Understanding Preferred Robot Reaction Times for Human-Robot Handovers Supported by a Deep Learning System

Jan Leusmann
LMU Munich
Munich, Germany
jan.leusmann@ifi.lmu.de

Ludwig Felder
TU Munich
Heilbronn, Germany
ludwig.felder@tum.de

Chao Wang
Honda Research Institute EU
Offenbach am Main, Germany
chao.wang@honda-ri.de

Sven Mayer
LMU Munich
Munich, Germany
info@sven-mayer.com

***Abstract*—Human-human handovers are natural and seamless. To be able to do this, humans optimize towards many factors. One of them is the timing when receiving an object. However, the preferred robot reaction time in Human-Robot handovers is currently unclear. To understand the preferred robot reaction time, we trained an Space-Time-Separable Graph Convolutional Network (STS-GCN) model using motion capture data of human-human handovers. We deployed this system on a robotic arm with live depth camera data. We conducted a user study (N=20) with five robot reaction times. We found that users perceived an early prediction as preferred. Furthermore, we found that designers can adapt this timing to their needs based on six sub-components of user perception. We contribute a ready-to-deploy handover classification model, a preferred handover time for our system, and an approach to determine the preferred robot reaction time for robotic systems.**

***Index Terms*—human-robot interaction, handover, human-to-robot handover, preferred timing, cobot, robot.**

I. Introduction

Humans can intuitively interact naturally, seamlessly navigating complex social cues and physical coordination, especially in simple yet fundamental tasks like handing over objects. Additionally, humans tend to understand systems as human partners expecting similar capabilities and interactions [1]. Thus, efficient human-robot teams should have similar abilities to create the same fluent collaborations, requiring a bi-directional understanding of each other's intentions [2]. Predictive robot behavior can improve effectiveness, fluency, and perception in human-robot interaction [3]. These skills would enable robots to assist humans in different environments, from handing over parts in an industrial working space to assisting people with limited reach in domestic environments [4]. Thus, we see a gap in robots conveying the intentions of their planned reactions to the human object hand approach.

Previous research shows that many factors (e.g., grip, load force, gaze, and eye contact) are essential for fluent handovers between robots and humans [4], [5]. Prior work on human interaction led to the observation that the giver in a handover is responsible for the safety and successful handover of the object. At the same time, the receiver is responsible for the handover timing [6]. From this, in a human-to-robot handover setting, the timing of the robot and, therefore, the reaction time are crucial for the success of a handover. This responsibility leaves the robotic partner with a challenging spatial and temporal precision task in handover interactions. Various previous works focus on forecasting and classifying human motion data using Machine Learning (ML), concluding handover intentions [7]–[10]. Those works show promising results in recognizing handover intentions between humans and robots. Previous work showed humans prefer faster behaviors over slower ones, leading to efficient handovers [11], [12]. Evaluating these interactions is also a well-researched topic, which nevertheless needs more generalizable metrics and benchmarks [13], showing the importance of user studies validating their applicability. While previous studies have shown the importance of timing, how humans perceive them, and how evaluations can look, the question of optimizing reaction timings remains open.

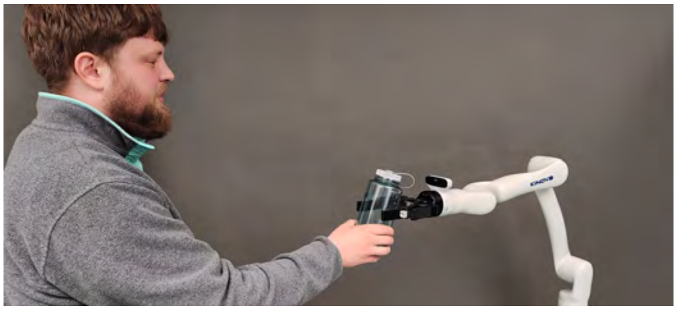

Fig. 1. A robot reacting to the users' handover approach.

This work aims to understand the preferred reaction times of robots taking an object from a user. First, we developed a deep learning system to detect human handovers. Second, we used the system to change the robot's reaction time to study which timing was preferred by users in a handover task. We used a publicly available dataset [14], [15] of human-human handover tasks to train a deep learning model to classify handovers based on extracted human poses. We achieved a 92.78% frame accuracy using a Space-Time-Separable Graph Convolutional Network (STS-GCN) [16]. Using this model, we conducted a user study in which we systematically studied the robot's reaction time with five different conditions. We used the RoSAS [17] to measure warmth, competence, and discomfort, a set of metrics evaluating fluency, trust, and working alliance [13] proposed by Pan et al. [11] for Human-Robot Collaboration, and the Too Little/Too Much Scale (TLTM) [18] to measure the preferred handover timing.

With this work, we contribute (1) a ready-to-deploy adapted STS-GCN model for classifying human handover motion using preprocessed RGB-D camera data and (2) a preferred robot reaction timing of 3.38 seconds after users (given) intent to pick up an object, and (3) an approach to determine the preferred handover time for any robot handover setup. In a deployment of our model, we show that humans share a preference for reaction speed and know what feels best for them and what reaction to their movement is too fast or slow. Participants also shared preferences on the timing they favored. Here, we found that the ideal reaction time for our system is 3.38 seconds after we gave users the intent to pick up an object via an audio signal. Future work will need to validate this approach. Furthermore, the user study and the handover were carried out in a carefully controlled environment, leaving future research to conduct user studies using an open environment to perform various handovers.

II. Related Work

In the following, we overview the literature on human-robot handovers, intent communication, the importance of timing in handovers, and machine learning for handover detection.

A. Human-Robot Handovers

Handovers are a form of joint action where collaboration between two agents, the *giver* and the *receiver*, is essential for the successful transition of an object [13]. These actions stand apart from individual tasks, requiring coordinated performance, an anticipating understanding of others' behaviors, adapting behavior for task compatibility, and precise action synchronization [19]. Both agents need to communicate on the *what*, *when(timing)*, and *where(location)* of the handovers between them [20]. Previous work has already applied insights from human-human handovers to develop models for human-robot handovers, which enable efficient and safe transfers of objects [7], [20], [21]. However, robotic systems performing handovers are currently considerably slower than human-human interactions [22], [23]. The scope of existing research predominantly concentrates on robot-to-human handovers [21], [24]–[28], with less emphasis on human-to-robot or bidirectional handovers [12], [29]–[31].

Literature found that a single handover has two distinct phases: The initial pre-handover phase and the physical handover itself [32]–[34]. They are initiated by either the receiver requesting an object or an agent prompting another to use it for a task, acting as the giver [13]. However, handovers encompass many aspects that affect each handover. Many of these aspects have already been investigated by previous work. Grasp planning, which focuses on how the robot will physically grasp the object, is a critical element of this process [35], [36]. Perception, involving the robot's ability to interpret its environment and the objects within it, is equally essential [37], [38]. The handover location is another significant factor, requiring careful consideration for efficiency and comfort for the human participant [29], [39]. Motion planning and control, which determine the robot's movements and responses during the handover, are prerequisites for the smooth execution of the task [39], [40]. In the physical handover phase, grip force modulation is critical to ensure a secure yet gentle object transfer [6], [34]. The object itself also affects the perceived danger in handovers [41]. At the same time, error handling is necessary to address any issues that might arise [42], [43]. All of these aspects, which instigate human-robot handovers, start with the initial communication between the agents that want to initiate the handover.

