Decoding the Unspoken: State of Art Review and Empirical Investigations on Empathy and Efficiency in Human-Robot Interaction through Intent Sharing

Tanmay Badhan[a]

[a]*Institute of Mechanical Engineering, TU Wien, Austria*

E-mail address: e12302474@student.tuwien.ac.at

Abstract

In the coming ages due to developments in generative AI, robots are likely to become socials actors. Henceforth it is important to study how can human robot communications be more empathetic, once such factor in sharing such empathy is intent detection without explicit communication. This paper reviews the current proceedings in intention detection, formulates a framework for intent detection through perception and then explores the implications and impact of humans sharing their intentions with robots before collaborative tasks, and its effects on overall efficiency and user experience. The later is conducted by a user study in a simulated environment using ROS Unity interface. At the end the impact of situational awareness context understanding for Collaborative task is outlined and a short user study is presented. The results reveal significant implications regarding empathy, boredom and intuition which forms a crucial input for designing future robotic interfaces.

Keywords: HRI; Prototype; Emperical Studies; Cobots; Sensing Systems; Perception Modelling; Human-Centered Robotics; AI; Shared-Situation awareness

1. Introduction

Robots are becoming increasingly complex with sensors and intelligent control systems. with sensors and sophisticated controls. This change has been even more evident in recent years, with modern robots developing increasing autonomy and capability to the point where they can actively interact with a human partner toward achieving a common goal. This has led to increasing research in the area of a Human Robot Interactions (HRI), this new field of research is aimed at maximising the performance though a common platform of understanding for robots as well as humans. HRI systems enable high-level task planning and flexibility achievable with humans at centre, while still leveraging the repeatability, precision and load capacity of robots. In HRIs social behaviours, interactions with natural human and fluent partners, communication modalities and cognitive load sharing are important facets.

While much of the previous literature focuses on safety in an HRI environment, we here have a view on agency and shared collaborative control. Here we are defining foundations of sharing of sense of agency and intentional awareness between Humans and Robots in situations where physical interactions between them is crucial, they are physically couples and are cooperating towards a shared goal. One interesting observation from the study is that the foundations of intent detection lie in rehabilitation where human intent is detected to control a robotic device that replaces lost capabilities, here its compensatory in nature, while in case of exoskeletons human intent is detected to enhance human capacity in many cases kinaesthetic capacity.

These four elements (Intent authoring, intent capture, task arbitration, and Feedback retrieval) can be used to model applications requiring physical human robot interaction. for example, Moving a heavy and bulky object in tandem with the collaborative robot.

1.1. Intent Authoring

We define Intent Authoring as a way in which human is communicating its intent. To answer the question on how to sense intent? We first had to go through the psychological analysis if what exactly is intent. High level intent is defined as the human's ability to communicate the trigger for a certain action. This is often binary ON/OFF or a YES/NO state intent. As we go further deep dimensions are added. For example, high level intent can be triggering a motion. A lower level would be the direction of motion and further lower it would be direction and

