

A Study of Reaching Motions for Collaborative Human-Robot Interaction

Sara Sheikholeslami¹, Gilwoo Lee², Justin W. Hart³, Siddhartha Srinivasa², and Elizabeth A. Croft⁴

Abstract In human-human interactions, individuals naturally achieve fluency by anticipating the partner’s actions. This predictive ability is largely lacking in collaborative robots, leading to inefficient human-robot interactions. Fluent meshing in human-robot collaboration requires the robot to make its intentions clear to its human collaborator. We propose a unified generative model of human reaching motions that allows the robot to **a)** infer human intent, and then **b)** plan its motion to be legible, or intent-expressive. We conducted a study on human reaching motion and constructed an elliptical motion model that is shown to yield a good fit to empirical data. In future studies, we plan to confirm the effectiveness of this model in predicting human intent and conveying robot intent for achieving fluency in human-robot handovers.

1 Introduction

We focus on single-arm reaching motions as a natural communication channel in tightly coupled physical human-robot collaborative tasks, like a human-robot handover scenario shown in Figure 1. Such collaboration requires legible coordination of the specific behavior of human-robot handovers: as the human reaches out to handover the water-bottle, the robot should be able to tell early on and reach out to receive the bottle (Figure 1).

Similar to human-human interactions, attaining fluency in human-robot interactions requires the collaborative robot to be able to infer the intentions of its human counterparts in order to determine the next appropriate action [1–3]. Likewise, in taking actions, the robot must balance two, often juxtaposed, objectives: **1)** moving in a *predictable* (expected) way that the human trusts and understands [4], and **2)** moving in a *legible* (intent-expressive) way that conveys its intent to the human - it enables the inference of intentions [5].

Few studies have sought to develop algorithms for the prediction of goal-directed human reaching motions [6], and generating legible robot motions [5]. These studies often model human-like motion as complex and time-expensive cost functions capturing different aspects of how the human observer(s) expects the robot to move [7, 8]. Such trajectory optimization techniques present two major challenges

¹Department of Mechanical Engineering, University of British Columbia, Vancouver, BC, Canada

²Paul G. Allen School of Computer Science & Engineering, University of Washington, Seattle, USA

³Faculty of Computer Science, University of Texas at Austin, Austin, Texas, USA

⁴Faculty of Engineering, Monash University, Australia

e-mail: sara.sheikholeslami@ubc.ca

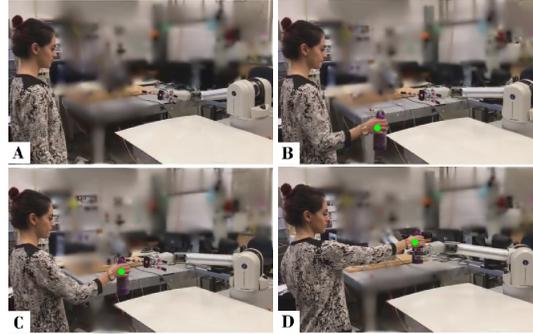


Fig. 1: An example of a human-to-robot handover in a shared-workspace collaboration: Human initial hand motion (**A,B**) is used to make predictions about the remainder of the path, generating a suitable trajectory for the robot's motion to grasp the bottle (**C,D**).

in high dimensional space: **(1)** learning the cost function based on context as the efficiency of robot motion has different interpretations for different observers [9, 10]); and **(2)** finding an efficient robot trajectory in real time that minimizes such non-convex cost functions, often subject to local minima [11]. In more ambiguous situations with many possible reach targets, such as a collaborative assembly tasks, the robot is faced with a much more challenging burden of conveying its intent. Modeling fast predictions of how the human observer expects the robot to move to be intent-expressive is an active research challenge that benefits from investigating some regularities of biological motion control [12].

Work in neuroscience [13, 14] and biomechanics [15, 16] suggests that the central nervous system (CNS) formulates motor commands for human arm reaching movements in terms of spatial trajectories of the hand rather than joint angles [14, 17]. Findings indicate that the hand follows a predictable path with smooth and symmetric bell-shaped velocity profiles [18], motivating our investigation of reaching motion geometry to model human reaching motions.

