

Using Spatial and Temporal Contrast for Fluent Robot-Human Hand-overs

Maya Cakmak¹, Siddhartha S. Srinivasa², Min Kyung Lee³, Sara Kiesler³, Jodi Forlizzi³

¹School of Interactive Computing
Georgia Inst. of Technology
801 Atlantic Dr., Atlanta, GA
maya@cc.gatech.edu

²Intel Labs Pittsburgh
4720 Forbes Ave., Pittsburgh, PA
siddhartha.srinivasa@intel.com

³Human Computer Interaction Inst.
Carnegie Mellon University
5000 Forbes Ave., Pittsburgh, PA
{mkleee,kiesler,forlizzi}@cs.cmu.edu

ABSTRACT

For robots to get integrated in daily tasks assisting humans, robot-human interactions will need to reach a level of fluency close to that of human-human interactions. In this paper we address the fluency of robot-human hand-overs. From an observational study with our robot HERB, we identify the key problems with a baseline hand-over action. We find that the failure to convey the intention of handing over causes delays in the transfer, while the lack of an intuitive signal to indicate timing of the hand-over causes early, unsuccessful attempts to take the object. We propose to address these problems with the use of *spatial* contrast, in the form of distinct hand-over poses, and *temporal* contrast, in the form of unambiguous transitions to the hand-over pose. We conduct a survey to identify distinct hand-over poses, and determine variables of the pose that have most communicative potential for the intent of handing over. We present an experiment that analyzes the effect of the two types of contrast on the fluency of hand-overs. We find that temporal contrast is particularly useful in improving fluency by eliminating early attempts of the human.

Categories and Subject Descriptors

I.2.9 [Artificial Intelligence]: Robotics; H.1.2 [Models and Principles]: User/Machine Systems

General Terms

Design, Experimentation

Keywords

Robot-human hand-overs, fluency

1. INTRODUCTION

Handing over different objects to humans is a key functionality for robots that will assist or cooperate with hu-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HRI'11, March 6–9, 2011, Lausanne, Switzerland.

Copyright 2011 ACM 978-1-4503-0561-7/11/03 ...\$10.00.



Figure 1: HERB (Home Exploring Robotic Butler) handing over a drink (a-b) during a public demonstration and (c-d) during our experiment for investigating effects of spatial and temporal contrast.

mans. A robot could fetch objects for elderly living in their homes or hand tools to a worker in a factory. While there are infinite ways that a robot can transfer an object to a human, including very simple ones, achieving this efficiently and fluently is a challenge.

Humans carry out seamless hand-overs on a daily basis with a variety of objects from credit-cards to drinks. Yet it is often difficult for us to remember these instances or identify how exactly we hand-over particular objects. This indicates that hand-overs are automatic and do not require much deliberation for humans. Furthermore, there is a remarkable coordination in the movements of the giver and the receiver during a hand-over [28, 15, 14]. This indicates that humans are good at anticipating the timing of a hand-over from the way an object is presented, as well as presenting it in a way that lets the other understand their intent and synchronize their movements. Our long-term goal is to reach this level of fluency in hand-overs between humans and robots.

Fluency in human-robot interactions has been studied in the context of collaborative task execution [13] identifying several quantitative measures of fluency that correlate with the human's *sense* of fluency. neither the robot nor the human have to wait for one another, resulting in an efficient

execution of the overall task. Furthermore possible inefficiency during the hand-over, such as unpredicted movements or failed attempts to take the object, must be eliminated to provide smooth hand-overs and avoid negative influences on the human’s *sense* of fluency.

Towards our goal of fluent robot-human hand-overs, we propose to use *contrast* in the design of a robot’s poses and movements for its hand-over interaction. We present two ways in which the fluency of a hand-over interaction can be improved. First, we believe humans will be more responsive to the robot if they can easily interpret its intentions. We propose to achieve this by making the robot’s hand-over poses distinct from poses that the robot might have during a different action with the object. We refer to this as *spatial contrast*. Second, we believe that the coordination of the hand-over can be improved by making the timing of the hand-over predictable for the human using an intuitive signal. We propose using the robot’s movements to signal the moment of hand-over by transitioning from a pose that is perceived as non-handing to a pose that is perceived as handing. We refer to this as *temporal contrast*.

In this paper we first present an observational study that led us to the proposed approach. This involves simple hand-overs of a drink bottle in unconstrained interactions during an all-day public demonstration of our robot HERB. Second, we present results from a survey that aims at identifying robot poses that are perceived as handing over. Finally, we present a human robot interaction experiment with 24 subjects, in which we investigate the effects of spatial and temporal contrast on the fluency of the hand-over. Our experiment demonstrates that temporal contrast in particular, can improve the fluency in hand-overs by effectively communicating the timing of the hand-over and eliminating early attempts by the human.

2. RELATED WORK

Different aspects of robot-human hand-overs have been studied within robotics, including motion control and planning [1, 14, 29, 15], grasp planning [21], social interaction [11, 18, 9] and grip forces during hand-over [25, 17]. A few studies involved human subject experiments with hand-overs between a robot and a human [18, 14, 11, 9].

