Explorations on Time-informed Human-Robot Interaction*

Michail Maniadakis[1], Emmanouil Hourdakis[1], Markos Sigalas[1], Maria Koskinopoulou[1] and Panos Trahanias[1]

***Abstract*— The sense of time is an essential capacity of humans, with a major role in many of the cognitive processes expressed in our daily lifes. So far, in cognitive science and robotics research, mental capacities have been investigated in a theoretical and modeling framework that largely neglects the flow of time. Only recently there has been a rather limited, but constantly increasing interest in the temporal aspects of cognition, integrating time into a range of different models of perceptuo-motor capacities. The current paper aims to review recent work conducted in the Computational Vision and Robotics Laboratory of FORTH-ICS, in a well focused attempt to provide robots with temporal cognition and thus facilitate their seamless integration into modern human societies.**

I. INTRODUCTION

Fluent multi-agent interaction assumes participating entities capable to consider the inherent temporal aspects of real world procedures in order to develop a common understanding on ordering, causality relationships and accomplish to synchronize effectively. Despite the crucial role that the sense of time has in human cognition, both in perception and action, the capacity of artificial agents to experience the flow of time is still poorly explored. The inability of existing systems to perceive time constrains and follow the heavily time-structured human course acts as an obstacle for the integration of robots into human environments [1]. As a result, the majority of existing robots either act asynchronously to the human needs, or assume the humans decide “which and when” actions should be triggered.

In order to fill this gap, we have implemented a series of models that address many of the temporal features involved in human-robot confluence in the long- and short-term. The current review paper aims to summarize the work conducted in the different research strands explored over the last years and thus reveal the impact that artificial temporal cognition may have in advancing human-robot interaction in daily setups.

The time-relevant components operate in parallel with other modules of a robotic system (control, navigation, perception, etc.), to assist situating robots not only in space, as currently implemented, but also in time, by considering the evolution of human-robot interaction over the past-present-future timeline. While embodiment focuses very much on the here and now of the world, the proposed “entiment” of artificial cognition additionally postulates that past experiences and future goals, together with their timing, are also very strong and important driving forces to shape the functionality of intelligent autonomous systems deployed in human environments. This is also particularly applicable for digital companies that aim to assist and advice humans along their largely time-centered course of activities.

Fig. 1. Graphical illustration of the symbolic representation of episodes stored in memory.

II. SYMBOLIC REPRESENTATION OF EPISODIC MEMORIES

To exploit and extract valuable information from the sequence of human-robot interaction sessions, we have implemented a symbolic, fast to save and search memory that stores experiences in the form of abstractly described, symbolic episodes. Past experiences are further recalled and processed in order to estimate human habits/intension and thus inform and improve robot’s decision making within the context of ongoing and upcoming HRI sessions.

The symbolic episodic memory module assumes an entity based representation of episodic memories in hierarchical multigraphs. The entities observed within a given HRI session are encoded in a temporally rich domain which enables the representation and time-stamping of episodes occurred at diverse past times. For example, Fig. 1 shows the symbolic episodic memory representation of the event breakfast preparation on a Sunday morning.

Entities and associations involved in episodes are stored in memory but may be forgotten after some time [2]. The lifecycle of memory items is defined by their importance in the relevant scenarios. Memory management assumes that entities and their associations are asynchronously updated,

---

*This work has been partially supported by the EU FET grant (GA: 641100) TIMESTORM - Mind and Time: Investigation of the Temporal Traits of Human-Machine Convergence and the EU FET grant (GA: 754423) EnTiment - Industrial Exploitation and Market Uptake of a Temporal Cognition Toolbox. for Commercial Robots

[1] All authors are with the Foundation for Research and Technology Hellas, Institute of Computer Science (FORTH-ICS), Heraklion, Crete, Greece.

adaptation.png

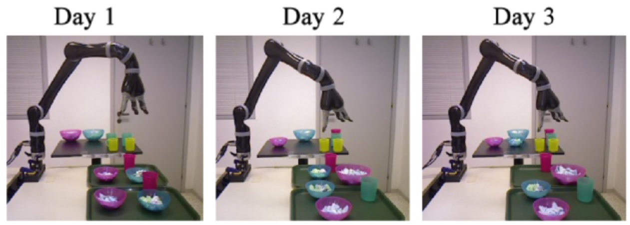


Fig. 2. Breakfast servings adapted to user preferences.

