Understanding Robots - Making Robotic Movements Clearer

Miguel Faria[1] Francisco S. Melo[1] Ana Paiva[1]

***Abstract*— The abstract goes here.**

I. Introduction

Robots are an increasingly more common sight in everyday lives in the most diverse capabilities, from toys like Anki's Vector to iRobot's Roomba vaccum cleaner to Paro the Therapeutic Robot. Thus, research in the field human-robot interaction and its possible applications has been important for a correct integration of robots and adapting them to correctly assist in tasks like healthcare [1]–[3], education [4], [5], entertainment [6], amongst others.

However, to fully integrate robots in society, they need to be able to correctly interact and communicate with humans. The need for correct communication is essential for interactions between humans and robots to be safe [7], natural [8] and efficient. [9], [10]

Humans communicate in a myriad of different ways: some explicitly convey intention, like speech or haptic communication; others implicitly convey intention, like body movements or proactively executing an action or task, requiring interpretation by the receiver on the sender's intentions. [11]

The combination of explicit and implicit communication allows humans to easily interact with each other, even in situations where no explicit signals are traded: like in collaborative tasks where all parties work together towards the same goal, without always explicitly communicating their current objectives or intentions. Robots involved in human-robot interactions must then also leverage explicit and implicit means of communication for those interactions to feel normal by humans. Thus, taking advantage of the humans' natural inclination to resort to implicit communication cues to help in understanding intentions. [12]

In fact, leveraging both implicit and explicit communication has been shown to have positive impacts in human-robot interactions. The case of Breazeal et al. [13], showed how the use of a combination of implicit and explicit communication, in a collaborative task of guiding a robot through lighting a sequence of buttons, erred less and finished the task in less time than using only explicit communication. Likewise, in [14], Liang et al. applied an AI agent that used implicature, an implicit communication mean where the utterances imply information beyond that the words literally convey, to a collaborative game achieving interactions considered as more natural by humans. In [15], [16] robots used signaling with lights and body movements to convey intentions, thus overcoming limitations created by their embodiment.

Nonverbal communication through movement is an essential mean of implicit communication, cross-cultural and most humans comprehend such communications without being previously explicitly taught their meaning. [17] Thus, also an essential communication technique used by robots in natural and efficient interactions with humans. As Saunderson et al. discuss in their 2019 review on the impacts on humans of nonverbal communication by robots [18], the use body movement by a robot can lead to people considering it more socially engaging, warm, friendly and empathetic; as well as more expressive when conveying different emotions. Saunderson et al. also discuss also the correct use of body movement can improve task performance, causing less errors and better completion times.

Other examples of the positive influence of the use of communication with movement are [15], [19]–[21]. In [19], [20], expressive movements, called legible movements, were used to allow the human partners in the task to understand what the robot's objectives were at the time and collaborate accordingly: in [19] the human partner would gather the correct order and in [20] the task partners would coordinate themselves to help the robot to fill the cups with water. In [15], expressive movements mimicking those of home pets improved a Sphero robot communication capabilities, allowing it to lead the humans it was interacting with from the room the interaction was taking place to another room with a bowl with sweets the humans could take sweets from. In [21], explicit communication and body movements were combined to allow a robotic mobile platform to navigate through an environment shared with humans without colliding with humans while reducing human effort to predict the robot's intentions and increasing the human's trust in the robot.

Expressive movements are a powerful tool for robots to better communicate with humans and one type of such movements are legible movements. Legible movements, fist proposed by Dragan et al. in [22], are movements that augment a robot's expressiveness by making their movements easier to read by users through an exaggeration of the robot's trajectory in a manner that removes doubt about what the robot is reaching for or what its objective is.

Legible movements have, thus, been explored to understand their true impacts and advantages in human-robot interactions. In [19], [20], legible movement was explored as a means to improve the efficiency of a team in human-robot collaboration tasks. In [23], the authors explored how a robot can improve it's legibility by learning during interacting

[1]M. Faria, F.S. Melo and A. Paiva are with INESC-ID and Instituto Superior Técnico, University of Lisbon, Portugal. E-mail: `miguel.faria@tecnico.ulisboa.pt` and `{fmelo,ana.paiva}@inesc-id.pt`.

with humans in order to improve the human's prediction capabilities about the robot's objectives. Legible motion is also powerful because it increases the transparency of a robot's state, as evidenced in [24] where a robot used legible movements to express its limitations and incapabilities.