2 Research Goals and Hypotheses

Our overarching goal is to create a single, unified model for describing reaching motions that can be solved in real-time for both motion prediction and motion production of collaborative robots in various shared human-robot collaborative work-space paradigms. This requires **(a)** understanding and predicting how people coordinate their reaching motions, **(b)** modeling these mechanisms, and **(c)** incorporating the models into robots to produce trajectories that synchronize naturally with human motion. This work primarily addresses **(a)** and **(b)**, setting out to better understand and construct models of human reaching motions.

To this end, we proposed the following Goal (**G1**) and Hypothesis (**H1**):

- **G1:** Understanding and modeling unconstrained goal-directed human reaching movements.
- **H1:** For human reaching movements, the hand trajectory follows a predictable ellipse-shaped path with smooth velocity profiles.

If **H1** holds true, an elliptical fit to the motion early in the path can be predictive of the remainder of the path, and, subsequently, models of velocity and acceleration can be used to determine the final timing and location of the reach trajectory early in the process. This predictive ability is necessary to attaining fluent meshing in human-robot team activities

3 A Study of Legible Motion

To better understand and construct models of human reaching motions, a high-fidelity Vicon motion capture system was used to track human participants reaching movements in a simulated box packing task. The task was designed to represent a possible warehouse scenario with the goal of generating a sample of reaching movements that would be naturally occurring in the application domain.

***Central Insight:** Human unconstrained reaching motion tends to follow an arc shaped trajectory that can be modeled by an ellipse.*

3.1 Task

Participants were asked to fulfill the role of a warehouse packer in a shipping operation while we recorded their movements. Each participant sat at a desk. Surrounding them were a ramped shelf with five columns of colored ball supplies, a ball holder with a single space for intermediate ball placement, and a rack consisting of a 3x3 square grid with a total of 9 possible spaces for final packing, Figure 2. The experimental task proceeded in the following phases:

- P1. Picking**, in which products are staged and scanned to be packed. This phase involves unshelving and staging a ball by removing a ball from the ramp and placing it on the ball-holder; followed by
- P2. Packing**, in which products are packed for shipment. This phase involves un-racking the current ball from the ball-holder and arranging it on an available spot on the rack.

The *Picking* and *Packing* tasks were repeated 9 times, such that the rack was full. Participants were instructed to pick and pack balls in any order they desired as the final arrangement of the balls on the rack was not important— The goal was to allow participants to focus their attention on completing the task at hand (see discussion of limitations in Section 6). After, in order to prepare for another round of testing, participants reset the setup by removing balls from the rack and placing them back onto the shelf.

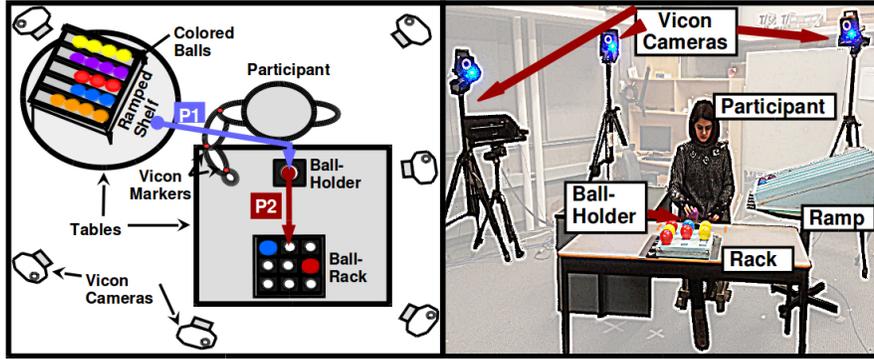


Fig. 2: Simulated Box Packing task designed with the goal of understanding and modeling point-to-point human reaching movements. **P1** corresponds to trajectories collected from the *Picking* phase, i.e. from ramped shelf to ball-holder, and **P2** corresponds to trajectories collected from the *packing* phase, i.e. from from ball-holder to ball-rack.