B. Intent Communication

Successful handovers require a lot of explicit and subconscious communication. Specifically, communication serves two purposes: it initiates the action by expressing the intent to start and facilitates the coordination of the action once underway [19]. Requesting an object and initiating the task forms the initialization of any handover process and, thus, requires effective communication [13]. Signaling strategies enhance coordination, enabling better prediction of each other's actions and reducing uncertainty [44]–[46]. Humans are inherently good at subconsciously understanding others' intents [47]. Thus, giving robots similar intent communication capabilities provides a natural way of using human knowledge for more seamless and successful human-robot handovers.

Humans can communicate requests through various means, often using speech. However, Masumoto et al. [48] discovered that dividing an agent's attention between different modalities can harm coordination during joint action. Between humans, non-verbal cues play an important role in social interaction and, therefore, in coordinating handovers between human agents [49]. Non-verbal communication can be divided into four main modes: kinesics, proxemics, haptics, and chronemics, as well as multimodal combinations of these modes [50], which can all be transferred to robots. Human-like gaze cues in robot-initiated handovers can enhance these interactions' timing and perceived quality [51]. Furthermore, grip and load forces are vital factors in coordinating handovers [6], [52]. Several studies have explored spatial behaviors and dynamics in handovers, examining where handovers occur in space [39], [53] and analyzing joint and limb motion of both the giver and receiver [54], [55]. The studies identified common kinematic characteristics in humans performing handovers, such as a rapid increase in arm velocity at the handovers' beginning [54].

C. Importance of Timing in Handovers

Timing plays a vital role in a wide range of HRI scenarios. From collaborative tasks to interactive performances, the precise timing of a robot's actions significantly impacts both the efficiency and the overall user experience of these interactions [56]. Different timings also influence the social perception of the robot [11]. For robots to become valuable partners, they must be capable of fast and seamless handovers. Handing over an object, therefore, requires careful timing between the giver and the receiver [57]. In human-to-robot handovers, timing mainly refers to the robot's reaction after communicating a handover intention. Previous work shows that the receiver is responsible for the timing, while the giver is responsible for the safety and success of handovers [6], [52]. Controzzi et al. [58] studied the preferred handover timing for handovers with humans as the receiver in two experiments: human-human handovers and robot-human handovers, finding the preferred reaction time for receivers when handed an object from a robot arm. While this brings understanding to the preferred receiver reaction timing, the preferred reaction timing of the robot's arm, when the robot arm is the receiver and the user the giver, is still unknown.

Admoni et al. [59] study the effect of introducing delays to emphasize robot non-verbal communication such as gaze. The timing of non-verbal communication can be quantified to the non-verbal communication happening around 1.2 seconds before the handover [39]. These findings enable robots to anticipate and predict rather than react to handovers for more fluent and seamless interaction [60], [61], reducing waiting times and enhancing collaboration. Koene et al. [62] demonstrated that human-robot handover interactions have a notable adaptation by human participants to robotic movements, resulting in decreased delays and improved prediction of handover points, suggesting a speed-accuracy trade-off where users prefer faster interactions with robots, even at the expense of reduced precision. They, therefore, show that temporal aspects are of greater importance than spatial aspects. Pan et al. [12] studied the effect of robot speed and reaction time on perceived interaction quality and found that humans prefer human-level timing. However, their study only varied the timing with varying delays and not in the predictive direction. As human-robot collaboration should be efficient [63], predicting the human motion to anticipate a handover could lead to faster interaction, which could be another trade-off to consider.

D. *Machine Learning to Understand Handovers*

Generally, there exist two approaches to understanding the handover approach: (1) forecasting the human skeleton based on the past motion sequence (e.g., [16], [64]–[66]), and (2) classifying the past motion sequence (e.g., Pan et al. [7]). This work will solely focus on the second approach, as classification approaches are still more robust. By combining LSTM networks and feature selection techniques, we can classify handovers from both the giver's and receiver's perspectives [67]. Spatial-Temporal Graph Convolutional Networks (ST-GCN) offer significant advantages in human action recognition by effectively modeling dynamic skeleton sequences over time [8]. Drawing on the breakthroughs in self-attention mechanisms within the field of natural language processing (NLP), as highlighted in studies [68], [69], [70] introduced the transformer architecture [9] into human motion forecasting. They emphasize the concepts of time and space by designing a spatio-temporal transformer. Inspired by this approach, Mascaro et al. [10] proposed a model where the temporal channel explores these relationships in each time frame. In contrast, the spatial channel identifies intraframe relationships of the skeleton. STS-GCNs improve upon ST-GCNs by including both temporal evolution and spatial joint interaction within one network [16].

III. Handover Recognition Model

We used a publicly available data set to train our deep learning model to recognize handover approaches. We adopted an STS-GCN model that predicts future pose sequences, for which we then trained a binary classification model predicting whether the pose sequence includes a handover or not.

A. *Dataset*

We utilized 1200 handovers from the *Handover Orientation and Motion Capture Dataset*[1] created by Chan et al. [14] recorded at 300Hz using a Vicon motion capture system. The handover tasks in the dataset contain 20 participants working in pairs in four different scenarios: natural handovers, giver-comfort-focused, and receiver-comfort-focused, involving.

B. *Preprocessing*

The preprocessing pipeline is visualized in Figure 2. The datasets provide us with skeleton data of the people involved in the handover task in step A. In step B, we used the point in time where the distance between the giver's and receiver's hands was minimal to determine this transition in the recorded handover data. This calculation allowed us to determine in which frame the pre-handover phase ended and where the physical handover phase began. Next, we split the data into the *giver* and *receiver* while we subsequently excluded the *receiver* data as we only focused on human-to-robot handover (Step C). In step D, we performed the joint selection. In the following steps, we will only focus on the head, left hand, left elbow, left shoulder, right hand, right elbow, right shoulder, and torso joints. Next, we performed normalization (Step E). We discarded orientation data. Thus, having all *givers* into the same direction and normalized the joins. The normalization involved setting the torso's coordinates to zero in the initial frame of each recording, with subsequent coordinates of all joints adjusted relative to this reference point. Finally, we performed data augmentation (Step F). Here, we rotated the

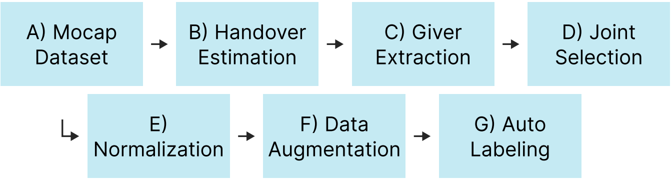

Fig. 2. Overview of the dataset preprocessing pipeline for forecast and classification dataset.

