2012 IEEE RO-MAN: The 21st IEEE International Symposium on
Robot and Human Interactive Communication.
September 9-13, 2012. Paris, France.

Learning the Communication of Intent Prior to Physical Collaboration

Kyle Strabala[1], Min Kyung Lee[2], Anca Dragan[1], Jodi Forlizzi[2], Siddhartha S. Srinivasa[1]

***Abstract*— When performing physical collaboration tasks, like packing a picnic basket together, humans communicate strongly and often subtly via multiple channels like gaze, speech, gestures, movement and posture. Understanding and participating in this communication enables us to predict a physical action rather than react to it, producing seamless collaboration. In this paper, we automatically learn key discriminative features that predict the intent to handover an object using machine learning techniques. We train and test our algorithm on multi-channel vision and pose data collected from an extensive user study in an instrumented kitchen. Our algorithm outputs a tree of possibilities, automatically encoding various types of pre-handover communication. A surprising outcome is that mutual gaze and inter-personal distance, often cited as being key for interaction, were not key discriminative features. Finally, we discuss the immediate and future impact of this work for human-robot interaction.**

I. Introduction

In this paper, we explore communication that occurs just prior to collaborative physical tasks. For example, when a roommate is helping their friend pack a picnic basket (Fig. 1, a running example in this paper and the focus of our user study in Section II), there is a flurry of communication, from posture, to gaze, to speed of motion, that occurs much before the physical act of lifting an arm to handover an object. It is this communication, occurring via multiple channels at multiple time-scales, and its response that enables the roommate to be *predictive* of the handover instead of being *reactive* to the handover. Our goal is to to automatically learn the key discriminative features hidden in the data that enable the prediction of the intent to handover.

Predicting this intent enables *seamless* collaboration, without pauses or confusion, and empowers a robot assisting humans to understand and participate in the communication. In the latter case, it gives the collaborating robot the ability to control the interaction, by selectively choosing if and when to communicate. Furthermore, it reduces false positives in detection: humans often wave their arms and it is the communication of intent that can discriminate a handover.

We are also motivated to understand human-human collaboration. There is a rich vein of research in interaction theory [1] that scrutinizes interaction and formalizes the key channels and features. Our method helps automate this process by learning from data.

We first conducted a study (Section II, Fig. 1) where 27 pairs of human subjects performed handovers in an instrumented kitchen, outfitted with several calibrated cameras and microphones. We labeled attributes like gaze, inter-personal distance, and object locations at 10 Hz, as well as the physical actions of the handovers such as reaching out and the completion of the handover, for ground truth.

To interpret this data, we employ machine learning which is used to analyze sequential data in domains such as human activity recognition [2], economics [3], fault prediction in mechanical systems [4], and genetics [5]. Although our domain (learning the communication of intent) is vastly different, we share a common goal: extracting key features hidden in large complex multi-channel time-series data and using these features to understand new experiences.

We use a classifier [2], [6]–[8] that learns to map features of the data to high-level behavior, by training on data labeled with the correct high-level behavior. A key challenge is to extract meaningful features from the data that are indicative of the high-level behavior. In our method, we automatically extract a large number of features using simple operations on the data, and then use *feature selection* [9]–[12] to adaptively remove non-informative features, honing in on the most meaningful features (Section III).

Our application is a significant departure from current work on handovers, which analyze important features of the handover process, like posture, interpersonal distance, arm dynamics, and grip forces [13]–[18]. Our results complement this work by focusing instead on the communication of intent prior to the physical act of the handover.

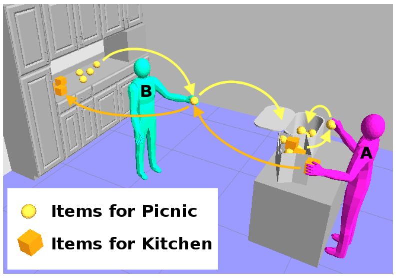
In our experiments (Section IV), decision trees outperformed other classifiers, with an average accuracy of 80%, peaking at 89%. Our final output is a tree of possibilities encapsulating types of handovers (Fig. 4). The tree identifies some commonsense features, for example, that to give an object the giver must face the receiver and must have an object. Although intuitive, it is reassuring that our algorithm was able to extract these features automatically.