Each trial was repeated 15 times, yielding a total of 135 unshelving and racking motion samples (**P1**), and 135 unracking and packing motion samples (**P2**) per participant. For the study, 10 participants were recruited.

3.2 Data Annotation

The Vicon motion capture system used in this study included six cameras that performed data processing to create real-time raw *2D* marker data. To minimize distortion of the movement and improve tracking accuracy (i.e., fill in any gaps in the trial data), a total of 19 markers were attached to each participant’s upper body to measure the kinematics of their motion, and construct a calibrated *3D* model of their upper-body skeletal structure¹.

The results herein are based on the reconstructed motion of the wrist marker, placed above the head of the ulnar bone, which we take as representing the movement of the arm in space. Two different categories of reach trajectories were collected from each trial: **P1**) corresponding to motion samples from the ramp to the ball-holder during *Picking*, and **P2**) corresponding to motion samples from the ball-holder to the rack during *Packing*. The start and end of each reach sample was determined through simple heuristics such as the wrist position, velocity, and distance from the target.

4 Building Elliptical Models of Human Reaching Motion

Ellipses arise as second degree curves generated by the intersection of a plane and a cone. Therefore, to justify an elliptical fit to the reach data, the reach trajectory must be planar (Section 4.1). After verifying planarity, an ellipse is fit to the motion points projected onto the best fit plane; the fit error is measured as the sum of the squared orthogonal distances between the found ellipse and the motion points (Section 4.2).

¹ See Vicon Nexus Documentation at <https://docs.vicon.com>

4.1 Modeling Planarity of Reaching Motion

To compute the best fit plane to a motion trajectory, principal component analysis (PCA) can be applied to each motion trajectory, providing an intuitive visualization and approximation of a new orthogonal basis along which the motion points are mainly distributed (cf. Figure 3). In the context of this study, the covariance matrix Σ is a 3 by 3 positive-definite or positive-semidefinite matrix constructed from the 3D coordinates of a single reach sample as [19]:

$$\Sigma = \frac{(\mathbf{X} - \mathbf{X}_{centroid})(\mathbf{X} - \mathbf{X}_{centroid})^T}{n} \quad (1)$$

where n is the number of the points collected in one reaching motion, and \mathbf{X} is a 3 by n matrix with all the 3D coordinates of the points row-stacked. $\mathbf{X}_{centroid}$ is the mean coordinate, i.e., the centroid of the reach points.

Eigenvectors of Σ (i.e. the principal components (PCs)) point to 3 mutually perpendicular directions along which the reach points, n , are mainly distributed, and subsequently, construct the basis of the best fitting plane. The corresponding Eigenvalues of Σ indicate the mean spread of the points along each PC. Therefore, in practice, due to the presence of noise in laser scanning data, a reach trajectory can be assumed planar only if the smallest PC, approximating the mean deviation of the points normal to the best fit plane surface, is found statistically insignificant compared to the other two non-trivial PCs constructing the plane of the motion.

4.2 Modeling Reaching Motion as an Ellipse

The equation of an ellipse in cartesian coordinates is a polynomial of degree two with five conic coefficients, defined as $Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$. For a sample trajectory projected onto the best fit plane as described in Section 4.1, each transformed motion position (x, y) places one constraint on these conic coefficients. Therefore, the best fit conic to the motion points can be modeled as the null vector of the matrix obtained from stacking the constraints from n wrist positions as it moves through space [20].

The type of the conic section, whether elliptical or not, can be determined from the sign of the invariant discriminant of the conic section. For a conic to be elliptical, 2 conditions must satisfy [21]:

$$B^2 - 4AC < 0 \quad (2a)$$

$$A \neq C \quad (2b)$$

For elliptical conics, determined by the above conditions, The fit error is measured as the mean of the squared orthogonal distances between the found ellipse and the motion points (Section 5.2).

