

# Graphically representing child-robot interaction proxemics

Marie D. Manner<sup>1</sup>, Maria Gini<sup>1</sup>, and Jed Elison<sup>2</sup>

<sup>1</sup>Department of Computer Science and Engineering

<sup>2</sup>Institute of Child Development

University of Minnesota

manne044, gini, jtelison @umn.edu

## Abstract

This paper discusses the current analysis of a large set of child-robot interaction experiments, in which a 2–4 year old child plays a set of games with a small humanoid robot, with the goal of detecting early signs of autism in toddlers based on the interactions. Our first goal in this paper is to condense and display these child-robot interactions as multi-channel time series, starting with the distances between the child, robot, parent, and experimenter. Our second goal is to use these data displays to compare and contrast different children, with the aim of clustering children with similar interaction patterns. Using a ceiling-mounted camera, we record the interaction between a child and a robot which performs different games and dances. After analyzing the video footage for the locations of all people and the robot in the room over the variable length of the interaction, we condense the interactions into simplified, quantifiable, scale-invariant data. We show how the distances between the child and robot, experimenter, and caregiver can be discretized into a few location zones and compared across children using classic similarity measures. Proxemics (social distances) between the child, robot, caregiver, and experimenter during a child-robot interaction show differences between participants and hence can provide additional information for behavior analysis.

## 1 Introduction

Our work is part of a larger experimental project in which we seek to automate the analysis of the reaction of a child while s/he plays games with a robot. This paper shows how we condense and display the movements of each child with respect to the robot, the child’s caregiver, and the experimenter throughout the interaction, and compare all participants to each other. The overarching goal is to assist in the detection of abnormal development by leveraging the interest of children with autism in robots, as established in (Scassellati 2007). The work shown here is with neurotypical children, giving us a baseline with which to compare future participants. Symptoms of some pervasive developmental disorders, such as autism spectrum disorder (ASD), include differences in personal space between individuals and objects or people (Asada et al. 2016), eye contact, physical contact, and a longer delay or non-response when called by name,

Copyright © 2018, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

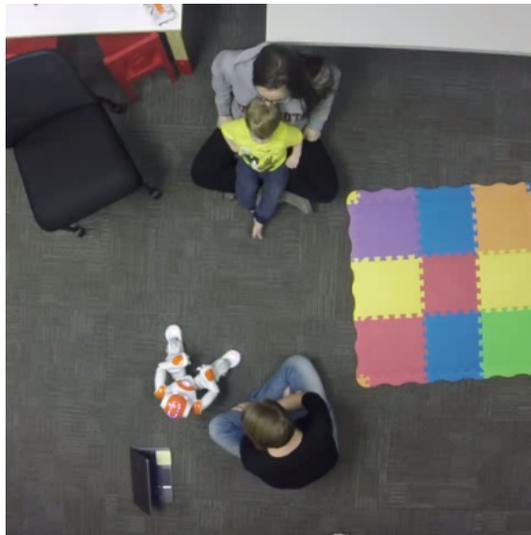


Figure 1: Sample overhead view of the human-robot interaction experiment.

among other differences (Zwaigenbaum et al. 2005). These behaviors are thus considered early markers for autism. We propose methods to generate a portable, reproducible, automated behavior analysis that can help detect these early markers.

The children we work with may be on the autism spectrum or have another pervasive developmental disorder, so we want to avoid adding fiducial markers to the child while maintaining a structured and reproducible play interaction. We thus record all interactions from several different perspectives; the only perspective that can reliably capture the participant’s location at all times is from the ceiling, so we mount a camera on the ceiling in the center of the room. Fig. 1 shows this vantage point, taken from a GoPro; distortion exists but still allows for identifying an individual’s position over time.

The work we present here has two goals, framed as questions. First, how can we compactly represent the interaction of a child and robot as a time series? Second, how can we compare participants to each other on the basis of the interaction? By answering these questions, we provide a method

to cluster participants based solely on proxemics throughout the interaction. The results for our neurotypical participants will provide a baseline with which to compare results from children at high risk for autism or even children with clinical diagnoses. The responses of high risk children, when compared to these neurotypical participants, could end up being outliers or could be similar to an identifiable subset of children reactions. In these cases, the results will provide a potentially clinically useful identifier, which could then be matched to other data collected on the participants and added to the autism phenotype. If instead the responses of high risk children look very similar to neurotypical children, we will know that children in this age group behave similarly for these interactions. In this case, future research will need to carefully consider how autonomous behaviors of robots affect reaction for this age group.

This paper contributes several ways of organizing and displaying child-robot interactions as time series data. We show how difference scores between interactions can be displayed graphically, how individual interactions can be displayed as interaction zones, how interaction zones simplify the visualization of how child comfort levels change over time, and how the interaction zones compare across children. We begin with related work in Section 2, give a high-level description of the human-robot interaction experiment and raw data taken therefrom in Section 3, describe the preliminary analysis in Section 4, and end with a summary in Section 5.