We are particularly interested in how the problem of choosing hand-over poses and trajectories has been addressed in the literature. One approach is to optimally plan the hand-over pose and trajectory using an objective function. A hand-over motion planner that uses *safety*, *visibility* and *comfort* in the value function is presented in [29]. A hand-over motion controller that adapts to unexpected arm movements of a simulated human is presented in [1]. A different approach is to use human evaluation. [18] analyzes human preferences on the robot’s hand-over behaviors in terms of the approach direction, height and distance of the object. User preferences between two velocity profiles for handing over is analyzed in [14] in terms of participant’s rating of human-likeness and feeling of safety.

Hand-overs between two humans have also been studied in the literature, some with an eye towards implications for robot-human hand-overs [23, 28, 3, 14, 15]. Trajectories and velocity profiles adopted by humans both in the role of giver and receiver are analyzed in [28]. Simulation results for a controller that mimics the characteristics of human hand-overs are presented in [15]. [14] analyzes the efficiency of

hand-overs in terms of the durations of three phases during a hand-over, and compares human-human hand-overs with robot-human hand-overs. The social modification of pick-and-place movements is demonstrated in [3] comparing velocity profiles for placing an object on a container versus another person’s palm. [2] analyses human approach and hand-over and observe a preparatory movements of lifting the object before the hand-over, which might play an important role in signaling the timing of the hand-over.

We believe that communicating the robot’s intent is crucial to the fluency of hand-overs. Expressing intentions of a robot has been addressed in the literature using gaze [24], speech [12], facial expression [27] and body movements [26, 16]. Expressivity has also been addressed in computer animation, mostly within the context of gestures [6, 22].

Our notion of contrast is closely related to *exaggeration* in computer animation. This refers to accentuating certain properties of a scene, including movements, by presenting it in a wilder, more extreme form [19]. The role of exaggerated movements in communication of intent is supported by psychological evidence for mothers’ modification of actions to facilitate infants’ processing, referred to as *motionese* [5].

3. APPROACH

In this section we describe the framework of our studies, define fluency and describe our approach for using contrast.

3.1 Platform

Our research platform is HERB (Home Exploring Robot Platform) (Fig.1) developed at Intel Labs Pittsburgh for personal assistance tasks in home environments [30]. HERB has two 7-DoF WAM arms, each with a 4-DoF Barrett hand with three fingers. The WAM arms provide position and torque sensing on all joints. Additionally their stiffness can be set to an arbitrary value between 0 (corresponding to maximally passive by means of actively compensating for gravity) and 1 (corresponding to maximally stiff by means of locking the joints). The sensing for objects being pulled from HERB’s hand is based on end effector displacements detected while the arm has low stiffness. HERB has a mobile Segway base and is capable of safe, autonomous navigation.

3.2 Hand-over actions for robots

We refer to an action triggered by the robot to satisfy the goal of transferring an object to a human as a *hand-over action*. A hand-over action on HERB is implemented as a sequence of three phases:

- **Approach:** The robot navigates towards the receiver with the object in its hand while its arm is configured in a *carrying* pose. It stops when it reaches a certain position relative to the receiver.
- **Signal:** The robot moves its arm from the *carrying* pose to a *hand-over* pose to signal that it is ready to hand-over.
- **Release:** The robot waits until it senses the object being pulled and opens its hand to release it. The robot then moves its arm to a neutral position and closes its hand.

We assume that the object is handed to the robot by someone prior to the hand-over action and that the arm is configured in a carrying pose before starting to approach.

In this framework, variations of the hand-over action are obtained by changing the *carrying* and *hand-over* poses. The *hand-over* pose determines the *spatial* characteristics of the

hand-over since the object is intended to be transferred at this pose. The transition from the *carrying* pose to the *hand-over* pose determines the *temporal* characteristics which can be manipulated by changing the *carrying* and *hand-over* poses. In this study, all trajectories between poses are obtained using the path planning algorithm described in [4]. The speed of the arm during transitions is kept constant.

3.3 Fluency in hand-overs

A hand-over ideally happens as soon as the robot is ready to release the object. If the human is not ready to take the object at that moment the robot will need to wait for the human. The opposite can also happen. The human can stop what they are doing in order to take the object while the robot is not ready to release the object. As a result the human will need to wait until the robot is ready. A fluent hand-over minimizes both the robot's and the human's waiting durations. This notion of fluency resembles *functional delay* defined in [13].

3.4 Using Contrast to Design hand-overs

We propose using *contrast* in the poses and the movements of the robot in order to improve fluency of hand-overs.

- **Spatial contrast** refers to the distinctness of the pose with which the object is presented to the person as compared to other things that the robot might do with an object in its hand. A *hand-over* pose with high spatial contrast is a distinct pose that conveys the intent of handing over.
- **Temporal contrast** refers to the distinctness of the pose with which the object is presented to the person as compared to the robot's previous pose. A transition to the *hand-over* pose has high temporal contrast if the *carrying* pose is distinctly different from the *hand-over* pose.

4. OBSERVATIONS ON FLUENCY

We first present an observational study on fluency in robot-human hand-overs during the demonstration of our robot HERB at the *Research at Intel day, 2010*. In this demonstration the robot's hand-over action has neither spatial nor temporal contrast. We present observations that motivate the need for both types of contrast.

