RESEARCH ARTICLE

Analysis of implicit robot control methods for joint task execution

Lena Guinot[1]* , Kozo Ando[1], Shota Takahashi[1] and Hiroyasu Iwata[2]

Abstract

Body language is an essential component of communication. The amount of unspoken information it transmits during interpersonal interactions is an invaluable complement to simple speech and makes the process smoother and more sustainable. On the contrary, existing approaches to human–machine collaboration and communication are not as intuitive. This is an issue that needs to be addressed if we aim to continue using artificial intelligence and machines to increase our cognitive or even physical capabilities. In this study, we analyse the potential of an intuitive communication method between biological and artificial agents, based on machines understanding and learning the subtle unspoken and involuntary cues found in human motion during the interaction process. Our work was divided into two stages: the first, analysing whether a machine using these implicit cues would produce the same positive effect as when they are manifested in interpersonal communication; the second, evaluating whether a machine could identify the cues manifested in human motion and learn (through the use of Long-Short Term Memory Networks) to associate them with the appropriate command intended from its user. Promising results were gathered, showing an improved work performance and reduced cognitive load on the user side when relying on the proposed method, hinting to the potential of more intuitive, human to human inspired, communication methods in human–machine interaction.

Keywords Human-robot interaction, Human–machine cooperation, Machine learning, Implicit control

Introduction

With the progress of robotics, the gap between humans and intelligent machines is rapidly shrinking. Individuals are becoming used to these artificial agents being increasingly present in their household or workplace environments, and interacting with them on a daily basis. Furthermore, the boom of research in the field of "Robot Autonomy", underlined the enthusiasm for autonomous robots [1, 2] capable of task completion with the same dexterity level as a human in a wide variety of fields such as rescue, medical, space exploration, marine research...

[3, 4]. With robots becoming more and more intelligent, capable of a higher level of understanding, a need has been created for a shift in communication methods and the way artificial agents and biological agents interact.

Several studies have tried defining Human–Machine Cooperation. For example, in their paper, Flemish et al. explain it as the balance between human and automation, and extensively studied the specifics and differences in various instances of cooperation and shared control [5, 6]. Similarly, Gervasi et al. define the collaboration between a human and a robot as form of direct interaction aiming at combining the skills of both parties to perform a task, and presented a framework to evaluate the collaboration while taking into account all aspects of the interaction [7]. Finally, Music et al. see Human-Robot Cooperation as a way to combine the complementary capabilities of humans, such as reasoning and planning in unstructured environments, with those of robots,

*Correspondence:
Lena Guinot
guinotl@moegi.waseda.jp
[1] Department of Modern Mechanical Engineering, Waseda University, Tokyo, Japan
[2] Faculty of Science and Engineering, Waseda University, Tokyo, Japan

Guinot *et al. ROBOMECH Journal* (2023) 10:12

Page 2 of 21

which include performing tasks repetitively and with high degree of precision. They also study the question of optimal control sharing methodologies design to fit both participating parties [8].

In the field of service robotics, recent studies have explored ways to reinvent the way robots and individuals cooperate on task execution within a shared workspace. One of the proposed concepts is the idea of an "Augmented Human", an concept related to the idea of human expansion. Professor Jun Remikoto of the University of Tokyo defines human expansion as "a technology that freely enhances and expands human capabilities through technology". Some of the more famous examples include Massachusetts Institute of Technology's Supernumerary Robotic Limbs (SRL) [9], and Tokyo University's "MetaLimbs" [10] and "Fusion" [11]. While traditional methods such as direct teaching or action coding [12] may still work for most of the current applications of robotics, they are highly context-dependant with a high task specific dependency, and hence become very limited when used in uncontrolled environments, common in the field of service robotics. This implies that the user has to focus his/her attention and energy on controlling or supervising the actions executed by the robot instead of collaborating with. A collaboraton between both agents would rather refer to each independently working on different elements for the comlpetion of a common task.

In the present project, we analysed the potential of an "implicit" communication method between a robot and its user during cooperative work. Similarly to the Clever Hans Effect [13], the robot becomes capable of understanding the action expected of it by responding to the involuntary and unconscious cuing, translated into postural adjustments, that occurs in human communication.

Throughout this paper we refer to the use of such ideomotor reactions ("implicit commands") to communicate with a robot as "implicit communication". First, in a simulated environment, we verified the validity of this theory and whether these cues would be perceptible to an individual when used by an artificial agent. Then, we verified the ability of the robot to learn and differentiate these cues, along with the usability of the method in an uncontrolled environment.

Concept and related works
Ideomotor phenomenon in popular culture
"Implicit" means that the robot understands the intended instruction based on imperceptible cues embedded in the natural motion of the user. The present study was inspired by the Japanese concept of "Aun breathing". In modern Japanese culture, the term if often used to describe a perfect synchronization between two individuals without relying on verbal communication to achieve it. In a previous study on the concept of "Aun", Ueda et al. [14] focused on the coordination of the puppeteers during Bunraku performances. Bunraku is a form of Japanese traditional puppet theatre requiring three puppeteers to work together on the operation of a single puppet. Findings showed that the main puppeteer relied on implicit signals, in this paper referred to as "Zu", to communicate with the two others and for the group to stay synchronised, signals that were perceptible only to the three puppeteers. In another study, Shibuya et al. [15], found that performers with extensive stage experience (performing together for 31 years) used such signals to unconsciously synchronise their breathing when starting a performance, something groups with fewer experience (13 years) would not do. Based on these gathered observation, it could be said that the Japanese concept of "Aun" falls under the umbrella of the ideomotor principle.

In the Western culture, one of the most famous studies that revealed the existence of these unconscious cues translating human intention is that of the German psychologist Oskar Pfungst on Wilhelm von Osten's horse, Hans. Pfunst was recruited by the German board of education commission to investigate whether math teacher von Osten's horse possessed the intelligence to perform arithmetic calculations as claimed by its owner. After extensive trials and observations, Pfungst made two discoveries:

- The horse would only get the answer right if the person doing the asking knew the answer. More importantly, if the person was mistaken, the horse would make the same error
- The horse was only capable of answering when the questioner was visible

According to findings reported in his book [13], Hans' behaviour was directly linked to the subtle and unintentional cues it would pick up on when observing the questioner. As the horse approached the answer, the questioner would unconsciously, ever so slightly adjust his posture or change his facial expression. The horse had simply learnt to understand these cues as an instruction to stop tapping its hoof, which was its way of giving the answer to the mathematical problem.

The subtle motion cues in this section have since then been given the name of "ideomotor effect" and are well-studied in the fields of psychology and neurosciences [16–18]. It is now widely admitted that they have communicative value to an observer. Allowing machines to learn from them would therefore appear to be a natural idea to pursue in Human–Machine Interaction (HMI).

Guinot *et al. ROBOMECH Journal*   (2023) 10:12

Page 3 of 21

Related works on user motion and intention estimation

There has naturally been extensive research done on the use of body language to operate intelligent machines in the field of Human-Machine Interaction. However, many of these studies chose to map different specific motions or non-verbal cues to a command. The user then performs the specified movement for the machine to execute the corresponding action. For example, Faria et al. [19] designed an interface based on user facial expression, with different expressions mapped to control signals used to operate a wheelchair. This means that the user had to memorise the mappings defined by the person who designed the interface and to translate his/her intention into the matching expression. Several other studies have developed similar approaches to Human–Machine communication [20–23]. As a result, they all rely on the mapping of body language to specific commands and require the user to memorise and translate his/her intentions to effectively control the system.

It has also been established that intention perception as a control method is one that feels most natural as it closely relates to interpersonal communication. This control method enables the artificial agent to behave according to the perceived intention of the user. Most studies that used the intention perception approach to Human–Machine Interaction are based on the extraction of biological signals, such as electromyography (EMG) and electroencephalography (EEG) and the analysis of their relationship to human psychology and intention manifestation [24, 25].

The main difference between the mentioned previous studies and the work presented in this paper is the focus put on having a "two-way dialogue" between the robot and the user. Indeed, in this study, we analyse the capacity of the robot to not just understand user motion cues or intention but also to provide information to the user through the use of human-like behaviour.

Analysis of individual behaviour

Task setting

Requirements

For the first part of the analysis, data collection, Japanese rice cake making was selected as a collaborative activity satisfying the following two elements:

- Two individuals working together
- Each individual's actions depend on the response of the other

Japanese rice cake, also called "mochi", is made by having one person repeatedly pounding the dough once, while the other quickly kneads in between hits, flipping over the dough every so often to ensure an homogeneous texture. Because of the texture of the rice cake the turning over of the rice cake takes substantially more time than kneading (the activity can be seen performed in the first few seconds of [26]). Not only does it requires perfect synchronisation for optimal work and safety, but the dangerous pound-knead, pound-turn rhythm of the two participants requires deep trust in one another. In case of an eventual change in the pace of one of the two parties, the other needs to be able to anticipate and adapt his/her own behaviour accordingly.

Data collection set-up

Using the Unity game engine, we built a method for measuring the potential implicit cues used during Japanese rice cake (mochi) making. Figure 1 is a representation of the designed environment. Movement of the pestle was measured and recorded by attaching VIVE trackers, a motion tracking accessory, calculating its position based on infrared signals emitted by virtual reality base stations [27], to its handle and head. VIVE trackers were also attached to each glove worn by the person kneading the dough. The mortar and dough were respectively represented by a stool with a height of 0.5 m and a seat of 0.4 m in diameter, and a disk shaped sponge with a radius of 0.3 m. Participants were asked to position themselves on opposite sides of the stool, facing each other.

Virtual reality environment

The second portion of this analysis was done in the virtual reality (VR) environment shown in Fig. 2. The system requires a computer, a Head Mounted Display (HMD), and a pestle with VIVE trackers attached to it (identically to the ones mentioned in previous paragraphs). This time, the kneading of the dough is performed by a robot arm in the virtual reality environment (Fig. 2). Here, the participant works in tandem with the robot arm on the making of the rice cake. The goal was to determine if identified implicit cues could be used for human–machine instance.

Experiment utline

Using the environment described in subsection "Data collection", collaboration between two individuals was analysed to detect eventual implicit cues or ideomotor reactions essential for the individuals to stay in sync during the execution of the task.