## 2 Related Work

Socially Assistive Robotics is a recent area of robotics research aimed at helping populations with special needs; it includes research for children or adults, such as robots as tools for children with pervasive developmental disorders or robots for adults as tools, companions, or helpers. Socially assistive robotics research in autism is over a decade old, yet does not currently meet standards of psychology and child development researchers (Diehl et al. 2012; Scassellati, Matarić, and Admoni 2012; Pennisi et al. 2016). Robotics research with children with autism stems from the fact that afflicted children tend to especially enjoy autonomous (or seemingly autonomous) robots (Dautenhahn and Werry 2004), and researchers have used a wide variety of robot appearances and abilities in this area (Scassellati, Matarić, and Admoni 2012). While the reason for this high level of interest is unknown, researchers clearly have the potential to leverage robotics for autism diagnosis or treatment (Scassellati 2005). We now give a brief overview of related work for this paper, detecting autism traits through automated video analysis and proxemics.

Automatically detecting autism or autistic traits is a current research area in computer vision, and much work uses as much data as possible. For example, Hashemi et al. (2012) analyzed non-intrusive camera footage using a GoPro placed on a table, two to four feet from a clinician-child pair in which the clinician was testing the child with a disengagement of attention task and a visual tracking task. The authors went even further in (Hashemi et al. 2014), in which they analyzed interest sharing and atypical motor behavior by estimating head motions from facial features and motor

behavior by arm symmetry. Fasching et al. (2015) automatically coded activities of people with obsessive-compulsive disorders from overhead video footage in a structured lab, tracking how many times participants touched various objects. These objects are statically located, such as faucets and handles, and easier to locate in a static environment. In contrast, our laboratory works with very young children in a play-based interaction, which adds difficulties in instrumenting the room and reliably tracking an active, potentially non-cooperative child.

Much research in socially assistive robotics studies child-robot interactions. Feil-Seifer and Matarić (2011) created a short free-play interaction with children and a robot, with the future intention of allowing a robot to adjust its own behavior based on the child’s reaction. This work tracked the child in relation to the robot to automatically determine if the child was having a positive or negative reaction to the robot. The authors manually coded for the child avoiding the robot, interacting with the robot or playing with bubbles the robot generated, staying still, being near parent, being against the wall, or none of those. Results showed that children with a positive reaction to the robot spent over 80% of time interacting, whereas children with a negative reaction spent less than 20% of time interaction with the robot. Mead et al. (2013) also investigated proxemics (the study of social distances), by placing a participant and researcher in discussion about a static humanoid robot. Using a video camera and depth data, they studied body pose during the experiment, training Hidden Markov Models on sensory experiences (such as voice loudness and a variety of distances to other people and environment objects) to correctly annotate initiation and termination of conversation.

## 3 Research Method

### 3.1 Experimental Paradigm

The overarching goal of the robot interaction study with toddlers (the age group of roughly 2–3 years old) is to identify children at high risk for autism spectrum disorder (ASD). Children were recruited from a laboratory-maintained database at the University of Minnesota’s Institute of Child Development. The participants in this study are very low to medium risk for ASD, but mostly low risk. Thus, the data in this paper represents a mostly neurotypical set of children, providing us with a reactionary baseline for future comparisons with very high risk or diagnosed toddlers. Written informed parental consent was ensured in advance of all testing; all research was approved by the university’s Institutional Review Board. We collected multiple data sets from each participant, including standardized and novel assessments and video footage, but for this paper we review and discuss only video footage taken from an overhead perspective throughout the duration of the interaction with the child. We wrote software to track the actors of interest throughout the video, usually the child, the robot, experimenters, and/or caregivers, and the analysis presented here is generated from the raw coordinates of each actor in every frame of the video.

During the experiments, we introduce the child to Robbie

the Robot (a NAO from Softbank Robotics). Robbie plays different games such as looking games, imitation games, and dances. The games include “I Spy” (a looking game that encourages the child to find objects in the room), “Simon Says” (a behavior imitation game that encourages the child to copy motions possible with gross motor skills like clapping and waving), and several dances set to music. The set of games is in the same order for every child. The experimenter controlling the robot imitates some of the robot’s movements and plays along during some of the looking games, encouraging the child to do the same. The interaction is recorded from up to four perspectives, most notably from a GoPro mounted on the ceiling. The GoPro is the only camera that is always located in the same place, is impossible to reach by participants, and from which we can almost always see all actors in the room.

We attempted to keep the same location for all actors in the room across experiments. The experimenter that controlled the robot (hereafter called simply the experimenter) sat next to the robot slightly off-center in the room. If the child did not need comfort or attention from their caregiver, they were seated or standing on the floor near the robot, facing the robot. In this case, the caregiver sat near the edge of the room with another researcher, answering questions from a development assessment. If the child needed constant or frequent attention from the parent, the child might be seated on or near their parent during all or part of the interaction, usually closer to the robot than the parent would be if the child did not need attention. Fig. 1 shows part of an overhead video frame; the child faces the robot, seated on their caregiver, and the experimenter is next to the robot.

### 3.2 Data

In all, 65 participants were recruited for this study, of which 60 contributed video footage of a robot interaction. In one case, the experimenter, a second researcher, and the parent all sat on the floor and attempted to draw the child’s attention to the robot; as this was a highly unusual configuration, this video was not analyzed for this paper, leaving us with 59 videos. These 59 participants, 31 males, 28 females, were aged 25 to 45 months with a mean of 32.9 months and standard deviation of 4.6 months. The interactions and therefore videos range from roughly nine to 15 minutes long, depending on the child’s willingness or ability to continue interacting with the robot. The original video is slightly distorted, thus we first perform an un-distortion and use the undistorted video for later analysis.