But, the tree also reveals some surprises. *Mutual gaze*, often cited as critical for interaction [19], was not a key discriminative feature. In fact, a classifier based just on mutual gaze performed significantly worse compared to ours. This reinforces the fact that humans use many channels, not all obvious, to effect communication.

The tree also provides a recipe for a robot to engage in effective handover communication, illuminating some key perceptual and action triggers a robot must be able to automatically detect and effect.

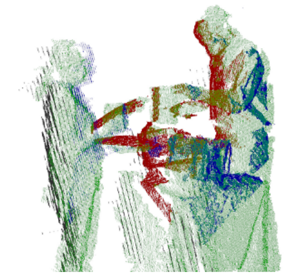
We discuss the implications of our analysis to human-robot physical collaboration, and the limitations of our approach in Section V. A key limitation is that as the classifier is focused solely on discrimination, it will, by design, ignore features that are common to both positive and negative examples,

[1]The Robotics Institute, Carnegie Mellon University
[2]Human-Computer Interaction Institute, Carnegie Mellon University

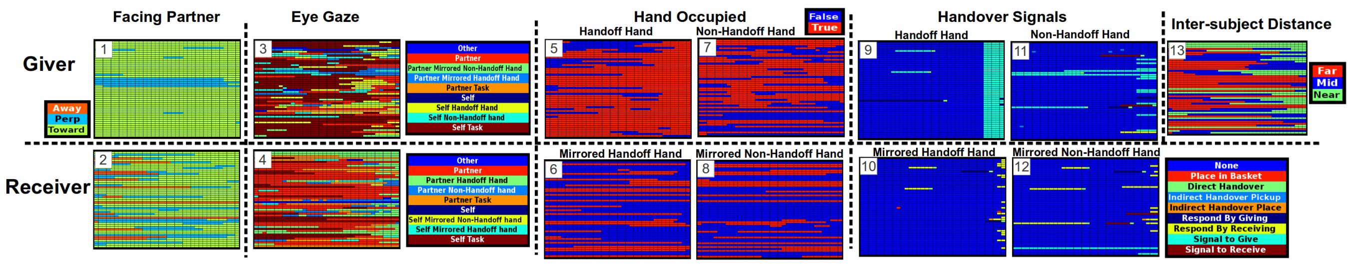


(a) Schematic of experiment

(b) Six time-synchronized camera views



(c) Aggregated 3D points



(d) A 3.5 second time window of the 13 variables coded for 50 handovers

Fig. 1. The handover user study.

features that are redundant, and features that are less discriminative. A potential worry is that these ignored features may include important social cues. We are excited about the potential of our approach to enable seamless human-robot collaboration by enabling the robot to participate not just in the physical interaction but also in the social interaction that precedes it.

II. User Study

Our first step towards learning the cues people use to initiate object transfer coordination in everyday tasks was to acquire data of natural interactions. To do so, we designed a user study that tracked pairs of participants as they handed over objects to each other. To avoid drawing attention to the handovers themselves, we placed the participants in a setting where handovers would occur naturally, without letting them know the study was focused on handovers. We chose a task that gave both participants equal opportunities to give and receive objects.

A. Participants

We recruited participants through a university experiment participant recruiting sites. We had 27 pairs of participants with an average age of 29 (SD=11)). 14 pairs were mixed-gender, 8 were female-female, and 5 were male-male. The participants were paired in the order they arrived in the lab. 7 pairs knew each other already and 20 pairs did not.