4.1 Description

In this demonstration HERB hands a drink bottle to a human as part of a drink delivery task (Fig. 1(a-b)). The robot stands near a table on which drinks are made available. It starts by grabbing a drink from the table and turns 90° towards the side where the demonstrator solicits visitors. Then, it says "Please take the drink" and starts waiting for a pull on the arm holding the object. This is a simpler version of the hand-over action described in Sec. 3.2 where the arm movement signal is replaced with a vocal signal. If the object is not pulled from the robot's hand for 10sec, the robot turns another 90° to drop the drink in a bin.

Before the robot starts the task the visitors are briefed about the what the robot will be doing. They are told that the robot plays the role of a bartender and that it can give them a drink if they want it. If they do ask for a drink they are told to pull the drink when the robot presents it to them.

4.2 Analysis

HERB's interactions with visitors are recorded from two different camera views. Hand-over attempts by the robot

Table 1: Distribution of HERB's hand-over attempts during the demonstration. Refer to text for a description of the categories.

Time-out	Experienced	Novice		
		Early	Prompt	Success
28	90	15	7	7

are separated into four groups: (i) ones in which there is an error or no visitor is present in the vicinity of the robot, (ii) ones in which *time-out* occurs and the drink is dropped in the bin, (iii) ones in which the *experienced* demonstrator takes the drink from the robot and (iv) ones in which the *novice* visitor takes the drink. Within the hand-over attempts that fall into the (ii) and (iii) we look for reasons why the robot cannot induce a reaction from the visitor. Within the hand-overs in (iv) we identify (a) the ones in which the visitor attempts to take the drink *too early*, (b) the ones in which the visitor is *prompted* by the demonstrator to take the object, and (c) the rest which we label as *successful*.

4.3 Observations

Table 1 gives the categorization of 147 hand-over attempts. We make the following observations.

Pose not conveying intent. Even though visitors are told that HERB will give them a drink, when the drink is presented several of them did not attempt to take the drink on their own. Note that in some cases the visitors might not have heard or understood the robot's verbal signal as they were engaged in a conversation. However even when they direct their attention towards the robot afterwards, they do not get a sense that the robot is trying to hand them the drink. Often they take the drink after the demonstrator prompts them by saying "You can take the drink now" and pointing to the drink. This indicates that the posture of the robot does not give the impression that the robot is trying to hand the object.

Ambiguous boundary between carry and hand-over. In some cases the receiver is paying close attention to the robot throughout the execution of the task and attempt to take the object too early, while the robot is still moving or before the verbal signal. As the robot turns toward the person, the object becomes more and more reachable to the receiver. Before and during the verbal signal, the object is already at its final hand-over pose. We believe this is the main cause for the early attempts by the receiver. In addition to affecting fluency by requiring more of the human's time, this results in *failed attempts* to take the object which may be frustrating for the human.

Overall we observe that the baseline hand-over action has several issues in terms of fluency. The failure to convey the intention of hand-over causes delays and time-outs or require prompting. To overcome this issue we propose using spatial contrast. The lack of an intuitive signal that indicates when the robot is ready to hand-over, causes early failed attempts. To overcome this issue we propose using temporal contrast. In addition we observe that whether the receiver is paying attention to the robot or not has important implications on how fluent the hand-over will be.

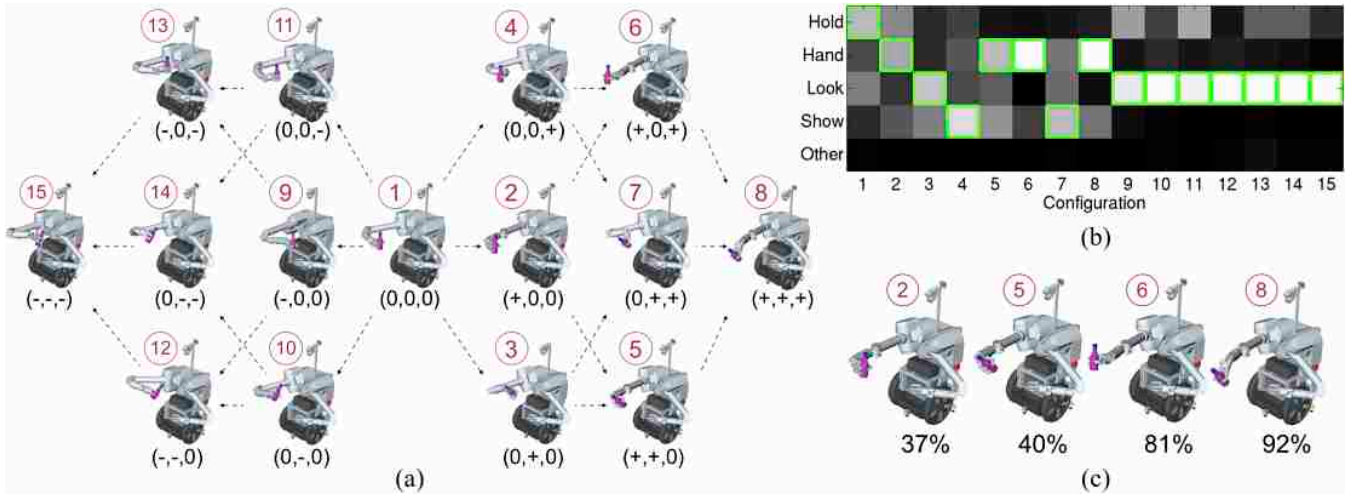


Figure 2: (a) Poses used in the survey to identify distinct hand-over poses. Poses are obtained by varying three features (*Arm extension*, *Hand position*, *Object tilt*). Possible values are 0:Neutral, +:Positive, -:Negative. (b) Responses by 50 participants for 15 poses. Light colors indicate high frequency of being chosen. The pose that got chosen more than the others are indicated with squares. (c) Poses that were labelled as *handing* more than other choices and the percentage of subjects who labelled this pose as *handing*.