User	Health	Coffee	Tea	Beverage converge	Strawberries	Cherries	Food converge
A1	✗	✓	✓	**1** iter	✓	✓✓	**5** iters
A2	✓	✗	✓	**2** iters	✗	✓	**5** iters
A3	✗	✓	✗	**2** iters	✓✓	✓	**6** iters
A4	✓	✗	✓	**1** iter	✓	✗	**5** iters
A5	✗	✓	✓	**1** iter	✗	✓✓	**3** iters
A6	✓	✓	✗	**3** iters	✓✓	✗	**8** iters

Fig. 3. Summary of symbolic episodic memory usage in uncovering user preference.

while a time-correlated importance factor modulates the merging, forgetting or refreshing of memories, in order to facilitate future search of the stored information. More specifically, information update regards examining whether an entity has been observed again in the past and, thus, should be merged with the previous occurrences and whether it is of low importance for a given task and the corresponding task-specific associations should be forgotten.

Past experiences are exploited to predict future states and develop expectations within HRI sessions, or identify abnormalities during the scenario unfolding, in order to properly adjust robot activities. This is implemented following an HMM-based probabilistic inference approach that builds on existing knowledge and the statistical correlation across the involved entities [3]. In particular, separate HMMs are developed on the fly, associated with queries and the relevant time-stamped data retrieved from the memory. The recalled observations are used to train the given HMM, infer scenario-relevant information and guide robot behavior.

To assess episodic memory performance and its ability to improve the ongoing HRI experience of end-users, the implemented model is embodied in the six-joint JACO-Kinova robot to be used in past-informed breakfast serving (Fig. 2). According to the examined scenario [3], the robotic servant needs to infer user preferences by experimenting with serving different menus (food+beverage), as users try different breakfasts day after day. Figure 3 summarizes the preferences of the six users assumed in our experiments and the iterations used by the system to identify their preferences. The exploitation of episodic memories by HMM greatly facilitates the inference of “hidden” information and enables the robot to adjust its behavior and properly serve each user according to his preferences.

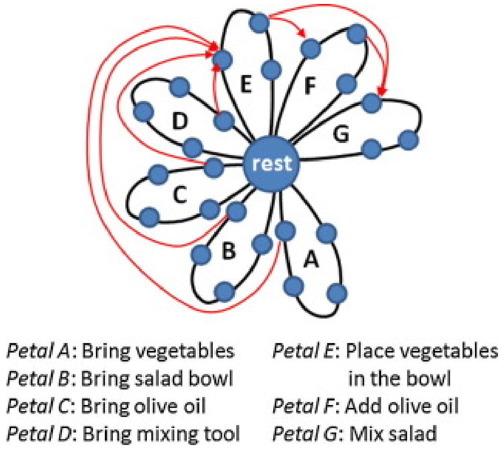


Fig. 4. A daisy graph representing the interaction of tasks that should be executed for the preparation of salad. Petals A, B, C and D represent tasks that can be implemented at any order. Petal E can be executed only after A, B, C, D, petal F can be executed only after A, B, C, D, E and petal G can be executed only after A, B, C, D, E, F. See text for details.

III. TIME-INFORMED PLANNING IN MULTI-AGENT SETUPS

Multi-agent collaboration in unstructured environments (e.g. a home) may be often interrupted by unexpected events that disturb the execution of actions (e.g. while cooking, the baby may cry, or the phone may ring). In such cases it is rather impractical to devote resources into the estimation of global plans that describe a long sequence of actions. It seems more effective to adopt a progressive, short-term planning that can directly adapt to the unfolding of task execution.

In line with the above described HRI framework, we have developed a new planning approach for multi-agent interaction that can effectively and flexibly drive human-robot synergies, Fig. 4.The planning approach is based on the short-term and incremental attribution of tasks to the available agents, therefore enabling the easy and smooth recovery of collaboration from unexpected disturbances and the on-the-fly adaptations on the synthesis of the collaborating team [4], [5].

The planner assumes the daisy-like representation of complex scenarios separated in tasks that can be implemented in parallel by different agents. Moreover, the fuzzy number representation of duration facilitates the processing of temporal information and the direct use of time in mathematical calculations. Interestingly, the latter enables combining temporal information with the quantitative description of the skills and properties of the participating agents to develop complex, multi-criteria measures which facilitate the detailed analysis of alternative plans in order to take better and time-informed planning decisions. In short, at any planning decision moment, the planner considers the agents that are not charged with any task and makes a new assignment for them by choosing the task that best matches their own skills and characteristics in order to maximize the benefit of the team. Fig. 5 shows an exemplar application of the