[1] Available at https://bridges.monash.edu/articles/dataset/Handover_Orientation_and_Motion_Capture_Dataset/8287799

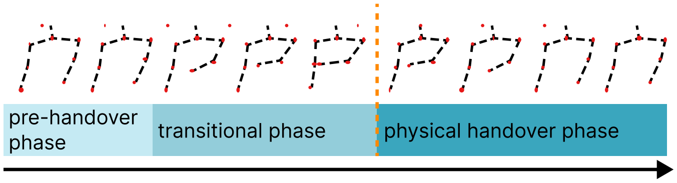

Fig. 3. Example sample for the model training.

recorded handovers in steps of 90 degrees around the x-axis to achieve data of handovers in four different directions. We loaded the data for several sample rates to mitigate the effect of handovers performed at various speeds but with the same motion. In addition to the default sample rate of ten, which reduced the data set from 300hz to 30, we also used six, eight, and twelve to load the data. We reduced the sampling rate to 30fps for the final training to support typical depth cameras, e.g., Intel RealSense Depth Camera D455. We used a sliding window approach to counteract downsampling, allowing us to use the frames that the downsampling would have skipped.

Finally, we labeled the data (Step G). Here, we utilized the point in time where the pre-handover phase ended, and this physical handover phase started to categorize the data into three distinct classes: pre-handover phase, transitional phase, and physical handover phase. This phase consisted of 450 frames, equal to 1.5 seconds, immediately preceding the physical transfer of the object. A resulting segment can be viewed as depicted in Figure 3. This resulted in 2.312.028 total labeled samples (poses), which we used for the following training phases.

C. Model Architecture

For sequence classification, we adopted using an STS-GCN [16] as the encoder and two dense layers for decoding and thus classification, see Figure 4. The STS-GCN graph captures the intricate interactions of body joints over time, leveraging a factored space-time adjacency matrix into separate spatial and temporal components to control the space-time interplay effectively. This model architecture allows for joint-joint, time-time, and joint-time interactions while bottlenecking the joint-time cross-talk for focused learning. We used two fully connected layers for the classification with a dropout layer, batch normalization, and activation function. For the full set of parameters, please see the full source code at https://osf.io/gn49s/. We used sequences of 10 frames, equivalent to 300 ms, as model input. In total, the model has only 47,263 trainable parameters.

D. Training

We implemented and trained the model in Pytorch. For training, we split the full dataset into training, testing, and validation sets of 70% (1.618.419 poses), 15%, and 15% (346.804 poses), respectively. We adopted an end-to-end training approach, following the practice outlined by Sofianos et al. [16]. We trained the model with a batch size of 32, a learning rate of 0.0001, and 50 epochs, after which it reached convergence, ensuring thorough learning and adaptability to classification tasks. We used ADAM [71] as optimizers.

E. Model Evaluation

The final classification training accuracy was 92.86% with a validation accuracy of 92.76%. This is the percentage of correctly classified poses in the predicted pose sequence. This is the first indication that no overfitting occurred. Finally, the classification achieved accuracy on the test set of 92.78% gain, confirming a stable model with this dataset. As a next step, we will also need to confirm if the model can be deployed in a real working system with new data from different devices, such as a depth camera for human pose estimation.

IV. User Study

With the handover classification model, we can now build a full robotic system to react early and late to a handover task. The user study was reviewed and approved by the University's Ethics Committee.

A. Study Design

For our within-subject design study, we used five reaction times as conditions: *Very Early*, *Early*, *Average*, *Late*, and *Very Late*. This allows us to study the preferred timing to start the robot's movement to take an object from a user for our study setup. Five reaction times allow us enough prediction points. Every participant did each condition twice, resulting in 10 tasks. Thus, we have 10 measurements per participant (two per condition). Each task consisted of 4 handovers. We determined the task order via a 10 × 10 Latin square design [72], to mitigate any learning effects. We used a 1.0L Nalgene Bottle as a handover object, see Figure 1 and Figure 5.

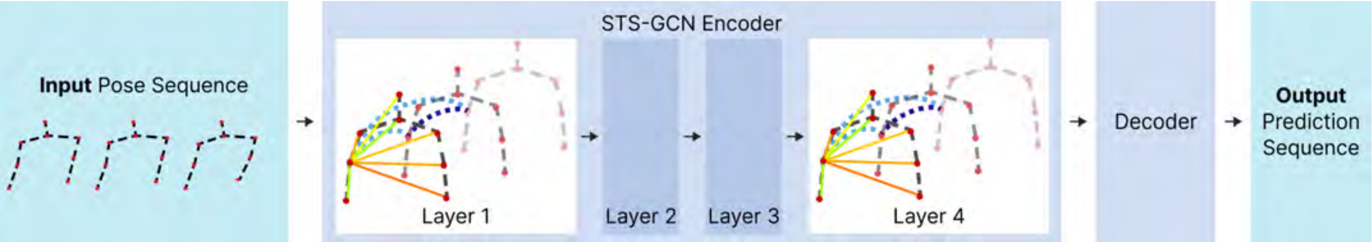

Fig. 4. The classification architecture using an STS-GCN.

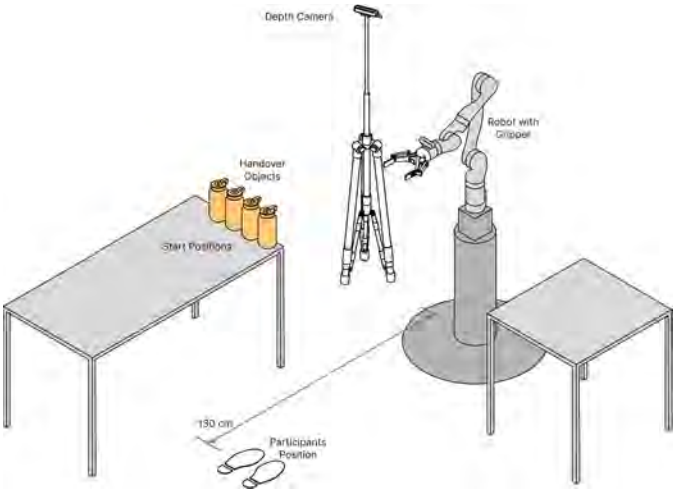

Fig. 5. Schematic diagram of the experimental setup for the user study. The diagram shows the ideal system in the home position with an additional measurement of the participant's distance to the robot.

B. Apparatus

We used the KINOVA Gen3 6DOF Robotic Arm for our user study and mounted it on a stand 110 cm above ground level. It has a maximum reach of 89.1 cm and a maximum speed of 50 cm/s. We attached a two-finger gripper to the end effector. We used an Intel RealSense Depth Camera D455 to capture participants' motion. We mounted the camera on a tripod adjacent to the robot arm so that the camera could see participants' movement as un-occluded as possible, c.f. Figure 5. Furthermore, we used two neutral acoustic signals to inform participants about the robot's state and reactions. (1) We gave participants the intent to hand over an object through a long two-second chime. This indicated that the robot started recording and awaiting the participants' actions. (2) The robot used the same long two-second chime when approaching the user to communicate its intent of getting ready to receive the object. (3) The robot played a short one-second chime when closing the gripper to communicate this to the participant[2].

We used the native Kortex API from KINOVA to control the robot's movements and actions. Our main reason for not using the ROS implementation was the need for real-time interaction with the robot. For human pose tracking, we used Nuitrack[3], a 3D skeleton tracking middleware taking color and depth input from the D455 directly. We operated the entire system from a single computer, with the robot and camera directly connected. The model classifies human motions into handovers. Furthermore, we perform vector analysis of the joints from the right hand and elbow relative to the origin.