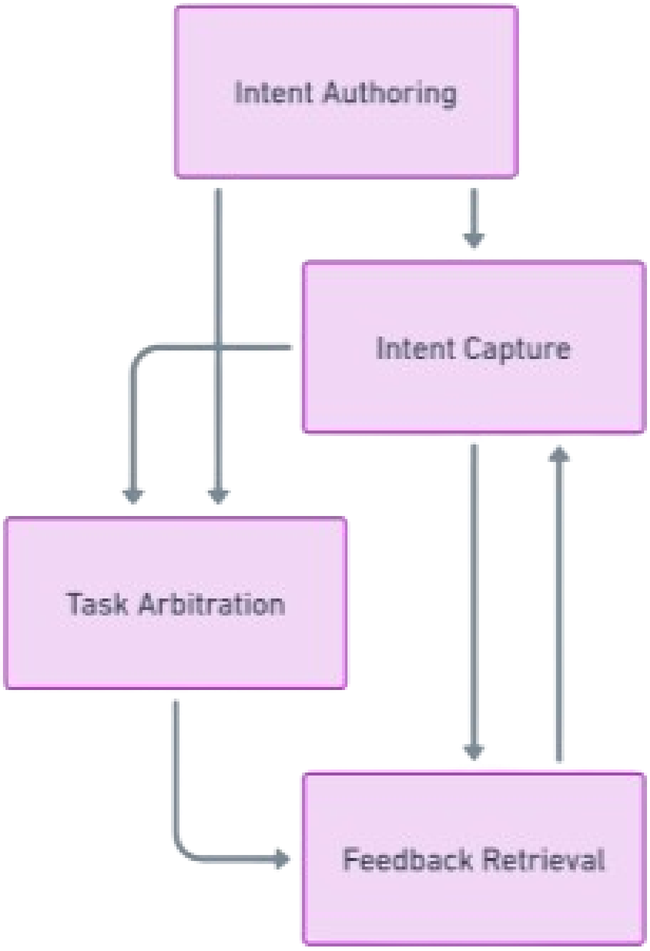

Fig. 1. The four elements of Intent Sharing.

magnitude of motion. Then direction, magnitude, velocity and so on.

1.2. Intent Capture

The component which deals with recognising an action as an intent is intent capture. In many cases the signals and the intent are not necessarily the same things but share a common frame. The design of the shared control architecture defines the quotient of intent captured based on measured signal values. For ex in ref. [1] The signal is measured from the 6 DoF force-torque sensor and intent is interpreted based on rotational and transnational movement of the robot joint. While pattern recognition approaches are appealing for their ability to learn arbitrary mappings between myographic signal features and intent information, they are most commonly used to select from a relatively small number of discrete control states. [2] The interaction forces measured in the active exoskeleton are fed into a model of the exoskeleton dynamics to predict the walker's forward path over a short time horizon. The user's state is a discrete variable representing possible walking modes, e.g., "go straight forward" or "turn to the right".

1.3. Intent arbitration

Types of role arbitration include co-activity, master-slave, teacher-student, and collaboration, which differ in the cost functions, objectives, and expectations of the human and robot agents.
Dynamic role arbitration is the ability to switch or adapt roles based on the situation, the human's intent, performance, or affective state, or the robot's confidence or prediction.
Machine learning and performance metrics are two common tools for determining when and how to change role arbitration, using data driven or task-based approaches.
Algorithms for role arbitration are methods for implementing the role arbitration mechanism, such as linear discriminant analysis, support vector machines, artificial neural networks, radial basis function neural networks, recursive least squares, hidden Markov models, and game theory.
These algorithms can be used to classify, estimate, predict, or optimize the human and robot roles based on the available sensory information and the desired task outcomes.

1.4. Feedback Retrieval

Feedback is the information provided by the robot to the human about the state of the coupled system, the characteristics of the environment, or the suggested actions or strategies for the task.
Feedback design involves choosing the [2] appropriate feedback modality, content, timing, and intensity for the given task and user. Feedback design should consider the user's preferences, abilities, and attention, as well as the task requirements, complexity, and context. Feedback design should also balance the trade-off between providing enough information to enhance the user's performance and satisfaction, and avoiding information overload or distraction.
Feedback evaluation is the process of assessing the effectiveness and impact of the feedback on the user and the task. Feedback evaluation can use objective measures, such as task completion time, accuracy, or error rate, or subjective measures, such as user satisfaction, trust, or workload. Feedback evaluation can also use physiological measures, such as heart rate, skin conductance, or brain activity, to capture the user's emotional and cognitive responses to the feedback.

2. Review of Studies on Impact of Intent Recognition or Anticipation on overall productivity

SAT model: [3]The Situation awareness-based Agent Transparency (SAT) model, a three-level model that organizes the information requirements for optimal human agent team performance. The model suggests that the agent should provide information about its current state, goals, intentions, actions, reasoning, constraints, projections, consequences, likelihood, and uncertainty to the user. This SAT model can help the user to better understand the agent's behaviors, develop shared situation awareness, and calibrate trust in the agent.