The experiment consisted in two participants, one pounding the "rice cake" and the other performing the kneading and turning over of the "dough". In the early stage of the experiment, an auditory signal was used to help participants know when or how often to "turn over the dough"
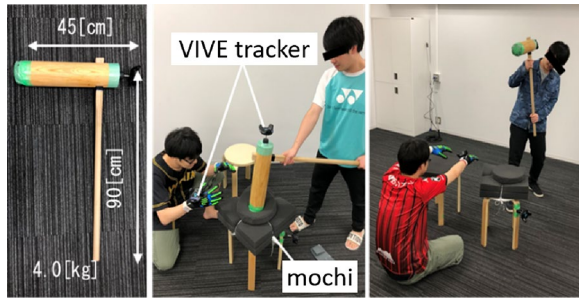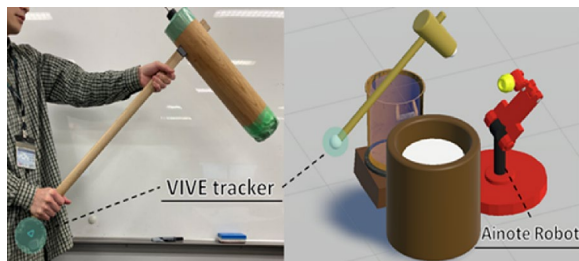
Fig. 1 Real environment system overview

Fig. 2 Virtual reality environment system overview

Table 1 Qualitative evaluation questionnaire items

Q1	I was able to anticipate the behavior of the other party
Q2	I was able to predict the action required for the situation
Q3	I was able to modify my own behavior according the other party

(first after 13 s, a second time after 26 s). No other hints were provided to assist participants in finding their pace.

Participants, 20 in total (Male/ Female: 13/7, age 22 to 35), were divided into two groups. Participants put into the first group were in charge of the pestling, participants in the second group were in charge of the kneading. Each participant from the first group performed the task 6 times for 60 s with each of the members from the second group (between-subjects study design). Evaluation was done both qualitatively and quantitatively. The former was done using a Likert scale based questionnaire with questions shown in Table 1.

For the quantitative evaluation, we focused on the relative distance between the hands of the individual kneading and the pestle using the index in Eq. 1. Here x_a represents the position of the hand and x_k is the coordinate of the pestle.

$$\Delta x = \sqrt{(x_a - x_k)^2} \qquad (\Delta x > 0) \qquad (1)$$

As shown in Fig. 3, the center of the mortar is set as the reference point ($x = 0$) with the pestle and hands moving
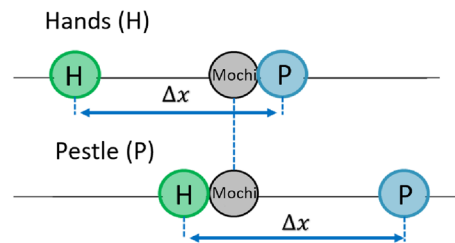

Fig. 3 Relative Distance between the Hands and the Pestle during task. Top: Pestling phase. Bottom: Kneading phase

with respect to that point and generating their respective amplitude (x_a and x_k). As mentioned, the action of turning over the dough takes more time than simply kneading. At this time, the value of Δx remains constant for an "extended" period of time, despite the hands performing various motions during that time period.

From Eq. 1 it can be said that the less Δx varies from one motion cycle (kneading-pounding) to the next, the more it is an indicator that the two individuals are synchronised and are capable of adjusting to motion changes of the other party. Therefore, this Δx index was chosen to quantitatively evaluate the smoothness and synchrony of the interaction. Since the average value of the relative distance differs from person to person, comparison between instances of the experiment was done using the "coefficient of variation".

$$C.V. = \frac{\sigma}{\Delta \bar{x}} \qquad (2)$$

C.V. is equal to the standard deviation σ divided by the arithmetic mean $\Delta \bar{x}$.

Results and observations

Qualitative evaluation results

For each collaboration pair, answers to the survey questions in Table 1, from both individuals were collected and response averages are displayed in Fig. 4. As can be seen, the answers are divided into two charts. The chart on the left represents average scores recorded for pairs that were able to complete the experiment without relying on any form of explicit communication throughout the task. On the other hand, the chart on the right corresponds to the scores reported by pairs that had to rely on explicit (vocal) communications to complete the task (wait, slow down...) In this paper, we refer to the individuals that could complete the task without any form of explicit communication as the "implicit pairs" or "implicit collaboration" and those that had to rely on such explicit communication as the "explicit pairs/collaboration". From Fig. 4 it seems the implicit collaboration scenarios received more positive responses from both parties than
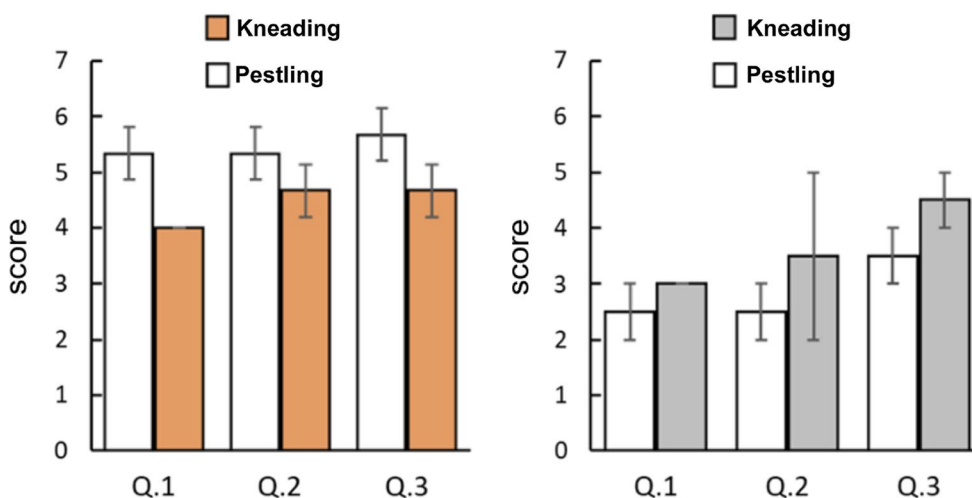
Fig. 4 Qualitative evaluation questionnaire results. Left: implicit pairs. Right: explicit pairs

the explicit collaboration. Focusing on the individuals in in charge of the pestling, who participated in collaborative work with all kneaders, responses when cooperating without explicit indication were much higher than when having to rely on explicit communication.

Regarding the quantitative comparison of the quality of the synchronization between individuals during the two collaborations, Fig. 7 shows the different averaged coefficient of variation (C.V.) of each pair. From this, we gathered that explicit pairs clearly had more difficulty keeping a stable rhythm, and therefore an harmonious collaboration throughout the experiment. This observation underlined the extent to which the communication elements used in implicit pairs' collaborative work were positively impacting work quality.

To verify our hypothesis of implicit communication through unconscious cueing, we chose to analyse and compare participant motion data at points where the smoothness of the collaboration was most likely to be disturbed (and therefore mutual understanding between participants seemed most crucial).

Synchrony of cooperation

As mentioned, the interval during which the rice cake is turned over is the main element that disrupts the established punch-knead, punch-knead of the collaboration (the pace established up to that point). Attention was therefore focused on methods used by participants to proceed with this action as smoothly as possible, with minimal impact on the overall task rhythm. To observe the periodic change caused by the turning of the rice

cake, the overall motion of the person performing the kneading was divided into three phases:

- T_{touch} (T_t): period during which the person kneads
- $T_{before\,reverse}$ (T_{br}): the kneading-pounding cycle preceding the turning over of the dough
- $T_{reverse}$ (T_r): the period during which the person turns over the dough

When paying close attention to the cycles of each participant, a major difference was noticed. While for some participants, the action cycle remained constant, increasing only for the turning over, for others, the kneading cycle preceding the turning over (T_{br}) was slightly shorter.

Figure 6a and 6b show the comparison between the group averages of the time length of T_r, T_{br} and T_t for implicit and explicit pairs. As can be seen, depending on the technique used by the kneading participant, the rhythm of the person pounding the rice cake was affected.

Additionally, the difference between the two collaboration styles was further emphasized when calculating the ratio E (Entrainment Rate) between the two variables T_{br} and T_t.

$$E = \frac{T_{br}}{T_t} \qquad (3)$$

Using the average of the mesured T_{br} and T_t of the kneaders, results of this ratio were as follows:

$$E = \frac{T_{br}}{T_t} = \begin{cases} 0.85 & (\textit{implicit communication pairs}) \\ 0.96 & (\textit{explicit communication pairs}) \end{cases}$$
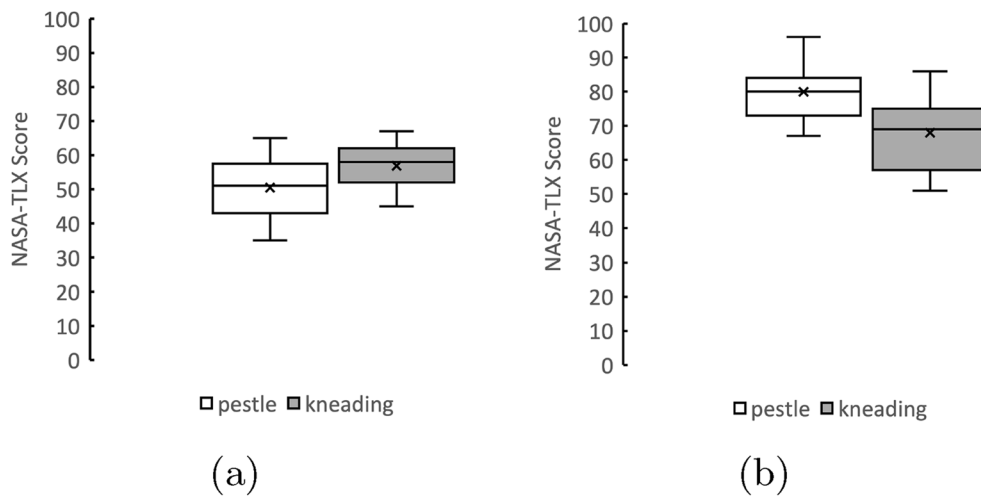
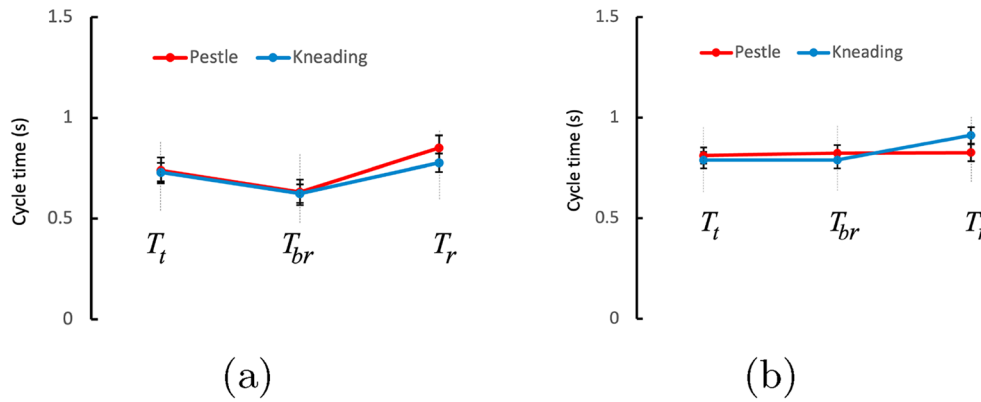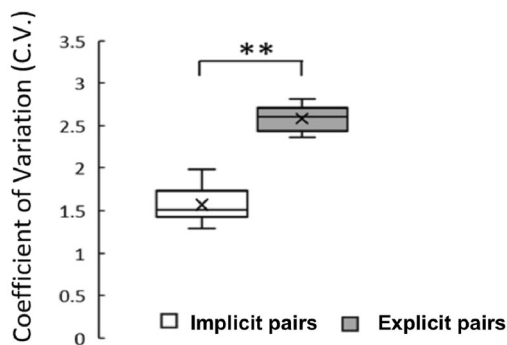**Fig. 5** Reported NASA-TLX Scores for **a** implicit pairs and **b** explicit pairs