B. Procedure

We conducted the experiments in a lab that simulates a home kitchen (Fig. 1a), with a cabinet, sink, refrigerator, and table. On the table was a bag of groceries and an empty picnic basket. The kitchen lab was instrumented with 4 depth cameras, 3 color cameras, and 2 microphones, so that we could record the coordination process from various view points. Fig. 1b shows the camera images from six of the cameras, and Fig. 1c shows an example aggregated point cloud from the depth cameras.

Upon the participants arrival, an experimenter told the pair of participants that they were roommates returning from grocery shopping and that they needed to pack a picnic basket. Then, the experimenter randomly assigned a role, Person A or Person B, to each participant. Person A was asked to stand at the table to unpack the grocery bag and pack the picnic basket while Person B was asked to help Person A by taking objects (e.g. canned goods, soda, etc.) to and from the cabinets and refrigerator. Each of them was also given a list with different items they would like to bring; these items were originally located in either the grocery bag, cabinets, or refrigerator, giving both Person A and Person B equal opportunities to give and take objects from each other.

Each experiment lasted around 250 seconds, during which the participants performed an average of 9.2 handovers per experiment, leading to a total of 248 handovers. The duration of the arm motions corresponding to each handover was 1.5 ± 0.9 seconds. All participants successfully completed the tasks, putting the right set of objects in the picnic basket; no one dropped objects while transferring them.

C. A Taxonomy of Handovers

Handovers can be direct or indirect. While in direct handovers the object is passed from the giver's hand to the receiver, in indirect handovers it is set down (e.g. on the table), waiting to be picked up by the receiver. Direct handovers are usually initiated by the giver himself, but they can also be initiated by the receiver: he can extend his hand to ask that for an object to be given to him. In this study, we focus on the former: handovers initiated by the giver. A first analysis of the recording revealed that this type of handover was consistent with our previous findings [20]. Handovers have distinct coordination phases: a signaling phase in which the giver reaches out his hand holding an item, a transaction phase in which the object is exchanged, and finally a termination phase in which both the giver and the receiver retract their arms. We found that the receiver starts extending their arm almost immediately after the giver starts moving his arm, indicating that the two communicated their intent to start the physical handover action. This communication is the focus of this paper: we will be analyzing the phase occurring right before the signal to give phase starts. Our analysis of the recordings also revealed that verbal communication happens only when the participants communicate what objects they want from each other. Therefore, we will be focusing on non-verbal communication behaviors [21].

D. Data Coding

First, we manually coded our recordings from the study at 10 Hz. A single coder annotated where the participants were looking (e.g. face, hands, torso, table, basket, etc.), what their location in the environment was (x, y, θ), where the objects were, and when the handover actions occurred[1]. Next, we extracted the following discrete-valued variables for both the giver and the receiver involved in a handover:

1-2. Orientation: facing towards, perpendicular to, or away from the partner.
3-4. Eye Gaze: looking at partner, partner's handover hand, partner's non-handover hand, partner's task area, own handover hand, own non-handover hand, own task area, or other.
5-8. Hand occupancy: occupied (true), or not occupied (false).
9-12. Handover signals: signal to give and signal to receive (the extension of the arm to offer or ask for an object), the response to give and to receive (the partner's extension of the arm as a response), a few indirect handover types, or none.
13. Inter-Subject Distance: near ($< 1.5m$, the distances a direct handover is possible), mid ($\in [1.5m, 2.75m]$), or far ($> 2.75m$, the distances where one subject is at the table and the other is within reach of the cabinets).

Fig. 1d shows the values for each of the variables during the time leading up to 50 handovers. Each variable box has a row for each of the 50 selected examples. The columns denote time at $10Hz$ and the entire row represents a 3.5 second time window starting 3 seconds before the signal to give phase begins.

E. Generating Sequences

Our goal is to analyze the phase right before the signal to give (i.e. the reaching) in handover interactions: the phase where we hypothesize the communication of intent happens. Our approach is based on learning what distinguishes this phase from the phase before other interaction signals (e.g. indirect handovers), and from phases that do not lead to any interaction signal. The first step towards this is to create example data sequences from both labels: signal to give vs. not. These sequences comprise of 3 seconds of data, either right before a signal, or sampled at random (and without a signal at the end).

