Online Prediction of Robot to Human Handover Events Using Vibrations

Harmeet Singh, Marco Controzzi, Christian Cipriani
Artificial Hands Area
The BioRobotics Institute
Scuola Superiore Sant'Anna
Pisa, Italy
h.singh@santannapisa.it

Gaetano Di Caterina, Lykourgos Petropoulakis, John Soraghan
Electronic and Electrical Engineering
University of Strathclyde
Glasgow, Scotland
Gaetano.di-caterina@strath.ac.uk

***Abstract*—One of the main issues for a robotic passer is to detect the onset of a handover, in order to avoid the object from being released when the human partner is not ready or if some impact occurs. This paper presents the methodology for a robotic passer, that is potentially able to estimate the interaction forces by the receiver on the object, thus to achieve fluent and safe handovers. The proposed system uses a vibrator that energizes the object and an accelerometer that monitors vibration propagation through the object during the handover. We focused on the machine-learning technique to classify between four states during object handover. A neural network was trained for these four states and tested online. In experimental trials an accuracy of 85.2% and 93.9% were obtained respectively for four classes and two classes of actions by a neural network classifier.**

Keywords—Autonomous; Handover; Machine Learning; Neural networks

I. INTRODUCTION

One of the main goals of robotic research is to enable robots to work alongside humans in different fields such as in industrial [1][2] and domestic settings [3]. For example, a human can ask for a tool from a robot in one scenario and for a workplace assistance in another [4]. Within this framework object handover between robots and humans is among the fundamental collaborative tasks that should occur in a safe and fluent manner: notably, these are two opposite requirements. In some cases, humans can ask for the object while not facing the robot. Therefore, a robot should know when the receiver is ready to receive the object and hand it over without dropping it. A robot should be able to distinguish between a mild touch, proper grasp or just a contact (hit), which could be accidental, by a human. Otherwise, the robot can make an incorrect decision which can harm the human and/or the object.

Humans hand over objects to one another efficiently and fluently as part of daily life, while robot-to-human handovers are unintuitive and often inadequate [5]. The way in which an object is released by the hand of a robotic passer plays a fundamental role in the perceived fluency of the handover action by the human receiver. Behavioral studies in humans have highlighted how, during object handover, the passer and the receiver regulate their grip forces in complementary fashion in order to accomplish a successful transfer [6]. To transfer an object, the passer must decrease the grip force on the object while the receiver increases it. During this time, both partners must share the production of adequate grasp forces on the object to counteract gravitational and inertial load forces, in order to prevent the object from slipping. Finally, when the receiver has produced grip forces strong enough to maintain stable control on the object, the passer can release the grasp [6][7]. The instant when the receiver makes initial contact with the object is known to be a salient event of the handover, because it alerts the passer to start the load transfer. Between humans, however, not all the contacts trigger a handover, e.g. when unwanted collisions occur or when the contact is due to a mild touch by the receiver. These cases are mastered with apparent ease by humans, which determine whether to start the handover based on their sensory inputs (including non-verbal communication) and on their previous experience. In particular, the passer is in charge for the safety/fluency of the handover, by adopting context-dependent safety margins on the grip force [6].

In a scenario where the passer is an autonomous robot with restricted cognitive capabilities and sensory inputs and the receiver is a human, there should be an in-built mechanism in the robot to identify whether the human has grasped the object properly and thus is ready for the handover. The robot should be able to distinguish between *mild touch* and *hit* by the receiver. Traditionally this issue has been tackled using force/torque sensors in the arm and end-effector, in order to monitor the reaction forces produced by the actions on the object by the receiver. In addition, systems that use computer vision to predict the partner's intentions have been developed [8].

Many studies have been carried out on various aspects of human-robot handovers. Several methods have been proposed that help the robot determine when to release the object during object handover. *Edsinger and Kemp* [9] proposed to trigger the handover at a fixed delay with respect to the first touch, as detected by tactile sensors in the gripper. However, it has been shown that the initial phase of the handover, i.e. immediately after the object contact, is crucial for the subjective impression of the robot behavior's quality, and a perceived latency in the control after this event is generally not acceptable [10].