Fig. 6 T_t - T_{br} - T_r Evolution for **a** implicit pair and **b** explicit pair

(Wilcoxon signed-rank test, N=6, *: p <0.05, **:p<0.01, N.S: not significant)

Fig. 7 Coefficient of Variation of the Relative Distance Δx

According to Fig. 6a, the kneading and pestling rhythms of implicit pairs seem to adapt to each other, both slowing down when reaching the "turning over" cycle. On the other hand, regarding the explicit pairs, (Fig. 6b), despite the increase in time taken for the turning over of the dough compared to the kneading, the rhythm of the pestling remained constant throughout the execution of the task. It seemed that the main reason why some pairs had to rely on vocal communication was that the person in charge of pestling failed to follow or adapt to the pace variations of the kneading-turning cycles. On the other hand, when the kneading participants relied on implicit signals and unconsciously increased their pace before turning over the dough, the subtle difference in the kneading pace was, just as unconsciously, noticed by

Guinot *et al. ROBOMECH Journal*   (2023) 10:12

Page 7 of 21

the individual pestling, and enough for him/her to understand the meaning of said signal.

Discussion

The authors believe that the increase in kneading pace observed in the behaviour of individuals, in the cycle preceding the turning over of the cake was directly correlated to the awareness of the participant that the following action would require more time. The participant therefore unconsciously used this as a signal to the individual pounding the cake to slow down his/her pace during the next cycle.

Another major difference noticed between participants was their answer when asked about their experience during the experiment. After the execution of the task, for each collaboration instance, the person in charge of kneading was asked several questions to analyse how much effort had been needed to appropriately match the rhythm of the pestle. When asked how they proceeded, individuals that had completed the task without explicit communication, all answered that they focused on watching the movement of the pestle to determine the appropriate time to knead the dough. On the contrary, individuals who had had to rely on explicit communication, were more likely to answer that they focused on looking at the rice cake and kneading as quickly as they could after the pestling. This revealed that these "cues" were only manifested when the person kneading was paying more attention to optimizing the quality of the collaboration rather than focusing on his/her own task alone. When looking at Fig. 4, it can be noted that, on average, the responses provided by the pestling group in the implicit cooperation returned more positive results than the kneading group in the same cooperation category. On the contrary, in the explicit cooperation responses from the kneading group, were more positive than that of the pestling group. We believe this observation to be due to the kneader being the primary instigator of the signaling while the pestling side "responds" to these signals. Hence the kneading side has to wait for this "response" to know that the signaling was correctly received and understood by the pestling side leaving room for uncertainty. The consistency in timing and expression of the cues greatly reduces the uncertainty that come with sudden cycle timing changes for the individual doing the pestling. On the other hand, in the explicit cooperation, the pestling side's behaviour is highly dependant on the timing of the vocalization of indications from the kneader. These observation seem to coincide with the reported felt cognitive demand of the task, on the NASA-TLX survey conducted together with the questionnaire [36]. As can be seen from Fig. 5, while the average score was lower for

both kneading and pestle sides in the implicit collaboration pairs, the kneading side reported a higher cognitive load compared to the pestle side. On the contrary, in explicit collaborations, people in charge of pestling would have answers on the survey reflecting a higher cognitive demand than the kneaders.

The aforementioned results and observations, suggested that the increased kneading pace during T_{br} served as "preliminary indication". It seemed that this "preliminary indication" was used as a way to implicitly communicate with the other party despite changes in the context. The absence of such communication method during the explicit collaboration explains the failure of the pestling person to adapt to sudden changes in the rhythm of the kneading person's motion.

Virtual reality human-robot collaboration
Experiment outline

Using the VR environment to analyse the validity of the previously revealed "preliminary indication" and applicability to Human-Robot Cooperation, we asked participants to take part in a second rice cake making simulation. In this scenario, the participant was in charge of pestling the rice cake while the robot in the simulated environment performed the kneading and turning over of the rice cake (Fig. 2).

During the early phase, the robot was meant to adapt to the motion of the pestle until the participant had found a comfortable rhythm. Once the robot and participant had achieved synchronized work on a simple pounding-kneading cycle, the experiment was divided into two cases. One in which the robot used the "preliminary indication" method to communicate the timing of the turning over of the rice cake, the other without it. Participants performed each experiment trial (with and without preliminary motion) for 45 s with a 3 min break in between trials for fatigue management. For this section, the participant pool consisted of 6 adult males, age 22 to 24. Future work includes broadening the participant pool for better reliability of the results. The robot was programmed to receive the information of the VIVE sensors positioned on the pestle and perform its part of the task with appropriate timing in the VR environment. As initial setting, participants were asked to perform the pounding potion once. Doing so allowed for the recordint of Δ_{xmax} the maximum distance between the "dough" and the pestle. The robot was then set to adapt its working speed to maintain this Δ_x distance.

Table 2 Human–Machine collaboration qualitative evaluation survey

Q1	I was able to predict the action required for the situation
Q2	I was able to modify my behaviour from that of the other party

Guinot *et al. ROBOMECH Journal*   (2023) 10:12

Page 8 of 21

The probability of each cycle being one in which the robot would "turn over the dough" was randomised with a probability of p = 0.5 Once again, qualitative evaluation of the experiment was performed using a 7-grade Likert scale survey. Survey questions are shown in Table 2.

Results

Answers collected from the survey are displayed in Fig. 7a. As can be seen in Fig. 7a, for both questions, the experiment scenario where the robot used implicit signal received much better ratings than when the robot did not rely on it.

Figure 7b shows the measured difference in the coefficient of variation of the relative distance Δx both with and without the use of preliminary indication. The Wilcoxon signed rank test was used to evaluate the difference between the result pair of each participant. As shown in Fig. 7b a significant difference ($p < 0.05$) was found in

the coefficient of variation between the two experiment variants. Results suggested that the use of the preliminary indication facilitated the synchronization between the kneading and pounding and therefore allowed for a more stable Δx with less variation from one motion cycle to the next.

Work performance

Work performance was first (Fig. 8) evaluated by observing the number of times the rice cake making process was completed within the imparted 45 s. This was done by recording the number of times the rice cake was pestled by the participant. From Fig. 9b, it can be seen that using the implicit indication resulted in a higher performance, with the participant pounding the rice cake, on average, an additional 7 times. This improvement was apparently due the quicker response/reaction time from
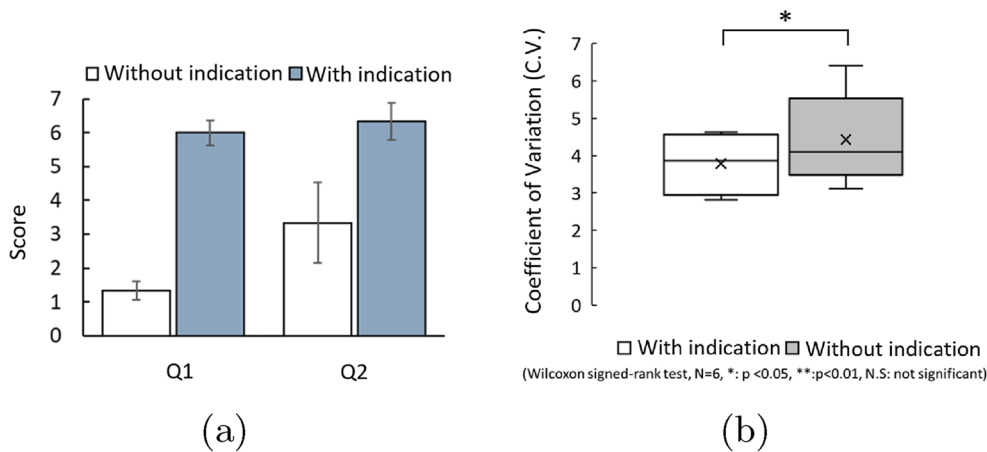

Fig. 8 Qualitative and quantitative evaluation results. **a** Survey answer scores. **b** Coefficient of Variation with/without the use of Preliminary Indication
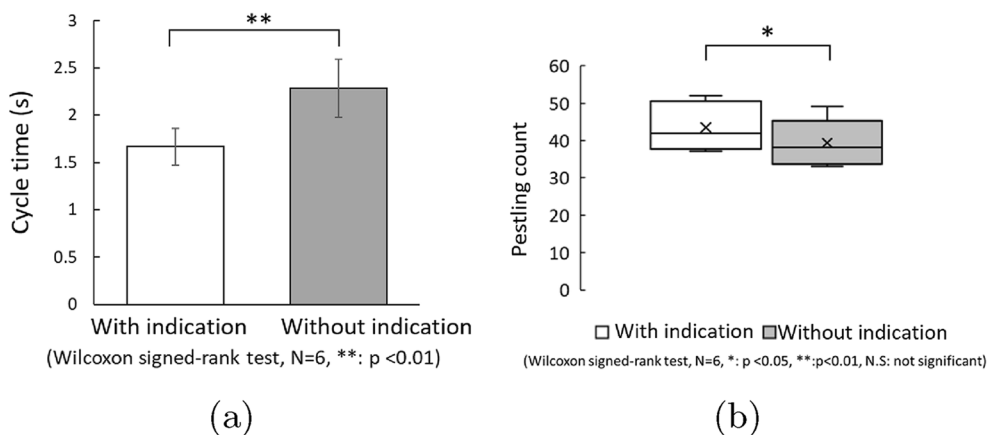

Fig. 9 Performance evaluation results **a** Cycle time comparison when asked to match a specific pace of 1.666 s . **b** Difference in number of time the dough was hit within a work instance

the participant. With the use of implicit cues, not only was the participant able to anticipate potential changes in the work rhythm, but also became more confident and less fearful of any unexpected behaviour. In situations where the robot did not use the implicit cues, the user became more uncertain and slowed his/her pace down as a cautionary measure (to compensate for any unexpected movement or behaviour from the robot).