5. DISTINCT HAND-OVER POSES

To better convey the robot’s intention, we propose using hand-over poses that are distinct from other things that the robot might be perceived to be doing when it has an object in its hand. We turn to the users for identifying such poses, since the primary objective is recognizability of the intent by the user. We present results from an online survey aimed at identifying such poses and investigate which variables of the pose are most effective in conveying the hand-over intent.

5.1 Survey design

The survey consists of 15 forced-choice questions asking the participant to categorize a pose of the robot holding a drink into an action category. The categories are: (i) Holding or carrying the object, (ii) Handing over or giving the object to someone, (iii) Looking at the object, (iv) Showing the object to someone and (v) None of the above, something else. Participants are shown images of the simulated robot taken from an isometric perspective in each pose. To avoid context effects the image contains nothing but the robot. To give a sense of the size of the robot a picture of the robot next to a person is included in the instructions. The order of images is randomized for each subject. All questions are available in one page such that the participant can change their response for any pose before submitting.

The poses are generated by changing three variables that we expect will effect the perception of the pose as *hand-over*. For each variable we use a *neutral*, *positive* and *negative* value. These are obtained based on our prediction of how each variable will affect the communication of the *hand-over* intention. These variables and their values are as follows:

- **Arm extension:** In the *neutral* pose, the object is about 50cm away from the robot in a comfortable position. In the *positive* pose the arm is fully extended and the object is about 80cm away. In the *negative* pose the object is about 20cm away.
- **Tilt:** In the *neutral* pose, the object is in an upright posi-

tion. In the *positive* pose the object is tilted away from the robot by 45° (towards a potential receiver). In the *negative* pose the object is tilted towards the robot.

- **Grasp location:** In the *neutral* case the robot holds the object from the side, in the *positive* case from the back (as to expose the object to a potential receiver) and in the *negative* case from the front (as to obstruct the object from a potential receiver).

The 15 poses consist of the following combinations of property values: 1 pose in which all properties have the neutral value, 6 poses in which one property has a positive or negative value, 6 poses in which two properties are both positive or both negative and 2 poses in which all properties are positive or negative. These poses are shown in Fig. 2(a).

5.2 Results

The distribution of choices by 50 participants over the 15 images are shown on Fig. 2(b) indicating the choices that were preferred more than the others. In all four poses that were tagged mostly as handing, we observe that the robot’s arm is extended. A chi-square feature analysis [20] (between handing versus all the other choices) supports the observation that arm extension is the most important feature for communicating the hand-over intention, followed by hand position ($\chi^2=155.60$ for arm extension, $\chi^2=100.51$ for hand position, $\chi^2=46.41$ for object tilt).

6. EXPERIMENT

We performed an experiment to analyze the effects of spatial and temporal contrast as well as the effect of the receiver’s attentional state on the fluency of hand-overs.

Experimental setup.

In our experiment HERB hands a drink bottle to the subject from the side while they are sitting on a tall chair in front of a computer screen (Fig. 1(c-d)). The robot starts facing away from the person and takes the drink bottle from

the experimenter. It configures its arm in the *carrying* pose, turns 180° and moves a certain distance towards the person. It then moves to the *hand-over* pose and waits for a pull. The object is always presented at the same location from the right side of the subject. Therefore if the arm is not extended in the hand-over pose, the robot gets closer to the person. After the bottle is taken by the subject the robot moves to a neutral position and goes back to the starting point to deliver the next drink. The grasp of the bottle is exactly the same in all cases – it is a power grasp at the bottom of the bottle.

Experimental design.

Our experiment aims at analyzing the effects of using *spatial* and *temporal* contrast in designing hand-overs. We consider hand-overs with different combinations of whether or not each type of contrast exists. This results in four conditions which differ in whether the *hand-over* pose is distinct or not (*spatial* contrast) and whether the transition to the *hand-over* pose is distinct or not (*temporal* contrast). We refer to the four conditions as follows (Fig. 3): spatial contrast – temporal contrast (CC), spatial contrast – no temporal contrast (CN), no spatial contrast – temporal contrast (NC), no spatial contrast – no temporal contrast (NN).

