

# Execution Memory for Grounding and Coordination

Stephanie Rosenthal, Sarjoun Skaff  
Bossa Nova Robotics  
Pittsburgh PA USA  
{stephanie,sarjoun}@bnrobotics.com

Manuela Veloso  
Carnegie Mellon University  
Pittsburgh PA USA  
veloso@cmu.edu

Dan Bohus, Eric Horvitz  
Microsoft Research  
Redmond WA USA  
{dbohus,horvitz}@microsoft.com

**Abstract**—As robots are introduced into human environments for long periods of time, human owners and collaborators will expect them to remember shared events that occur during execution. Beyond naturalness of having memories about recent and longer-term engagements with people, such execution memories can be important in tasks that persist over time by allowing robots to ground their dialog and to refer efficiently to previous events. In this work, we define execution memory as the capability of saving interaction event information and recalling it for later use. We divide the problem into four parts: *salience filtering* of sensor evidence and saving to short term memory, *archiving* from short to long term memory and *caching* from long to short term memory, and *recalling* memories for use in state inference and policy execution. We then provide examples of how execution memory can be used to enhance user experience with robots.

## I. INTRODUCTION

As robots are deployed long-term in human environments, they will increasingly have roles in which they interact with people on a continuing basis. People collaborating with robots will expect robots to remember who they met, what conversations they had, and what actions they took throughout their deployment in order to ground future interactions. People will come to expect that robots recall and refer to past engagements in the same way that people encode and share memories about events and activities [1]. Additionally, they will expect robots to learn about their users, task, and environment using remembered data (e.g., through socially-guided machine learning for robots [2], [3]). We call this capability of saving and remembering interaction events *execution memory*.

Beyond the naturalness of encoding and sharing memories with people about short- and long-term histories of activities and engagements, execution memories can be crucial in effective collaborations on tasks that persist over time and span multiple sessions. Shared memories are important to ground dialog and allow human and robot to refer to one or more aspects of a previous event with shorthand or gesture. Such memories can also enable interrupted, incomplete, or persisting tasks to be efficiently refined, extended, and continued. We see opportunities for endowing robots with execution memory, including knowledge about the ways that people selectively encode and recall events about situations and activities.

Work to date on learning and using models of human memory in human-computer interaction includes efforts to learn and harness models of human memory from data in search and retrieval and in reminding ([4], [5], [6]). We see these ideas as framing directions with inferences and

uses of execution memory in human-robot interaction. Unlike active learning [2] and learning by demonstration [7], these techniques focus on remembering specific instances of events in order to ground actions, similar to what humans recollect [8].

In this work, we formalize execution memory, breaking it down into four main parts: *salience filtering*, *archiving*, *caching*, and *recall*. Then, we describe the use of execution memory to enhance the interactions with our robots.

## II. EXECUTION MEMORY

Similar to human memory, we break down execution memory into two parts: the short-term task-related memory and the long-term lifelong memory. The short-term memory can be used to recall information including filtered salient features from the sensor data as well as cached data from long-term memory. Once a task is over, short-term memories can be archived to long-term memory where they may later be cached back into short-term memory. We define four distinct functions over execution memory – salience filtering, archiving, caching, and recall – that input to and output from short-term and long-term parts. We provide an overview of each function. We then provide examples of execution memory on our own robots.

**Salience Filtering:** The salience filtering function determines whether to save the current sensor evidence and/or state or policy inferences into short-term memory. Salient filtered information might include users’ answers to questions, face recognition data, or the robots’ current location.

**Archiving:** We assume that robots’ current task time is much shorter than its total deployment time, and that remembering every detail of every task is neither efficient nor necessary to ground interactions with humans. For example, it may not be necessary to remember exactly what a person said during a previous task but only the last time the interaction occurred. The archiving function determines which short-term memories should be transferred to long-term memories.

**Caching:** The caching function determines which long-term memories could be useful for the current task. In order to both archive and cache effectively, we must identify when current events will likely be relevant in future activities and when past memories are relevant to current tasks, respectively.

**Recall:** Finally, robots will need to recall the cached memories and incorporate them with current sensor data in their task. Determining which memories, if any, will be important for the

current state inference or policy execution depends highly on user expectations and what they remember themselves [5].

### III. EXAMPLES OF EXECUTION MEMORY

**mObi**, developed by Bossa Nova Robotics, is designed to interact with people over extended periods of time. mObi is approximately 1.5m tall, visually detects people, and can talk with them about a number of topics. In order to avoid repetitious dialog, mObi remembers what it has already said during the task and changes its current dialog accordingly. For example, mObi can answer questions about itself and how it works without providing redundant information. To do this, mObi's filtering function remembers which topics it has spoken about in the current dialog. When a person asks a similar question to one it has already answered, it recalls this information, reminding the user that it has already answered their question and recommends other topics to ask about.