Most works that explore the uses and capabilities of legible movements focus in single-user scenarios [19], [23]. However, for a robot to be fully integrated in society and help humans in their daily lives, it must be able to be expressive in single and multi-user scenarios, as would happen in a surgery scenario, like Kaplan et al. in [2] refer to, or in an educational scenario in which a robot would have to deal with multiple learners. This way, we find necessary to explore how legible movements would impact multi-user interactions. There are evidences that support the use of legible movement in multi-user scenarios, as in [20], but these evidences are for specific tasks and it is not certain they can be generalized.

In this work we propose an extension to how we have looked at legibility and how to generate legible movements. We propose generating legibility but instead of looking at how legible a movement is for each person involved individually, we evaluate the legibility of the movement as the average legibility for all people involved in the task or interaction.

By considering a movement's legibility as the average of the legibility of that movement for each individual participant, a robot when executing a legible movement will guarantee that the movement is always legible for all users, instead of legible for one of the users but possible not legible for another user involved in the interaction.

In the following sections we will first, in section II, define what is legible movement and the principles behind legibility, as well as defining how we can generate legibility considering each person's point of view, contrary to the omniscient view in [25]. In section III we will explore how to consider legibility in multi-user scenarios, the drawbacks of considering only single users and how to generate legible movements that take into consideration multiple users. In section IV, we will show the results of the two different validation tests executed: results for a simulation test where we compared the legibility metric of the multi-user approach versus the traditional single-user approach in a simple multi-user interaction scenario are showed in section IV-A and the results of the user study we conducted on M-Turk, using the same scenario as in section IV-A, are showed in section IV-B.

II. Legible Motion

Legible motion was presented by Dragan et al. in [22] as type of movement that allowed a robot to be more expressive during interaction with humans, drawing on the fact that humans interpret an action's objective in a *goal-oriented* manner. [26], [27] This means that humans, when observing an agent performing an action, they consider the agent to be logical and thus will execute the action in the most efficient way possible. Leveraging on this fact, a legible movement tries emphasize the movement's objective by performing a movement that brings the robot as close as possible to the objective and at the same time as far away as possible from other possible objectives as soon as possible.

To achieve this emphasis on the objective, legible movements leverage animation principles to create trajectories that are readable [28] and that encourage "anticipatory" [29] movements, creating trajectories that convey the most relevant information for goal prediction in the beginning of the movement. A legible movement is then a movement that, when observed by a human, allows the human to infer the objective G_R as soon as possible given the observed movement ξ:

$$Legibility(\xi) = \frac{\int P(G_R|\xi_{S\to\xi(t)})f(t)dt}{\int f(t)dt}, \tag{1}$$

where $P(G_R|\xi_{S\to\xi(t)})$ is the probability of achieving objective G_R given the observed trajectory $\xi_{S\to\xi(t)}$. The term $P(G_R|\xi_{S\to\xi(t)})$ in equation 1 models how a human attributes the probability of the goal G_R being reached by trajectory ξ observed until timestep t, this probability is obtained by

$$P(G_R|\xi_{S\to\xi(t)}) \propto \frac{\exp(-C(\xi_{S\to\xi(t)}) - V_{G_R}(\xi(t)))}{\exp(-V_{G_R}(S))}P(G_R), \tag{2}$$

with $C(\xi_{S\to\xi(t)}$ the cost of the observed trajectory from the start until timestep t, $V_{G_R}(Q)$ the best possible cost of a trajectory that starts in Q and finishes in the goal G_R and $P(G_R)$ is a prior on G_R that can be uniform in the absence of prior knowledge.

The cost function used in $C(x)$ models how a human expects the robot to move in an efficient manner, with lower costs meaning more efficient and predictable trajectories. Following previous works that use legible movement [25], one possible cost function is the sum of squared velocities $C(\xi) = \frac{1}{2}\int \|\xi'(t)\|^2$ that encourages smooth trajectories that go straight to the goal.

Using the definition of legibility in equation 1, it is possible to generate legible trajectories using a gradient ascent approach that in each iteration improves the legibility score of the trajectory ξ:

$$\xi_{i+1} = \xi_i + \frac{1}{\eta}M^{-1}\nabla Legibility(\xi), \tag{3}$$

M is used to measure the norm of a trajectory, $\|\xi\|_A^2 = \xi^T A\xi$.