There are seven total parts to the interaction (games and dances), which we call presses for attention, or presses. Each press varies in time, and there is a one minute buffer between each press to give the child time to re-engage if they were not interested or took a break from playing for some reason (e.g. requested a snack). If the experimenter did not trigger the next press, after one minute passed the robot started the next press anyway. Some interactions were not completed due to equipment malfunction or participant choice, giving us some children who did not complete all seven presses. Where possible, these data are included in the analyses.

The data used here stems from the raw coordinates of each

actor in the room. First, we find the Euclidean distance between actors, giving us three channels of data: the distances between the child and robot, the child and caregiver, and the child and experimenter. In choosing Euclidean distances between actors over time, we start on our first goal, to represent an interaction over time. An example of this data is shown in Fig. 2; the distance between the child and caregiver, shown in a solid green line, starts at 0 feet, which is also the frame shown in Fig. 1. Around minute 4 of the interaction, the child began to move away from the caregiver, shown in Fig. 3, and away from the caregiver, robot, and experimenter around minute 6 of the interaction, shown in Fig. 4. For sake of continuity, most of the graphs or figures in this paper that show a single child participant’s data comes from the same child; the only exception is in Fig. 11.

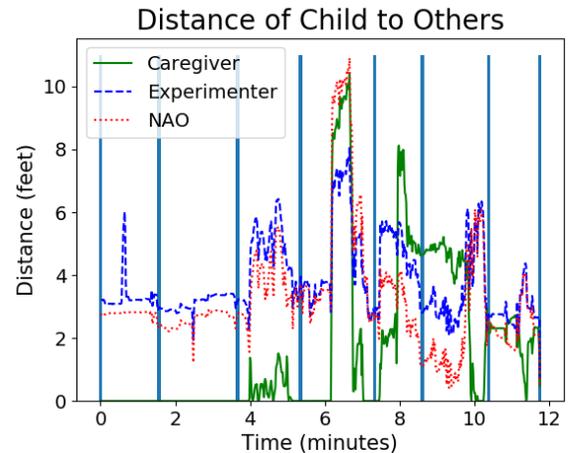


Figure 2: The Euclidean distances between the child and parent, child and robot, and child and experimenter (smoothed by averaging over every second). Blue vertical bars indicate the beginning of a press for social interaction.

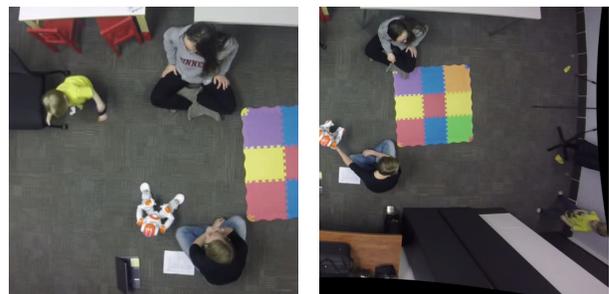


Figure 3: The child moves away from caregiver around minute 4. Figure 4: The child moves away from caregiver around minute 6.

In some experiments, two caregivers were present during the interaction; the minimum distance between the child and either parent was used in our data, ensuring that we can reasonably compare children with one or two caregivers present. The interactions, which are recorded at 33

frames/second, were reduced by averaging the Euclidean distances in one second windows to smooth the data slightly; this averaged dataset is what we used for data analysis.

#### 4 Data Analysis and Results

Our first data reduction method reduces the three time series per child into one, essentially normalizing the data.

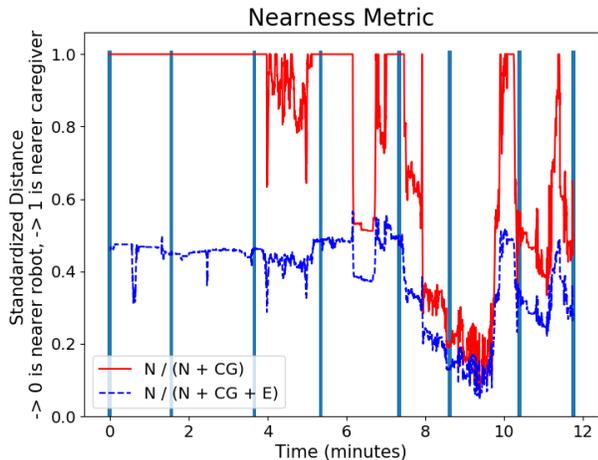


Figure 5: Condensing three time series to a single metric, with and without considering the distance to the experimenter (smoothed by averaging over every second). Blue vertical bars indicate the beginning of a press.