[1]The coding of the annotations was not subjective (e.g., did not require interpreting a meaning of sentences) and we did not see disagreement when we came up with the coding schemes and in our training session with the coder.

III. Learning from Sequential Data

Our learning pipeline consists of the three following steps.
1. Definition of Sequence Features. Our key insight is to generate features that capture *patterns* of events that occur one after the other, independent of their duration. For example, let A be a variable with values $A1, A2, A3$. If $A1$ represents 'facing away from partner', $A2$ represents 'facing perpendicular to partner' and $A3$ represents 'facing towards partner', then we consider the two sequences A1-A2-A3 and A1-A1-A2-A2-A3 to indicate the same pattern, turning to face your partner.

We generalize *compositional sequence features* to capture this behavior. In previous works [5], these features evaluate to 0 or 1 if the pattern is absent or present. The disadvantage of this is that similar sequences do not evaluate to similar values, as shown in Fig. 2, which makes them susceptible to noise in coding or in human behavior. In such cases, we want the feature to evaluate neither to 0, because it is a close match, nor to 1, because it is not a perfect match. To address this issue, we soften the evaluation to return a number between 0 and 1, based on the strength of the pattern in the data, as shown in Fig. 2.

Although different handovers take different amounts of time, they share one common time point: the *trigger*, or the end of the sequence which is immediately followed by a handover action in the positive examples. We utilize this structure to generate *trigger features*, which are identical to compositional features, except we enforce that the pattern must end at the trigger.

In addition, univariate features fail to capture important interactions between channels. Connections between the same participant's channels could be important: for example, the gaze and the orientation on their own might not be predictive of his intent, but the fact that he is looking at his partner *and* is facing him might be a strong indication of the intent to interact. Similarly, connections between channels from each partner could be important: mutual gaze might be a better predictor than each participant's gaze independently. To enable detecting such connections, we generalize these features to multivariate data, generating *multivariate compositional* and *multivariate trigger* features.
2. Feature Selection. An exhaustive enumeration of every possible pattern would result in an astronomical number of features (in our case, about 10^{20} just for features of length 1). To make this problem tractable, we use an iterative scheme for expanding and pruning the feature set based on the work by [5].

We start by creating and evaluating univariate sequence features with a single entry for each variable. We expand the highest ranking features, either by adding one more

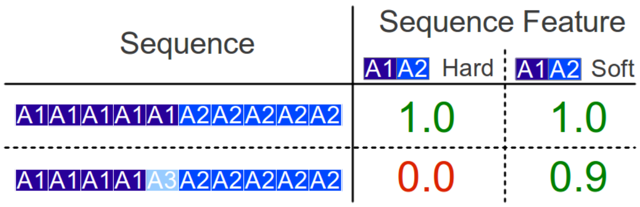


Fig. 2. Instead of searching for an exact match, we use a soft evaluation algorithm that evaluates between 0 and 1 depending on the strength on the match.

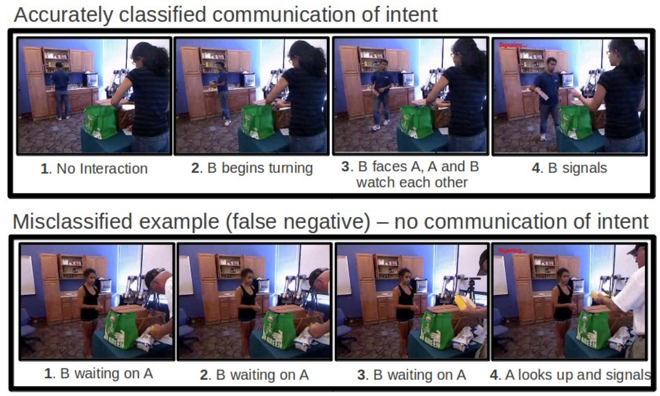


Fig. 3. Two examples that led to handovers. The top example is a typical example where the decision tree accurately predicts intent. The bottom example was misclassified and after examination it was found that the giver did not have the intent to handover until he signaled which meant there could not have been communication of intent before the signal.

entry to the pattern, or by combining multiple features into a multivariate sequence feature. These new features are then evaluated and ranked, iterating the process until we reach a desired classification accuracy. We use a number of feature ranking measures, including Information Gain, Pearson Correlation, Relief, and Positive Ratio [22].