C. Task

Each task consisted of four handovers from the participant to the robot. Before each handover, the robot signals that it is ready to receive an object via an audio signal. The participant then initiated the handover by grabbing an object and handing it to the robot. When the robot detected a handover, it would move with a predefined motion towards the handover position, closing its gripper at that position to receive the object, to then move the object to another table, see Figure 5. Participants were done with that task after they handed over four objects to the robot.

D. Procedure

First, the study conductor gave an overview and basic explanation of the user study before participants gave their written consent to participate. The consent form included permission to record audio, video, and motion capture data from the experiment. After participants signed the consent form, the study conductor explained the study procedure in detail and performed one handover in the *Average* condition with the robot while explaining the process to the participant. We then instructed participants to stand on the marked position, maintain a relaxed posture facing the robot, and perform three trial handovers with their right hand once in the three conditions *Very Early*, *Average*, and *Very Late*.

In the main study part, participants performed ten tasks, with each of the five conditions occurring twice. Each task consisted of four handovers from the participant to the robot. After each task, we asked them to answer a questionnaire on a separate PC. We took inspiration to use these questionnaires from prior work correlating social aspects with human-robot handovers [11], [12], see the full questionnaire via https://osf.io/gn49s/. First, we asked them about the dimensions of warmth, competence, and discomfort by Carpinella et al. [17]. Second, they had to rate their perceived fluency, trust, and working alliance by Ortenzi et al. [13]. Finally, we asked about the appropriate timing through the *TLTM scale* [18], which has been shown to effectively reveal curvilinear effects by finding a balance between extremes. In the meantime, the study conductor moved all objects back to their original positions. Finally, after participants had done all ten tasks, we concluded the study with a semi-structured interview. Participants were then reimbursed with 10€ for their participation.

E. Participants

We recruited 20 participants (9 females, 10 males, and 1 non-binary) aged 18-37 years ($M = 25.5$, $SD = 4.2$). Nineteen of the participants are students or are currently pursuing a PhD. Twelve participants had never interacted with a robotic system, five 1-3 times, one participant 4-7 times, and two participants more than seven times. All participants were right-handed. The ATI scale [73] yielded a mean score of 4.05 ($SD = 0.39, \alpha = 0.91$), indicating a high overall affinity towards technology. Participants took 54 minutes on average to complete the study.

V. Results

In the following, we present the results of our user study with 20 participants. Each participant performed 5 conditions

[2] Without an audio signal, participants were unsure when the robot would close the gripper, leading to confusion.

[3] Nuitrack™ skeleton tracking software, see https://nuitrack.com

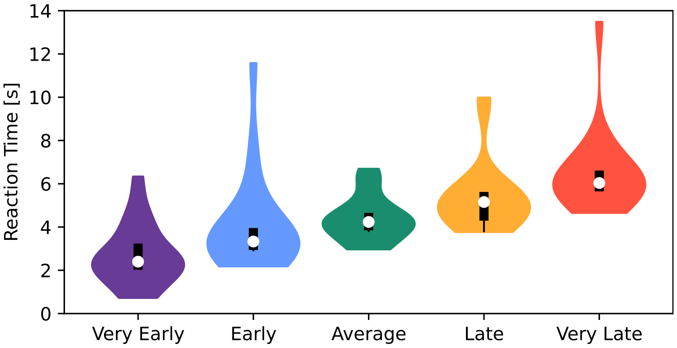

Fig. 6. Distribution of time taken between users starting the task and the robot starting to move for each condition.

× 2 repetitions = 10 tasks, resulting in 10 tasks × 20 participants = 200 total questionnaire measurements. Each participant performed 4 handovers × 10 tasks = 40 handovers, resulting in 40 handovers × 20 participants = 800 total handovers. We gathered subjective ratings for the following questionnaires: RoSAS [17], handover evaluation questions by Ortenzi et al. [13], and a TLTM question regarding timing [18].

A. *Timing Quality*

We first extract the average timings for the robots' *Reaction Time* with respect to the start of the handover, see Table I. We define the Robot Reaction Time as the time between the audio cue for the user to start the handover (intent) until the time the robot starts moving, see Figure 6. We found significant differences between the five Reaction conditions (*Very Early*, *Early*, *Average*, *Late*, *Very Late*), see Table II. Moreover, pairwise post hoc Wilcoxon signed-rank tests showed that all comparisons are significantly different ($p < .05$) but *Early* × *Average* with $p = .083$. This shows that the manipulation, aka reacting differently to the users' handover, using the neuronal network classification worked consistently.

We used Shapiro-Wilkinson to test the *TLTM scale* for normality and found the results to be not normally distributed ($p < .001$ for all). We then conducted Friedman tests to determine if the Reaction condition influenced the *TLTM scale*. We found a significant main effect of Reaction ($p < .001$). Pairwise post hoc Wilcoxon-signed rank tests with Bonferroni correction showed significant differences between *Very Early* × *Average* ($p = .003$), *Very Early* × *Late* ($p = .005$), *Very Early* × *Very Late* ($p = .002$), *Early* × *Very Late* ($p = .036$), and *Average* × *Very Late* ($p = .014$). Furthermore, we performed linear regression to find the intersection of the *TLTM scale* with 0 ("right time"), see Figure 7. The regression equation was significant, $R^2 = .280$, $R^2_{Adj.} = .277$, $F(1, 198) = 77.14$, $p < .001$. The regression shows that the "right" time is at 3.38 seconds as indicated by the red line in Figure 7. This means that the most preferred timing for participants was 3.38 seconds after the task started.

TABLE I
The time (in seconds) of the robot took after users started the task to move for the first time.

Reaction	**Seconds**
Very Early (VE)	2.398
Early (E)	3.33
Average (A)	4.235
Late (L)	5.156
Very Late (VL)	6.035

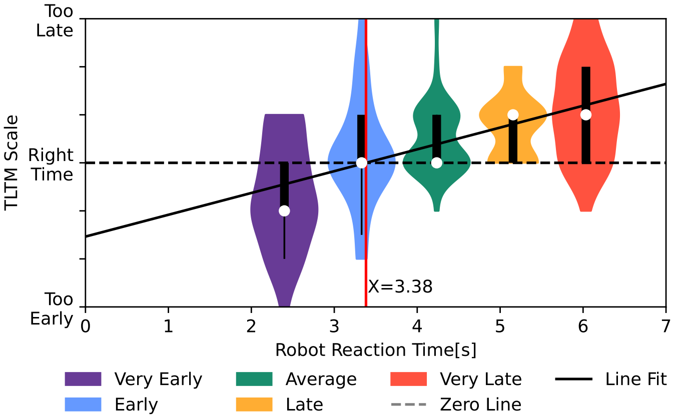

Fig. 7. The relation between robot reaction time and TLTM score. The five violins indicate the distribution of the TLTM score for the different Reaction conditions. The solid black line indicates the linear trend line, which describes the change in the TLTM score across the robot's reaction time. Finally, we can intersect the trend line with the "Right Time" (indicated by the dashed black line) to derive the red line, which is the perceived preferred robot reaction time at 3.38 sec.

B. *Handover Quality*

Figure 8 depicts the results for each subscale. We used Shapiro-Wilkinson to test for normality and found no results of the six sub-scales to be normally distributed ($p < .001$ for all). We then conducted Friedman tests to determine if the robots' Reaction influenced the *warmth*, *competence*, *discomfort*, *fluency*, *trust*, or *working alliance*.