Legible movement as described in [22] assumes humans have an omniscient view, that is, consider that the human partner has an omniscient view of the workspace. This assumption can lead to movements that go outside the field-of-view of a human, passes through human blind spots or through obstructed parts of the workspace, creating trajectories that are not as legible as expected.

To solve this problem, Nikoladis et al. in [30] propose an extension to the original approach to generating legible movements. In this approach the cost function $C(x)$ becomes a function $\overline{C}(x) = C(\bar{\xi}) = C(T(\xi))$, where $T(\xi)$ is the transformation that combines the transformation of trajectory ξ from world coordinates to coordinates in the human referential and then projects the 3D points into 2D points in the human point-of-view.

With the extension, the legibility metric loses the omniscient approach and instead, before calculating the legibility of the movement, the trajectory is projected to the point-of-view of the human partner and then the legibility is calculated on that trajectory. The transformation into the human's viewpoint allows the optimization procedure to improve the legibility of the trajectory to create trajectories that are always inside the user's field-of-view. This improves the human partner's ability to correctly predict the robot's objective, as showed by Nikoladis et al. in [30].

III. Multi-User Legible Motion

As we addressed in section II, legible motions were defined as a way to allow robots to be more expressive by leveraging the human perceiving an agent executing an action in a rational way and combining it with animation principals to create movements that convey intentions as soon as possible. These motions have been shown in [19], [23], [30] to increase human confidence and ability to predict a robot's intentions, but they have been tested in single-user scenarios, that is scenarios where a robot interact with one human at a time. However, there are several example of scenarios where a robot must interact with multiple humans at the same time. Example of these are the scenario described in [6] where a robot plays a game of cards with three other people, or in [2] where a robot would be inserted in a surgical team to support the surgical staff, or in [20] where a robot has to sequentially fill cups of water to different human partners. In all of these scenarios, for a robot to use legible motions it must be able to generate legible movements that are simultaneously legible for all the partners involved, otherwise it could optimize the legibility for one partner but reduce the legibility for another if the other partner's field-of-view is not adequate.

In order for legible motions to be legible to multiple users simultaneously, we propose an extension to the model of Nikoladis et al. [30] where instead of the legibility metric only consider one user, it will equally consider all users when evaluating the legibility of a trajectory.

To extend the model presented in [30], we propose that the legibility metric is an average of the legibilities of the legibility for each user. So, when we improve the legibility, for example via a gradient ascent approach, we are looking for a trajectory that is not perfect for one of the partners involved in the task, but instead we are looking for the trajectory that maximizes the average legibility of the group. With this principle, we assume that for the group to perform adequately and understand more easily the robot, it is better that the legibility for each individual partner is lower but that the legibility for the overall group is higher, leading to better cooperation within the group.

A. *Definition of Multi-User Legibility*

Under this new paradigm of multi-user legibility, we define the legibility of a trajectory ξ as follows:

$$Legibility_{MU}(\xi) = \frac{\sum_{n=1}^{N} Legibility_{Hn}(\xi)}{N}, \tag{4}$$

where N is the number of human partners and $Legibility_{Hn}(\xi)$ is the legibility of the trajectory ξ for one of the human partners, i.e., the legibility of the trajectory in the point-of-view of one of the human partners as defined in [30].

Using this new definition of legibility, the expression for the update in optimization procedure also changes to

$$\xi_{i+1} = \xi_i + \frac{1}{\eta} M^{-1} \nabla Legibility_{MU}(\xi), \tag{5}$$

with

$$\nabla Legibility_{MU}(\xi) = \frac{\sum_{n=1}^{N} T_n^{-1} \cdot \nabla Legibility_{Hn}(\xi)}{N}, \tag{6}$$

where T_n^-1 is the transformation from the human partner to the coordinate space where ξ was defined and

$$\nabla Legibility_{Hx}(\xi) = \nabla \mathcal{L}(\xi(t)) \cdot \nabla_{3D}^{2D} T_x(\xi(t)), \tag{7}$$

where $\nabla \mathcal{L}(\xi(t))$ is as defined in [25] and $\nabla_{3D}^{2D} T_(\xi(t))$ is the gradient for the transformation of the trajectory ξ to the perspective of the human partner in question.