Distinct and indistinct *hand-over* poses are obtained based on the results of the survey explained in Section 5. In order to keep the position and orientation in which the object is presented fixed across conditions we choose hand-over poses that differ only in *arm extension* and *hand position*. As the distinct hand-over pose, we use a positive *arm extension* and *hand position* (Fig. 2(a)). As an indistinct hand-over pose we use neutral values for both variables. High temporal contrast is produced using a distinct non-hand-over pose as the *carrying* pose. This *carrying* pose has negative values for *arm extension* and *hand position*. Low temporal contrast is produced using a *carrying* pose in which the end-effector is moved 10cm towards the robot from the *hand-over* pose.

In order to account for whether the person is paying attention to the robot during the hand-over we perform our experiment in two groups. The *available* group is asked to pay attention to the robot while it is approaching. The *busy* group is asked to perform a task throughout the experiment such that they do not pay attention to the robot while it is approaching. To keep the subjects *busy* we use a *continuous performance task*. We use an open source implementation¹ of Conner’s continuous performance test [10]. This involves responding to characters that appear on a black screen by pressing the space bar on the keyboard, except when the character is an ‘X’. The frequency with which characters appear is varied between 1.2 and 1.4sec.

As a result we have a mixed factorial design experiment with three factors. *Spatial* and *temporal* contrast are repeated measure factors, while attentional state of the receiver (*available* or *busy*) is a between groups factor. Each subject carries out a hand-over in the four conditions twice, resulting in a total of 8 hand-overs per subject. The order of four conditions is counter balanced across subjects.

Procedure.

Prior to the experiment subjects are given some experience of taking the object from the robot such that they know

¹<http://pebl.sourceforge.net/battery.html>

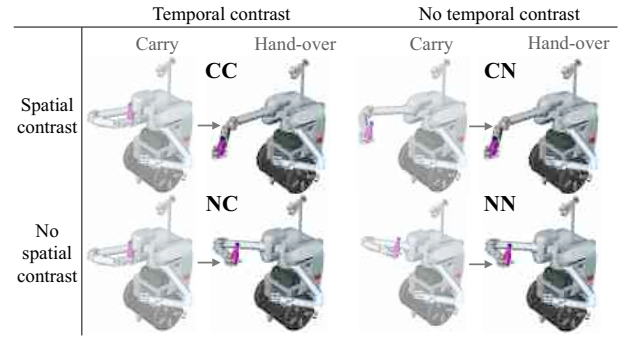


Figure 3: Four conditions for testing *spatial* and *temporal* contrast.

how much to pull the object. During these trials the robot says “Please take the object” to indicate when it is ready to hand-over. The subject is told that during the experiment the robot will not use this verbal signal so they need to decide when to take the object. Subjects in both groups are told to take the object as soon as possible. Subjects are asked to use their right hand while taking the object from the robot. Subjects in the *busy* group are told to use their right hand also for pressing the space bar and to not use their left hand to press the space bar at any time.

Evaluation.

We evaluate hand-overs in different groups and conditions in terms of their fluency. Timing of two events are determined from video recordings of the interactions: the moment their hand starts moving to take the object from the robot (t_{move}) and the moment they contact the object (t_{touch}). Other timing information is obtained from the logs of the robot’s internal state: the moment the robot starts moving its arm towards from the *carry* pose to the *hand-over* pose (t_{signal}), the moment that the robot starts waiting for the pull (t_{ready}), and the moment that the person takes the object ($t_{transfer}$). Our main measures of fluency are the waiting durations by the robot ($t_{transfer}-t_{ready}$) and the human ($t_{transfer}-t_{touch}$). In our analysis we use the second 4 interactions out of the 8 in order to exclude the effects of unfamiliarity in the very first interaction. In addition subjects are given an exit survey including the question: *Did you notice any difference in the way that HERB presented the object to you? Please explain.*

Hypotheses.

We expect that the intention of handing over can be communicated better with distinct *hand-over* poses and reduce the time that the robot waits for the person to take the object. We also hypothesize that by using temporal contrast the intended moment of transfer can be communicated better and reduce the time that the person waits for the robot to give the object and avoid unsuccessful attempts.

7. RESULTS

Our experiment was completed by 24 subjects (9 female, 15 male, between the ages of 20-45). Subjects were equally assigned to *available* and *busy* groups. The average robot and human waiting times for each condition individually, and collapsed for each factor are given in Fig. 4. We per-