We did not find a significant main effect of *warmth*, *discomfort*, *fluency*, and *working alliance* on Reaction. We found a significant main effect of *competence* on Reaction

TABLE II
Summary of the statistical tests.

	Shapiro-Wilk		Friedman test		
Measurement	**W**	**p**	χ^2	**p**	**Ken. W**
Robot Reaction Time	.923	<.001	60.28	<.001	.754
TLTM Scale [18]	.948	<.001	38.294	<.001	.479
Warmth [17]	.913	<.001	.477	.975	<.006
Competence [17]	.946	<.001	13.46	<.01	.168
Discomfort [17]	.919	<.001	8.643	.071	.108
Fluency [13]	.945	<.001	4.389	.356	.055
Trust [13]	.919	<.001	10.134	.038	.127
Working Alliance [13]	.901	<.001	8.481	.075	.106

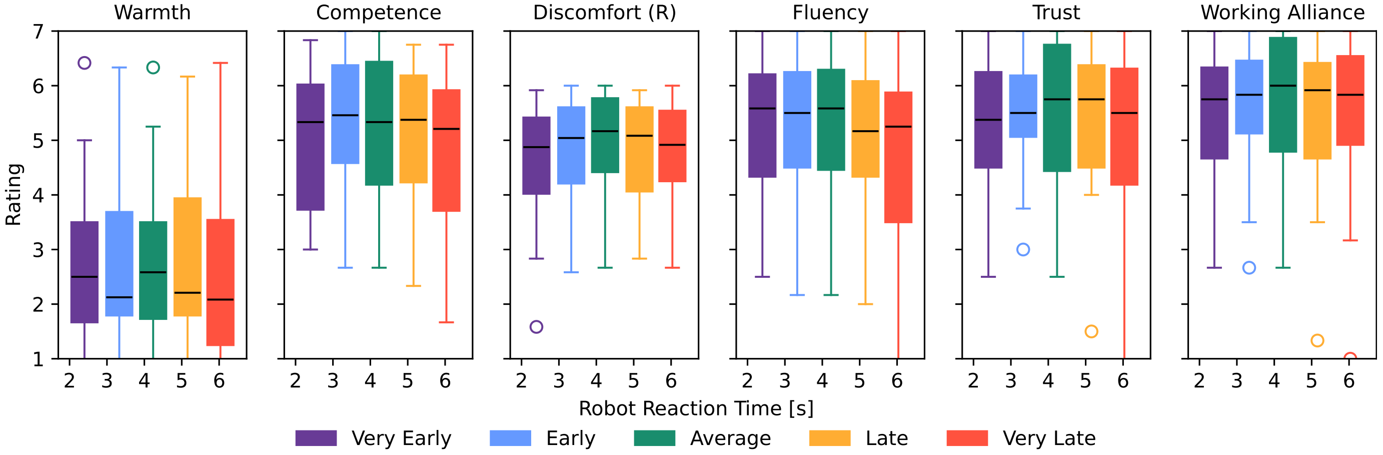

Fig. 8. Questionnaire results as boxplots for the RoSAS and Handover Questionnaire subscales for each condition (VE = Very Early, E = Early, A = Average, L = Late, VL = Very Late).

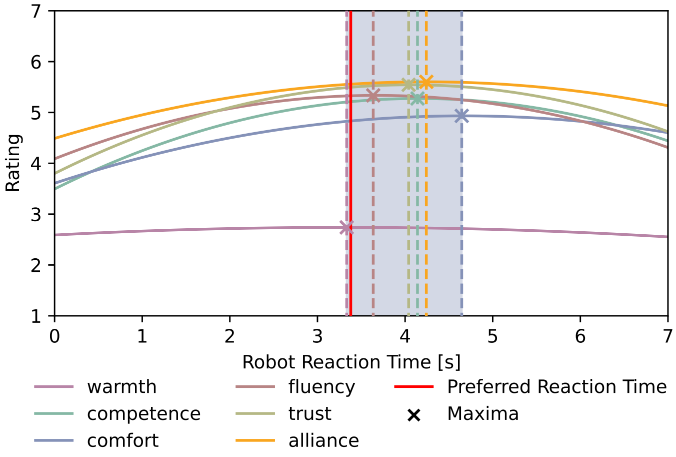

Fig. 9. The quadratic trend lines for each of the six sub-components of the user's rating. The dashed lines indicate the maxima for each curve, indicating when the maximum occurred. The red line shows the preferred robot reaction time form Figure 7. The filled area depicts the range in which designers can adapt the robot's timing to prioritize different sub-components.

($p < .001$). Pairwise post hoc Wilcoxon-signed rank tests with Bonferroni correction did not show significant differences. We found a significant main effect of *trust* on REACTION ($p = .038$). Pairwise post hoc Wilcoxon-signed rank tests with Bonferroni correction showed no significant differences.

Next, we fitted a quadratic curve through the means of each level from REACTION for each sub-scale and calculated the maxima, see Figure 9. In line with our findings on the *TLTM scale*, all sub-scales maxima are in the range from 3.333 seconds and 4.646 seconds, meaning around the *Early* and *Average* REACTION, c.f. Table I.

VI. DISCUSSION

In the following, we discuss the applicability of our implemented system, our findings regarding understanding the preferred reaction time for robots in handover tasks, and future challenges.

A. Robustness of our Handover Detection System

We trained an STS-GCN model on motion capture data from human-human handovers and used it to predict human-to-robot handovers using live depth camera data. We show that the model works very accurately, as the distribution of reaction time from the robot for each level of REACTION is low (Figure 6), and the five levels are perceived as significantly different from each other except for the *Early* and *Average* levels (Section V-A). We note that our classification accuracy of 92.78% could be transferred accurately into the real world. Furthermore, our live pose data came from a depth camera, and training data came from a motion capture dataset.

With the successful deployment of our model, we provide the model structure, training protocol, and system pipeline for public use via https://osf.io/gn49s/.

B. Designing Preferred Robot Reaction Times

We used the TLTM scale [18] to understand the preferred human-robot handover timing. Our results show that the preferred time to start the robot is 3.38 seconds after the task begins. Accounting for the 4.62 seconds of the robot movement until the handover is performed, this is a total of 8 seconds after the user starts to reach for the object. At this point, the participants are typically in the pre-handover phase, as they have already grabbed the object, and the arm is approaching the handover position. The appropriateness measured by the TLTM scale is only one way of looking at the user's perception. The other subscales give a more detailed view of warmth, competence, discomfort, fluency, trust, and working alliance. Here, we found that only competence and trust were significantly influenced by the REACTION conditions. However, we could model the trend through quadratic functions, allowing us to understand better how the other components [13], [17] correlate with the appropriate timing measured through the TLTM scale. The identified maxima are between 3.33 seconds and 4.646 seconds (Figure 9), which is generally between the *Early* and *Average* REACTION condition. Our results indicate that designers have room to tune the robots toward which other components are most important in their specific use case. For example, if comfort is important, then the design can choose a slower-than-preferred reaction time. This will be especially important when handling potentially dangerous objects [41].

C. Limitations and Future Challenges

A key limitation in building more robust systems is the lack of datasets. While some handover datasets exist, labeled data of movements similar to, but not handovers, is needed for better differentiation. Despite this, our approach using an STS-GCN [16] demonstrated effective classification of human handovers. Accurate forecasting of human motion could further improve robustness, but better datasets are essential for this progress.