5 Results

Results presented in this section are based on **P1** and **P2** movements collected from 9 participants. For each participant, the number of sampled trajectories is indicated

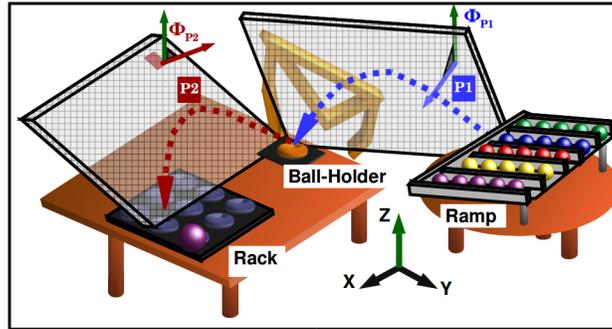


Fig. 3: The 3D view of the box packing experiment. For each phase, (**P1**, and **P2**) , ϕ represents the angle between the normal vector of the best fit plane to each reach sample, and the Z-axis of the global coordinate system.

as N , and correspond to the sample of trajectories that we were able to reconstruct from the scanned data due to the noise inherent in the Vcon system (e.g., Figure 4).

5.1 Analyzing Planarity of Human Reaching Motions

Principal component analysis (PCA) was performed to each of the collected reaching trajectories from **P1** and **P2** for all participants to model the orthogonal bases defining the plane of the motion (Figure 4). Motion planarity is approximated by the smallest principal component which corresponds to the mean offset of the fit motion plane with respect to the horizontal X-Y plane of the camera.

A. The PCA Solution: The covariance matrix Σ , defined in Section 4.1, has two degrees of freedom comparing the spread of points along the transformed basis of the fit plane: the ratios $\{eigenvalue3 : eigenvalue2 : eigenvalue1\}$, or equivalently the 3 principal components (PCs) less one for scale. Figure 4 provides a visualization of the mean of the smaller of the two ratios, comparing the minor planar distribution of motion (i.e. along the vertical axis of the fit plane, PC_2) to its out of plane spread (PC_3).

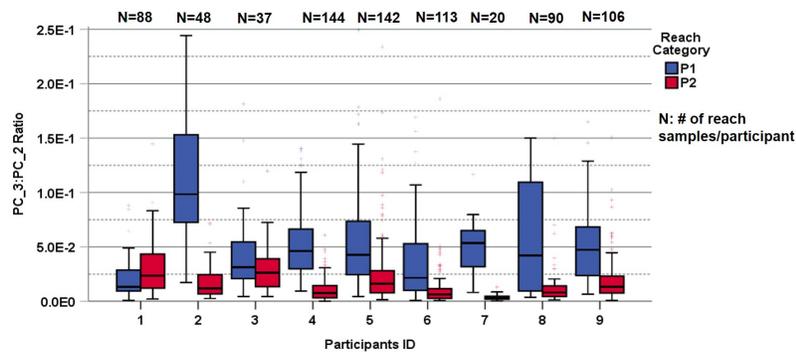


Fig. 4: Motion planarity is evaluated from the PCA solution, providing a comparison of the motion distribution along the minor axis of the modeled motion plan, approximated by PC_2 , to the deviation of the motion perpendicular to the plane (PC_3).

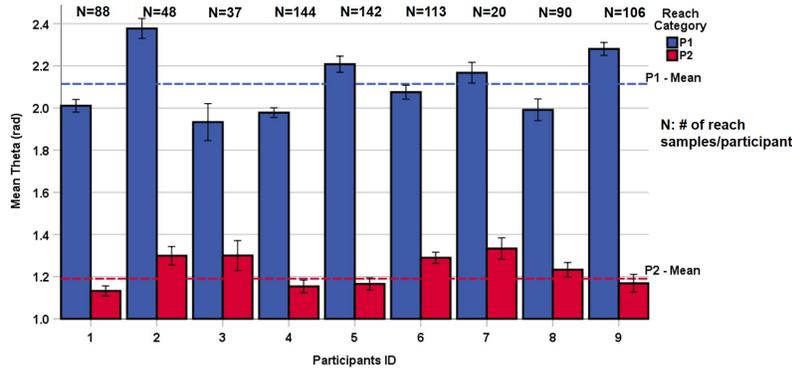


Fig. 5: θ represents the angle between the normal vector of the best fit plane to each reach sample, and the Z-axis of the global coordinate system.