3. Classification. We use the selected sequence features to train a set of standard machine learning classifiers, including Support Vector Machines (SVMs) with linear and radial kernels, k-Nearest Neighbors with a range of values for k, and Decision Trees with various settings for the ranking methods Information Gain (IG), IG Ratio, Gini, and Relief [22].

We partition the data into three sets: a training set, on which the various classifiers are trained, an evaluation set which is used to rank and select the best among the classifiers, and a test set which is used to evaluate true performance.

IV. Results

In our implementation of the method described above, we stopped the feature expansion process at combining 2 variables, and having a maximum length of 3. We combined the best sequence features from each expansion step into a set containing 1371 features. The average classification accuracy on the evaluation data (across different classifiers) was 80%, with a decision tree using Relief (either with $k = 5, m = 40$ or with $k = 10, m = 130$) performing the best: 89% success rate. We verified that using sequence features leads to better results than using PCA features based on the raw variable data, which are not robust to time warping between examples. Indeed, the average accuracy using PCA features was only 65%, peaking at 80%.

Fig. 4 shows a depiction of the best performing decision tree on the evaluation data. On the test data, this tree kept a high success rate of 82%, proving that the result generalizes well. The advantage of the decision tree, aside from its high accuracy, is its *amenability to interpretation*. The path from the root of the tree to the signal-to-give prediction gives us insight into the cues that must or must not be present in setting the stage for a handover. Although there is one main path to the signal-to-give class, this path corresponds to 240 different types of pre-signal sequences, due to the alternatives available at each node. The tree suggests that there are four important features that make the difference between phases that precede signals-to-give, and ones that precede other or no signals:

Previous signals: This feature cannot be true if a signal-to-give is about to occur. It implies that for a signal-to-give to occur, the sequence cannot not contain actions in which the giver is receiving an object with his handover hand, or in which either person is performing indirect handovers instead of directly interacting with one another.

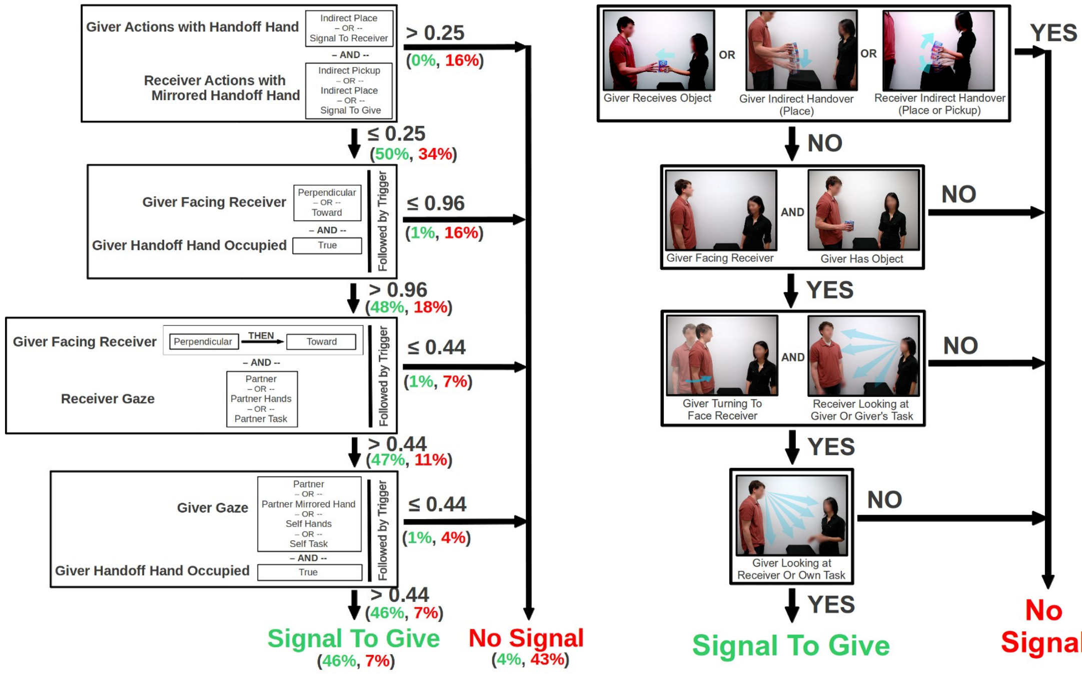
Giver orientation and hand occupancy: At the end of the sequence, the giver must have an object in his handover hand, and cannot face away from his partner.