In our study, we demonstrated that for our static setup with a robot arm, the preferred reaction timing for a robot receiving an object from a human is 3.38 seconds, and we presented an approach to determine the preferred reaction time. Future work should investigate whether our findings are generalizable to different robots in a more ecologically valid task using a larger sample size and in a more dynamic scenario, where users potentially have a secondary task and choose to hand over objects to the robot itself. We found that classifying human handovers works very well and is robust for the REACTION levels we chose. However, when including very early predictions of human handovers, there will always be a trade-off between people preferring faster reaction times and movement being detected wrongly as handover attempts. With knowledge about the user groups, designers can use our findings to determine the trade-off for their specific case. However, the results of our study show that the preferred perceived handover timing lies somewhere on the spectrum of our tested conditions, as the intercept of the trend line is within these timings, see Figure 9.

Lastly, our preferred timing for human-to-robot handovers contrasts with earlier findings on robot-to-human handovers [58], [74]. Controzzi et al. [58] reported a shorter preferred reaction timing of 429 - 626 ms when humans act as *receivers* of objects from robots, compared to our longer *giver* reaction time. This difference may arise from variations in study setup, robot speed, path, or increased safety as *giver* [6], [52]. Future work should explore this contrast between preferred timings for humans as receivers and givers. Nonetheless, in both cases, participants favor predictive robot behavior.

VII. CONCLUSION

In this work, we developed a classification model based on an STS-GCN to classify human handovers at different times. With this model, we conducted a user study in which participants had to hand objects to a robotic arm to understand the preferred handover timing from a user perspective. We measured the appropriate reaction timing with a TLTM question [18] and six sub-scales of user perception of the robot's movement [13], [17]. Our contribution is threefold. (1) We confirmed that STS-GCNs work well to classify human handovers. (2) We found that the preferred robot's reaction timing should be early, which in our case was 3.38 seconds after the handover task started. (3) We contribute a general approach to find the preferred reaction timing of robots when receiving objects from humans. Furthermore, we found that designers have a trade-off to adapt this timing based on their needs for warmth, competence, discomfort, fluency, trust, and working alliance for their system.

REFERENCES

[1] B. Reeves and C. Nass, "The media equation: How people treat computers, television, and new media like real people," *Cambridge, UK*, vol. 10, no. 10, 1996.

[2] A. William Evans, M. Marge, E. Stump, G. Warnell, J. Conroy, D. Summers-Stay, and D. Baran, "The future of human robot teams in the army: Factors affecting a model of human-system dialogue towards greater team collaboration," in *Advances in Human Factors in Robots and Unmanned Systems*, P. Savage-Knepshield and J. Chen, Eds. Cham: Springer International Publishing, 2017, pp. 197–209. doi: 10.1007/978-3-319-41959-6_17

[3] G. Hoffman and C. Breazeal, "Effects of anticipatory action on human-robot teamwork efficiency, fluency, and perception of team," in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '07. New York, NY, USA: Association for Computing Machinery, 2007, p. 1–8. doi: 10.1145/1228716.1228718

[4] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, p. eaat8414, Jun. 2019. doi: 10.1126/science.aat8414

[5] H. Duan, Y. Yang, D. Li, and P. Wang, "Human–robot object handover: Recent progress and future direction," *Biomimetic Intelligence and Robotics*, vol. 4, no. 1, p. 100145, 2024. doi: 10.1016/j.birob.2024.100145

[6] W. P. Chan, C. A. Parker, H. M. Van der Loos, and E. A. Croft, "Grip forces and load forces in handovers: implications for designing human-robot handover controllers," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 9–16. doi: 10.1145/2157689.2157692

[7] M. Pan, V. Skjervøy, W. Chan, M. Inaba, and E. Croft, "Automated detection of handovers using kinematic features," *The International Journal of Robotics Research*, vol. 36, Feb. 2017. doi: 10.1177/0278364917692865

[8] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," Jan. 2018. doi: 10.48550/arXiv.1801.07455

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc., 2017.

[10] E. V. Mascaro, S. Ma, H. Ahn, and D. Lee, "Robust human motion forecasting using transformer-based model," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. New York, NY, USA: IEEE, Oct. 2022, pp. 10 674–10 680. doi: 10.1109/iros47612.2022.9981877

[11] M. K. Pan, E. A. Croft, and G. Niemeyer, "Evaluating social perception of human-to-robot handovers using the robot social attributes scale (rosas)," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. New York, NY, USA: Association for Computing Machinery, Feb. 2018, pp. 443–451. doi: 10.1145/3171221.3171257

[12] M. K. Pan, E. Knoop, M. Bacher, and G. Niemeyer, "Fast handovers with a robot character: Small sensorimotor delays improve perceived qualities," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. New York, NY, USA: IEEE, Nov. 2019, pp. 6735–6741. doi: 10.1109/iros40897.2019.8967614

[13] V. Ortenzi, A. Cosgun, T. Pardi, W. P. Chan, E. Croft, and D. Kulić, "Object handovers: A review for robotics," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1855–1873, 2021. doi: 10.1109/TRO.2021.3075365

[14] W. Chan, M. Pan, E. Croft, and M. Inaba, “Handover orientation and motion capture dataset,” 6 2019. doi: 10.26180/5d09946e4e64d

[15] W. P. Chan, M. K. X. J. Pan, E. A. Croft, and M. Inaba, “An affordance and distance minimization based method for computing object orientations for robot human handovers,” *International Journal of Social Robotics*, vol. 12, no. 1, pp. 143–162, Jan 2020. doi: 10.1007/s12369-019-00546-7

[16] T. Sofianos, A. Sampieri, L. Franco, and F. Galasso, “Space-time-separable graph convolutional network for pose forecasting,” in *2021 IEEE/CVF International Conference on Computer Vision*, ser. ICCV’21. New York, NY, USA: IEEE, 2021, pp. 11 189–11 198. doi: 10.1109/iccv48922.2021.01102

[17] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, “The robotic social attributes scale (rosas): Development and validation,” in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI ’17. New York, NY, USA: Association for Computing Machinery, 2017, p. 254–262. doi: 10.1145/2909824.3020208

[18] J. Vergauwe, B. Wille, J. Hofmans, R. B. Kaiser, and F. D. Fruyt, “The too little/too much scale: A new rating format for detecting curvilinear effects,” *Organizational Research Methods*, vol. 20, no. 3, pp. 518–544, 2017. doi: 10.1177/1094428117706534

[19] C. Vesper, S. Butterfill, G. Knoblich, and N. Sebanz, “A minimal architecture for joint action,” *Neural Networks*, vol. 23, no. 8-9, pp. 998–1003, Oct. 2010. doi: 10.1016/j.neunet.2010.06.002

[20] K. Strabala, M. K. Lee, A. Dragan, J. Forlizzi, S. S. Srinivasa, M. Cakmak, and V. Micelli, “Toward seamless human-robot handovers,” *Journal of Human-Robot Interaction*, vol. 2, no. 1, pp. 112–132, Feb. 2013. doi: 10.5898/JHRI.2.1.Strabala