Fig. 5 shows two ways of doing this. The first reduction considers the distance to the robot, denoted  $N$ , and the distance to the caregiver, denoted  $CG$ , in Eq. 1.

$$N/(N + CG) \quad (1)$$

The second reduction also considers the distance to the experimenter, denoted  $E$ , in Eq. 2, so as to retain data on some cases in which participants sought the experimenter’s company while interacting with the robot and not the caregiver.

$$N/(N + CG + E) \quad (2)$$

In either case, we have reduced multiple series of variable magnitude to a single, unit-less number scaled  $[0, 1]$ , where a 1 value means closest to caregiver and a 0 value means closest to the robot, regardless of the absolute distance between child and robot or child and caregiver. We will return to our first goal of representing a single child’s interaction with a robot over time, but we move now to our second goal, to compare and contrast all participants with each other, with these raw Euclidean distance data.

Recall that each interaction varies in length due to potential buffer time between presses in a single interaction. The buffer time, lasting up to one minute between presses as needed, was included in the last press that occurred. For example, if Child A needed a 40 second break after a one minute press, but Child B didn’t need a break and only used two seconds after a one minute press, the press lasted for 100 seconds for Child A but only 62 seconds for Child B.

This flexibility in interaction time naturally raises the question of how to compare these variable length data. The time difference between presses is capped at 60 seconds, and any time between presses is used to draw the participant’s attention back to the robot. By and large, the excess time was spent by the participant getting a snack or toy, playing with other items in the room, or talking to their caregiver; none of the buffer time was spent interacting with the robot while the robot was moving autonomously. Therefore, we consider the time between presses to be noise.

To remove this noise, we choose a simple method of aligning all presses in all interactions, and we truncate each press to the length of the shortest occurrence of that press over all participants. For example, say Child A took 60 seconds during Press 1 with a 30 second break, then 180 seconds during Press 2 with a 10 second break. Say Child B took 60 seconds during Press 1 with a 2 second break, then 180 seconds during Press 2 with a 2 second break. It should be noted that in most cases the experimenter manually starts the next press for attention with the robot, so short breaks of 1-3 seconds is simply the time taken to reach over and push buttons on the robot (or occasionally, to first re-orient the robot towards the child if they shifted position). Our first data exploration only compares participant reactions to the robot while the robot is actively moving or speaking; thus, Press 1 is truncated to 62 seconds for both participants and Press 2 is truncated to 182 seconds for both participants. Alternatively, we could have timed the presses off-line and truncated participant data with those timings, but in practice, some participants progressed through the interactions in immediate succession and this off-line timing was unnecessary.

The effect of such data loss, i.e. 28 seconds after Press 1 and 8 seconds after Press 2 in the above example, admittedly contains some distance data. Either the participant didn’t move between presses or approached the robot again from somewhere else in the room. In theory, a participant might have outlasted the one minute buffer time and started the next press further from the robot than they were at the (truncated) end of the previous press. However, in practice, this did not happen; only the first two cases occurred. Thus, minimal interaction information was lost due to this truncation. Comparisons of these extraneous portions of time that may include redirection, however, are left as future work.

Following this alignment and truncation procedure for every press, we next used several classic similarity measures for comparing interactions. With the shortened, three channel time series, we applied four classic similarity measures as given in (Cassisi et al. 2012): mean similarity, root mean square similarity, peak similarity, and Pearson’s correlation function. In any case where a participant did not finish all seven presses, we compared only the presses he did finish with other participants. The mean similarity and root mean square similarity uses the similarity between two numbers as defined in Eq. 3.

$$numSim(x, y) = 1 - \frac{|x - y|}{|x| + |y|} \quad (3)$$

Let two time series,  $X = x_1, x_2, \dots, x_n$  and  $Y = y_1, y_2, \dots, y_n$ . Then we define mean similarity in Eq. 4, root

mean square similarity in Eq. 5, peak similarity in Eq. 6, and the cross-correlation or Pearson’s correlation function in Eq. 7. In all cases,  $n$  is the length of the time series  $X$  and  $Y$ ; in Pearson’s correlation function,  $l$  allows a shifted comparison of positions within the second time series.

$$tsim(X, Y) = \frac{1}{n} \sum_{i=1}^n numSim(x_i, y_i) \quad (4)$$

$$rtsim(X, Y) = \sqrt{\frac{1}{n} \sum_{i=1}^n numSim(x_i, y_i)^2} \quad (5)$$

$$psim(X, Y) = \frac{1}{n} \sum_{i=1}^n \left[ 1 - \frac{|x_i - y_i|}{2max(|x_i|, |y_i|)} \right] \quad (6)$$

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{X})(y_{i-l} - \bar{Y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (y_{i-l} - \bar{Y})^2}} \quad (7)$$

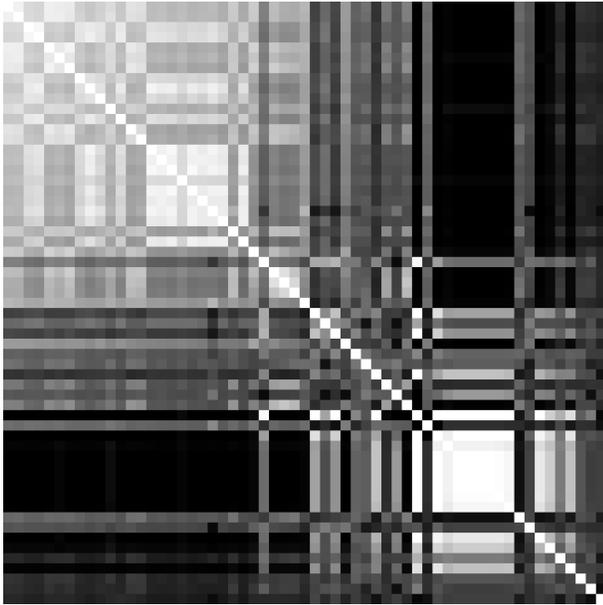


Figure 6: Mean similarity between the 59 children, with each number represented as the darkness of a grey-scale pixel. White pixels indicate total similarity; black pixels represent complete dissimilarity. This is only the distance to caregiver, sorted in order of average intensity of each child participant’s similarity scores to the other participants.