Since arbitration and the idea of shared control over a collaborative task is a central issue to human–machine interaction [28, 29], attention was also paid to the effect of the presence or absence of implicit cues when the control authority switched from user to robot during the task. As reflected in Fig. 9a, participants were also able to correctly match a sudden work speed increase demand by the robot (in this example 1.66 s between two punches) when it used motion cues. On the contrary, when these cues were not used, participants tended to be surprised by the sudden behavioural change of the robot (change in working pace) and would instead slow down their own rhythm as preemtive measure (again, in case of any other unexpected changes).

In the future, it would be interesting to pay closer attention to the division of control between the human and the robot in order to ensure that the artificial agent, in this type of task, is capable of adapting to demands of the user just as much as the user can adapt to the robot.

Discussion

Although the designed system did show evidence of improved performance and human–machine cooperation quality, two main issues were identified.

1. The participant sometimes kept swinging at a constant rhythm, not slowing down for the "turning over of the cake" phase
2. The participant would realize that he had moved with excessive speed and would completely stop his motion until the end of the "turning over of the cake" phase

Issue (1) seemed to be due to the user getting too used to swinging the pestle according to a certain rhythm and forgetting to alter it despite the signaling of the implicit indication. On the other hand, Issue (2) seems to be due to the user not being used to the time taken by the robot to turn over the rice cake and therefore not knowing how to time his own motion accordingly. It seemed that future experiments would require an element that gradually guides the pestle rhythm over the "turning over of the rice cake", during the earlier stage of the experiment. It would also be wise to consider extending the length of the experiment to account for this "adaptation" period.

Implicit interface design

The first part of this study focused on verifying the manifestation of these implicit cues during collaborative work and analysing their effect on the quality of work and
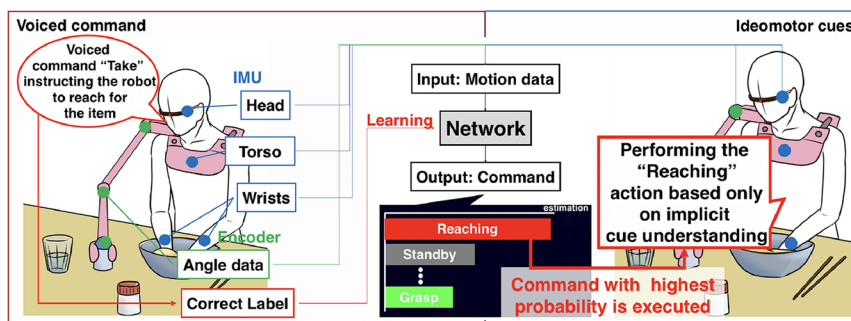
(a)
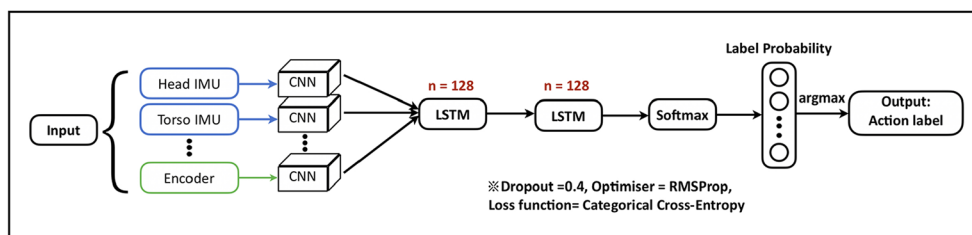
(b)

Fig. 10 Concept and Network Representation. **a** Overall flow with use case example. **b** Detail of the network architecture

Guinot *et al. ROBOMECH Journal*　　(2023) 10:12

Page 10 of 21

performance in a collaboration instance between an artificial and a biological agent, especially when used by the former. During the second half, the authors focused on determining whether an artificial agent would be capable of identifying these cues and associate them with the appropriate command (based on the user's intention).

Learning of cues

As mentioned, the goal was to have the robot autonomously learn the implicit cues in the same way Hans the horse does in the Clever Hans theory. Figure 10a shows an overview of the flow of proposed method. For example, the operator first turns his/her face to the object to be picked-up by the robot arm. The three-dimensional coordinates of the target point are measured, and the instruction is transmitted to the robot arm by voicing the command (e.g. "get!"). The robot arm then moves as instructed and grasps the target object. Throughout this study, using voiced commands to control the robot is also referred to as using "explicit instructions". This first phase using explicit instructions was used to gather user motion data and corresponding labels, to be used in the training of the neural network. Once the system has learnt to recognise the motion cues and to correctly estimate user intention, robot operation is done using the flow represented in the right, blue area of Fig. 10a, using exclusively these implicit cues. As the user turns his/her face towards the target object, the system recognises the motion cues corresponding to the instruction ("take!") and moves accordingly. In other words, it becomes possible to control the robot arm exclusively using implicit naturally occurring motion cues, identified as relevant and labelled during the initial training process.

To have a sustainable model for long term Human–Machine Collaboration, two conditions were considered as essential:

- Task and environment independence
- Motion data gathered using fewest possible number of sensors

To satisfy the first requirement, no task-related information was provided to the system. In addition, we avoided the use of any kind of image/visual data as input data to the system, to ensure minimal context dependence. Regarding the second requirement, to prevent the motions of the user from being restricted or obstructed by heavy data collection equipment, a minimally invasive sensing system was used. Therefore, the placement of Inertial Measurement Unit (IMU) sensors was limited to strategic locations. The following four locations were

used: head (eyeglasses), torso and wrists. The IMU sensors used were Bluetooth 9-axis inertial sensor TSND151 [30], with 3 axes of acceleration, 3 axes of angular velocity, and 4 axes of posture (quaternion), for a total of 10 dimensions.

Structure of model

A simple network (Fig. 10a) was designed for the robot to learn the implicit cues and associate them with intended commands. As can be seen on Fig. 10b, the network is composed of concatenated Convolutionnal Neural Networks (CNN, 5 layers) followed by two layers of Long Short-Term Memory (LSTM) Network. The "CuDNNL-STM" in the Keras library was used as an accelerator for the LSTM. The IMUs were used to collect user motion data from 4 different locations (attached to the individual's body). For the robot arm, data input into the network consisted of the angle data on each axis of the arm (0° to 180°, 4 degrees of freedom). Overall data format before formatting operations was of 44 dimension. Sensor data normalization was done with equation 4 with resulting values between 1 and -1

$$Y = 2\frac{X - x_{min}}{x_{max} - x_{min}} - 1 \tag{4}$$

With X the original sensor data, Y the same data after normalization, x_{min} and x_{max} as the maximum and minimum values recorded for the sensor over the period of time. Collected training data was divided by time steps and shaped into three dimensional input. During collection of the training data, matching label for each movement of the users was collected by having them vocally express which action they wished for the robot arm to perform at that instant (in the designed experiment, 5 options were avalable: reach, grasp, release, return, wipe).

Validation on human-robot collaboration task
Experiment setup and method
Setup and requirements

To verify the capacity of the robot to learn and identify the implicit cues in real time, a final experiment was conducted. For the present study experiments and data collection were performed using a static robot, as shown in Fig. 11. For this part of the study, the experiment was conducted over the span of 3 days (each participant had to participate for three consecutive days, Fig. 13). On the first day, participants performed tasks together with the robot arm by explicitly expressing commands (voicing them), while wearing the IMUs at the locations indicated on Fig. 11. The second day, participants performed the
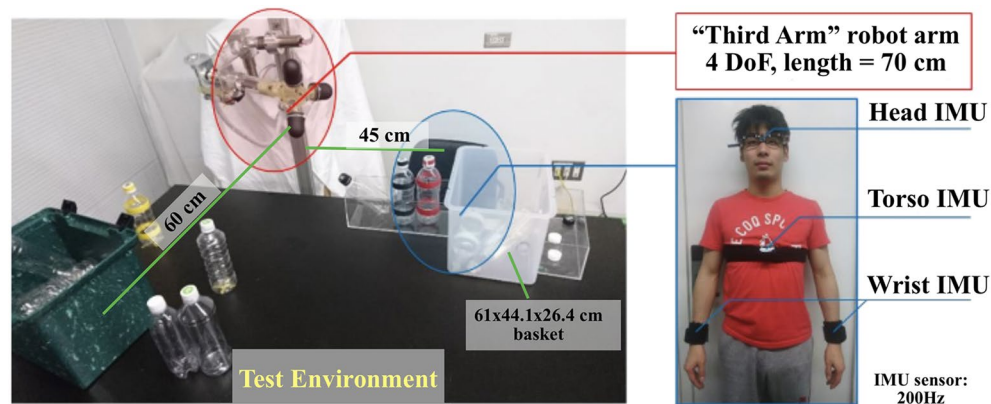
Guinot *et al. ROBOMECH Journal*   (2023) 10:12

Page 11 of 21

Fig. 11 Experiment setup

same tasks as the previous day, only this time, whether the collaboration would be done using voiced (explicit) commands or only the user motion data (implicit cues) was randomly determined. On the third day, whichever operation method (implicit or explicit) had not been used the previous day was used for the collaboration. Each day, the participants performed a task for 10 min, since their were a total of four task, the experiment lasted 40 min per person. Data collected from the experiments was used as training data for the system at the end of each day.

Task setting

The tasks used for this phase of the experiment were designed to mimic daily chores that an individual may perform, while remaining relatively simple as to keep the number of commands and labels relatively low (4 different labels and a standby phase). 3 Tasks (Fig. 14, Task 1, Task 2, Task 3) were "periodic" with a constant label order, one task (Task 4) was "aperiodic" with the labels order randomly changing. The mochi making task was chosen for its repetitiveness and extremely simple mechanisms. Such a task made it easier to isolate the potential implicit cues and analyse how and when they surfaced by having several instances of the almost exact same scenario and very limited room for variation in how the participants interact with the environment. In this section however, the goal was to have a system capable of identifying the relevant cues in the appropriate situation. Indeed, despite the myriad of unconscious body language cues that we produce when interacting with our environment, we argue that some of these are consistent enough for a system to recognise them and behave appropriately. It was therefore necessary to have a more uncontrolled task environment

with a higher possibility of behavioural discrepancy between participants. Details of the tasks are as follows:

- Task 1 - Wiping Task The user lifts a basket of dimensions 61 cm x 44.1 cm x 26.4 cm (length x width x height) from the desk while pointing the head IMU towards the area to wipe. The robot arm (already holding a piece of cloth) is expected to move and start wiping the instructed area (horizontal back and forth motion over a 30 cm distance). As the person starts lowering the basket back on the table, the robot arm is retracts.

- Task 2 - Pick and Place Task The task consists in the robot arm grabbing 500 ml empty water bottles being handed by the user and placing them in a container box on the desk, out of reach to the user. While the robot is placing the bottle into the container, the user prepares the next one. The bottles are initially uncapped. For each bottle, the user has to fasten the cap on the bottle before handing it to the robot arm.