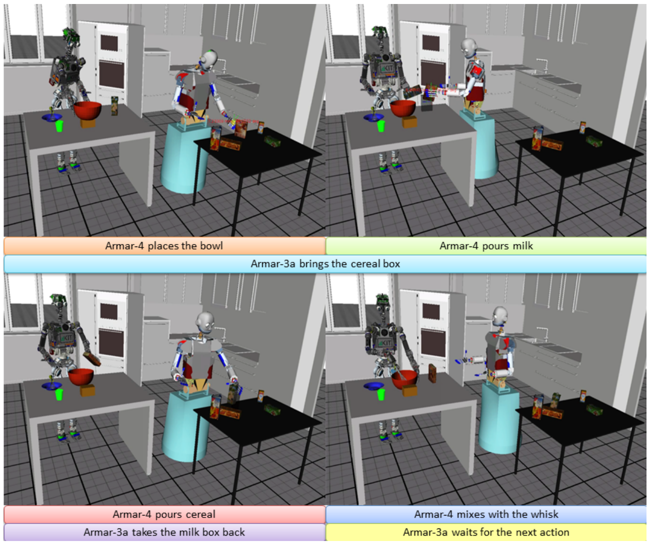


Fig. 5. The DP orchestrates the collaboration of the simulated ARMAR-3 and ARMAR-4 robots during cereal-milk preparation.

proposed approach. The examined scenario assumes two heterogeneous robots, to collaborate for the preparation of cereal milk breakfast. The planner considers the diverse behavioral capacities of the robots in order to effectively orchestrate collaboration between the two agents.

IV. EMOTIONALLY-MODULATED TIME PERCEPTION

The above described time-informed planner is enriched with the ability to consider how humans experience the flow of time, accomplishing to effectively prioritize servings in setups where a single robot interacts with multiple humans. For example, in the field of domestic service robots it may be often the case that two inhabitants ask help from the (single) robotic assistant that supports house activities. To decide which of the two help requests will be served first, the robot should ideally consider both the emotional state of humans and the estimated completion time of the two individual requests [6]. It is well known that emotions affect human time perception. In particular, time flies when we are happy or in a high arousal state, and to drag when we are sad or bored. Considering the domestic robot example discussed above, the emotional effects on time perception imply that a happy person in high arousal state would only be satisfied with the direct realization of his request. In contrast, a person in low arousal or sad state would thoughtlessly accept being served second in the queue (given that the completion time of the other task is not very long).

We have exploited experimental data from human studies investigating the effect of emotions on human time perception in naturalistic daily conditions. These data are used to train an artificial neural network that predicts "how fast" humans experience the flow of time, given their current emotional state. The planner exploits human-time-judgment predictions in order to effectively prioritize robot activities by considering how delays are experienced by the participants. In particular, the planner aims to minimize their average experienced delay and in that way improve the composite level of satisfaction for the interacting humans.

Fig. 6. The DP is embodied in the JACO-Kinova robot to prioritize food serving between users.

The effectiveness of the proposed approach has been validated in a hypothetical multi agent interaction setup where randomly paired lab members at various predeclared emotional states are served meals taking different times to prepare by a JACO robotic arm (see Fig. 6). Obtained results showed that DP effectively considers the emotional state of humans to successfully prioritize planning and thus improve the HRI experience of participants.

V. TIME EFFECT IN THE LATENT SPACE REPRESENTATION OF ACTIONS

Contemporary Learning from Demonstration (LfD) methods have been widely employed for robotic imitation of human action behavior. Previously, we had developed an LfD framework, termed IMFO (IMitation Framework by Observation), that is based on the compact, low-dimensional representation of both human and robot arm motions, which are properly associated to facilitate learning. Interestingly, the compressed representation of actions, in the so-called latent space, preserves the significant properties of the actions' spatial trajectories and facilitates mapping between an observed human action and the reproduced robotic one. However, the role of temporal information in the computational representation and reproduction of actions remains poorly understood. To address this issue, we have investigated how speed changes shape the low dimensional latent representation of the demonstrated actions.

In short, the proposed method regards augmenting the algorithm that implements the transformation from the full configuration space to the compact latent space, with temporal information that affects execution speed. The compressed representation of similar actions with different spatio-temporal characteristics shows that speed plays a major role in the derived latent representation, effectively separating similar arm motions that are executed at different speeds. Accordingly, the latter actions assume unambiguous latent space representations when only speed of execution varies, allowing thus the accurate reproduction of acts with different velocities.