We first considered each data stream individually, comparing each child to every other child (using the shorter of the two data streams if a participant did not finish all presses). The similarity measure for each child to every other child was multiplied by 255 and turned into a three-part tuple, which was then used as the grey-scale pixel color in a 59x59 image where a row and column correspond to the similarity measure between two children. Fig. 6 shows the similarity measures between 59 participants using the distance to

caregiver as the only data stream, using the mean similarity measure. The participants have been ordered by the intensity of the average of the entire vector of differences per child.

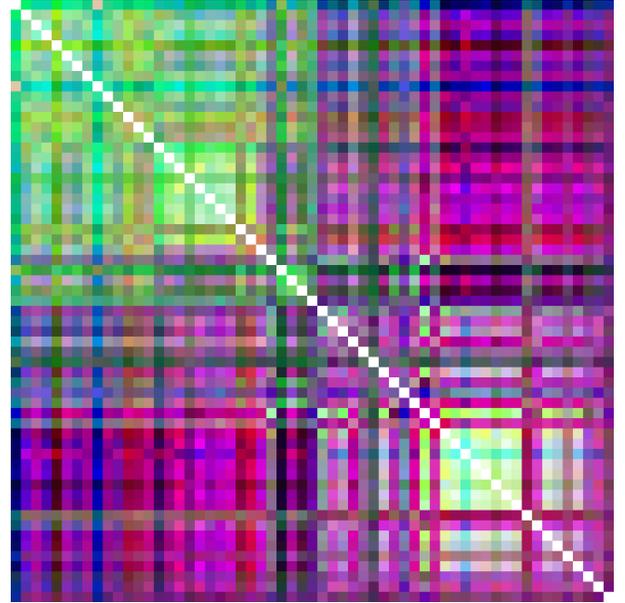


Figure 7: Best viewed in color. Mean similarity between the 59 children, in three color channels: caregiver in green, experimenter in red, and robot in blue. White pixels indicate total similarity; black pixels represent complete dissimilarity. This represents all channels of the three-channel time series that form a participant’s proxemics, sorted in order of average intensity of distance to caregiver of each child participant’s similarity scores to the other participants, resulting in the same sorting order as the single channel in Fig. 6.

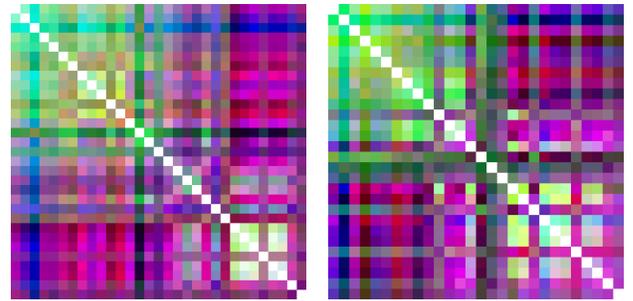


Figure 8: The combined mean similarity measure between the 32 male participants, sorted by average distance to caregiver. Figure 9: The combined mean similarity measure between the 28 female participants, sorted by average distance to caregiver.

Figures 7, 8, and 9 show combined similarity measures between all 59 participants, just the 32 males, and just the 28 females, respectively. These images use all three channels, again using the mean similarity measure; the similarity

measure between the participants' distance to robot was assigned to the blue color channel, caregiver distance was assigned green, and experimenter distance was assigned red. Each image has been ordered by the average intensity of the caregiver channel of the time series, meaning the order of participants in Fig. 6 and Fig. 7 are the same. Note that the images look similar, but the variation in distance to robot and distance to experimenter adds further characterization of the participants. In Figures 6 – 9, the order of the children is the same across the rows and columns, resulting in a symmetric image split from top left to bottom right, sorted by the summation of the intensity of differences per child. In these images, total similarity comes out as white, and total dissimilarity comes out as black.

Thus far, we have used absolute and continuous distances, even in our distance metrics in Eq. 1–2. Consider that shorter distances between a participant and another actor probably indicate more comfort or interest, and longer distances indicate less comfort or interest. A participant sitting two feet away from the robot and five feet away from the caregiver probably indicates strong comfort with or interest in the robot (e.g. Fig. 10). However, if a participant is seated on their caregiver and the caregiver is seated two feet from the robot, they probably need comfort from their caregiver while they watch the robot (e.g. Fig. 13). Two feet of distance between the participant and the robot can indicate very different levels of comfort and interest. We therefore must create a system which considers relative, rather than absolute, distances to quantify and discretize the levels of interest or comfort the child has in the robot.