**CoBot** is an autonomous mobile robot that performs tasks for people at Carnegie Mellon University [9]. One such task is escorting visitors to meetings around one building throughout the day, giving tour information along the way [10]. For this task, CoBot remembers the locations it has passed and the information it shared with its visitor (a competency that would be expected by a person in the same situation). Later, when it passes the same location, CoBot recalls what it has already said throughout its tour to avoid repeating information. Instead, it provides more detailed information about each location each time it passes by the location, extending information that it knows it has already shared. For example, near the Robot Soccer Lab, it might introduce the location and the researchers, later share motivation for the work, and finally talk about recent research results.

**The Personal Assistant** at Microsoft Research is a prototype developed as part of a study in physically situated spoken dialog [11]. The system serves outside the door of its owner as an automated secretary to help visitors and the owner with calendar and scheduling. The Assistant engages people in dialog to identify good times for dropping by, schedules new meetings, relays messages from visitors, and sends information about its encounters to the owner. When visitors wait for the owner (after engaging) or leave and return to the area and then re-engage the system on their meeting goals, they expect shared memories of their recent encounters. Given the system's competencies in dialog and inference, lack of short- and long-term memory would be a jarring and salient cognitive deficiency.

To make re-engagements more natural and effective, we implemented execution memory. In addition to dialog, the Assistant uses face recognition and tracking to remember the salient features of times when visitors come and go from its proximity. The system considers the durations of waiting and whether visitors are on the owner's schedule. If a visitor moves away after engaging with the Assistant to wait for the owner to arrive or become available, the system archives their presence and engagement. If the visitor reappears later, the system caches previous recent engagements recognizes that they are

likely part of the same task of finding the owner, and shares its common ground about the period of the waiting and persisting goals. At re-engagement, the system recalls this data and uses it to modify its dialog to provide revised information about the owner's status, considering explicitly the updated state and goals of the visitor. Finally, unlike most dialog systems that only remember information for the current task, the Assistant also uses cached presence information to inform new non-merged tasks over longer time frames to share and ground its long-term memories with the visitor. When a person walks up, the system caches the last known presence and recalls the memory of a more distant collaboration, saying, *e.g.*, "Nice to see you again" or "Long-time no-see," depending on the recency of the last interaction.

### IV. CONCLUSION

As robots such the Assistant, CoBot, and mObi are introduced into human environments, people will expect them to remember shared events and encounters experienced during execution. We see opportunities for providing robotic systems with knowledge about how people selectively encode and recall events into their memories and for leveraging such shared memories in grounding conversations and activities. Encoding and harnessing memories shared with human collaborators about prior joint experiences can enable robots to refer to those experiences more naturally as well as to enrich competencies for collaborating over time in short, intra-session and longer-term cross-session engagements. We present execution memory as an important direction for research for robots that are in human environments long-term, splitting the problem into four distinct processes: salience, archiving, caching, and recall. We believe that each of these components warrants attention to understand what humans and robots find natural and useful to remember about activities and encounters.

### REFERENCES

- [1] H. H. Clark, M. F. Schober, J. M. Tanur, and J. M. Tanur, *Asking questions and influencing answers.*, 1992, pp. 15–48.
- [2] C. Chao, M. Cakmak, and A. L. Thomaz, "Transparent active learning for robots," in *HRI 2010*, 2010, pp. 317–324.
- [3] A. L. Thomaz and M. Cakmak, "Learning about objects with human teachers," in *HRI 2009*, 2009, pp. 15–22.
- [4] E. Kamar and E. Horvitz, "Jogger: Investigation of principles of context-sensitive reminding," in *AAMAS 2011*, 2011, pp. 1089–1090.
- [5] E. Horvitz, S. Dumais, and P. Koch, "Learning predictive models of memory landmarks," in *CogSci 2004*, 2004, pp. 1–6.
- [6] M. Ringel, E. Cutrell, S. Dumais, and E. Horvitz, "Milestones in time: The value of landmarks in retrieving information from personal stores," in *Interact 2003*, 2003, pp. 184–191.
- [7] B. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robotics and Autonomous Systems*, vol. 57, no. 5, pp. 469–483, 2009.
- [8] N. M. Bradburn, L. J. Rips, and S. K. Shevell, "Answering autobiographical questions: The impact of memory and inference on surveys," *Science*, vol. 236, pp. 157–162, 1987.
- [9] M. Veloso, J. Biswas, B. Coltin, S. Rosenthal, T. Kollar, C. Mericli, M. Samadi, and S. B. R. Ventura, "Cobots: Collaborative robots servicing multi-floor buildings," in *IROS 2012*, 2012.
- [10] S. Rosenthal and M. Veloso, "Mixed-initiative long-term interactions with an all-day-companion robot," in *AAAI Fall Symposium on Dialog on Robots*, 2010.
- [11] D. Bohus and E. Horvitz, "Dialog in the open world: Platform and applications," in *ICMI-MLMI 2009*, 2009, pp. 31–38.