Giver orientation and receiver gaze: At the end of the sequence, the giver must turn from facing perpendicular to his partner to facing toward him. While this happens, the receiver must be looking toward the giver (his face, torso, or hands).

Giver gaze and hand occupancy: At the end of the sequence, the giver is either looking at his hands (or the object he is holding), or at the receiver.

We found the sequences that the decision tree misclassified just as insightful as the ones classified correctly, an example of each is shown in Fig. 3. While one way to hand on object over is by first communicating this intention with the partner, a far less common way is to simply start the physical action and wait for the partner to react. This happened when the giver extended his arm and continued doing his task while waiting for the partner to take the object, when the receiver extended his arm and waited for the giver to bring over the object, and when the receiver was ready to interact but the giver did not have the intent to handover yet. The decision tree, which is focused on the communication cues before the signal happens (i.e. before arm extension), incorrectly classified these instances as negative examples.

Overall, our method gives us some intuitive, interesting and even surprising insight into the communication before handovers, all automatically, by learning what distinguishes signal-to-give phases from other interactions. That the giver must be holding an object and must not be facing away is



(a) Sequence Feature Decision Tree (b) Interpreted Decision Tree

Fig. 4. Decision tree classifier based on the training set. To save space, the smaller branches were not expanded.

a very intuitive requirement of this phase. That he turns and faces his partner is a reinforcement that orientation is an important communication cue for physical interactions. Similarly, it is very intuitive that someone who had just received and object will usually not hand it back (there was one exception to this rule in our data). The role of the receiver's gaze is more interesting: gazing toward the giver is possibly used to indicate that he is ready to receive the object. What is surprising is not in the features that are used by the classifier, but in the ones that are not: distance and *mutual* gaze are not considered important by the classifier, and we discuss this in the next section.

V. Discussion

A. Proxemics

Research on proxemics [23] suggests that different distances between people have varying meanings and functions in human-human interaction. Research in human-robot interaction proposes that a human coming in close proximity to a robot suggests an intention to engage in interaction with the robot [24]–[26]. Our initial hypothesis was that a giver being close enough that the receiver can reach out to grab an object (approx. 1.2 meters) will be a strong indicator of the initiation of the handover. Our results contradict this: distance was not part of the features used for classification. A giver reached out his hand even when he was not close to the receiver (37% of the signals to give happened when their distance was between 1.5 and 2.75 meters, and 17.7% when it was greater than 2.75 meters). This suggests that the reachable distance might not be the only factor to consider, and the dynamic gaze pattern between the robot and the person is important. This also suggests that the robot should be sensing communication cues from a person even when she/he is not in the immediate interaction zone.