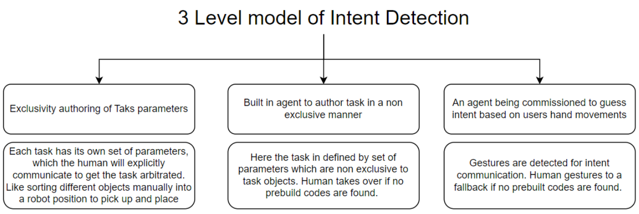

Fig. 2. Intent sharing levels defined in the task

This study [4] how intent sharing, or the communication of goals and plans, affects human-robot collaboration and trust. The authors compare three types of intent sharing: no intent, low-level intent (actions), and high-level intent (goals). They find that high-level intent sharing leads to higher team performance, situation awareness, and trust than low-level or no intent sharing . This research sheds light on the critical role of intent communication in enhancing collaboration and trust between humans and robots.
By sharing intentions effectively, we can unlock the potential for more efficient and empathetic interactions in human-robot teams. Trust is evaluated using subjective ratings and physiological measures such as heart rate variability.
The results reveal that high-level intent sharing significantly improves team performance by enhancing coordination and reducing conflicts. Participants also report higher trust in robots that transparently communicate their goals. Interestingly, individual differences, such as prior experience with robots and personality traits, moderate the impact of intent sharing. For instance, participants with high openness to experience tend to benefit more from high-level intent sharing, while those with low trust in automation show greater reliance on low-level intent information.

3. Experiment Design

Based on the above studies a 3 level experiment was designed and in person study was conducted.

1. Explicit Command: Where tasks have specific parameters that a human communicates explicitly to the robot—for instance, sorting objects for the robot to pick up and place.
2. Prebuilt Task Execution: Here, tasks are defined by parameters that are not exclusive to the objects. The human intervenes only if the robot encounters uncertainty and prebuilt responses are inadequate.
3. Gesture-Based Intent Communication: The robot guesses intent based on the user's hand movements, with gestures being used for intent communication and to signal fallback actions if prebuilt codes are not sufficient.

By observing their ability to perform the task under both conditions, we aimed to gather data on the system's impact on user experience, efficiency, and ease of use. The contrast between the experiences was fundamental to our evaluation of the project.

Level 1: Explicit Command In the first level, participants are instructed to arrange objects in specific patterns which the robot then replicates. This stage assesses the robot's ability to execute tasks with hardcoded parameters based on direct human input.

Level 2: Prebuilt Task Execution The second level introduces an element of robot autonomy. Participants interact with a control interface to signal the robot to perform tasks with some level of prebuilt flexibility. The robot utilizes predefined algorithms to place objects according to the type signaled by the human operator, with the human stepping in only if the robot encounters uncertainty.

Level 3: Gesture-Based Communication The third and most advanced level tests the robot's capability to interpret human gestures. This scenario requires no physical intervention from the human, relying entirely on the robot's ability to decipher and act upon non-verbal cues.

3.1. Simulation of the Experiment

General Environment: The experimental environment for intent detection in human-robot interaction is set up with a robot in a simulated workspace where colored objects are randomly populated. The robot is programmed for pick-and-place tasks and is equipped to handle objects of two colours: Green, and Blue while the human has to intervene in case of other colours. The participants were asked to sort 10 cubes and the timing to accomplish the task in each intent detection level was noted.

Experiment Level 1: The participant engages in a task requiring spatial reasoning and manual interaction with a simulated robotic environment. The task involves clearing the workspace of potential obstacles (represented by white cubes) and sorting colored cubes green to the green target area, blue to blue, and red cubes to be discarded while avoiding collisions. Each action is initiated by the participant pressing a 'move' button, signaling the robot to execute the task. The participant has to wait until the action is completed and the robot is disengaged. This level assesses human involvement in error prevention and task efficiency in a controlled setting.

Experiment Level 2: Participants communicate their intent to the robot through specific keystrokes corresponding to different colored cubes. Initially, tasks mirror the prior setup with participants moving white cubes to prevent robot collisions. Upon encountering a familiar coloured cube, participants are asked to communicate there intent by pressing 'P' for green and 'O' for blue. The robot then comes to the position of cube grabs it and places it at target. No need to manually move cube to desired pickup target. The participant retains manual control in scenarios involving the red cube, directly intervening in the robot's task sequence. This setup explores the efficacy of keyboard-based intent signaling in directing robotic action and the human's role in overriding automated processes when necessary.

