar dynamics of reach motion in normal circumstances [27]. This would allow us to consider the manipulation of empty cups as *not careful* behaviours, and full cups as *careful* behaviours.
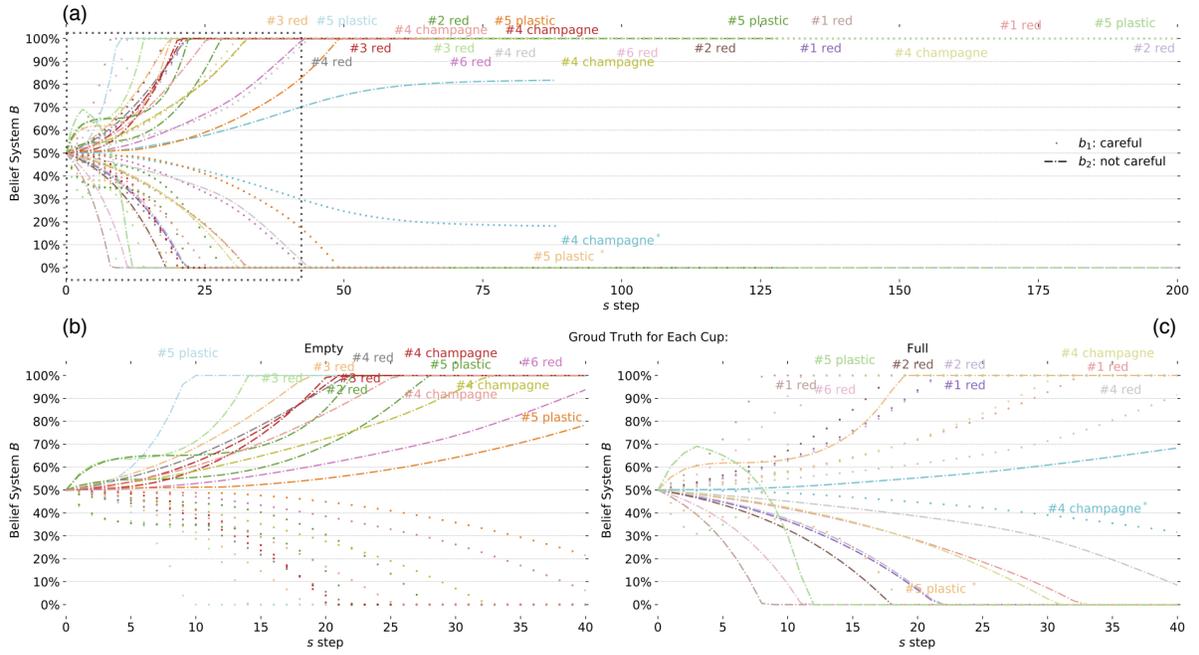
Fig. 5: The belief system B output. (a) full length of HHI trajectories. #n ${cup}$ is the label of the n participant and the type of cup grasped. Additionally, the position of the label marks the classification result of the final s step of the handover trajectory. (b) and (c) are the highlighted region in (a) for the two conditions: empty, and full, respectively. The $*$ represents the trajectories with a wrong classification.
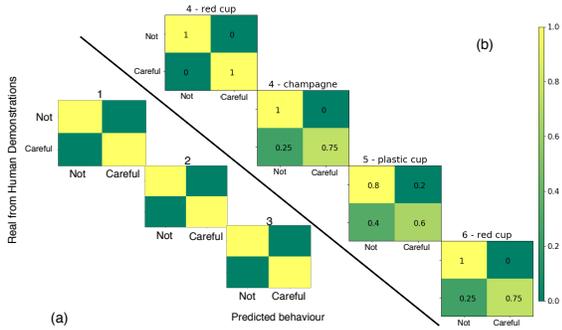

Fig. 6: Confusion matrices: (a) training set; (b) testing set.