IV. Model applied

We validated our proposed model in two stages: first we defined a simulation scenario where we compared the legibility of a trajectory optimized using our model, versus the legibility of a trajectory that used the model defined in [30]; second, we used the same scenario as in the simulation tests to create videos of trajectories executed by a robot and through a questionnaire, distributed over Amazon's M-Turk, tested how the trajectories obtained with the different models differed in the eyes of real humans.

A. *Simulations*

The scenario used for the simulations was a simple interaction in which a robot had to grab one of three objects on a table and three humans had to predict which object the robot was going to grab.

We defined different configurations of object positioning and human poses, under the restriction that the human orientation was such that their field of view captured both the robot and the objects on the table. Also, since we were modeling a real world scenario, the human were oriented as if looking in a slightly downward manner as they would in real life; this downward angle would vary between 10^{o} and 15^{o} depending on the distance between the human and the objects on the table.

The different configurations are showed in Figure ***ADD FIGURE REFERENCE HERE*** and aimed at representing both configurations that one could encounter in usual interactions - configurations 1 to 4 - as well as configurations that would present a challenge to execute a legible movement given the distance and placement of both the humans and the objects - configurations 5 and 6.

*1) **Results**:*

B. User Study

For the user study we used configuration 2, since it was the one that would best represent a normal interaction with the robot and the three human parts around a table and the three objects in the middle of the table. To create the videos used in the study we used the WeBots simulator and replicated configuration 2 of Figure ***ADD FIGURE REFERENCE HERE***.

We generated 12 different trajectories: 4 different trajectories for each object and for each object 1 trajectory was optimized using the multi-user model and 1 trajectory was optimized for each one of the human's points-of-view. Then, each of the 12 trajectories was recorded from the point-of-view of each of the 3 humans resulting in 36 different videos of 20 seconds each.

To test how the different trajectories differed in human's eyes, we split each of the 36 videos in videos with 6, 12 and 18 seconds, in order to compare how fast and how confident a human was in predicting the robot's target under each optimization model.

We then conducted a between-subjects study, where each participant would watch 3 sets of 6, 12 and 18 seconds videos for different movements all optimized using the same optimization model, this way each participant evaluated the legibility of one optimization model. After each video the participant was asked to predict the object the robot is going to grab and rate how confident they are in their prediction; at the end of each set the participant would watch the full 20 seconds of video and was asked if the movement matched their prediction. After watching the 3 sets of videos the participant would have to rate how clear the three movements were.

*1) **Results**:*

V. Conclusion

Acknowledgments

References

[1] F. S. Melo, A. Sardinha, D. Belo, M. Couto, M. Faria, A. Farias, H. Gambôa, C. Jesus, M. Kinarullathil, P. Lima *et al.*, "Project inside: towards autonomous semi-unstructured human–robot social interaction in autism therapy," *Artificial intelligence in medicine*, vol. 96, pp. 198–216, 2019.

[2] K. E. Kaplan, K. A. Nichols, and A. M. Okamura, "Toward human-robot collaboration in surgery: performance assessment of human and robotic agents in an inclusion segmentation task," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 723–729.

[3] C. A. Huijnen, M. A. Lexis, R. Jansens, and L. P. de Witte, "Mapping robots to therapy and educational objectives for children with autism spectrum disorder," *Journal of autism and developmental disorders*, vol. 46, no. 6, pp. 2100–2114, 2016.

[4] S. Chandra, P. Alves-Oliveira, S. Lemaignan, P. Sequeira, A. Paiva, and P. Dillenbourg, "Children's peer assessment and self-disclosure in the presence of an educational robot," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2016, pp. 539–544.

[5] E. Yadollahi, W. Johal, A. Paiva, and P. Dillenbourg, "When deictic gestures in a robot can harm child-robot collaboration," in *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 2018, pp. 195–206.

[6] F. Correia, P. Alves-Oliveira, T. Ribeiro, F. S. Melo, and A. Paiva, "A social robot as a card game player," in *Thirteenth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2017.

[7] G. Michalos, S. Makris, P. Tsarouchi, T. Guasch, D. Kontovrakis, and G. Chryssolouris, "Design considerations for safe human-robot collaborative workplaces," *Procedia CIrP*, vol. 37, pp. 248–253, 2015.