- Task 3 Pick and Place with Wiping Task It is a compound task of Task 1 and Task 2. Before preforming the wiping as in Task 1, the robot arm has to grab the cloth being handed by the user. Similarly, once the wiping action is over, the robot arm has to place the cloth into the container (same set-up as Task 2). The user first hands the cloth to the robot, then lifts the basket from the desk, and lowers it back onto the desk to end the wiping action. He then prepares the next cloth.

- Task 4 - Unknown Task The user is free to choose to perform any combination of Task 1, 2 and 3 in any order he wishes. The robot arm has no way of knowing which task he will be asked to perform next.
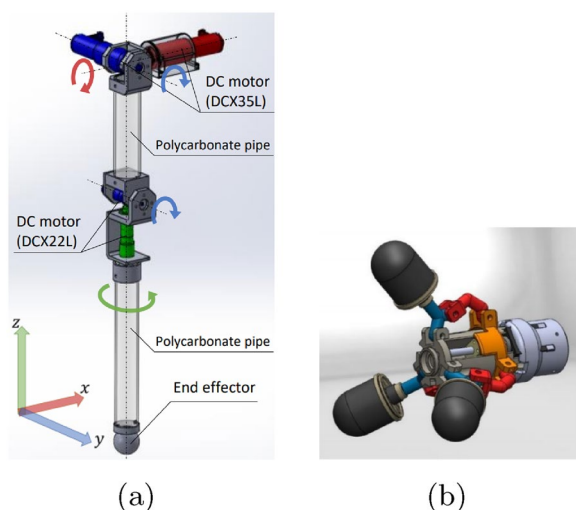
Guinot *et al. ROBOMECH Journal* (2023) 10:12

Page 12 of 21

Fig. 12 Robot Arm used for experiments. **a** Arm details and specificities. **b** End effector used in experiments

All experiments for this part of the study were conducted using the robot arm shown in Fig. 12 called "Third Arm" [31, 32] with 12b as end effector. Participant pool consisted in 15 people (Male/Female: 9/6) with ages ranging from 20 to 25 years old. Future work includes further trials with a more diversified participation pool.

Estimation results

Estimation accuracy of the model (a single model was used for the four different tasks) is displayed on Fig. 15. As expected, highest estimation accuracy was obtained on the task with the fewer labels. As the number of labels increased, accuracy decreased. Although Tasks 2 and 4 have the same number of labels (4), as explained above, the lower F1 score on the last task is due to its aperiodic nature.

When paying closer attention to the results, particularly confusion matrices of each task, it was noted that most of the errors in Tasks 1, 2 and 3 were due to the implicit cues being labelled as 0, the label for the "standby". This meant that the primary problem during the collaboration was that the robot arm would sometimes fail to detect the implicit cues. If the cues were detected, however, they were always correctly matched to the appropriate command and therefore followed by the robot behaving according to the user's intention. Despite the occurrence of errors, the conducted experiment showed promising results regarding the ability of the system to recognise the implicit cues regardless of the task being executed (no

prior information about the task or the context was input to the neural network). Indeed, results showed evidence that the robot was capable of recognising the implicit cues embedded in the user's motion enough to understand intended commands. This would point towards the idea that the implicit ideomotor cues (or "zu") referred to in the "Clever Hans Effect" and Japan's "Aun Breath" could be used as a communication method in a Human-Robot Collaboration. Because results of this study are achieved using training data acquired over a very short time period, it is assumed that the system could benefit from additional data and training. Nevertheless, since the aim of the overall study was to analyse the limitations of this communication method, this was also considered as an indicator to how much effort is needed before it becomes useable.

Regarding Task 4, as can be seen from the confusion matrix (Fig. 16d) the lower accuracy was found to be due to the lack of consistency in the order of the labels. Not only does the instability of the context prevent the system from relying on any past information, but the fact that there were no clear instructions made to the participant regarding the order of the commands, may have hindered the quality of the ideomotor reactions (no clear idea of what action to execute next, no clear idea of target point...)

When comparing obtained results on the methods used in the present studies to similar ones, results were encouraging. For example, Hayakawa et al. [33] used a "Self-organizing Map" (unsupervised learning) to estimate the intention of the operator and have the robot assist in the task. Although the method was designed to be used on a single task (assembly), reported accuracy was of 70%, a lower results than obtained in the present paper for a model designed for estimation on four different tasks. In the results, the task with the lowest estimation accuracy had an F1 score of 79%. Furthermore, the accuracy on Task2, inspired by [33] was of 90%. The increased estimation accuracy in the present study is believed to be due to the user motion data collection method. The IMU and higher number of data collection points (4 points: head, both hands and torso), provided a higher level representation with more sensitivity to minor changes in user motion compared to the camera-based method tracking three locations (head and both hands) used in the study by Hayakawa et al. Similarly, on the task with highest accuracy results (Task 1), the designed model showed a 1.5% increase in estimation accuracy compared to the LSTM RNN method presented by Nicolis et al. [34], with their system unable to adapt to changes in goal/target point during estimation.
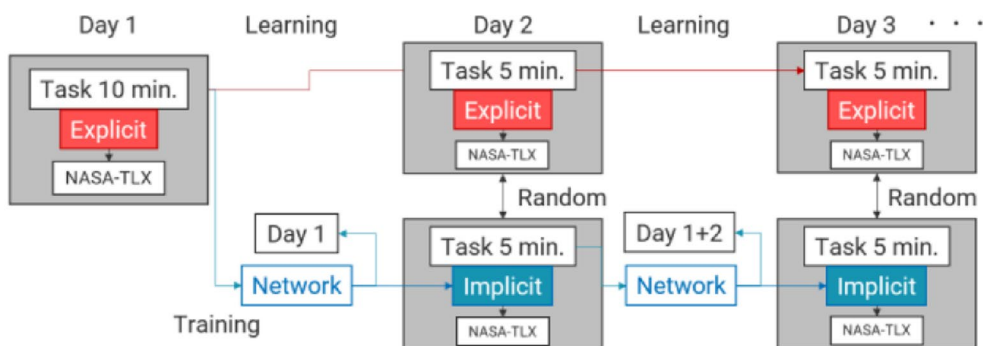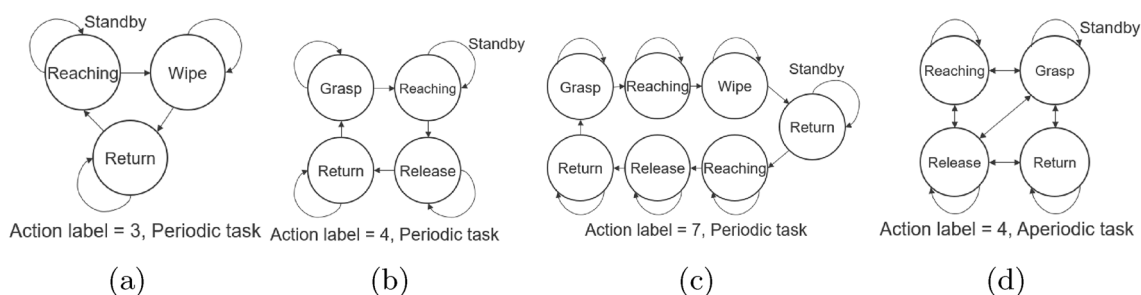
Fig. 13 Experiment flow

Fig. 14 Command pattern for each task. **a** Task 1 - wiping a table with held cloth. **b** Task 2 - Pick and Place. **c** Task 3 - pick cloth up, wipe and put it back. **d** Task 4 - unknown task (decided by the user)
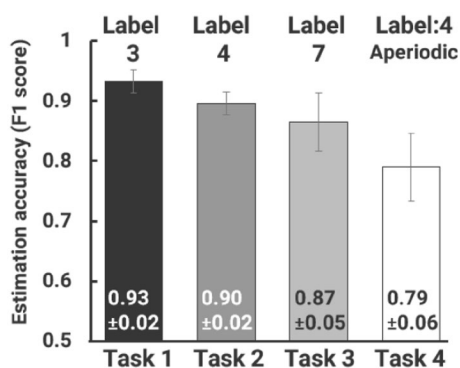
Fig. 15 F1 Score on Task Estimation

Additionally, Nicolis et al. having trained their model on artificially generated trajectory instead of data directly measured from human motion may have impacted overall performance when used in a real world scenario.

Cognitive load

The tasks designed for this part of the study (Task 1 to 4), rely heavily on allocation control and effective allocation of user attention (between his own task and robot control). Allocation control means the task is divided into two subtasks, with the individual in charge of one and the robot or machine in charge of the other ([35]).

Since in our study, the individual is, although implicitly, actually instructing the robot while performing his/ her subtask, we decided to pay attention to the mental burden placed on the user. We performed a qualitative evaluation using the NASA-TLX evaluation method [36] with six different scales: mental demand (intellectual burden), time pressure, physical demand, work performance, effort and frustration. Participants were asked to answer a questionnaire after performing the task by giving a score from 0 to 100 for all of the six load scales (everyday for 3 days as shown in Fig. 13).

Figure 17 shows the comparison of the evolution of the overall workload score (NASA-TLX score) of the NASA-TLX survey for an instruction method based either on implicit cues or voiced instructions from Day 1 to Day 3. Despite a high standard deviation, by the end of Day 3, the overall workload score of the implicit cues-based instruction method had significantly improved. These results suggest that, after a short adaptation period for the user, the designed body language-based method has