This work was partially funded by the European Research Council under the MYKI project (ERC-2015-StG, Grant no. 679820) and partially by INAIL (the Italian national workers' compensation) under the PPR AS 1/2 - CECA 2020 project.

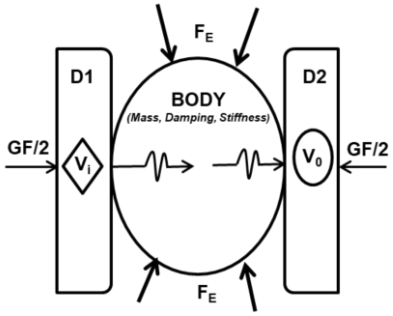


Fig. 1. Representation of the proposed method. GF is the Grip Force, Vi and Vo are the input (forced) and output vibration,respectively. FE are external forces

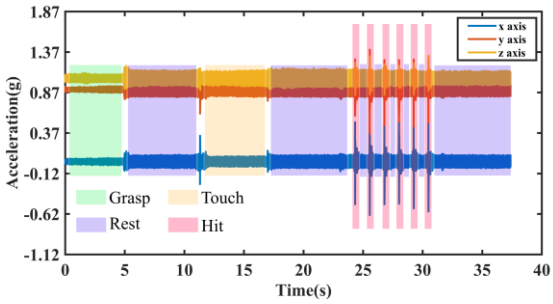


Fig. 2. Time series recorded with the experimental setup.

However, in some situations the robot can drop the object if the receiver is not ready after that fixed delay. Likewise *Choi et al.* [11] used force/torque sensors in the digits and a simple threshold technique. This method requires less force, but manual tuning of hardware-dependent thresholds is needed and the robot can drop the object due to an unexpected collision/hit. *Bohren et al.* [12] proposed to monitor the position in space of a compliant robot arm in order to derive the pulling force applied by the receiver on the object, and to use such information to trigger the load transfer. In this approach there is less chance of a drop, but the user has to pull very hard. Finally, *Chan et al.* [13] proposed a releasing algorithm based on the weight experienced on the end-effector as measured by a load cell in the wrist. Once the controller detects the slightest pull from the receiver, it releases the object.

In general the above mentioned methods are unable to effectively distinguish between the three possible conditions, namely *mild touch*, *proper grasp* and *hit*. However, we argue that the reason for these limitations actually resides in the sensing approach used, which is inappropriate for measuring the grip force applied by the receiver. Contrary to the conventional load/torque sensors, we propose a new method for a robotic passer, which is potentially able to estimate the interaction forces exerted on the object by the receiver and, therefore, able to implement acceptable fluent and safe handovers.

In this paper, we introduce a novel approach which can help the robot release the object at the right moment with less complexity. The general idea is to energize the grasped object with a mechanical vibration, and to monitor how such vibration propagates through the object due to its interaction with the environment. To find out the intensity of the grip force applied by the receiver on the object during the handover, we focus on machine-learning techniques to classify between different states

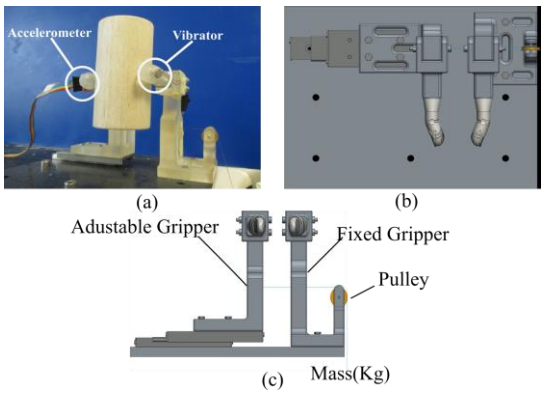


Fig. 3. Experimental setup: (a)Gripper with accelerometer on one digit and vibrator on other digit (b) upper view (c) fornt view of the setup.

during object handover rather than to use thresholds to detect just the touch by the receiver. The remainder of this paper is organized as follows. Sections 2 discusses the methodology used in the proposed design, Sections 3 and 4 discuss the data recording and data analysis respectively. The experimental results are reported and analyzed in Section 5 while Section 6 concludes the paper.