Experiment Level 3: This is very similar to level 2 with the only difference being the keystrokes here are replaced by an IMU sensor. The participants are asked to communicate its intent by moving an IMU based toy (Sphero mini). The toy registers the IMU positions and then simulated the keystrokes P, O and I. These gestures direct the robot to handle the colored cubes accordingly, with 'I' being the new input for red cube intervention. The human still has to take over in case white ball is populated at a random.

Using Sphero a realistic feeling of perception input is simulated. They movement are used to track the IMU and map them to keystrokes with an 80 % accuracy. To communicate to sphero a ROS 2 node is running in background which is just simulating keystokes and it is not connected either to Unity nor to ROS, its an independent perception input. The experiment owner manually assigns the listener ON/OFF states to avoid fluctuation in Unity environment.

3.2. NASA Task Load Index (NASA-TLX)

In order to gain a valuable understanding of the cognitive demands imposed by our evolutionary prototype, we utilized the NASA-TLX questionnaire which is an extensively validated and acknowledged instrument for assessing subjective workload. In this subsection, we briefly explain the details of our approach to measuring participant workload.

In our pursuit of a thorough understanding of participant workload experiences, we implemented a post-interaction assessment with this instrument.

Following each participant's interaction with the prototype, we sought to gather their respective subjective opinions. To streamline simplicity and ensure anonymity, participants were presented with a QR core. Scanning this code with their devices granted them access to the questionnaire.

The responses from the questionnaire fed into Google Forms, our chosen platform for data collection. This straightforward integration allowed us for a quick and user-friendly analysis of participant feedback.

Leveraging the simplicity of Google Forms, we were able to simply visualize and interpret the data, gaining insights into the user experiences from our prototype demonstration. The results from the visualized on fig 3.

4. Results

4.1. Positive and negative effects:

+ *Enhanced Accessibility for Individuals:* While the projected reality system is very beneficial for novice cooks, it also significantly improves kitchen accessibility for people with disabilities.
+ *Accuracy and efficiency improvement*: By precise instructions, the cooking process is streamlined and the cooking results will get consistently better.
+ *Minimizing food waste:* Accurate ingredient usage instructions can help reducing food waste, as cooks are less likely to use more than what is needed.

- *Dependency on Technology:* Could lead to cooks following monotonous work steps, not thinking what they are doing and therefore losing creativity and intuitive cooking skills.

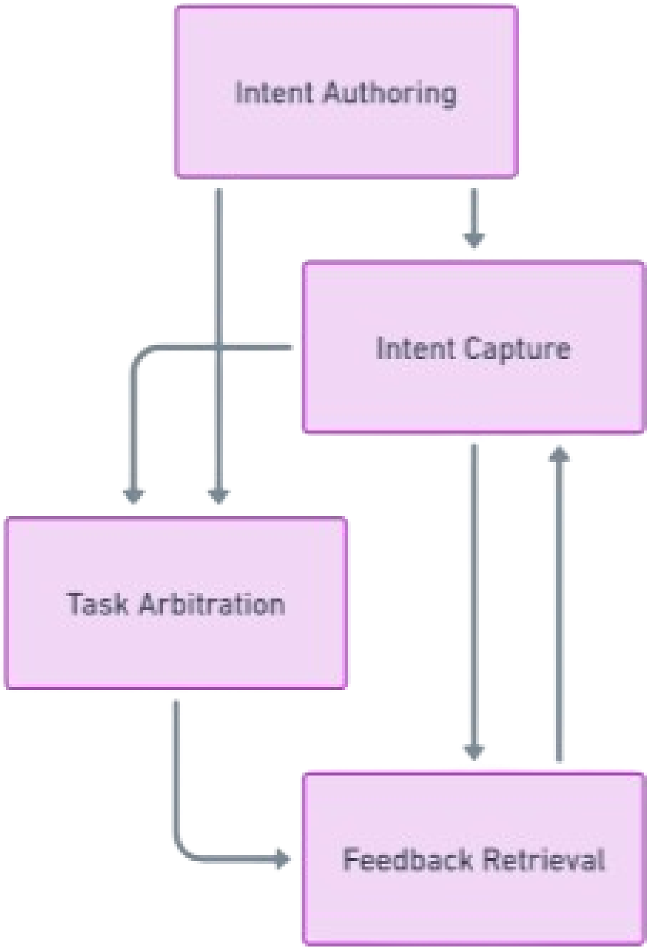

Fig. 3. Results from NASA-TLX questionnaire normalized to 7 point scale

- *Costs:* The initial costs to implement such a system are very high and not all companies have the resources to integrate it into the kitchen workflow.
- *Reduced agency:* Due to dependency on technology for food production the sense of agency is somewhere shifted away from the user, this might result in misjudgement of kitchen workflow for customised dishes or medical specified menu.