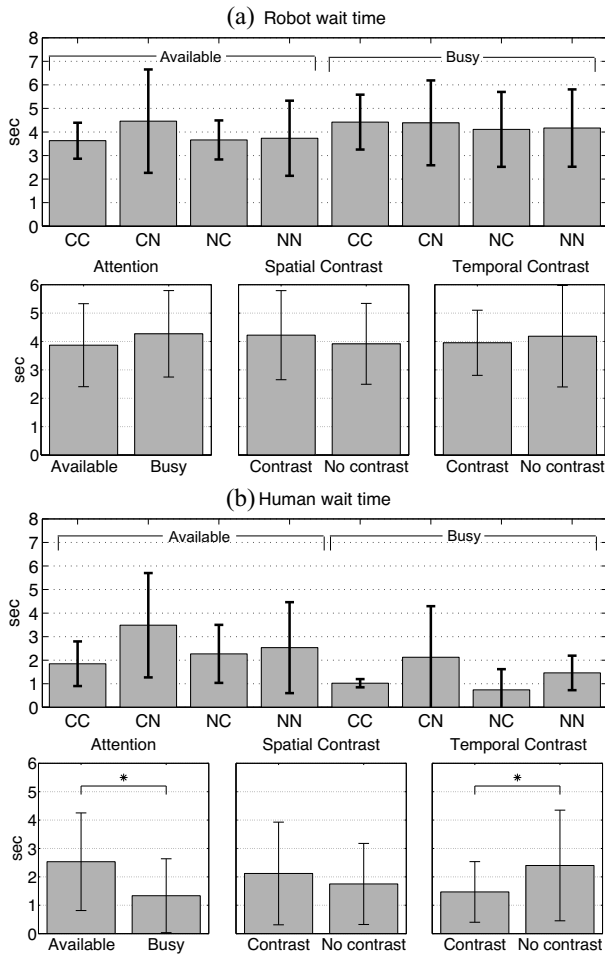


Figure 4: Average robot and human waiting times for each of the 8 conditions ($2 \times 2 \times 2$) and collapsed into two groups for each factor.

form a mixed factor three-way ANOVA on robot and human waiting durations with two repeated measure factors (temporal and spatial contrast in hand-overs) and one between subjects factor (attention to robot) [?]. The results are given in Table 2. We find supporting evidence for our hypothesis about temporal contrast; whereas our hypothesis about spatial contrast cannot be supported by our experiment. These results are summarized as follows.

Effect of temporal contrast.

We find that temporal contrast significantly reduces the waiting time of the human (Fig. 4(b), Table 2).² This means that temporal contrast lets the receivers correctly time their attempt to take the object and avoid early attempts.

Waiting duration of the human is highest for the CN condition. We observe that 9 subjects in this condition attempted to take the object too early. In the *available* group 6 subjects tried to take the object while the robot was navigating towards the person. They kept holding the object until they obtained it. Snapshots from two such incidents

²Same statistical results are obtained using $t_{\text{transfer}} - t_{\text{move}}$ as the measure of human waiting time.

Table 2: Results of the mixed factor three-way ANOVA for robot and human waiting durations. The three factors are attention (A), spatial contrast (SC) and temporal contrast (TC).

	Robot wait time	Human wait time
A	$F(1,22)=3.24, p>.05$	$F(1,22)=14.55, p<.005^*$
SC	$F(1,22)=0.97, p>.05$	$F(1,22)=1.16, p>.05$
TC	$F(1,22)=0.82, p>.05$	$F(1,22)=9.05, p<.005^*$
A \times SC	$F(1,22)=0.03, p>.05$	$F(1,22)=0.59, p>.05$
A \times TC	$F(1,22)=0.14, p>.05$	$F(1,22)=0.05, p>.05$
SC \times TC	$F(1,22)=0.13, p>.05$	$F(1,22)=1.54, p>.05$

are given in Fig. 5. In the *busy* group 3 subjects moved their hand to touch the object, went back to the attention task after they realize they cannot take it, and tried again later when the robot stopped moving. One subject in the *busy* group describes this in the survey saying the he tried to take the object when “the drink appeared in [his] peripheral vision, but HERB was not yet ready to hand over [so he] gave up to go press space again”. The same problem was observed on 3 subjects in the NN condition (all in the *available* group) and never observed on conditions with temporal contrast. These instances further motivate the benefit of temporal contrast.

The timeline of events for a subject in the *available* group is illustrated in Fig. 6. The subject starts moving her hand before the robot’s arm starts moving in both conditions with no temporal contrast (CN, NN). In the CC condition the subject’s hand moves towards the bottle after the robot’s arm stops moving. The NC condition demonstrates an instance where the person adapts movement speed as to reach the object around the time that the robot’s arm stops moving. This indicates that temporal contrast might be helpful in letting the human anticipate the point of hand-over.

The time wasted by the subjects in conditions with no temporal contrast (CN, NN) is reflected in their performance on the attention task in the *busy* group. We see that subjects miss an average of 2.54 (SD=1.32) stimuli in conditions with temporal contrast, while they miss an average of 3.05 (SD=1.41) stimuli in conditions with no temporal contrast.

These observations also demonstrate the issues related to carrying the object in a pose that is perceived as handing. Although the interaction between temporal and spatial contrast is not significant (Table 2), we see that the CN condition is more problematic than the NN group due to the carrying pose. In other words, spatial contrast in the absence of temporal contrast might be harmful to the interaction.

Effect of spatial contrast.

There was no significant effect of spatial contrast on robot waiting time. Our hypothesis was that spatial contrast would help the robot communicate its intention of handing the object and reduce the waiting time of the robot. We believe that our experiment was not suited for testing this hypothesis as the subjects were explicitly instructed to take object from the robot and the robot was not doing anything other than delivering the drink. Thus subjects did not need to distinguish the robot’s handing intention from other inten-



Figure 5: Two examples of early attempts by a subject in the *available* group, in CN and NN conditions.

tions. We believe that a setting where the person does not expect the hand-over and the robot is doing multiple actions will be more suitable for testing this hypothesis. Note that temporal contrast might also help reduce robot waiting time by functioning as an attention grabber in situations where the person is busy.