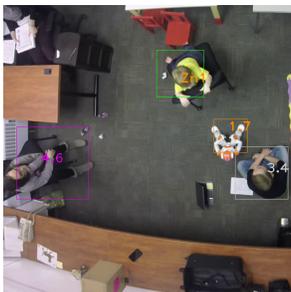


Figure 10: Zone 1 – the child is closest to the robot.

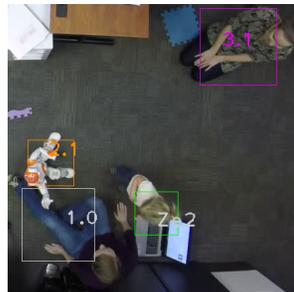


Figure 11: Zone 2 – the child is nearest the experimenter.

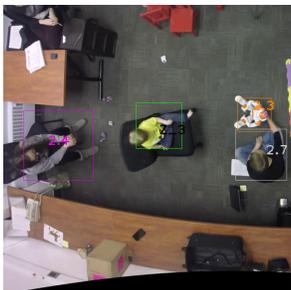


Figure 12: Zone 3 – the child is not close to anyone.

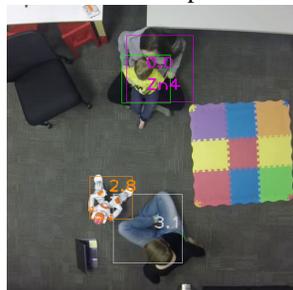


Figure 13: Zone 4 – the child is nearest the caregiver.

There are four basic places a child can be during an inter-

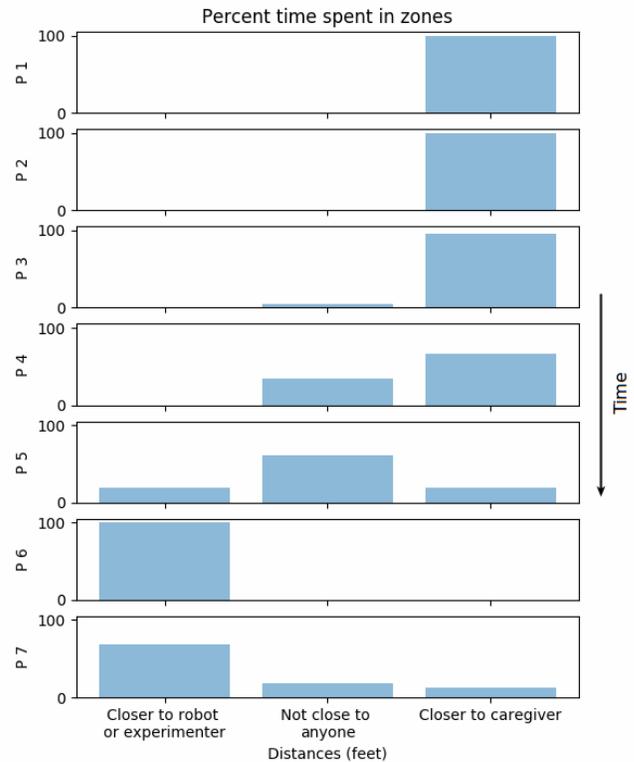


Figure 14: Percent of time spent in each zone for one participant for each press, from P1 to P7. Note the progression from Zone 4, closest to caregiver, to Zone 1, closest to robot. This can indicate greater comfort with the robot or less need for comfort from a caregiver, or both, as time progressed.

action: closest to their parent, closest to the robot, closest to the experimenter, or not close to any of these. If the child is exactly mid-way between the robot and caregiver, we cannot say there is a clear preference for one or the other. Therefore, we want to qualify definite spaces that indicate a preference for robot (or experimenter) or caregiver, with a buffer space between them. The simplest method of doing so is to split the distance between the robot and the caregiver into thirds, giving the caregiver, robot, and the buffer an equal share of the distance. If the child is within the third of distance closest to the caregiver, we say they have a preference for the caregiver (Fig. 13); if the child is within a third of that distance to the robot, we say they have a preference for the robot (Fig. 10). If they are much closer to the experimenter than the robot while within that distance (which happens very infrequently), we say they have a preference for the experimenter (Fig. 11). If none of those situations apply, the child is near no-one (Fig. 12). We call these locations 'zones,' and number them from 1 to 4, in order of preference from robot (Zone 1), experimenter (Zone 2), no-one (Zone 3), to caregiver (Zone 4). When averaging zones across time, we use the mode zone during that time.

We first show a sample breakdown of percent time spent

in each zone over an entire interaction, separated by press, in Fig. 14 (for the same participant as in Fig. 1, 3, and 4). This figure shows the presses in order from 1 to 7, top to bottom, as the percent time in each press spent in each zone. This child spent most of the time in Press 1 and Press 2 near their caregiver; in Press 3 and Press 4 they began to explore the room and spent time away from (or in between) the robot and the caregiver. In Press 5 more time was spent away (or between) the robot and caregiver, until Press 6 was entirely spent nearest the robot or experimenter, and Press 7 was more mixed. This concludes our first goal, representing a single child’s interaction proxemics.