A repeated measures ANOVA, with a Greenhouse-Geisser correction, to account for the violation of sphericity assumption, was applied to the PCA solution of 1574 reconstructed trajectories from 9 participants (collected from Section 3) to evaluate the planarity of the sampled motions. Results of this analysis determined that the mean eigenvalue ratios were statistically significantly different ($F(1.368, 3233.32) = 12425.89, p < 0.0001$). Further, for each motion category, **P1** and **P2**, a Bonferroni analysis was performed to provide pairwise comparisons of the motion distribution along each principal direction. This post hoc test revealed that both motion categories, the distribution of the points perpendicular to the surface of motion, approximated as smallest eigenvalue of Σ (PC_3), is significantly negligible compared to the distribution of the points along PC_1 ($p < 0.0001$), or PC_2 ($p < 0.0001$), corresponding to the major and minor axis of motion plane, respectively. These experimental results indicate that human reaching motions can be accurately fitted to a plane, regardless of the constraints imposed during the experimental task (Section 3).

B. Modeling the Motion Plane: The smallest PC represents the vector normal to the plane of the motion, forming an inclination angle ϕ with the vertical Z-axis of the camera frame (Figure 4). our experimental results indicate that the fit planes to both **P1** and **P2** reaching movements are at an oblique angle to the vertical Z-axis of the camera frame, represented as inclination angles ϕ_1 and ϕ_2 , respectively (Figure 4). Analysis of Independent-samples *t*-tests was performed to the measures of ϕ from 1574 trajectories across **P1** ($N = 788$) and **P2** ($N = 788$) motion categories; ϕ_1 was found to be significantly smaller than ϕ_2 for all participants ($t(1492.61) = 83.69, p < 0.0001$). This analysis indicates that the participants consistently moved more horizontally to the camera X-Y plane to perform P2 reaching movements compared to P1 reaching movements (Figure 5).

5.2 Evaluating the Fit of Elliptical Models to Human Reaching Motion

Analysis of the planar reach trajectories, transformed using PCA (Section 4.1), indicates that human reaching movements can be modeled as an ellipse 89% of the

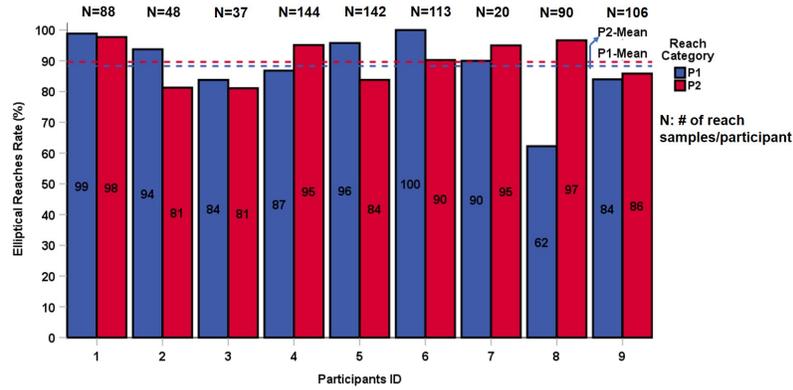


Fig. 6: Rate of reach trajectories for both **P1** and **P2** data (Section 3.1 that fit to an Elliptical Conic collected for the box-packing task across the participants.