Guinot *et al. ROBOMECH Journal*   (2023) 10:12

Page 14 of 21
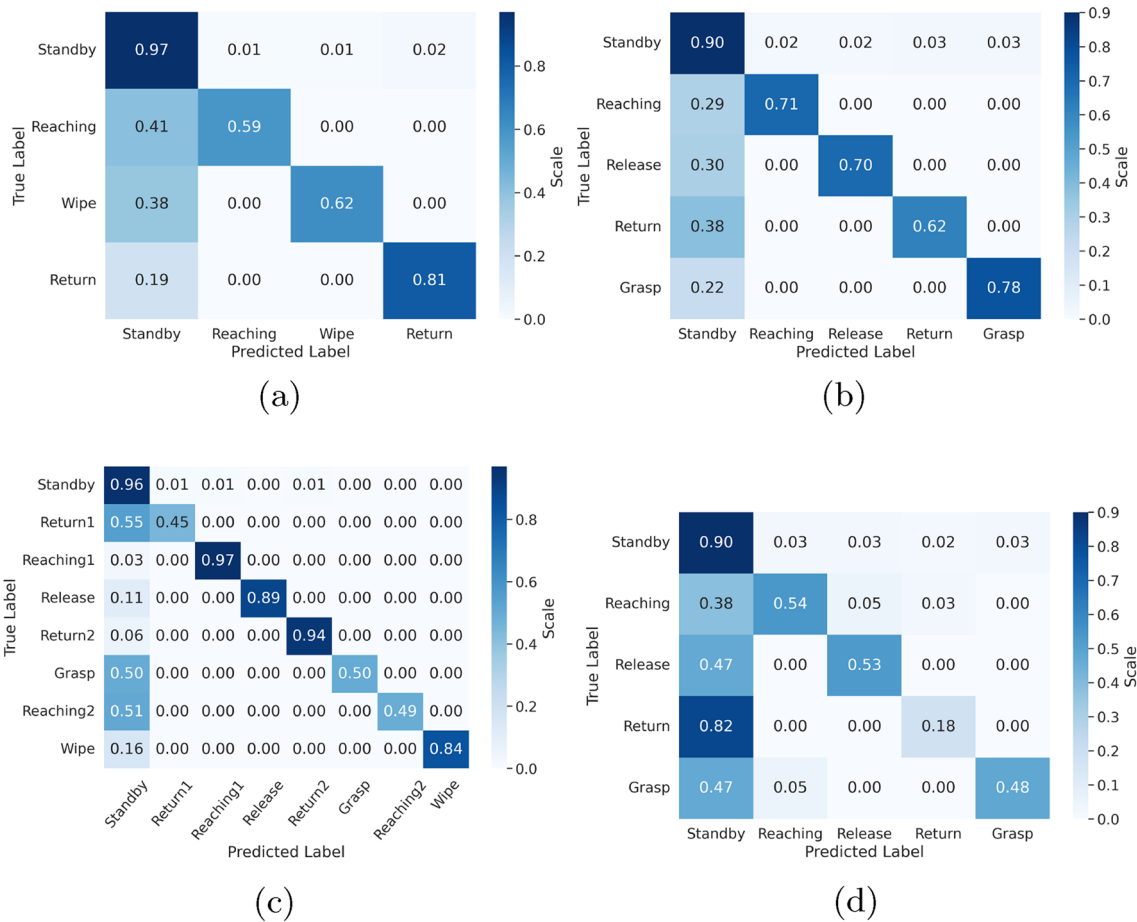
Fig. 16 Confusion matrix for each task. **a** Task 1. **b** Task 2. **c** Task 3. **d** Task 4
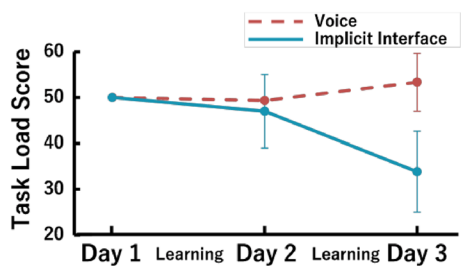
Fig. 17 Cognitive load evaluation results

a lower cognitive burden than the more explicit control method of voicing commands. We believe the main cause of this cognitive load reduction is that when using naturally occurring motion cues to control the robot, users no longer have to memorise all the commands (which can be difficult when they become numerous, such as in Task 3 [37]).

Conclusion

If we hope to continue using Artificial Intelligence to expand human capabilities (both cognitive and physical), we need to design sustainable forms of communications for human–machine interaction. Indeed, the future of Artificial Intelligence highly depends on cooperation between individuals and intelligent machines. The approach to human–machine interaction that appears to be the most viable is one that closely adheres or mimics the principles underlying interpersonal communication.

The present study focused on the analysis of the potential of implicit cues in human–machine cooperation and collaboration, through two scenarios:

- Machine to human information transmission: The goal was to determine if the implicit cues identified during 2-person collaboration instances could be mimicked by a robot and still produce the same effect (nonverbal communication/understanding) during a human–machine collaboration instance.

Guinot *et al. ROBOMECH Journal*   (2023) 10:12

Page 15 of 21

- Human to machine information transmission: The goal was to analyse the ability of an artificial agent to autonomously recognise ideomotor cues as commands (manifestation of the desire of the user for the robot to perform a specific action) and behave accordingly.

Experiments conducted during the first half of this study suggested that if the robot used the same implicit cues as an individual would (unconsciously) in an identical situation, the meaning of these cues were understood by the user. The user would then adapt his/her motion or behaviour accordingly. Quantitative evaluation results not only showed that the use of the cues by the robot would not only allowed for more stable consistent work with reduced variation, it also increased work quality, reducing working speed by 28%, and improved work performance. The second half of the study introduced a body language approach for users to teach their machines using whatever body language cues they produce during interaction. The designed model returns a promising average implicit cues estimation accuracy of 79% across 4 different tasks with an accuracy of up to 93% on individual task estimation. In addition, qualitative evaluation showed a progressive decrease of the cognitive burden, compared to more direct/explicit robot control methods such as speech, with participants reporting a halved cognitive load compared to that of using explicit indications after 3 days of continuous use of the presented system.

In this paper, we studied an approach for detecting intention with application to the robotic domain. So far, it seems this problem has not been sufficiently addressed, despite the ability to infer other individuals' intentions being essential for effective communication and collaboration. We believe it should be an essential component of a robot's cognitive system. The main contribution of this study is two-fold. In the first part of this study we show that, that some of the cues use in human-human interaction, when correctly identified could be copied by a robot used to produce identical results. Hence, we showed that despite differences between human-to-human and human–machine interactions, the implicit cues could be used as a common way to expressing intention, not just

from the "human" side but also from the "machine" side. In the second part, we introduced a prospective body language approach, which allows people to teach artificial agents based on the cues naturally manifested during the process of interaction. Not only is our system capable of recognising cues between users, it was also able to adapt to four different tasks. We can expect that the body language approach presented in this paper can be easily extended to many real-world human–machine interaction scenarios such as self-driving cars or prosthetics.

The present study primarily focused on verifying the validity of the communication of information in a two-way fashion but by treating each information flux independently. During collaboration situations, and in order to further the transition from people communicating through technology to communicating with technology, addressing simultaneous information transmission becomes essential.

Future work includes further investigation of two elements that would allow for a two-way dialogue between the human and the artificial agent. First, in human-human interaction, cues are produced not just to express intention but also as feedback as an expression of acknowledgement of having received the information. The absence of acknowledgement or feedback from the machine can leave the user puzzled as to whether the system is processing their request or stuck. Similarly, work needs to be do for the robot on the receiving end of the feedback to gain an awareness of when the cues it produced were not understood by the user. This issue becomes all the more relevant when the robot and the user share a workspace. Second, to be used during a collaborative task, the robot needs to be capable of listening to the user even when it is already in the process of expressing intention. It is therefore necessary to design a system that can properly address communication overlap with both the user and the robot expressing intention, and understand the hierarchy of these signals so that the robot can adapt its behaviour accordingly.

Appendix 1
See Tables 3 and 4.

Table 3 Average measured T_t, T_{br} and T_r cycles in all implicit collaboration pairs for both the pestle and the kneading

	kneading				pestle			
	T_t (s)	T_{br} (s)	T_r (s)	Standard Deviation	T_t(s)	T_{br} (s)	T_r (s)	Standard Deviation
Pair 1	0.7376	0.6095	0.7814	0.0066	0.7294	0.6025	0.8909	0.0030
Pair 2	0.7524	0.681	0.7932	0.061	0.7520	0.6200	0.7594	0.0062
Pair 3	0.7085	0.5699	0.7679	0.0045	0.7213	0.5638	0.8946	0.0020
Pair 4	0.7353	0.6883	0.7887	0.0020	0.7331	0.6855	0.8504	0.0047
Pair 5	0.7382	0.6158	0.7561	0.0053	0.7500	0.6438	0.7723	0.0048
Pair 6	0.6458	0.6364	0.6859	0.0066	0.6571	0.6480	0.6774	0.0048
Pair 7	0.7891	0.6583	0.8023	0.0035	0.8026	0.6638	0.8601	0.0027
Pair 8	0.7237	0.6008	0.7239	0.0026	0.7120	0.6076	0.7553	0.0067
Pair 9	0.6590	0.6296	0.6611	0.0022	0.6619	0.6237	0.6855	0.0062
Pair 10	0.7224	0.7032	0.8217	0.0026	0.7153	0.6906	0.9001	0.0028
Pair 11	0.5928	0.5744	0.6072	0.0062	0.5997	0.5808	0.7391	0.0068
Pair 12	0.6418	0.6239	0.6617	0.0051	0.6371	0.6180	0.7210	0.0055
Pair 13	0.7249	0.6110	0.7447	0.0054	0.7367	0.6004	0.7652	0.0034
Pair 14	0.6625	0.6370	0.6636	0.0043	0.6758	0.6382	0.7151	0.0045
Pair 15	0.6151	0.5912	0.7066	0.0060	0.6053	0.5564	0.8322	0.0056
Pair 16	0.7715	0.6781	0.7891	0.0032	0.7670	0.6893	1.0685	0.0024
Pair 17	0.7563	0.6171	0.7787	0.0028	0.7600	0.7384	0.8339	0.0015
Pair 18	0.8527	0.6223	0.9136	0.0035	0.8609	0.6212	1.0610	0.0058
Pair 19	0.7816	0.5130	0.8963	0.0023	0.7924	0.5057	1.0160	0.0044
Pair 20	0.6370	0.6062	0.6723	0.0052	0.6123	0.6044	0.6606	0.0047
Pair 21	0.7197	0.6177	0.7606	0.0027	0.7572	0.6283	0.7997	0.0047
Pair 22	0.6928	0.6895	0.7228	0.0027	0.7357	0.6939	0.7525	0.0065
Pair 23	0.5986	0.5922	0.6288	0.0033	0.5925	0.5787	0.6320	0.0053
Pair 24	0.6168	0.6091	0.6401	0.0023	0.6157	0.5942	0.7119	0.0019
Pair 25	0.7600	0.6692	0.9745	0.0063	0.7728	0.6742	1.1779	0.0060
	kneading				**pestle**			
	T_t (s)	T_{br} (s)	T_r (s)	Standard Deviation	T_t(s)	T_{br} (s)	T_r (s)	Standard Deviation
Pair 26	0.6422	0.6257	0.6586	0.0028	0.6519	0.6276	0.6700	0.0063
Pair 27	0.7731	0.6453	0.7870	0.0063	0.7589	0.6344	0.8751	0.0031
Pair 28	0.6907	0.6534	0.7964	0.0037	0.6906	0.6464	0.8573	0.0039
Pair 29	0.6916	0.6063	0.7305	0.0060	0.7054	0.6009	0.9063	0.0014
Pair 30	0.8105	0.7331	0.8328	0.0056	0.8556	0.7309	0.8816	0.0018
Pair 31	0.5363	0.5235	0.6507	0.0027	0.5494	0.5424	0.8425	0.0023
Pair 32	0.8413	0.6984	0.8875	0.0065	0.8498	0.6984	0.9582	0.0040
Pair 33	0.7106	0.6965	0.7126	0.0070	0.7206	0.7085	0.7637	0.0038
Pair 34	0.7299	0.6030	0.7398	0.0050	0.7180	0.6600	0.7472	0.0033
Pair 35	0.7980	0.6441	0.8513	0.0049	0.7961	0.6530	0.9244	0.0024
Pair 36	0.8241	0.6126	0.8491	0.0052	0.8341	6220	0.8596	0.0033
Pair 37	06808	0.5655	0.7209	0.0044	0.6875	0.5554	0.8082	0.0044
Pair 38	0.8060	0.7206	0.8525	0.0034	0.8528	0.7148	0.8708	0.0055
Pair 39	0.8068	0.5903	0.9504	0.0035	0.9093	0.6941	0.9523	0.0068
Pair 40	0.6923	0.6523	0.7151	0.0022	0.6974	0.6383	0.7684	0.0044
Pair 41	0.6764	0.6111	0.8531	0.0027	0.6615	0.6150	0.9167	0.0032
Pair 42	0.7181	0.6319	0.7702	0.0033	0.7177	0.6211	0.8204	0.0035
Pair 43	0.6731	0.6052	0.6950	0.0039	0.6635	0.6271	0.6984	0.0029
Pair 44	0.7236	0.6826	0.7323	0.0062	0.7132	0.6155	0.8273	0.0017