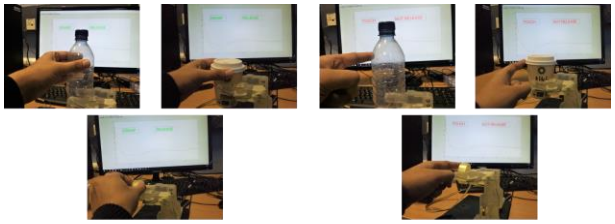
II. METHODOLOGY

There are several potential machine-learning techniques such as support vector machine (SVM), artificial neural network (ANN) and linear discriminate analysis (LDA) that have been used in the last decade [14]. We have used ANN for the online validation of our system.

We have designed an experimental setup with a small vibrator on one digit and an accelerometer on another digit of a gripper to hold objects by applying constant external force. Then we have recorded the variation when the human hand touches, hits or grasps the object. In our experiment we have used the gripper as passer a human as a receiver and 3 objects. Then, based upon the vibration analysis, the algorithm is able to classify the strength of the receiver grasp and thus to safely and efficiently trigger the beginning of the handover. The basic concept in this work is to provide a forced vibration to the system "end-effector – object" and to measure the propagated vibration modes using a sensor embedded on the robot. Every material vibrates at some frequency and its vibration changes when someone touches that vibrating material. The method here proposed measures these changes and helps the robot to differentiate between a *mild touch*, *proper grasp* and *hit* by a human. As illustrated in the simplest configuration in Fig. 1, the forced vibration is provided by a motor with rotating eccentric mass, placed on one digit, and is sensed by an accelerometer placed on the other digit.

In stationary conditions, the mechanical vibration of the system as sensed by the accelerometer V_o, will be ruled by a function *F*, based on a multitude of variables, which include: the grip force (GF), the features of the forced vibration (Vi), the dynamic properties of the system (mass *m*, damping *c*, stiffness *k*), and the external interacting forces (F_E), as in (1):

$$V_o = F\,(GF, V_i, m, c, k, F_E) \qquad (1)$$



(a) (b)

Fig. 4. Subject performing (a) Grasp (b) Touch with the three objects.

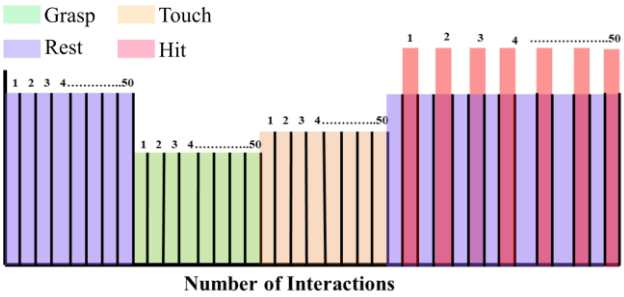


Fig. 5. Sequence of the interactions

As illustrated in Fig. 2, when the system is perturbed, i.e. the external force F_E is not null or the dynamic features (m, c, k) are modified due to the interaction with external bodies, the vibration picked up by the accelerometer will change accordingly. At first approximation changes may pertain to the vibration amplitude and frequency, or a combination of these. For example, one can expect that a *mild touch* will slightly reduce the amplitude of the picked up vibration, whereas an actual *grasp* (two counteracting forces on the object by an external agent) will attenuate the V_o to a larger degree. Finally, a *collision/hit* will produce an impulse in the measured vibration as illustrated in Fig. 2.

III. DATA RECORDING

A. Experimental Setup

The experimental setup was developed to prove the working principle of the proposed method, and comprises a bi-digital gripper as illustrated in Fig. 3. In the handover scenario, the gripper acts as the passer, holding the object in place, while the subject, i.e. a human, acts as the receiver. The subject had to perform different interactions with the object either by touching, grasping or hitting it. The gripper was able to hold objects with diameter ranging between 20-90 mm and can produce constant grip force (GF). One digit was equipped with a miniature vibrator (Model 308-107, Precision Microdrives Ltd) and, the other digit with an analogue accelerometer (ADXL 335, Analog Devices Inc.). Data was recorded with a PC through a data acquisition board (NI-USB 6009, National Instruments Corp.) with the rate of 1 kHz.