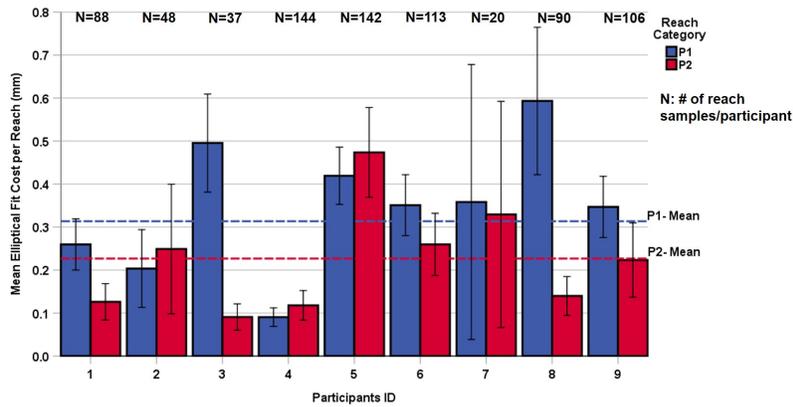


Fig. 7: The fit error measured as the sum of the squared orthogonal distances between the found ellipse and the motion points. The bars indicate the margin of error for a 95% confidence interval.

time, for both **P1** and **P2** motion categories ($N : 788, M : 88.35\%, SD : 11.50\%$, and $N = 788, M : 89.64\%.SD : 6.76\%$, respectively), see Figure 6. Supporting **H1**, evaluation of the performance of an ellipse on approximating the motion path determined that the model can fit the sampled data within a mean accuracy of $\pm 0.31mm$ for **P1**, and $\pm 0.23mm$ for **P2** motion trajectories, see Figure 7.

Figure 8 illustrates two different views of the 3D coordinates of a sample trajectory collected from a random participant overlaid by the corresponding predictions of the elliptical model. The predicted ellipse closely fits the reach.

6 Discussion

This work proposes a mathematically simple and very efficient elliptical motion model of human movements that involves geometric fitting to the motion points,

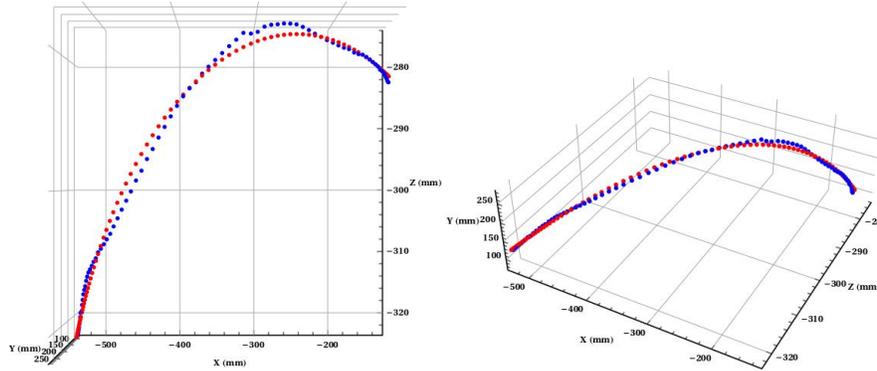


Fig. 8: A demonstration of the current state of the developed model applied to a sample reaching trajectory collected from the Vicon cameras: Top (*Left*) and Isometric (*Right*) views of the cartesian position of the wrist marker, in *mm*, as it moves through space (blue), and the corresponding elliptical model approximated to the wrist trajectory (red).

given the entire sampled trajectory. The analysis of our experimental data indicates that such model exists, and is able fit the motion within a mean accuracy few millimeters (Section 5.2).

A. Implications: Collaborating agents often share spaces, parts, tools and equipment and, in the normal course of their work, encounter conflicts when accessing such shared resources. To resolve such conflicts, each agent’s selection of anticipatory actions must have a temporal dependency on the actions of their collaborating counterpart, and be based on **early** and **fast** predictions of the future state of their collaborator’s motion. Existing motion planning techniques for these contexts primarily focus on learning and optimizing complex cost functions of biological motion control, rendering online and fast predictions of human-motion very challenging. The proposed elliptical motion model (**H1**) has potential to alleviate some of these issues due to its simple formulation. Future studies will evaluate the effectiveness of the elliptical model in predicting human intent to robot, and generating legible motion conveying robot intent for achieving fluency in human-robot handovers.

B. Limitations and Future Work: The presented model is preliminary, and developed based on the limited data sampled from 9 individuals. To test the validity of the elliptical model (**H1**), the experiment was designed to minimize the influence of any external constraints, such as visual information or proprioceptive disturbances, on determining the model of an ongoing movement. Our analysis only focused on the hand positional data, and did not include other types of motion, nor other important aspects, such as orientation, grasp condition, or the relative coordination of the gaze, head and hand.