4.2. Potential risks:

- *Overcrowding of visual space:* Displaying too much information and complex instructions in kitchens could lead to confusion and errors.
- *Reliability:* The system has to work for different conditions in a condition like lightning or kitchen surfaces. The projection has to be consistent in various conditions.

Hypothesis	Pre-Study		Post-Study		t-Value	p-Value
	Mean	Std Dev	Mean	Std Dev		
Improvement in Instruction Quality	1.67	0.91	4.60	1.06	6.63	0.0001
Improved Accuracy	2.73	4.48	5.87	1.20	2.14	0.0462
Improved Hygiene	1.33	0.98	5.37	1.06	8.85	0.0001
Quick Learning	2.24	0.75	6.37	1.06	10.06	0.0001
Frustration	1.39	0.98	5.87	0.99	10.16	0.0001

Fig. 4. Comparison of Pre-Study and Post-Study Results

4.3. *Observational Insights:*

Some users get frustrated if some extremely simple tasks are hinted. Every one has there own style or method of cooking and the assistant systems doesn't always well align with it.

4.4. *Hypothesis Validation:*

At the beginning of the experiment we had three hypothesis to Validate
Due to time constraints and the experimental complexity, we were not able to collect data for the study without projector hence, statistical data from previous state of art was considered for hypothesis testing. We found a similar dataset from a study conducted by Ghasemi et.al [**?**], the mean SUS score from their questionnaire was normalised to a 7 point scale and similar questions were compared. *Although the experimental setup was similar the environment in which study was conducted may have been different so we add a mean delta of 10 percent for the differences
(H1) Improvement in product quality by a new chief.
To test this hypothesis we considered the following questions:
-Instructional Quality: How effectively did the task teach you kitchen skills?
-Accuracy: How accurately could you perform tasks?
Analogous Questions from the dataset:
-Do you think you would use the system ?
-How Successful were you in accomplishing the task given to you?
(H2) Improved Kitchen Hygiene.
Questions Considered:
-Do you feel that You kept the kitchen clean?
Analogous Question:
-How cumbersome was it to use the system?
(H3) Reduced cognitive load arising due to unknown semantics in the kitchen space.

Questions Considered:
-How quickly were you able to learn and adapt to?
-Frustration: How insecure or frustrated did you feel while doing the task?
Analogous Question:
-How quickly did you learn the system?
-How insure, discourages and annoyed you felt while using the system?

Figure 4 and 5 shows that The t-values are positive, indicating that the post-study mean scores are higher than the pre-study mean scores for all hypotheses. The p-values for all hypotheses are below the standard threshold of 0.05, indicating statistically significant differences in those categories, confirming the effectiveness of the assistance systems.

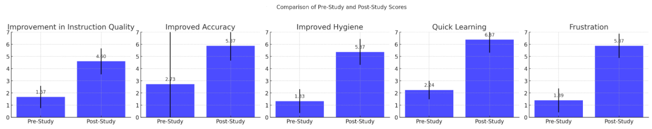

Fig. 5. Visual comparison for Pre and post study results

5. Discussion

In this section the aim is to compare the findings of our study with those from CounterIntelligence: Augmented Reality Kitchen by Bonanni, Lee, and Selker [**?**]. It is important to note that the "Counterintelligence" project used a wider range of technologies like FridgeCam to see on the door what is inside, RangeFinder to measure the surface temperature of food, Augmented Cabinetry help find items in cabinets, and heat sink to show water temperature. Our study focused on a simpler implementation with projection limited to ingredients and recipes due to time and resource constraints within the project.

5.1. *Efficiency and Task Completion:*

"CounterIntelligence" Study: This was measured with time for a participant to find an ingredient. The AR system showed a slight advantage in locating items but not a significant impact.
Our Study: Participants showed high satisfaction with ease of use, performance, and response time. Although it is not comparable to the task completion time, it indicates a positive impact on efficiency in our simpler AR setup.