As the damping due to the introduced vibration is a function of the mass and the compliance of the object, three different objects with widely different properties were considered: an empty plastic bottle (very soft, long, 30 grams), coffee cup (compliant, medium, 10 grams) and white tape (stiff, small, 5 grams).

TABLE I. TOTAL NUMBER OF TRIALS

Object	Trails	Grasp	Touch	Hit
Bottle	10	500	500	500
Cup	10	500	500	500
Tape	10	500	500	500
	30	**1500**	**1500**	**1500**

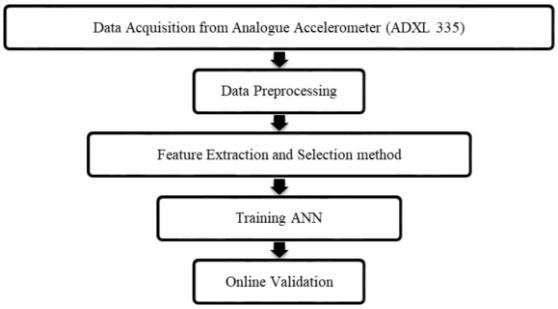


Fig. 6. Flowchart purpose illustrating the classification methodology.

We applied a constant GF of 5N on the gripper during the whole experiment.

B. Procedure

For the grasp condition, the subject had to grasp the object in the same manner as in real handover scenario of taking the object from another human. For this condition the subject was allowed to use all fingers as shown in Fig. 4(a).For the touch condition, the subject was allowed to use only one finger to touch the object as shown in Fig. 4(b), and in the hit condition the subject was instructed to hit the object repeatedly. All three interactions were performed in the following sequence: 50 grasps, 50 touches and 50 hits. There was a random rest phase i.e. when there was no interaction with an object occurring between hits. The trial started with the rest phase and then followed by the interactions. In one trial 50 grasps, 50 touches and 50 hits, as shown in Fig. 5, were recorded. For each object there were 10 trials, with a total of 30 trails for all three objects. The whole dataset comprises of 1500 grasps, 1500 touches and 1500 hits as shown in TABLE I.

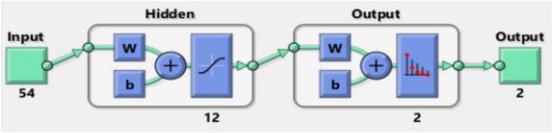
IV. DATA ANALYSIS

A. Preprocessing and feature extraction.

A flowchart of the methodology carried out in this study is shown in Fig. 6. The raw data was filtered using high pass and low pass third order Butterworth filters, to remove frequencies lower than 5Hz and higher than 300Hz. Then the filtered signal was processed according to different window sizes and overlaps. We selected window sizes from 100ms to 300ms with an increment of 100ms. For each window size, overlaps of 0%, 25%, 50%, 75% and 95%. From our experiments we have selected the window size 300ms for online validation of our system, as results indicated that this window size gives the best classification results in this case.

In this study all the features are extracted from the three channels, i.e. axes, of the accelerometer.

(a)



(b)

Fig. 7. ANN for (a) four classes (b) two classes

The fundamental purpose of feature extraction is to emphasize the important information in the measured signal, while rejecting the noise and irrelevant data.

The time domain features that are extracted include [15]:

- Standard Deviation (STD);
- Mean absolute value (MAV);
- Peak-Peak amplitude (PPA);
- Median absolute value (MEDAV);
- Autoregressive model (AR), 4 coefficients;
- Waveform length (WL);

Time–frequency representation (TFR) can localize the energy of the signal both in time and in frequency, thus allowing a more accurate description of the physical phenomenon. On the other hand, TFR generally requires a transformation that could be computationally heavy. In this study Discrete Wavelet Transform (DWT) has been used as TFR [15].

Feature selection is the process of selecting a subset of relevant features for the use in model construction. The main objectives of the feature selection technique are: (1) to enhance generalization by reducing overfitting, (2) to provide faster and cost effective models (3) to avoid issues arising due to dimensionality. For Wavelets, we decomposed the signal till the 5^{th} level. But there was no useful information in the first and the last levels. So we have used only the 2^{nd}, 3^{rd} and 4^{th} levels for the processing of the data. For each level three features were extracted, namely MAV, Singular Value Decomposition (SVD) and WFL. Therefore, for each axis there were 9 features (3 features x 3 DWT levels), with a total of 27 features for the three axis. So we performed classification with combinations of DWT and time domain features (54 features in total).