B. HRI scenario of Pick and Place

1) Setup: We used a KUKA LBR iiwa 7-DoF manipulator ($14\,\text{kg}$ payload)[3] and the Robotiq 2F-85 2-finger gripper[4] is attached to robot's end-effector to perform grasping and manipulation. The HRI scenario is as follows: a human picks a cup from a table and places it on a shelf, as shown in Fig. 7. The cup's position is provided to the pipeline as the input during human manipulation. The cup's position is set as reference, instead of the human wrist, to simplify the experiments with different participants. The human motion is extracted from the cup's motion and the classifier predicts whether we are in the presence of a *careful* or *not careful* motion. Afterwards, the controller from Section

[3]https://www.kuka.com/en-ch/products/robotics-systems/industrial-robots/lbr-iiwa/

[4]https://robotiq.com/products/2f85-140-adaptive-robot-gripper/

IV-D adapts the robot's motion and gripper to the desired behaviour. Depending on the output of belief system, the robot controller manipulates the cup keeping the gripper's orientation fixed, to prevent spilling (*careful*), or allows for an unrestricted movement (*not careful*). The supplementary material provides video demonstrations.

2) Experimental Results: The HRI experiments involved 4 participants picking and placing the same cup onto the shelf under two conditions: (i) empty, and (ii) full cup. Each participant performed 10 trials for each condition. The cup used was not present in the HHI dataset. The results of the classification In Table I for new subjects, with a new cup, proves that our pipeline can correctly distinguish *careful* and *not careful* manipulation of cups by only varying one underlying condition: empty vs full of water.
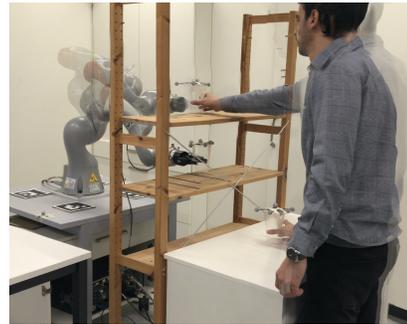

Fig. 7: Setup outside perspective for the Pick and Place task.

The pipeline invokes the robot to behave according to the desired "carefulness" (i.e. the type of object and filling con-

TABLE I: "Carefulness" level predicted on unknown people.

True \ Predicted	Not Careful	Careful
Empty	**0.80**	0.20
Full	0.07	**0.93**

dition), taking longer in the *careful* case. In the *not-careful* behaviour, the robot grasps the cup and places it a manner that is simpler for the robot's configuration, taking less time to complete the action. The downside is that the robot spills its content. The approach developed allows for real-time HRI where the robot can adapt the behaviour of manipulating a cup according to the human motion behaviour.

VI. CONCLUSION

We hypothesise that human motion behaviour is dependent on object properties, specifically on object's weight, liquid fillings, or their overall structural integrity. In this paper we collected and analyzed a human-to-human handover dataset in an attempt to identify the crucial features that affect the motion of the manipulation given the "carefulness" required. We identified that the velocity is one of the main factors and that for cups filled with water the human motion got slower. Moreover, the differences were clearer in the start of the human motion which allowed us to construct a model-based classifier that is capable of predicting whether the human is *careful* when manipulating a cup or *not careful*. Furthermore, our results enabled us to develop a pipeline capable of utilizing these knowledge to adapt the robot's grasp and manipulation controller during cup manipulation. This gave a successful collaboration where action understanding from non-verbal cues in human motion object manipulation translated into an adaptive controller for action execution of robot object manipulation. As future work we intend to explore features, such as the cup's orientation, and extend our pipeline to other HRI scenarios.

REFERENCES

[1] N. Mavridis, "A review of verbal and non-verbal human–robot interactive communication," *Robotics and Autonomous Systems*, vol. 63, pp. 22 – 35, 2015.

[2] H. C. Mayer and R. Krechetnikov, "Walking with coffee: Why does it spill?," *Physical Review E*, vol. 85, no. 4, p. 046117, 2012.

[3] M. Iacoboni and M. Dapretto, "The mirror neuron system and the consequences of its dysfunction," *Nature Reviews Neuroscience*, vol. 7, no. 12, p. 942, 2006.

[4] K. Alaerts, P. Senot, S. P. Swinnen, L. Craighero, N. Wenderoth, and L. Fadiga, "Force requirements of observed object lifting are encoded by the observer's motor system: a tms study," *European Journal of Neuroscience*, vol. 31, no. 6, pp. 1144–1153, 2010.

[5] J. Rosen, J. C. Perry, N. Manning, S. Burns, and B. Hannaford, "The human arm kinematics and dynamics during daily activities - toward a 7 dof upper limb powered exoskeleton," in *ICAR '05. Proceedings., 12th International Conference on Advanced Robotics, 2005.*, pp. 532–539, 2005.