[21] C.-M. Huang, M. Cakmak, and B. Mutlu, “Adaptive coordination strategies for human-robot handovers,” in *Robotics: Science and Systems XI*. Robotics: Science and Systems Foundation, Jul. 2015. doi: 10.15607/rss.2015.xi.031

[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)*. New York, NY, USA: IEEE, Nov. 2016, pp. 324–331. doi: 10.1109/humanoids.2016.7803296

[23] S. Meyer zu Borgsen, J. Bernotat, and S. Wachsmuth, “Hand in hand with robots: Differences between experienced and naive users in human-robot handover scenarios,” in *Social Robotics*, ser. Lecture Notes in Computer Science, A. Kheddar, E. Yoshida, S. S. Ge, K. Suzuki, J.-J. Cabibihan, F. Eyssel, and H. He, Eds. Cham: Springer International Publishing, 2017, pp. 587–596. doi: 10.1007/978-3-319-70022-9_58

[24] A. Kupcsik, D. Hsu, and W. S. Lee, “Learning dynamic robot-to-human object handover from human feedback,” Mar. 2016. doi: 10.48550/arXiv.1603.06390

[25] A. Sidiropoulos, E. Psomopoulou, and Z. Doulgeri, “A human inspired handover policy using gaussian mixture models and haptic cues,” *Autonomous Robots*, vol. 43, no. 6, pp. 1327–1342, Aug. 2019. doi: 10.1007/s10514-018-9705-x

[26] J. Mainprice, M. Gharbi, T. Simeon, and R. Alami, “Sharing effort in planning human-robot handover tasks,” 2012. [Online]. Available: https://laas.hal.science/hal-01976706

[27] M. Huber, M. Rickert, A. Knoll, T. Brandt, and S. Glasauer, “Human-robot interaction in handing-over tasks,” in *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*. New York, NY, USA: IEEE, Aug. 2008, pp. 107–112. doi: 10.1109/roman.2008.4600651

[28] J. Laplaza, A. Pumarola, F. Moreno-Noguer, and A. Sanfeliu, “Attention deep learning based model for predicting the 3d human body pose using the robot human handover phases,” in *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. Vancouver, BC, Canada: IEEE, Aug. 2021, pp. 161–166. doi: 10.1109/ro-man50785.2021.9515402

[29] M. K. X. J. Pan, E. A. Croft, and G. Niemeyer, “Exploration of geometry and forces occurring within human-to-robot handovers,” in *2018 IEEE Haptics Symposium (HAPTICS)*. New York, NY, USA: IEEE, Mar. 2018, pp. 327–333. doi: 10.1109/haptics.2018.8357196

[30] M. Bianchi, G. Averta, E. Battaglia, C. Rosales, M. Bonilla, A. Tondo, M. Poggiani, G. Santaera, S. Ciotti, M. G. Catalano, and A. Bicchi, “Touch-based grasp primitives for soft hands: Applications to human-to-robot handover tasks and beyond,” in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. New York, NY, USA: IEEE, May 2018, pp. 7794–7801. doi: 10.1109/icra.2018.8463212

[31] K. Strabala, M. K. Lee, A. Dragan, J. Forlizzi, S. S. Srinivasa, M. Cakmak, and V. Micelli, “Toward seamless human-robot handovers,” *J. Hum.-Robot Interact.*, vol. 2, no. 1, p. 112–132, feb 2013. doi: 10.5898/JHRI.2.1.Strabala

[32] M. R. Cutkosky and J. M. Hyde, “Manipulation control with dynamic tactile sensing.”

[33] C. L. MacKenzie and T. Iberall, *The grasping hand*, ser. The grasping hand. Amsterdam, Netherlands: North-Holland/Elsevier Science Publishers, 1994.

[34] A. H. Mason and C. L. MacKenzie, “Grip forces when passing an object to a partner,” *Experimental Brain Research*, vol. 163, no. 2, pp. 173–187, May 2005. doi: 10.1007/s00221-004-2157-x

[35] V. Ortenzi, M. Controzzi, F. Cini, J. Leitner, M. Bianchi, M. A. Roa, and P. Corke, “Robotic manipulation and the role of the task in the metric of success,” *Nature Machine Intelligence*, vol. 1, no. 8, pp. 340–346, Aug. 2019. doi: 10.1038/s42256-019-0078-4

[36] J. Bohg, A. Morales, T. Asfour, and D. Kragic, “Data-driven grasp synthesis—a survey,” *IEEE Transactions on Robotics*, vol. 30, no. 2, pp. 289–309, Apr. 2014. doi: 10.1109/tro.2013.2289018

[37] W. Yang, C. Paxton, A. Mousavian, Y.-W. Chao, M. Cakmak, and D. Fox, “Reactive human-to-robot handovers of arbitrary objects,” New York, NY, USA, pp. 3118–3124, 2021. doi: 10.1109/icra48506.2021.9561170

[38] Z. Li and K. K. Hauser, “Predicting object transfer position and timing in human-robot handover tasks,” 2015. [Online]. Available: https://www.semanticscholar.org/paper/Predicting-Object-Transfer-Position-and-Timing-in-Li-Hauser/da35aba40ab38273e62b2c929ffbceba55c75c97

[39] P. Basili, M. Huber, T. Brandt, S. Hirche, and S. Glasauer, “Investigating human-human approach and hand-over,” in *Human Centered Robot Systems: Cognition, Interaction, Technology*, H. Ritter, G. Sagerer, R. Dillmann, and M. Buss, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 151–160. doi: 10.1007/978-3-642-10403-9_16

[40] 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*. New York, NY, USA: IEEE, May 2013, pp. 1629–1636. doi: 10.1109/icra.2013.6630788

[41] J. Leusmann, C. Oechsner, J. Prinz, R. Welsch, and S. Mayer, “A database for kitchen objects: Investigating danger perception in the context of human-robot interaction,” in *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA ’23. New York, NY, USA: Association for Computing Machinery, 2023. doi: 10.1145/3544549.3585884

[42] M.-J. Davari, M. Hegedus, K. Gupta, and M. Mehrandezh, “Identifying multiple interaction events from tactile data during robot-human object transfer,” *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp. 1–6, Oct. 2019. doi: 10.1109/ro-man46459.2019.8956306

[43] A. G. Eguíluz, I. Rañó, S. A. Coleman, and T. M. McGinnity, “Reliable object handover through tactile force sensing and effort control in the shadow robot hand,” in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. Singapore, Singapore: IEEE Press, May 2017, pp. 372–377. doi: 10.1109/icra.2017.7989048

[44] G. Pezzulo, F. Donnarumma, and H. Dindo, “Human sensorimotor communication: A theory of signaling in online social interactions,” *Plos One*, vol. 8, no. 11, p. e79876, Nov. 2013. doi: 10.1371/journal.pone.0079876

[45] J. Leusmann, C. Wang, M. Gienger, A. Schmidt, and S. Mayer, “Understanding the uncertainty loop of human-robot interaction,” 2023. doi: 10.48550/arXiv.2303.07889

[46] A. Belardinelli, C. Wang, and M. Gienger, “Explainable human-robot interaction for imitation learning in augmented reality,” in *Human-Friendly Robotics 2023*. Cham: Springer Nature Switzerland, 2024, pp. 94–109. doi: 10.1007/978-3-031-55000-3_7