B. Classification

The standard ANN used for function fitting was a two-layer feed-forward network, with Bayesian Regularization, as illustrated in Fig. 7. As hidden layer, 12 neurons was used, which gave the minimum training model error with Neural Network in MATLAB 2016a. To train the ANN model, 3 trails were used excluding the online validation trails.

TABLE II. CLASSIFICATION ACCURACY FOR WINDOW SIZE OF 300ms

Overlap (%)	Accuracy (%)
0	79.39
25	79.46
50	79.44
75	79.19
95	78.23

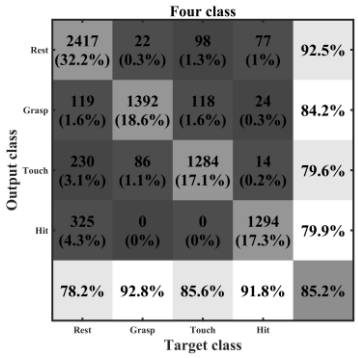


Fig. 8. Confusion matrix for four classes validation

To create the ANN model, 70% of 3 trails was used as training data, 15% as testing data and 15 % as validation data. The ANN model was trained using all three objects and then validated online.

The ANN was trained to classify the data into four classes (Fig. 7(a)), and then into two classes (Fig. 7(b)). The four classes were *rest*, *grasp*, *touch* and *hit*. Then we generalized these four classes into just two classes, i.e. "release" class (which includes the *grasp* state) and "not release" class (which includes the *rest*, *touch* and the *hit* states). The main rationale for generalizing the four classes was to improve the decision making of the system, because even if there are four possible classes, effectively the system makes a binary decision, either to release the object or not. So with just two classes higher accuracy could be obtained as compared to the four classes case.

V. RESULTS

We measured the online validation results for four classes and two classes separately.

A. Window size

TABLE II shows the classification results for a single window size (300ms) with different levels of overlaps. The results clearly indicate that the accuracy is not affected by the percentage of overlap. We have selected the 0% overlap for the online validation of our system. The main reason for selecting this overlap is that the online processing takes 400ms to process data, i.e.100ms more than the window size. Strictly speaking, the system is not real-time as the processing time is longer than the acquisition time; however, the delay of 100ms for every output form the ANN classifier is comparable with the delay associates to the time response of the tactile feedback in healthy humans [16].

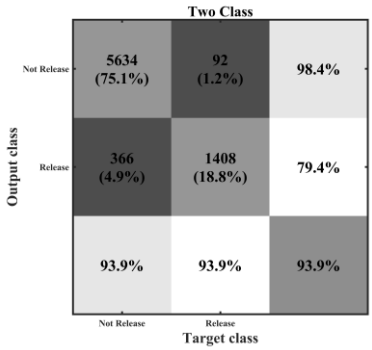


Fig. 9. Confusion matrix for two classes validation

B. Four class validation

The confusion matrix for four class validation is shown in Fig. 8. An overall accuracy of 85.2% was achieved. The maximum accuracy was of 92.8% for the grasp class. For the hit class, out of 1500 hits, the system was able to correctly classify 1409 hits, therefore achieving an accuracy of 91.8%. However for the hit class and the rest class the accuracy was less compared to the other two classes. There was a rest phase between each hit; that is why the classifier was confused between the *hit* class and the *rest* class. Nonetheless 91.8% of hits were classified correctly regardless of the 78.2% accuracy of the *rest* class.

C. Two class validation

The confusion matrix for the two class validation is shown in Fig. 9. The overall accuracy when compared to the four class validation has increased to 93.9%. In this case the following three classes *rest*, *touch* and *hit* were taken as only one class (i.e. the not release class), while the *grasp* class as the release class. Therefore system accuracy has increased as it only needs to decide whether to release the object or not release the object.

VI. CONCLUSION

This paper described the development of a new system that is able to effectively, and online detect different events like *rest*, *grasp*, *touch* and *hit* during human robot interaction. The main motivation was to reduce the complexity in the present system for detecting handover events. We obtained accuracy of 85.2% for four classes and 93.9% for two classes, with a delay of 100ms for the detection of the events.