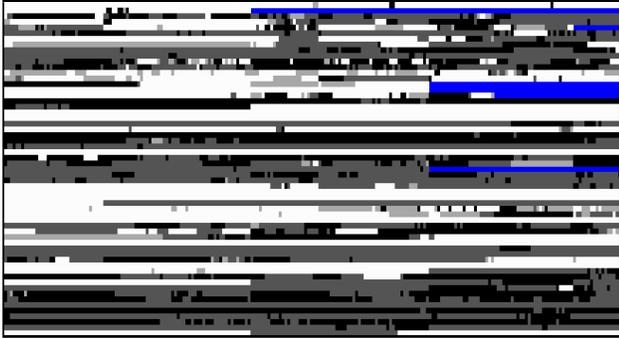


Figure 15: Zones throughout every child’s interaction, sorted top to bottom by increasing age. White indicates time closest to caregiver, and black indicates closest to robot.

Fig. 15 shows all children’s zones over time; the blue bar indicates a lack of press information (i.e. that the experiment had ended). White indicates the zone closest to the caregiver, light gray indicates not close to anyone, darker gray indicates closest to experimenter, and black indicates closest to the robot. The 59 children here are sorted from youngest to oldest, with the youngest at the top. A border has been included to make the top two rows more obvious – these children were always in Zone 4, or nearest the caregiver. At this juncture, it appears that as age increases, the zones tend closer towards the robot.

We show the same zone information sorted by age with just the male participants in Fig. 16 and with just the female participants in Fig. 17. Note that the tendency towards being closer to the robot is more visible in boys. We also graph the total percent of time per child in each zone as a three dimensional plot in Fig. 18. Note that there are two areas of concentration: 100% near caregiver, and 100% near robot. A three dimensional plot of age (in months) versus preference for robot, experimenter, or caregiver (over the entire interaction) is shown in Fig. 19. Note that the age range varies across the cluster of individuals that spent 100% of their time near the robot, but there is a slight effect of age on percent time spent with robot or caregiver. This concludes our second goal of comparing and contrasting all participants.

## 5 Conclusions and Future Work

In this paper we have shown several ways of collapsing our child-robot interaction proxemics data. We began by show-

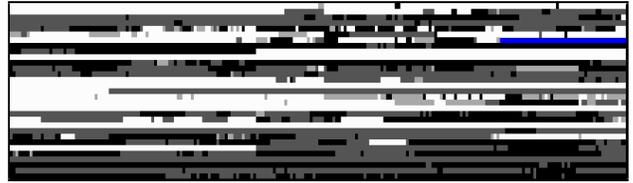


Figure 16: Zones throughout the male participants’ interaction, sorted top to bottom by increasing age.



Figure 17: Zones throughout the female participants’ interactions, sorted top to bottom by increasing age.

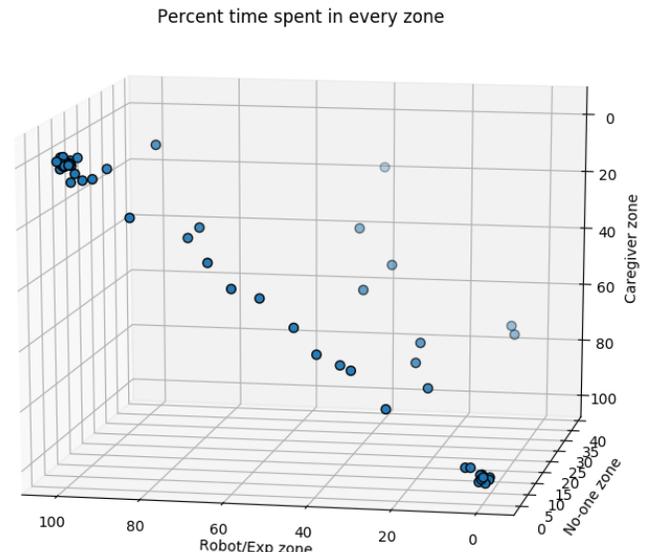


Figure 18: Percent of time spent in each zone for all participants; some jitter has been added to all points to make the clusters at both ends more visible.

ing the simple Euclidean distance between the child participants and other actors of interest, and normalizing that three-channel time series into a single metric over the same amount of time. We then employed similarity measures over the time series both as three separate series and a three-channel series, condensing variable length interactions to a single distance measure between participant pairs. We then discussed how using the space around the parent, robot, or experimenter could be simplified into a single zone at any point in time, and showed how this can demonstrate progression over time from one zone to another. We also succinctly showed all participants’ progressions over time.