5.2. *User Experience and Cognitive Learning:*

"CounterIntelligence" Study: Beside the time measurement, this study also used a pre- and post-test questionnaire. It found no significant differences in user experience compared to the normal kitchen workflow, except for illuminated drawers to locate items.
Our Study: High scores in user satisfaction scores like ease of use and engagement. Positive but not that significant responses were in learning experience and impact on learning.

5.3. *Implications for Future AR Implementations:*

It turns out that a much more advanced system like "CounterIntelligence" offers many more possibilities, but a simple system like ours can also lead to considerable improvements. This could highlight that optimization in future AR implementations in kitchens needs a good balance between technological complexity and user tasks.

5.4. Watching a simple instructional Video vs Augmented reality:

As discussed by Ghasemi [**?**], AR is more efficient that Video Based instructions, due to Hands free Interaction, Real time synchronisation with the Users pace, and multimedia segmentation principles can be used to loop a step until its accurately completed.

6. Conclusion and Future Directions

In conclusion, this paper shows the development and evaluation of an innovative assistance system with projected augmented reality, specifically designed for kitchen environments. Our findings demonstrate that the system significantly helps streamlining the kitchen processes and making them more efficient and worker-friendly. Such systems can offer great added value and new opportunities in the future of culinary practices. However, it is important to acknowledge that these systems are not fully developed and further research and work is needed. Future work should focus on refining these systems that they will perfectly aid when needed and never hinder or confuse people.

6.1. Design Implications for Future Directions

The Assistant. Making a Sandwich was a really simple task, it was just an assembly. But many a times cooking is more than just an assembly so the assistant should be tuned to each recipe, based on the experience of the chef, the assistant should adapt its functionality, viz. for a beginner the assistant should also provide instructions on using a food processor, but for an expert it should provide additional functionality ideas only. The assistant should also help for efficient task planning with respect to particular recipe and order log.
Multi modal Interaction. In the current system we still dint have user input for feedback, we have to integrate it. A combination of gestures and Voice based interfaces is very crucial for an hands free integration. for system input and user feedback other modalities should be included, like weighing scale on the table top, Computer Vision algorithms to monitor tasks, kitchen hygiene and way of presentation.For instance, sensors can monitor cooking temperatures and provide alerts if they deviate from the desired range
Flexible Navigation. Users should have the flexibility to navigate back and forth and choose what information they want to engage with, based on their needs and personal cooking process.

User-centered Pro activity. Over-reliance on AR interfaces may lead to distractions if workers spend excessive time interacting with the technology rather than focusing on their tasks. The right balance is crucial. AR can be integrated with the Internet of Things (IoT) and automation systems in the kitchen. This allows for seamless coordination between AR displays and appliances, further enhancing efficiency and reducing the risk of errors.
Respecting Tradition:Assistant systems should also respect culinary traditions and cultural aspects of cooking. Users may have specific traditions or family recipes that they want to preserve, and the technology should support these preferences.

List of Figures

References

[1] V. Q. Nguyen, "Intent recognition of rotation versus translation movements in human-robot collaborative manipulation tasks," 2016. [Online]. Available: https://scholarworks.umass.edu/masters_theses_2/439

[2] D. P. Losey, C. G. Mcdonald, E. Battaglia, and M. K. O'Malley, "A review of intent detection, arbitration, and communication aspects of shared control for physical human-robot interaction," *Applied Mechanics Reviews*, vol. 70, pp. 010 804–010 804, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:8023194

[3] K. E. Schaefer, E. R. Straub, J. Y. Chen, J. Putney, and A. Evans, "Communicating intent to develop shared situation awareness and engender trust in human-agent teams," *Cognitive Systems Research*, vol. 46, p. 26–39, Dec. 2017. [Online]. Available: http://dx.doi.org/10.1016/j.cogsys.2017.02.002

[4] "Survey of Emotions in Human–Robot Interactions: Perspectives from Robotic Psychology on 20 Years of Research - International Journal of Social Robotics — link.springer.com," https://link.springer.com/article/10.1007/s12369-021-00778-6, [Accessed 05-02-2024].