Guinot *et al. ROBOMECH Journal*   (2023) 10:12

Page 17 of 21

Table 3 (continued)

	kneading				pestle			
	T_t (s)	T_{br} (s)	T_r (s)	Standard Deviation	T_t (s)	T_{br} (s)	T_r (s)	Standard Deviation
Pair 45	0.8747	0.5991	0.8969	0.0029	0.8796	0.5981	1.0798	0.0026
Pair 46	0.7890	0.6093	0.8020	0.0034	0.7915	0.6054	0.8989	0.0022
Pair 47	0.7636	0.5957	0.7664	0.0046	0.7635	0.5958	0.8462	0.0068
Pair 48	0.9204	0.7108	0.9321	0.0058	0.9362	0.7293	0.9508	0.0064
Pair 49	0.7523	0.6559	0.7767	0.0041	0.7666	0.6621	0.8328	0.0036
Pair 50	0.8472	0.6374	0.9548	0.0066	0.8371	0.6391	1.0076	0.0028
	kneading				pestle			
	T_t (s)	T_{br} (s)	T_r (s)	Standard Deviation	T_t (s)	T_{br} (s)	T_r (s)	Standard Deviation
Pair 51	0.7092	0.6978	0.7199	0.0063	0.7278	0.7074	0.7698	0.0057
Pair 52	0.6024	0.4887	0.7065	0.0031	0.6169	0.4877	0.8537	0.0016
Pair 53	0.5972	0.4794	0.8091	0.0021	0.5840	0.4701	0.8855	0.0016
Pair 54	0.8170	0.6047	0.8289	0.0059	0.8440	0.6395	0.8637	0.0066
Pair 55	0.6999	0.5810	0.7805	0.0028	0.7100	0.6679	0.9439	0.0060
Pair 56	0.7058	0.6334	0.7556	0.0045	0.7532	0.6379	0.7834	0.0013
Pair 57	0.7249	0.6042	0.7283	0.0032	0.7340	0.6038	0.7808	0.0054
Pair 58	0.8436	0.7527	0.8751	0.0062	0.8305	0.7391	0.9343	0.0033
Pair 59	0.7898	0.6796	0.8686	0.0063	0.7937	0.6717	0.9286	0.0027
Pair 60	0.6719	0.6648	0.7323	0.0031	0.6847	0.6725	0.9894	0.0028
Pair 61	0.7183	0.7009	0.8217	0.0041	0.7228	0.7031	1.0099	0.0067
Pair 62	0.7993	0.6845	0.9361	0.0054	0.8110	0.6713	1.0430	0.0028
Pair 63	0.7785	0.6782	0.7852	0.0044	0.7869	0.6877	0.8010	0.0066
Pair 64	0.6984	0.5349	0.7505	0.0068	0.6911	0.5362	0.8020	0.0030
Pair 65	0.7787	0.6520	0.8363	0.0032	0.7766	0.6721	0.9060	0.0030
Pair 66	0.8196	0.7834	0.8691	0.0044	0.8072	0.7718	0.9527	0.0024
Pair 67	0.7770	0.6645	0.7984	0.0059	0.7877	0.6528	0.8697	0.0023
Pair 68	0.8582	0.6467	0.8749	0.0031	0.8737	0.6523	0.9230	0.0042
Pair 69	0.6729	0.4736	0.6744	0.0033	0.6615	0.4849	0.7316	0.0041
Pair 70	0.7409	0.5356	0.8065	0.0051	0.8305	0.5384	0.8527	0.0029
Pair 71	0.7675	0.6203	0.8080	0.0063	0.7812	0.6521	0.9004	0.0053
Pair 72	0.7499	0.6013	0.7718	0.0064	0.7615	0.6129	0.8200	0.0048
Pair 73	0.7618	0.6179	0.7730	0.0053	0.7605	0.6267	0.8745	0.0063
Pair 74	0.7153	0.6003	0.7342	0.0064	0.7105	0.6004	0.8521	0.0063
Pair 75	0.8052	0.6549	0.8412	0.0027	0.8446	0.6564	0.8733	0.0034
Pair 76	0.6573	0.4403	0.7135	0.0054	0.7177	0.5272	0.7465	0.0055
Pair 77	0.5310	0.4682	0.5720	0.0061	0.5335	0.4793	0.6723	0.0025
Average	0.7299	0.6238	0.7768		0.7389	0.6303	0.8501	
St.Dev.	0.0779	0.0644	0.0848		0.0840	0.0632	0.1089	

Table 4 Average measured T_t, T_{br} and T_r cycles in all explicit collaboration pairs for both the pestle and the kneading

	kneading				pestle			
	T_t (s)	T_{br} (s)	T_r (s)	St.Dev.	T_t(s)	T_{br} (s)	T_r (s)	St.Dev.
Pair 1	0.6241	0.6244	0.8902	0.0049	0.8293	0.8294	0.8460	0.0050
Pair 2	0.6953	0.6976	0.9822	0.0025	0.8061	0.8069	0.8002	0.0047
Pair 3	0.7033	0.7052	0.9100	0.0063	0.8078	0.8100	0.8114	0.0035
Pair 4	0.7166	0.7180	0.9656	0.0040	0.8199	0.8214	0.8198	0.0032
Pair 5	0.6972	0.6973	0.8035	0.0059	0.8928	0.8940	0.8873	0.0052
Pair 6	0.6886	0.6878	0.8037	0.0038	0.8482	0.8501	0.8457	0.0048
Pair 7	0.7936	0.7939	0.9500	0.0064	0.7326	0.7336	0.7317	0.0030
Pair 8	0.6459	0.6475	0.9089	0.0067	0.9073	0.9110	0.9018	0.0031
Pair 9	0.6955	0.6985	0.8723	0.0056	0.7843	0.7844	0.8854	0.0036
Pair 10	0.7275	0.7282	0.8436	0.0035	0.8297	0.8333	0.8298	0.0046
Pair 11	0.9454	0.9459	1.2228	0.0069	0.8694	0.8700	0.8667	0.0035
Pair 12	0.7489	0.7514	0.9085	0.0027	0.8712	0.8748	0.8803	0.0043
Pair 13	0.7964	0.7970	0.8807	0.0051	0.6984	0.7985	0.7953	0.0030
Pair 14	0.7053	0.7074	0.8244	0.0064	0.8065	0.8068	0.8153	0.0050
Pair 15	0.7778	0.7802	0.9136	0.0048	0.8386	0.8408	0.8329	0.0040
Pair 16	0.7130	0.7152	0.9367	0.0052	0.7768	0.7774	0.7766	0.0050
Pair 17	0.8247	0.8247	0.9358	0.0054	0.8367	0.8369	0.8441	0.0035
Pair 18	0.5313	0.5324	0.7256	0.0030	0.8045	0.8073	0.8081	0.0056
Pair 19	0.7202	0.7230	0.8857	0.0031	0.7582	0.7594	0.7529	0.0046
Pair 20	0.7247	0.7267	0.9351	0.0033	0.7681	0.7690	0.7718	0.0052
Pair 21	0.7993	0.7993	0.10178	0.0049	0.7781	0.7805	0.7757	0.0031
Pair 22	0.7305	0.7320	0.8358	0.0048	0.7387	0.7390	0.7554	0.0054
Pair 23	0.7851	0.7874	1.0872	0.0042	0.8557	0.8578	0.8644	0.0052
Average	0.7299	0.7313	0.9148		0.8113	0.8171	0.8217	
St.Dev.	0.0796	0.0794	0.7021		0.0524	0.0469	0.0473	

Guinot *et al. ROBOMECH Journal*   (2023) 10:12

Page 19 of 21

Appendix 2

See Tables 5 and 6.

Table 5 Measured Coefficient of Variation for each kneading-pestle implicit pair

	Coef. of Variation
Pair 1	1.5009
Pair 2	1.8660
Pair 3	1.5193
Pair 4	1.8355
Pair 5	1.4684
Pair 6	1.2254
Pair 7	1.7616
Pair 8	1.5746
Pair 9	1.0898
Pair 10	1.4047
Pair 11	1.4836
Pair 12	1.6855
Pair 13	1.3650
Pair 14	1.4084
Pair 15	1.5568
Pair 16	1.5351
Pair 17	1.7072
Pair 18	1.4202
Pair 19	1.5887
Pair 20	1.1051
Pair 21	1.5734
Pair 22	1.8047
Pair 23	1.4790
Pair 24	1.3608
Pair 25	1.3029
Pair 26	1.5692
Pair 27	1.7800
Pair 28	1.7799
Pair 29	1.3275
Pair 30	1.2787

	Coef. of Variation
Pair 31	1.3124
Pair 32	1.4753
Pair 33	1.7099
Pair 34	1.7687
Pair 35	1.5442
Pair 36	1.8244
Pair 37	1.2089

Table 5 (continued)

	Coef. of Variation
Pair 38	1.2407
Pair 39	1.3982
Pair 40	1.7363
Pair 41	1.5904
Pair 42	1.3787
Pair 43	1.5887
Pair 44	1.3656
Pair 45	1.8719
Pair 46	1.3751
Pair 47	1.9531
Pair 48	1.8948
Pair 49	1.4713
Pair 50	1.4742
Pair 51	1.3956
Pair 52	1.4830
Pair 53	1.6461
Pair 54	1.2358
Pair 55	1.8880
Pair 56	1.6570
Pair 57	1.3256
Pair 58	1.6719
Pair 59	1.8776
Pair 60	1.5979
Pair 61	1.0859
Pair 62	1.7791
Pair 63	1.7536
Pair 64	1.8290
Pair 65	1.3857
Pair 66	1.5578
Pair 67	1.9353
Pair 68	2.1366
Pair 69	2.1249
Pair 70	1.7344
Pair 71	1.5601
Pair 72	1.9105
Pair 73	1.8318
Pair 74	1.4781
Pair 75	1.4910
Pair 76	1.4561
Pair 77	1.5854
Average	1.5692
St.Dev.	0.2345

Guinot *et al. ROBOMECH Journal*   (2023) 10:12

Page 20 of 21

Table 6 Measured Coefficient of Variation for each kneading-pestle explicit pair

	Coef. of Variation
Pair 1	2.5280
Pair 2	2.3728
Pair 3	2.7654
Pair 4	2.6346
Pair 5	2.5112
Pair 6	2.6883
Pair 7	2.6040
Pair 8	2.4667
Pair 9	2.4463
Pair 10	2.4205
Pair 11	2.6208
Pair 12	2.6922
Pair 13	2.7227
Pair 14	2.4995
Pair 15	2.4016
Pair 16	2.5520
Pair 17	2.6902
Pair 18	2.8613
Pair 19	2.6914
Pair 20	2.3107
Pair 21	2.7827
Pair 22	2.4301
Pair 23	2.8477
Average	2.5887
St.Dev.	0.1566

Appendix 3
See Table 7.