The implementation of the model assumes its embodiment in a robotic arm, to develop correspondence with the human arm actions. Current experiments have used a six-joint arm

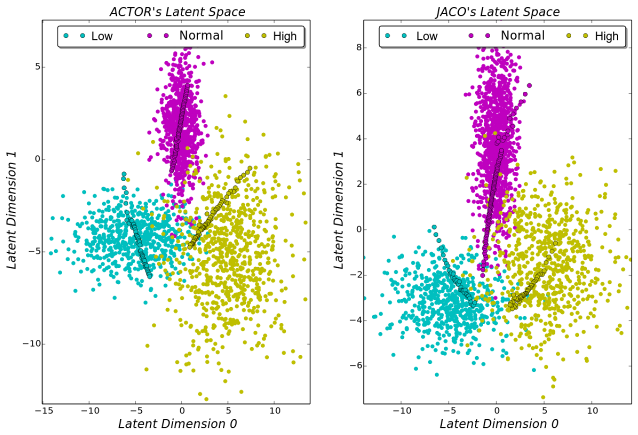


Fig. 7. The latent space representation of actions performed at different speeds by the human (left) and the robot (right).

manipulator, namely the JACO robotic arm, by Kinova Robotics. Thorough testing of the implemented model in a kinesthetic learning setup for actions executed at three different speeds (low, normal, high) revealed well-separated sectors in the latent space representation of actions, which correspond to the three implemented speeds (see Fig 7). Interestingly, the latter holds true for the latent representations of both the human and the robot actions.

At a second stage of our study, model performance is tested and validated in a realistic robot service scenario that assumes speed adaptations. Interestingly, the proposed spatio-temporal formulation of IMFO readily lends itself to integration with time-informed planning approaches to effectively address temporal constraint satisfaction in real-world scenarios. Following our previous work on latency estimation, the composite system is capable to cope with cases where completion of certain behaviors is expected to be delayed. The elimination of latency is accomplished through the estimation of a requested, reduced time for action completion. Taking advantage of learning from demonstration actions at different speeds, we select the action implementation that best matches the requested completion time, therefore facilitating the realization of the composite behavior within the predefined time limits.

The examined scenario is inspired by restaurant standing queues with customers served one at a time. The simplified serving considered here assumes two cups and one bowl to be placed on each tray. We consider varying times of requested tray filling, centered at 2.0 minutes (the average period of customer arrival). In short, when serving a customer is delayed, the system tries to compensate this latency by asking for faster filling of future trays. Following this formulation/scenario, the repetitive tray filling task must be implemented at varying time limits and hence robot action speeds.

We have developed a simple setup that enables tray filling in naturalistic conditions (see Fig. 8). For quantitative assessment of the action execution times, we conducted 20 repetitions of the tray filling task; cases where grasp failures where encountered were dropped out and the experiment was repeated. Accordingly, we ended up with 20 successful task completions for which we contrasted the actual execution times against the commanded ones by the planner. Time differences above 10% were regarded as failures. Interestingly, only 3 executions did not meet the latter criterion, and were marked unsuccessful. Given the complexity and variability of the studied scenario, the accomplished result is considered highly promising and indicative of the method's potential.

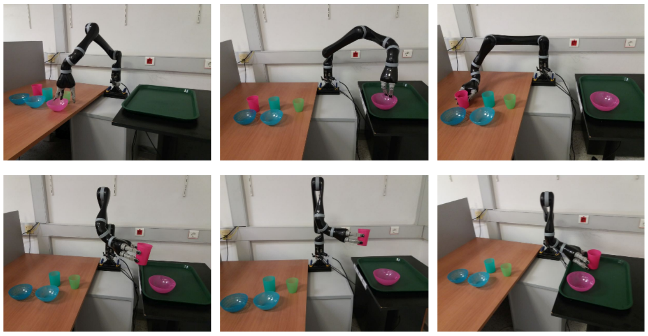


Fig. 8. Embodied assessment of the time-informed guided speed adaptation of robotic behaviours.

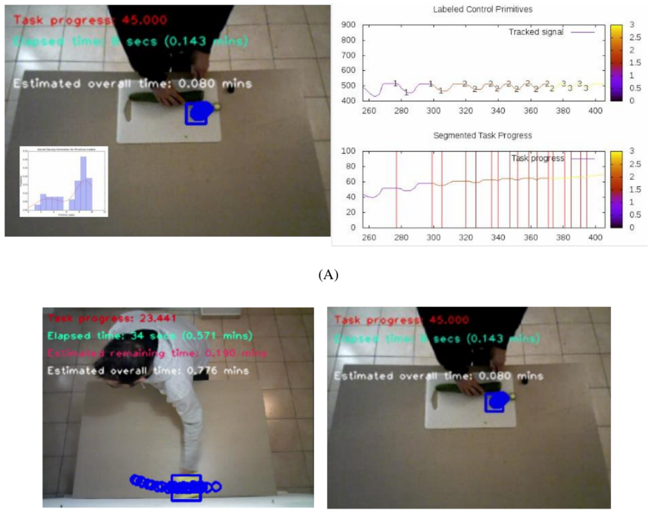
VI. GENERATIVE TIME MODELS

The ability to estimate and predict temporal properties of events is one of the core components in human cognition. It allows someone to plan his/her actions ahead, allocate effort and resources to tasks that are time-constrained or critical, and facilitates interactions with the environment, by predicting the expected event outcomes.