While our hypothesis on reducing robot waiting duration is not supported by our experiment, we believe there is evidence that spatial contrast served its goal of communicating the intention of handing over an object. We see that when the robot was approaching the person with an extended arm in the CN condition, several subjects made early attempts to take the object. Even though at that point subjects have had experience with the hand-over action, the extended arm of the robot induced a reaction from the human to take the object. This shows that the extended arm during the approach communicated a handing intention even though there was no signal from the robot to hand the object.

Note that the robot waiting time for the CN condition is relatively high. As this is the condition in which the human waiting time is highest, one would expect that the robot waiting time will be lower. However we observe two behaviors that result in the contrary. In some cases the subjects fail to obtain the object when they pull so they stop pulling, however keep holding the object and move along with the robot (Fig. 5). Only after the robot stops they attempt to pull again. In other cases, the subject unsuccessfully attempts to take the object and give up. To avoid another failed attempt they make sure to wait a sufficient amount of time, thus overcompensate for the failed attempt.

While describing the differences between the hand-overs in the survey, 5 subjects stated preference for either or both temporal and spatial contrast. One of them explained that “[he] liked it when HERB held the bottle close to itself and not with an outstretched arm while moving [and that this] helped [him] figure out when it was in the process of handing the bottle and when it was time for [him] to grab the bottle”. Another subject said that “[she] preferred when HERB was further away when it finished driving and started to move the arm, because when it moved closer [she] got worried that it was going to continue to drive into [her] or when it moved its arm that it would hit [her]”. This shows that temporal

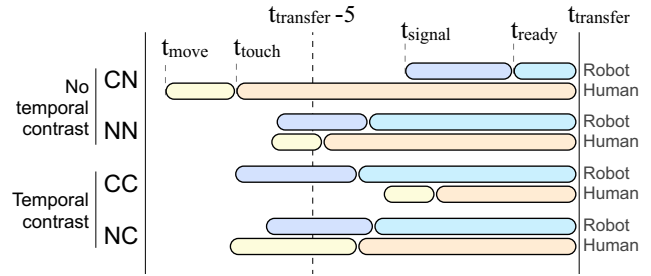


Figure 6: Sample timeline of events in four conditions from a subject in the *available* group.

and spatial contrast is not only desirable for fluency but might also be preferred by users and make them feel safer.

Effect of attention.

We find that the waiting time for the receiver is smaller when the subjects are performing the attention test. This is not surprising as these subjects are mostly not looking at the robot while it is approaching or while its arm is moving. 4 subjects in the *busy* group performed more than half of the hand-overs without turning their head away from the computer screen. Even though they are told to take the object as soon as possible, they often wait for the robot to come to a complete stop before they attempt to take the object. Consequently they get the object immediately when they try to take it and they do not need to wait. A side effect of this is the noticeable, but not significant, increase in the robot waiting duration when the subject is busy (Fig. 4(a)).

All subjects in the *available* group reported in the survey that they noticed a difference in the way HERB presented the objects. Their description of the differences referred to both types of contrast. In the *busy* group only half of the subjects noticed a difference in the way the object was presented. Their description of the difference was often limited to the distance of the robot being different.

There is no significant interaction between attention and temporal contrast (Table 2). The waiting time is higher for conditions with no temporal contrast whether the subject is available or busy. While early attempts occurred less in the *busy* group, the average waiting time of the human was also smaller for all groups. As a result the difference is preserved.

8. CONCLUSIONS

This paper is motivated from observations of unconstrained hand-over interactions between novice humans and our robot HERB during drink deliveries. We see that novices either do not recognize the robot’s attempt to hand them a drink, or they attempt to take the drink too early. To address these issues we propose using *contrast* in the robot’s actions. By making the robot’s hand-over pose distinct from other things that the robot might do with an object in its hand, the intent of the robot can be conveyed better (spatial contrast). By transitioning to the hand-over pose from a pose that is clearly non-handing, the timing of the hand-over can be communicated better (temporal contrast). We present results from a survey that aims to identify poses that are perceived as handing over. We find that all three features we proposed were useful in conveying the hand-over intention, while arm extension was the most effective. These

findings can guide the design of hand-over poses for a range of different robots and objects.

Finally we present an experiment that investigates the effects of spatial and temporal contrast. We find that temporal contrast improves the fluency of hand-overs by letting the human synchronize their taking attempts and by eliminating early failed attempts. This finding suggest that robots can greatly benefit from concealing the object from the receiver while carrying it and by transitioning to the hand-over pose when they are ready to release the object. While we don't see an effect of spatial contrast in this experiment, we believe that a different setup can capture the usefulness of spatial contrast. We plan to explore this hypothesis further in the next public demonstration of our robot as well as with an experiment that emphasizes recognition of intent.

Acknowledgments

This work is partially supported by NSF under Grant No. EEC-0540865. M. Cakmak was partially supported by the CMU-Intel Summer Fellowship. Special thanks to the members of the Personal Robotics Lab at Intel Pittsburgh for insightful comments and discussions.