Future research steps include experiments with a large number of objects and the implementation of this algorithm on a robotic system to test it in industrial and domestic environments.

REFERNCES

[1] C. Schou, J. Damgaard, S. Bogh and O. Madsen, "Human-robot interface for instructing industrial tasks using kinesthetic teaching," *2013 44th International Symposium on Robotics (ISR)*, Seoul, 2013.

[2] Stefanos Nikolaidis and Julie Shah, "Human-robot cross-training: Computational formulation, modeling and evaluation of a human team training strategy," , 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI). Tokyo, 2013.

[3] Q. Zhao, D. Tu, S. Xu, H. Shao and Q. Meng, "Natural human-robot interaction for elderly and disalbled healthcareapplication," *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2014.

[4] A. Graser, T. Heyer, L. Fotoohi, U. Lange, H. Kampe, B. Enjarini, S. Heyer, C. Fragkopoulos and D. Ristic-Durrant, "A Supportive FRIEND at Work: Robotic Workplace Assistance for the Disabled," *IEEE Robotics Automation Magazine*, vol. 20, no. 4, pp. 148-159, 2013.

[5] Maya Cakmak, Siddhartha S. Srinivasa,Min Kyung Lee,Sara Kiesler and Jodi Forlizzi, "Using spatial and temporal contrastfor fluent robot-human hand-overs," *Proceedings of the 6th international conference on Human-robot interaction*, Lausanne, 2011.

[6] Andrea H. Mason and Christine L. MacKenzie, "Grip forces when passing an object to a partner," *Exp Brain Res*, pp. 173-187, 2005.

[7] Mark R. Cutkosky and James M. Hyde, "Manipulation Control with Dynamic Tactile Sensing," *6th ISRR*, Hidden Valley, Pennsylvania, Oct. 2-5 1993.

[8] Andrea Zunino, Jacopo Cavazza, Atesh Koul, Andrea Cavallo, Cristina Becchio and Vittorio Murino, "Predicting Human Intentions from Motion Only: A 2D+3D Fusion Approach," *Proceedings of the 2017 ACM on Multimedia Conference*, 2017.

[9] Aaron Edsinger and Charles C Kemp, "Human-robot interaction for cooperative manipulation: Handing objects to one another," *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*, 2007.

[10] M Controzzi, M. Rank, I Strazzulla, C Peccia, AM Wing and C Cipriani, "Human-inspired release controller for natural robot to human handover tasks," *workshop on Behaviour Coordination between Animals, Humans, and Robots, within the Human-Robot Interaction international Conference*, March 2-5, 2015.

[11] Young Sang Choi, Tiffany Chen, Advait Jain, Cressel Anderson, Jonathan D. Glass, and Charles C. Kemp, "Hand It Over or Set It Down: A User Study of Object Delivery with an Assistive Mobile Manipulator," *Robot and Human Interactive Communication, RO-MAN 2009. The 18th IEEE International Symposium*, 2009.

[12] Jonathan Bohren, Radu Bogdan Rusu, E. Gil Jones, Eitan Marder-Eppstein and Caroline Pantofaru, "Towards Autonomous Robotic Butlers: Lessons Learned with the PR2," *Robotics and Automation (ICRA)*, 2011.

[13] Wesley P Chan, Chris AC Parker, HF Machiel Van der Loos and Elizabeth A Croft, "A human-inspired object handover controller," *The international Journal of Robotics Research*, pp. 971-983, 2013.

[14] F Lotte, M Congedo, A Lécuyer, F Lamarche and B Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces.," *Journal of Neural Engineering* , vol. 4, pp. R1-R13, 2007

[15] M. Zecca, S. Micera, M. C. Carrozza and P. Dario, "Control of Multifunctional Prosthetic Hands by Processing the Electromyographic Signal," *Critical Reviews™ in Biomedical Engineering,* vol. 30, no. 4-6, pp. 459-485, 2002.

[16] G.Westling and R. S. Johansson, "Signals in tactile afferents from the fingers eliciting adaptive motor responses during precision grip," *Experimental Brain Research*, vol. 66, no. 1, pp. 141-154, 1987.