Future investigations will aim to validate and refine the current model by generating a more exhaustive exploration of the application domain, incorporating other constraints and variations. Our current analysis is investigating the effect of introducing visual cues on the onset of the movement by incorporating data collected

from an analogous box-packing task (Section 3.1); however, instead of providing no guidelines to participants, the experiment was modified such that, during each phase, two computer monitors provided pictographic instructions on the completion of the task. Another limitation and exciting area of future work that our current analysis lacks is to evaluate and compare motion consistency among individuals based on the measures of the inclination angle of the computed motion plane, and/or the fit ellipse parameters, such as the axis ratio.

6.1 Summary

This work has aimed to understand and construct models of human reaching motions to facilitate prediction of the timing and location of the end of human arm reaching motions in tasks such as grasping or object handover. Incorporating the elliptical model (**H1**) into robotic devices can improve human-robot interaction fluency by (1) providing faster predictions of human motion, enabling the robot to make quick inferences, and (2) generating time-efficient and human-like robot trajectories that conveys intent, enabling the human collaborator to focus their attention on completing the task at hand, rather than controlling, or understanding the robot.

References

1. M. Tomasello, M. Carpenter, J. Call, T. Behne, H. Moll, *Behavioral and Brain Sciences* **28**(5), 675–691 (2005)
2. A.D. Dragan, S. Bauman, J. Forlizzi, S.S. Srinivasa, in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (ACM, 2015), pp. 51–58
3. M. Cakmak, S.S. Srinivasa, M.K. Lee, S. Kiesler, J. Forlizzi, in *Proceedings of the 6th International Conference on Human-robot Interaction* (ACM, New York, NY, USA, 2011), HRI '11, pp. 489–496
4. M. Beetz, F. Stulp, P. Esden-Tempski, A. Fedrizzi, U. Klank, I. Kresse, A. Maldonado, F. Ruiz, *Autonomous Robots* **28**(1), 21 (2009)
5. A.D. Dragan, K.C.T. Lee, S.S. Srinivasa, in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (2013), pp. 301–308
6. G. Hoffman, C. Breazeal, *IEEE Transactions on Robotics* **23**(5), 952 (2007)
7. J. Lasseter, in *ACM Siggraph Computer Graphics*, vol. 21 (ACM, 1987), pp. 35–44
8. T. Flash, N. Hogan, *Journal of neuroscience* **5**(7), 1688 (1985)
9. B.D. Argall, S. Chernova, M. Veloso, B. Browning, *Robotics and autonomous systems* **57**(5), 469 (2009)
10. P. Abbeel, A.Y. Ng, in *Proceedings of the twenty-first international conference on Machine learning* (ACM, 2004), p. 1
11. A.D. Dragan, G.J. Gordon, S.S. Srinivasa, in *Robotics Research* (Springer, 2017), pp. 309–326
12. A. Sciutti, G. Sandini, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **25**(12), 2295 (2017)
13. T. Flash, N. Hogans, *Journal of neuroscience* **5**, 1688 (1985)
14. C. Harris, D. M. Wolpert, **394**, 780 (1998)
15. G. Wu, F.C. van der Helm et al., *Journal of Biomechanics* **38**(5), 981 (2005)
16. E. Todorov, M. Jordan, *Nature Neuroscience* **5**, 1226 (2002)
17. P. Morasso, *Experimental Brain Research* **42**(2), 223 (1981)
18. H. Collewyn, C.J. Erkelens, R.M. Steinman, *The Journal of Physiology* **404**(1), 157 (1988)
19. A. Nurunnabi, D. Belton, G. West, *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Science*, Volumes I-3 pp. 269–274 (2012)
20. R.I. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd edn. (Cambridge University Press, ISBN: 0521540518, 2004)
21. A.B. Ayoub, *Mathematics Magazine* **66**(5), 322 (1993)