In robotics however, models that predict the time-related properties of an activity have not been yet explored, despite their obvious benefits. For example, models for these temporally salient primitives can be employed to predict temporal quantities of the task, answering all sorts of time-related questions, such as:

- How long will it take to finish this activity?
- Which behavior configuration will finish this task as fast as possible?
- I want to clean the table, during the 5 minutes of free time that I have. How clean will the table be after that?
- Am I performing the task in an efficient manner, given the spare time that I have?
- I have 3 minutes of spare time. What tasks can I perform during this period? The ability to answer these type of questions endows artificial systems with enhanced inference capabilities. Such temporal models can be useful in various disciplines, including process modeling, planning, perception and human-robot interaction, to name a few.

  In the current deliverable, we derive a principled formulation that enables artificial systems to estimate, by observation, the duration, and other temporal quantities, of an activity with minimal prior experience and



(A)

(B)

Fig. 9. GTMs used to predict the temporal properties of human actions. (A) The GTM is used to predict the duration for the cut the cucumber task. (B) Generalization of the GTM to more than one tasks.

training. We introduce the concept of Generative Time Models (GTMs), i.e. models that segment an observed activity into temporally salient intervals, and use them to predict time-related information.

In order to provide temporal predictions, a GTM analyzes an activity using two observation models: (1) task progress and (2) control [7]. The first one estimates the progress of the task, i.e. how much of the activity has been completed. The second one, identifies and records information about the primitive motions that appear during the activity. The model tracks the control system in task space, thus solving implicitly the embodiment correspondence problem. This makes a GTM compatible with different embodiments and users. To estimate time related information about an activity, a GTM observes and processes indicators of the activity progress. Due to its architecture, with segregated Control and Activity observation components, GTMs can generalize well, by using Control and Activity observation models across tasks.

In order to demonstrate the effectiveness of the proposed approach, we have applied the model on two different household activities, cleaning the table and cutting vegetables, with minimal changes on each GTM component (Fig 9). In both cases, the model has successfully and accurately predicted human-task completion times. This has been a valuable information for synchronizing robot functionality with human activities.

VII. CONCLUSIONS

Besides the above described robotic applications, in the context of intelligent personal assistants, the “entiment” of artificial cognition has the potential to support numerous functions and capacities such as enabling them to identify and exploit their spare time, provide recommendations after considering the temporal properties of tasks, predict and plan for distant future events, or adapt action plans based on the recall of past experiences.

Overall, it is clear that artificial temporal cognition is without doubt, not an optional extra, but a necessity for the development of cognitive machines and intelligent digital companions. The exaptability of relevant/such systems may significantly gain from the artificial time processing capacities that is expected to improve fluency in man-machine interaction as well as the willingness of end users to cooperate with personalized assistance.

ACKNOWLEDGMENT

This work has been partially supported by the EU FET grant (GA: 641100) TIMESTORM - Mind and Time: Investigation of the Temporal Traits of Human-Machine Convergence.

REFERENCES

[1] M. Maniadakis and P. Trahanias, “Temporal cognition: a key ingredient of intelligent systems,” *Frontiers in Neurorobotics*, vol. 5, 2011.

[2] M. Sigalas, M. Maniadakis, and P. Trahanias, “Time-aware long-term episodic memory for recurring hri,” in *ACM/IEEE Int. Conference on Human-Robot Interaction (HRI)*, 2017.

[3] ——, “Episodic memory formulation and its application in long-term hri,” in *26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2017.

[4] M. Maniadakis, E. Hourdakis, and P. Trahanias, “Time-informed task planning in multi-agent collaboration,” *Cognitive Systems Research*, vol. 43, pp. 291–300, 2017.

[5] M. Maniadakis, E. E. Aksoy, T. Asfour, and P. Trahanias, “Collaboration of heterogeneous agents in time constrained tasks,” in *IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, 2016.

[6] M. Maniadakis, S. Droit-Volet, and P. Trahanias, “Emotionally modulated time perception for prioritized robot assistance,” in *ACM/IEEE Int. Conference on Human-Robot Interaction (HRI)*, 2017.

[7] E. Hourdakis and P. Trahanias, “A robust method to predict temporal aspects of actions by observation,” *submitted to IEEE RA-L*, 2017.