B. Mutual Gaze

Human communication research [i.e.., [1]] suggests that people must coordinate in performing tasks together, and that this coordination requires communication. One way that people communicate their intent and confirm their agreement is through mutual gaze [19]. We initially hypothesized that mutual gaze would play a central role in the communication of intent for handovers. However, very surprisingly, our results suggest that mutual gaze is not one of the features used by the decision tree to predict the intent to handover. The receiver has to look in the givers direction, but there is no requirement, according to this classifier, that the giver looks back at all, let alone that he looks back simultaneously. Indeed, mutual gaze only occurred in 43% of the signal-to-give sequences. Although this is a much larger percentage than for randomly drawn sequences (only 11% of them including mutual gaze), it leads to an important observation - mutual gaze itself is not predictive of giver initiating the handover process; rather asynchronous eye gaze exchange that confirms each other's state and availability matters.

C. Implications for Robots

With the decision tree (Fig. 4), a robot can accurately recognize the intent to give in human-human interactions. This is useful for a robotic helper that is assisting multiple collaborating humans. The robot can recognize a human's intent to handover an object to another human, predict the handover between the humans, and start planning actions to help the humans in advance of the handover attempt. Most state of the art handover predictors use arm motion to classify intent to handover. However, 90% of handovers in our dataset contained events before the physical arm motions that the decision tree learned were signs of intent to handover. Therefore, a robot could use our results to greatly increase both the recognition time and classification accuracy of predictors based on arm motion, thus allowing it to respond sooner and with more confidence.

The decision tree not only informs us about the communication before handovers, but also has important implications for the design of robot behavior for human-robot handovers. Rather than being reactive to a human's handover motion after-the-fact, the robot can detect the human's intent to interact and anticipate the motion. If the human is holding an object, turns and looks at the robot, then the robot should look at the person in order to participate in the communication, and get ready (e.g. compute a motion plan) for the interaction. Aside from enabling more seamless interaction by participating in the communication, the robot can actually *control* the interaction. For example, the robot could purposefully keep its gaze away from the human in order to indicate that it is not ready to receive an object. When roles flip and the robot is the one handing over the object, it could make its intent clearer by explicitly turning toward the human and looking at him or at its own hand. The robot also should wait until the human glances in the direction of it before reaching out its arm to hand over the item.

We are also excited to use this method in future work. We believe it is possible to extend these sequence features into sequence prototypes that represent entire sequences. With these full length sequence prototypes, we could say a lot more about the events and communication leading up to a handover. And since the sequences in our data use minimally-processed variables that a robot can both perceive and enact, we could even directly generate robot behavior from these sequence prototypes.

D. Limitations

Like any studies, this work is limited in many ways. We collected data on human coordination in a lab, which may have influenced users' behaviors. Our number of subjects was limited. Our sequential analysis only accounts for patterns, and not for duration of each element in the pattern. Furthermore, we didn't look beyond 3 seconds before the handover, nor did we look at how the participants were holding the objects, which might also serve as a communication cue. A key limitation is that as the classifier is focused solely on discrimination, it will, by design, ignore features that are common to both positive and negative examples, features that are redundant, and features that are less discriminative. A potential worry is that these ignored features may include important social cues.