[6] A. d. C. Hamilton, D. Joyce, J. Flanagan, C. Frith, and D. Wolpert, "Kinematic cues in perceptual weight judgement and their origins in box lifting," *Psychological research*, vol. 71, no. 1, pp. 13–21, 2007.

[7] A. Sciutti, L. Patanè, and G. Sandini, "Development of visual perception of others' actions: Children's judgment of lifted weight," *PLOS ONE*, vol. 14, pp. 1–15, 11 2019.

[8] 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, pp. 22–29, March 2018.

[9] O. Palinko, A. Sciutti, L. Patané, F. Rea, F. Nori, and G. Sandini, "Communicative lifting actions in human-humanoid interaction," in *2014 IEEE-RAS International Conference on Humanoid Robots*, pp. 1116–1121, IEEE, 2014.

[10] N. F. Duarte, M. Raković, 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.

[11] G. Saponaro, P. Vicente, A. Dehban, L. Jamone, A. Bernardino, and J. Santos-Victor, "Learning at the ends: From hand to tool affordances in humanoid robots," in *2017 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, pp. 331–337, IEEE, 2017.

[12] L. Jamone, E. Ugur, A. Cangelosi, L. Fadiga, A. Bernardino, J. Piater, and J. Santos-Victor, "Affordances in psychology, neuroscience, and robotics: A survey," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 1, pp. 4–25, 2016.

[13] H. Kjellström, J. Romero, and D. Kragić, "Visual object-action recognition: Inferring object affordances from human demonstration," *Computer Vision and Image Understanding*, vol. 115, no. 1, pp. 81–90, 2011.

[14] Y. Yamani, A. Ariga, and Y. Yamada, "Object affordances potentiate responses but do not guide attentional prioritization," *Frontiers in integrative neuroscience*, vol. 9, p. 74, 2016.

[15] M. Hassanin, S. Khan, and M. Tahtali, "Visual affordance and function understanding: A survey," *arXiv preprint arXiv:1807.06775*, 2018.

[16] L. Yu, N. Duncan, and S. Yeung, "Fill and transfer: A simple physics-based approach for containability reasoning," in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 711–719, Dec 2015.

[17] R. Mottaghi, C. Schenck, D. Fox, and A. Farhadi, "See the glass half full: Reasoning about liquid containers, their volume and content," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1871–1880, 2017.

[18] D. A. Rosenbaum, K. M. Chapman, M. Weigelt, D. J. Weiss, and R. van der Wel, "Cognition, action, and object manipulation.," *Psychological bulletin*, vol. 138, no. 5, p. 924, 2012.

[19] H. Admoni, T. Weng, B. Hayes, and B. Scassellati, "Robot nonverbal behavior improves task performance in difficult collaborations," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, HRI '16, p. 51–58, IEEE Press, 2016.

[20] C. Della Santina, V. Arapi, G. Averta, F. Damiani, G. Fiore, A. Settimi, M. G. Catalano, D. Bacciu, A. Bicchi, and M. Bianchi, "Learning from humans how to grasp: a data-driven architecture for autonomous grasping with anthropomorphic soft hands," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1533–1540, 2019.

[21] C.-M. Huang, M. Cakmak, and B. Mutlu, "Adaptive coordination strategies for human-robot handovers," 07 2015.

[22] J. R. Medina, F. Duvallet, M. Karnam, and A. Billard, "A human-inspired controller for fluid human-robot handovers," in *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, pp. 324–331, Nov 2016.

[23] K. Strabala, M. K. Lee, A. Dragan, J. Forlizzi, S. Srinivasa, M. Cakmak, and V. Micelli, "Towards seamless human-robot handovers," *Journal of Human-Robot Interaction*, January 2013.

[24] 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)*, pp. 72–78, IEEE, 2019.

[25] S. M. Khansari-Zadeh and A. Billard, "Learning stable nonlinear dynamical systems with gaussian mixture models," *IEEE Transactions on Robotics*, vol. 27, no. 5, pp. 943–957, 2011.

[26] M. Khoramshahi and A. Billard, "A dynamical system approach to task-adaptation in physical human–robot interaction," *Autonomous Robots*, vol. 43, no. 4, pp. 927–946, 2019.

[27] A. Lemme, Y. Meirovitch, M. Khansari-Zadeh, T. Flash, A. Billard, and J. J. Steil, "Open-source benchmarking for learned reaching motion generation in robotics," *Paladyn, Journal of Behavioral Robotics*, no. 1, 2015.