Future work will include broader methods of participant

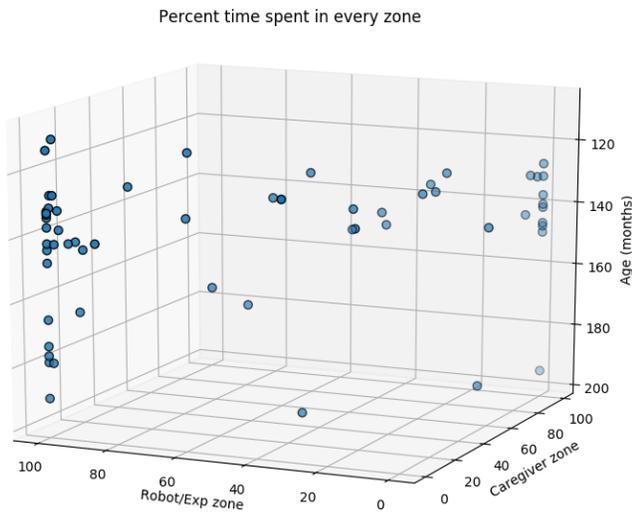


Figure 19: Percent of time spent in caregiver and robot zones (without any adjustment for time spent in no-one’s range), plotted against age in months. This combines robot and experimenter zones. Note a slight preference for older children for being closer to the robot.

comparison, participant clustering, and further validations of statistical significance between male and female participants and over age. First, while our initial method of equalizing participants’ time series is truncation of excess noise, it could be that the time lapse between presses holds key information— for example, longer time between presses potentially indicates a less interested child or a more active child that repeatedly needs their attention drawn back to the robot. Thus, other methods that include the entire interaction, such as dynamic time warping, on a press-by-press basis or on the normalized nearness metric across all participants, might show useful differences.

Second, we will explore how to group or cluster these participants; while straightforward methods such as k-means require a priori knowledge of how many clusters to use, other methods such as k-medoids may prove to give reasonable (if not optimal) solutions. We have also explored choosing clusters based on the visual data (e.g. the two green patches and the pink or blue strips in Fig. 6), but we need to explore potential relationships between participants’ similarity scores and participants’ location zones over the interactions. Lastly, while we have visually explored the sex and age differences of participants, it is not yet clear how much statistical significance the differences hold.

## 6 Acknowledgments

The authors thank the National Science Foundation “I/UCRC: Robots and Sensors for the Human Well-being,” [NSF IIP-1439728] and the MnDRIVE RSAM initiative at the University of Minnesota for partial support for this work, and the anonymous reviewers for their feedback.

## References

- Asada, K.; Tojo, Y.; Osanai, H.; Saito, A.; Hasegawa, T.; and Kumagaya, S. 2016. Reduced personal space in individuals with autism spectrum disorder. *PLoS One* 11(1).
- Cassisi, C.; Montalto, P.; Aliotta, M.; Cannata, A.; and Pulvirenti, A. 2012. Similarity measures and dimensionality reduction techniques for time series data mining. In Karahoca, A., ed., *Advances in Data Mining Knowledge Discovery and Applications*. InTech.
- Dautenhahn, K., and Werry, I. 2004. Towards interactive robots in autism therapy: background, motivation and challenges. *Pragmat. Cogn.* 12(1):1–35.
- Diehl, J. J.; Schmitt, L. M.; Villano, M.; and Crowell, C. R. 2012. The clinical use of robots for individuals with autism spectrum disorder: a critical review. *Res. Autism Spectr. Disord.* 6(1):249–262.
- Fasching, J.; Walczak, N.; Morellas, V.; and Papanikolopoulos, N. 2015. Classification of motor stereotypies in video. *IEEE Int. Conf. Intell. Robot. Syst.* 4894–4900.
- Feil-Seifer, D. J., and Matarić, M. J. 2011. Automated detection and classification of positive vs. negative robot interactions with children with autism using distance-based features. *Proc. 6th Int. Conf. Human-robot Interact. - HRI '11* 323–330.
- Hashemi, J.; Spina, T. V.; Tepper, M.; Esler, A.; Morellas, V.; Papanikolopoulos, N.; and Sapiro, G. 2012. Computer vision tools for the non-invasive assessment of autism-related behavioral markers. *2012 IEEE Int. Conf. Dev. Learn. Epigenetic Robot. ICDL 2012* 1–33.
- Hashemi, J.; Tepper, M.; Vallin Spina, T.; Esler, A.; Morellas, V.; Papanikolopoulos, N.; Egger, H.; Dawson, G.; and Sapiro, G. 2014. Computer vision tools for low-cost and noninvasive measurement of autism-related behaviors in infants. *Autism Res. Treat.*
- Mead, R.; Atrash, A.; and Matarić, M. J. 2013. Automated proxemic feature extraction and behavior recognition: applications in human-robot interaction. *Int. J. Soc. Robot.* 5(3):367–378.
- Pennisi, P.; Tonacci, A.; Tartarisco, G.; Billeci, L.; Ruta, L.; Gangemi, S.; and Pioggia, G. 2016. Autism and social robotics: A systematic review. *Autism Res.* 9(2):165–183.
- Scassellati, B.; Matarić, M. J.; and Admoni, H. 2012. Robots for use in autism research. *Annu. Rev. Biomed. Eng.* 14(1):275–294.
- Scassellati, B. 2005. Quantitative metrics of social response for autism diagnosis. *Proc. - IEEE Int. Work. Robot Hum. Interact. Commun.* 2005:585–590.
- Scassellati, B. 2007. How social robots will help us to diagnose, treat, and understand autism. *Robot. Res.* 552–563.
- Zwaigenbaum, L.; Bryson, S.; Rogers, T.; Roberts, W.; Brian, J.; and Szatmari, P. 2005. Behavioral manifestations of autism in the first year of life. *Int. J. Dev. Neurosci.* 23(2-3 SPEC. ISS.):143–152.