[8] A. Sauppé and B. Mutlu, "Design patterns for exploring and prototyping human-robot interactions," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014, pp. 1439–1448.

[9] B. Mutlu, A. Terrell, and C.-M. Huang, "Coordination mechanisms in human-robot collaboration," in *Proceedings of the Workshop on Collaborative Manipulation, 8th ACM/IEEE International Conference on Human-Robot Interaction*. Citeseer, 2013, pp. 1–6.

[10] M. A. Goodrich and D. R. Olsen, "Seven principles of efficient human robot interaction," in *SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance (Cat. No. 03CH37483)*, vol. 4. IEEE, 2003, pp. 3942–3948.

[11] A. Bauer, D. Wollherr, and M. Buss, "Human–Robot Collaboration: a Survey," *International Journal of Humanoid Robotics*, vol. 05, no. 01, pp. 47–66, 2008.

[12] N. Gildert, A. G. Millard, A. Pomfret, and J. Timmis, "The need for combining implicit and explicit communication in cooperative robotic systems," *Frontiers in Robotics and AI*, vol. 5, p. 65, 2018.

[13] C. Breazeal, C. D. Kidd, A. L. Thomaz, G. Hoffman, and M. Berlin, "Effects of nonverbal communication on efficiency and robustness in human-robot teamwork," in *2005 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2005, pp. 708–713.

[14] C. Liang, J. Proft, E. Andersen, and R. A. Knepper, "Implicit communication of actionable information in human-ai teams," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–13.

[15] M. Faria, A. Costigliola, P. Alves-Oliveira, and A. Paiva, "Follow me: Communicating intentions with a spherical robot," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2016, pp. 664–669.

[16] K. Baraka, A. Paiva, and M. Veloso, "Expressive lights for revealing mobile service robot state," in *Robot 2015: Second Iberian Robotics Conference*. Springer, 2016, pp. 107–119.

[17] P. Ekman and W. V. Friesen, "The repertoire of nonverbal behavior: Categories, origins, usage, and coding," *semiotica*, vol. 1, no. 1, pp. 49–98, 1969.

[18] S. Saunderson and G. Nejat, "How robots influence humans: A survey of nonverbal communication in social human–robot interaction," *International Journal of Social Robotics*, vol. 11, no. 4, pp. 575–608, 2019.

[19] A. D. Dragan, S. Bauman, J. Forlizzi, and S. S. Srinivasa, "Effects of robot motion on human-robot collaboration," in *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2015, pp. 51–58.

[20] M. Faria, R. Silva, P. Alves-Oliveira, F. S. Melo, and A. Paiva, ""me and you together" movement impact in multi-user collaboration tasks," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 2793–2798.

[21] Y. Che, A. M. Okamura, and D. Sadigh, "Efficient and trustworthy social navigation via explicit and implicit robot–human communication," *IEEE Transactions on Robotics*, 2020.

[22] A. D. Dragan, K. C. Lee, and S. S. Srinivasa, "Legibility and predictability of robot motion," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2013, pp. 301–308.

[23] B. Busch, J. Grizou, M. Lopes, and F. Stulp, "Learning legible motion from human–robot interactions," *International Journal of Social Robotics*, vol. 9, no. 5, pp. 765–779, 2017.

[24] M. Kwon, S. H. Huang, and A. D. Dragan, "Expressing robot incapability," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 2018, pp. 87–95.

[25] A. Dragan and S. Srinivasa, "Generating legible motion," 2013.

[26] G. Csibra and G. Gergely, "'obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans," *Acta psychologica*, vol. 124, no. 1, pp. 60–78, 2007.

[27] A. L. Woodward, "Infants selectively encode the goal object of an actor's reach," *Cognition*, vol. 69, no. 1, pp. 1–34, 1998.

[28] L. Takayama, D. Dooley, and W. Ju, "Expressing thought: improving robot readability with animation principles," in *Proceedings of the 6th international conference on Human-robot interaction*, 2011, pp. 69–76.

[29] M. J. Gielniak, C. K. Liu, and A. L. Thomaz, "Generating human-like motion for robots," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1275–1301, 2013.

[30] S. Nikolaidis, A. Dragan, and S. Srinivasa, "Viewpoint-based legibility optimization," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2016, pp. 271–278.