[47] C. Becchio, V. Manera, L. Sartori, A. Cavallo, and U. Castiello, “Grasping intentions: From thought experiments to empirical evidence,” vol. 6. doi: 10.3389/fnhum.2012.00117

[48] J. Masumoto and N. Inui, “Effects of speech on both complementary and synchronous strategies in joint action,” *Experimental Brain Research*, vol. 232, no. 7, pp. 2421–2429, Jul. 2014. doi: 10.1007/s00221-014-3941-x

[49] E. C. Grigore, K. Eder, A. G. Pipe, C. Melhuish, and U. Leonards, “Joint action understanding improves robot-to-human object handover,” in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. New York, NY, USA: IEEE, 2013, pp. 4622–4629. doi: 10.1109/IROS.2013.6697021

[50] S. Saunderson and G. Nejat, “How robots influence humans: A survey of nonverbal communication in social human–robot interaction,” *International Journal of Social Robotics*, vol. 11, no. 4, pp. 575–608, Aug. 2019. doi: 10.1007/s12369-019-00523-0

[51] A. Moon, D. M. Troniak, B. Gleeson, M. K. Pan, M. Zheng, B. A. Blumer, K. MacLean, and E. A. Croft, “Meet me where i’m gazing: how shared attention gaze affects human-robot handover timing,” in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. Bielefeld Germany: Association for Computing Machinery, Mar. 2014, pp. 334–341. doi: 10.1145/2559636.2559656

[52] I. Kim and H. Inooka, “Hand-over of an object between human and robot,” in *[1992] Proceedings IEEE International Workshop on Robot and Human Communication*, 1992, pp. 199–203. doi: 10.1109/roman.1992.253888

[53] M. Huber, A. Knoll, T. Brandt, and S. Glasauer, “Handing over a cube,” *Annals of the New York Academy of Sciences*, vol. 1164, no. 1, pp. 380–382, 2009. doi: 10.1111/j.1749-6632.2008.03743.x

[54] S. Kajikawa, T. Okino, K. Ohba, and H. Inooka, “Motion planning for hand-over between human and robot,” in *Proceedings 1995 IEEE/RSJ International Conference on Intelligent Robots and Systems. Human Robot Interaction and Cooperative Robots*, vol. 1. New York, NY, USA: IEEE, Aug. 1995, pp. 193–199 vol.1. doi: 10.1109/iros.1995.525796

[55] S. Glasauer, M. Huber, P. Basili, A. Knoll, and T. Brandt, “Interacting in time and space: Investigating human-human and human-robot joint action,” in *19th International Symposium in Robot and Human Interactive Communication*. New York, NY, USA: IEEE, Sep. 2010, pp. 252–257. doi: 10.1109/roman.2010.5598638

[56] G. Hoffman, M. Cakmak, and C. Chao, “Timing in human-robot interaction,” in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, ser. Hri ’14. New York, NY, USA: Association for Computing Machinery, Mar. 2014, pp. 509–510. doi: 10.1145/2559636.2560029

[57] W. P. Chan, C. A. Parker, H. M. Van der Loos, and E. A. Croft, “A human-inspired object handover controller,” *The International Journal of Robotics Research*, vol. 32, no. 8, pp. 971–983, Jul. 2013. doi: 10.1177/0278364913488806

[58] M. Controzzi, H. Singh, F. Cini, T. Cecchini, A. Wing, and C. Cipriani, “Humans adjust their grip force when passing an object according to the observed speed of the partner’s reaching out movement,” *Experimental Brain Research*, vol. 236, no. 12, pp. 3363–3377, Dec. 2018. doi: 10.1007/s00221-018-5381-5

[59] H. Admoni, A. Dragan, S. S. Srinivasa, and B. Scassellati, “Deliberate delays during robot-to-human handovers improve compliance with gaze communication,” in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, ser. HRI ’14. New York, NY, USA: Association for Computing Machinery, Mar. 2014, pp. 49–56. doi: 10.1145/2559636.2559682

[60] G. Hoffman and C. Breazeal, “Cost-based anticipatory action selection for human–robot fluency,” *Robotics, IEEE Transactions on*, vol. 23, pp. 952–961, Nov. 2007. doi: 10.1109/tro.2007.907483

[61] M. Huber, C. Lenz, C. Wendt, B. Färber, A. Knoll, and S. Glasauer, “Increasing efficiency in robot-supported assemblies through predictive mechanisms: An experimental evaluation,” in *2013 IEEE RO-MAN*. New York, NY, USA: IEEE, 2013, pp. 503–508. doi: 10.1109/RO-MAN.2013.6628554

[62] A. Koene, A. Remazeilles, M. Prada, A. Garzo, M. Puerto, S. Endo, and A. Wing, “Relative importance of spatial and temporal precision for user satisfaction in human-robot object handover interactions,” Apr. 2014.

[63] Y. Cheng, L. Sun, C. Liu, and M. Tomizuka, “Towards efficient human-robot collaboration with robust plan recognition and trajectory prediction,” *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2602–2609, 2020. doi: 10.1109/lra.2020.2972874

[64] Z. Cao, H. Gao, K. Mangalam, Q.-Z. Cai, M. Vo, and J. Malik, “Long-term human motion prediction with scene context,” in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham: Springer International Publishing, 2020, pp. 387–404. doi: 10.1007/978-3-030-58452-8_23

[65] J. Martinez, M. J. Black, and J. Romero, “On human motion prediction using recurrent neural networks,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. New York, NY, USA: IEEE, Jul. 2017, pp. 4674–4683. doi: 10.1109/cvpr.2017.497

[66] Y. Yuan and K. Kitani, “Dlow: Diversifying latent flows for diverse human motion prediction,” Jul. 2020. doi: 10.48550/arXiv.2003.08386

[67] I. A. Monir, N. El-Bendary, and M. W. Fakhr, “Human handover classification using a deep learning model,” in *2021 31st International Conference on Computer Theory and Applications (ICCTA)*. New York, NY, USA: IEEE, 2021, pp. 187–192. doi: 10.1109/ICCTA54562.2021.9916602

[68] N. Reimers and I. Gurevych, “Sentence-bert: Sentence embeddings using siamese bert-networks,” Aug. 2019. doi: 10.48550/arXiv.1908.10084

[69] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, “Language models are few-shot learners,” Jul. 2020. doi: 10.48550/arXiv.2005.14165

[70] E. Aksan, M. Kaufmann, P. Cao, and O. Hilliges, “A spatio-temporal transformer for 3d human motion prediction,” Nov. 2021. doi: 10.48550/arXiv.2004.08692

[71] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” Jan. 2017. doi: 10.48550/arXiv.1412.6980

[72] E. J. Williams, “Experimental designs balanced for the estimation of residual effects of treatments,” *Australian Journal of Chemistry*, vol. 2, no. 2, pp. 149–168, 1949. doi: 10.1071/CH9490149

[73] T. Franke, C. Attig, and D. Wessel, “A personal resource for technology interaction: development and validation of the affinity for technology interaction (ati) scale,” *International Journal of Human–Computer Interaction*, vol. 35, no. 6, pp. 456–467, 2019. doi: 10.1080/10447318.2018.1456150

[74] N. Gyory, W. E. Lawson, and J. Gregory Trafton, “Spiking neural networks for improved robot-human handoffs,” in *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*, 2024, pp. 944–950. doi: 10.1109/RO-MAN60168.2024.10731389