9. REFERENCES

- [1] A. Agah and K. Tanie. Human interaction with a service robot: Mobile-manipulator handing over an object to a human. In *Proc. of ICRA*, 575–580, 1997.
- [2] 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*, 151–160, 2009.
- [3] C. Becchio, L. Sartori, and U. Castiello. Toward you: The social side of actions. *Current Directions in Psychological Science*, 19(3):183–188, 2010.
- [4] D. Berenson, S. Srinivasa, D. Ferguson, and J.J. Kuffner. Manipulation planning on constraint manifolds. In *Proc. of ICRA*, 1383–1390, 2009.
- [5] R. Brand, D. Baldwin, and L. Ashburn. Evidence for ‘motionese’: modifications in mothers’ infant-directed action. *Developmental Science*, 5:72–83, 2002.
- [6] J. Cassell, H. H. Vilhjalmsson, and T. Bickmore. Beat: the behavior expression animation toolkit. In *Proc. of SIGGRAPH*, 477–486, 2001.
- [7] B.H. Cohen. *Explaining Psychological Statistics*. 2. New York: John Wiley & Sons, 2001.
- [8] D. Chi, M. Costa, L. Zhao, and N. Badler. The emote model for effort and shape. In *Proc. of SIGGRAPH*, 173–182, 2000.
- [9] Y.S. Choi, T.L. Chen, A. Jain, C. Anderson, J.D. Glass, and C.C. Kemp. Hand it over or set it down: A user study of object delivery with an assistive mobile manipulator. In *Proc. of RO-MAN*, 2009.
- [10] C. Conners. Conners continuous performance test. *Multi-Health Systems*, 1995.
- [11] A. Edsinger and C. Kemp. Human-robot interaction for cooperative manipulation: Handing objects to one another. In *Proc. of RO-MAN*, 2007.
- [12] E. Jee, Y. Jeong, C.H. Kim and H. Kobayashi. Sound design for emotion and intention expression of socially interactive robots. *Intelligent Service Robotics*, 3(3):199–206, 2010.
- [13] G. Hoffman and C. Breazeal. Cost-based anticipatory action selection for human-robot fluency. *IEEE Transactions on Robotics*, 23(5):952–961, 2007.
- [14] M. Huber, M. Rickert, A. Knoll, T. Brandt, and S. Glasauer. Human-robot interaction in handing-over tasks. In *Proc. of RO-MAN*, 107–112, 2008.
- [15] S. Kajikawa, T. Okino, K. Ohba, and H. Inooka. Motion planning for hand-over between human and robot. In *Proc. of IROS*, 193–199, 1995.
- [16] T. Kanda, H. Ishiguro, M. Imai, and T. Ono. Body movement analysis of human-robot interaction. In *Proc. of IJCAI*, 177–182, 2003.
- [17] I. Kim and H. Inooka. Hand-over of an object between human and robot. In *Proc. of RO-MAN*, 1992.
- [18] K. Koay, E. Sisbot, D. Syrdal, M. Walters, K. Dautenhahn, and R. Alami. Exploratory study of a robot approaching a person in the context of handing over an object. In *Proc. of AAAI-SS on Multi-disciplinary Collaboration for Socially Assistive Robotics*, 18–24, 2007.
- [19] J. Lasseter. Principles of traditional animation applied to 3d computer animation. *SIGGRAPH Comput. Graph.*, 21(4):35–44, 1987.
- [20] H. Liu. and R. Setiono. Chi2: Feature selection and discretization of numeric attributes. In *Proc. IEEE Intl. Conf. on Tools with AI*, 338–391, 1995.
- [21] E. Lopez-Damian, D. Sidobre, S. DeLaTour, and R. Alami. Grasp planning for interactive object manipulation. In *Proc. of the Intl. Symp. on Robotics and Automation*, 2006.
- [22] M. Mancini and G. Castellano. Real-time analysis and synthesis of emotional gesture expressivity. In *Proc. of the Doctoral Consortium of Intl. Conf. on Affective Computing and Intelligent Interaction*, 2007.
- [23] A. Mason and C. MacKenzie. Grip forces when passing an object to a partner. *Experimental Brain Research*, 163:173–187, 2005.
- [24] B. Mutlu, F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita. Nonverbal leakage in robots: communication of intentions through seemingly unintentional behavior. In *Proc. of HRI*, 69–76, 2009.
- [25] K. Nagata, Y. Oosaki, M. Kakikura, and H. Tsukune. Delivery by hand between human and robot based on fingertip force-torque information. In *Proc. of IROS*, 750–757, 1998.
- [26] T. Nakata, T. Sato, and T. Mori. Expression of emotion and intention by robot body movement. In *Proc. of the Intl. Conf. on Autonomous Systems*, 1998.
- [27] J. Schulte, C. Rosenberg, and S. Thrun. Spontaneous, short-term interaction with mobile robots in public places. In *Proc. of ICRA*, 1999.
- [28] S. Shibata, K. Tanaka and A. Shimizu. Experimental analysis of handing over. In *Proc. of RO-MAN*, 53–58, 1995.
- [29] E. Sisbot, L. Marin, and R. Alami. Spatial reasoning for human robot interaction. In *Proc. of IROS*, 2007.
- [30] S. Srinivasa, D. Ferguson, C. Helfrich., D. Berenson, A. Collet, R. Diankov, G. Gallagher, G. Hollinger, J. Kuffner, and M. Weghe. Herb: A home exploring robotic butler. *Autonomous Robots*, 2009.