Table 7 Number of times the "dough" was hit with and without the use of indication

	Without Indication	With Indication	Standard Deviation
Participant 1	49	52	1.06
Participant 2	34	38	1.12
Participant 3	44	50	1.14
Participant 4	33	37	1.12
Participant 5	38	45	1.18
Participant 6	38	39	1.03
Average	39	45	1.11

Appendix 4
See Table 8

Table 8 Coefficient of Variation with and without indication

	Without Indication	With Indication
Participant 1	6.4132	4.5526
Participant 2	3.5832	2.9897
Participant 3	4.4943	3.7402
Participant 4	5.2234	4.6305
Participant 5	3.7189	4.0037
Participant 6	3.1183	2.8166

References
1. Kase K, Suzuki K, Yang P-C, Mori H, Ogata T (2018) Put-in-box task generated from multiple discrete tasks by ahumanoid robot using deep learning. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 6447–6452. https://doi.org/10.1109/ICRA.2018.8460623
2. Yang P-C, Sasaki K, Suzuki K, Kase K, Sugano S, Ogata T (2017) Repeatable folding task by humanoid robot worker using deep learning. IEEE Robot Autom Lett 2(2):397–403. https://doi.org/10.1109/LRA.2016.2633383
3. Tzvetkova G (2014) Robonaut 2: Mission, technologies, perspectives. J Theor Appl Mech **44**. https://doi.org/10.2478/jtam-2014-0006
4. Khatib O, Yeh X, Brantner G, Soe B, Kim B, Ganguly S, Stuart H, Wang S, Cutkosky M, Edsinger A, Mullins P, Barham M, Voolstra CR, Salama KN, L'Hour M, Creuze V (2016) Ocean one: A robotic avatar for oceanic discovery. IEEE Robot Autom Mag 23(4):20–29. https://doi.org/10.1109/MRA.2016.2613281
5. Flemisch F, Abbink D, Itoh M, Pacaux-Lemoine M-P, Weßel G (2019) Joining the blunt and the pointy end of the spear: towards a common framework of joint action, human-machine cooperation, cooperative guidance and control, shared, traded and supervisory control. Cogn Technol Work 21. https://doi.org/10.1007/s10111-019-00576-1
6. Flemisch F, Heesen M, Hesse T, Kelsch J, Schieben A, Beller J (2011) Towards a dynamic balance between humans and automation:

Guinot *et al. ROBOMECH Journal*   (2023) 10:12

Page 21 of 21

Authority, ability, responsibility and control in shared and cooperative control situations. Cogn Technol Work 14:3–18. https://doi.org/10.1007/s10111-011-0191-6

7. Gervasi R, Mastrogiacomo L, Franceschini F (2020) A conceptual framework to evaluate human-robot collaboration. Int J Adv Manuf Technol 108. https://doi.org/10.1007/s00170-020-05363-1

8. Music S, Hirche S (2017) Control sharing in human-robot team interaction. Ann Rev Control 44:342–354. https://doi.org/10.1016/j.arcontrol.2017.09.017

9. Llorens -Bonilla B, Parietti F, Asada HH (2012) Demonstration-based control of supernumerary robotic limbs. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3936–3942. https://doi.org/10.1109/IROS.2012.6386055

10. Sasaki T, Saraiji M, Fernando C, Minamizawa K, Inami M (2017). MetaLimbs: Multiple arms interaction metamorphism. In ACM SIGGRAPH 2017 Emerging Technologies, SIGGRAPH 2017 [a16] (ACM SIGGRAPH 2017 Emerging Technologies, SIGGRAPH 2017). Association for Computing Machinery, Inc. https://doi.org/10.1145/3084822.3084837

11. Saraiji M, Sasaki T, Matsumura R, Minamizawa K, and Inami M (2018) Fusion: full body surrogacy for collaborative communication. In ACM SIGGRAPH 2018 Emerging Technologies (SIGGRAPH '18). Association for Computing Machinery, New York, NY, USA, Article 7, 1–2. https://doi.org/10.1145/3214907.3214912

12. Khan W, Abbas S, Khan M, Qazi W, Saleem Khan M (2020) Intelligent task planner for cloud robotics using level of attention empowered with fuzzy system. SN Appl Sci 2. https://doi.org/10.1007/s42452-020-2312-4

13. Pfungst O (1911) Clever Hans : the horse of Mr Von Osten. Holt, Rinehart and Winston, New York

14. Ueda K, Sakura T, Narita Y, Sawai K, Morita T (2012) Silent communication among bunraku puppeteers. In: Proceedings of the 29th Annual Meeting of the Japanese Cognitive Science Society

15. Shibuya T, Morita Y, Fukuda H, Ueda K, Sasaki M (2012) Asynchronous relation between body action and breathing in bunraku: Uniqueness of manner of breathing in japanese traditional performing arts. Cogn Stud 19:337–364

16. Jack R, Schyns P (2015) The human face as a dynamic tool for social communication. Current Biol 25:621–634. https://doi.org/10.1016/j.cub.2015.05.052

17. Gelder B, Hadjikhani N (2006) Non-conscious recognition of emotional body language. Neuroreport 17:583–6. https://doi.org/10.1097/00001756-200604240-00006

18. Adolphs R (2002) Recognizing emotion from facial expressions: Psychological and neurological mechanisms. Behav Cogn Neurosci Rev 1(1):21–62. https://doi.org/10.1177/1534582302001001003

19. Faria P, Braga R, Valgode E, Reis L (2007) Interface framework to drive an intelligent wheelchair using facial expressions. In: IEEE International Symposium on Industrial Electronics, Vigo, Spain, 2007, pp. 1791–96. https://doi.org/10.1109/ISIE.2007.4374877

20. Cruz F, Twiefel J, Magg S, Weber C, Wermter S (2015) Interactive reinforcement learning through speech guidance in a domestic scenario. In: 2015 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. https://doi.org/10.1109/IJCNN.2015.7280477

21. She L, Cheng Y, Chai JY, Jia Y, Yang S, Xi N (2014) Teaching robots new actions through natural language instructions. In: The 23rd IEEE International Symposium on Robot and Human Interactive Communication, pp. 868–873. https://doi.org/10.1109/ROMAN.2014.6926362

22. GALAN F, (2008) A brain-actuated wheelchair : asynchronous and non-invasive brain-computer interfaces for continuous control of robots. Clin Neurophysiol 119:2159–2169

23. Rechy-Ramirez EJ, Hu H, McDonald-Maier K (2012) Head movements based control of an intelligent wheelchair in an indoor environment. In: 2012 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 1464–1469. https://doi.org/10.1109/ROBIO.2012.6491175

24. Yin YH, Fan YJ, Xu LD (2012) Emg and epp-integrated human-machine interface between the paralyzed and rehabilitation exoskeleton. IEEE Trans Inf Technol Biomed 16(4):542–549. https://doi.org/10.1109/TITB.2011.2178034

25. Qiu S, Li Z, He W, Zhang L, Yang C, Su C-Y (2017) Brain-machine interface and visual compressive sensing-based teleoperation control of an exoskeleton robot. IEEE Trans Fuzzy Syst 25(1):58–69. https://doi.org/10.1109/TFUZZ.2016.2566676

26. GreatBigStory: Pounding Mochi With the Fastest Mochi Maker in Japan. Youtube. https://www.youtube.com/watch?v=tmSrULDVRPc

27. HTC Vive Tracker. Accessed on 01.11.2022 (2020). https://www.vive.com/us/accessory/tracker3/

28. Kelsch J, Temme G, Schindler J (2013) Arbitration based framework for design of holistic multimodal human-machine interaction. In: Contributions to AAET 2013, 6.-7. Feb. 2013, Braunschweig, Germany, ISBN 978-3-937655-29-1

29. Baltzer M, Altendorf E, Kwee-Meier S, and Flemisch F (2021) Mediating the Interaction between Human and Automation during the Arbitration Processes in Cooperative Guidance and Control of Highly Automated Vehicles: Base concept and First Study. In: Advances in Human Aspects of Transportation: Part I

30. ATR-Promotions: TSND151. Accessed on 01.02.2021. http://www.atr-p.com/products/sensor.html

31. Takahashi S, Iwasaki Y, NAKABAYASHI K, IWATA H (2017) Research on "third arm": voluntarily operative wearable robot arm: - development of face vector sensing eyeglasses -. The Proceedings of JSME annual Conference on Robotics and Mechatronics (Robomec) 2017, 1–208 (2017). https://doi.org/10.1299/jsmermd.2017.1P2-L08

32. Iwasaki Y and Iwata H (2018) A face vector - the point instruction-type interface for manipulation of an extended body in dual-task situations. In: IEEE International Conference on Cyborg and Bionic Systems (CBS), Shenzhen, China, 2018, pp. 662–66, https://doi.org/10.1109/CBS.2018.8612275.

33. Hayakawa Y, Ogata T, Sugano S (2003) Flexible assembly work cooperation based on work state identifications by a self-organizing map. In: Proceedings 2003 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM 2003), vol. 2, pp. 1031–1036

34. Nicolis D, Zanchettin AM, Rocco P (2018) Human intention estimation based on neural networks for enhanced collaboration with robots. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1326–1333. https://doi.org/10.1109/IROS.2018.8594415

35. Tee KP, Wu Y (2018) Experimental evaluation of divisible human-robot shared control for teleoperation assistance. In: TENCON 2018 - 2018 IEEE Region 10 Conference, pp. 0182–0187 (2018). https://doi.org/10.1109/TENCON.2018.8650436

36. Hart S. G. and Staveland E. L. (1988) Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Advances in psychology 52: pp.139–183.

37. Kobayashi K, Yamada S . (2005) Human-robot cooperative sweeping by extending commands embedded in actions. In: 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2977–2982 . https://doi.org/10.1109/IROS.2005.1545243

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.