REFERENCES

[1] H. H. Clark, *Using Language*. Cambridge University Press, 1996.
[2] J. Aggarwal and M. Ryoo, "Human activity analysis: A review," *ACM Computing Surveys*, vol. 43, no. 3, 2011.
[3] M. C. Lineesh, K. K. Minu, and C. J. John, "Analysis of nonstationary nonlinear economic time series of gold price: A comparative study," *International Mathematical Forum*, vol. 5, no. 34, 2010.
[4] S. D. Xin and C. B. Zhong, "Some practical application of sequential analysis to the fault prediction system of a main diesel engine," *Industrial Electronics, Control and Instrumentation*, vol. 3, 1991.
[5] R. Islamaj, L. Getoor, and W. J. Wilbur, "A feature generation algorithm for sequences with application to splice-site prediction," in *Knowledge Discovery in Databases*, 2006.
[6] T. G. Dietterich, "Machine learning for sequential data: A review," in *Structural, Syntactic, and Statistical Pattern Recognition*, 2002.
[7] M. W. Kadous, "Temporal classification: Extending the classification paradigm to multivariate time series," Ph.D. dissertation, University of New South Wales, 2002.
[8] H. Salamin and A. Vinciarelli, "Introduction to sequence analysis for human behavior understanding," in *Computer Analysis of Human Behavior*, A. A. Salah and T. Gevers, Eds. Springer London, 2011.
[9] B. Chakraborty, "Feature selection and classification techniques for multivariate time series," *Innovative Computing, Information and Control*, 2007.
[10] J. G. Dy and C. E. Brodley, "Feature selection for unsupervised learning," *The Journal of Machine Learning Research*, vol. 5, 2004.
[11] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, 2003.
[12] E. Keogh, S. Chu, D. Hart, and M. Pazzani, "Segmenting time series: A survey and novel approach," in *Data Mining in Time Series Databases*, vol. 57, 2004.
[13] 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 *AAAI-SS on Multidisciplinary Collaboration for Socially Assistive Robotics*, 2007.
[14] M. Huber, M. Rickert, A. Knoll, T. Brandt, and S. Glasauer, "Human-robot interaction in handing-over tasks," in *Robot and Human Interactive Communication*, 2008.
[15] A. Edsinger and C. Kemp, "Human-robot interaction for cooperative manipulation: Handing objects to one another," in *Robot and Human Interactive Communication*, 2007.
[16] Y. Choi, T. Chen, A. Jain, C. Anderson, J. Glass, and C. Kemp, "Hand it over or set it down: A user study of object delivery with an assistive mobile manipulator," in *Robot and Human Interactive Communication*, 2009.
[17] M. Cakmak, S. S. Srinivasa, M. K. Lee, S. Kiesler, and J. Forlizzi, "Using spatial and temporal contrast for fluent robot-human handovers," in *Human-Robot Interaction*, 2011.
[18] M. Cakmak, S. S. Srinivasa, M. K. Lee, J. Forlizzi, and S. Kiesler, "Human preferences for robot-human hand-over configurations," in *Intelligent Robots and System*, 2011.
[19] M. Argyle and M. Cook, *Gaze and Mutual Gaze*. Cambridge University Press, 1976.
[20] M. K. Lee, J. Forlizzi, S. Kiesler, M. Cakmak, and S. Srinivasa, "Predictability or adaptivity?: Designing robot handoffs modeled from trained dogs and people," in *Huamn-Robot Interaction*, 2011.
[21] R. A. Hinde, *Non-Verbal Communication*. Cambridge University Press, 1972.
[22] T. Mitchell, *Machine Learning*. McGraw-Hill, 1997.
[23] E. T. Hall, R. L. Birdwhistell, B. Bock, P. Bohannan, A. R. J. Diebold, M. Durbin, M. S. Edmonson, J. L. Fischer, D. Hymes, S. T. Kimball, W. La Barre, F. Lynch, J. E. McClellan, D. S. Marshall, G. B. Milner, H. B. Sarles, T. G. L., and A. P. Vayda, "Proxemics," *Current Anthropology*, 1968.
[24] R. Gockley, A. Bruce, J. Forlizzi, M. Michalowski, A. Mundell, S. Rosenthal, B. Sellner, R. Simmons, K. Snipes, A. Schultz, and J. Wang, "Designing robots for long-term social interaction," in *Intelligent Robots and Systems*, 2005.
[25] R. Mead, A. Atrash, and M. J. Mataric, "Recognition of spatial dynamics for predicting social interaction," in *Human-robot Interaction*, 2011.
[26] T. Tasaki, S. Matsumoto, H. Ohba, M. Toda, K. Komatani, T. Ogata, and H. G. Okuno, "Distance-based dynamic interaction of humanoid robot with multiple people," in *Innovations in Applied Artificial Intelligence*, 2005.