Towards More Human-Like Episodic Memory for More Human-Like Agents

Cyril Brom[1] and Jiří Lukavský[2]

[1] Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic
[2] Institute of Psychology, Academy of Sciences, Prague, Czech Republic

1 Introduction

Episodic memory (EM) is an umbrella term for memory systems that operate with representations of personal history of an entity. The content of EM is related to particular places and moments, and connected to subjective feelings and current goals.

Recently, it has been argued that EM is one of the key components contributing to believability of intelligent virtual agents (IVAs), at least when agents interact with humans for more than a couple of minutes, because it allows the user to understand better the agent's history, personality, and internal state: both actual state and past state [e.g. 2, 5]. Technically, the EM is merely a data structure for loss compression of the flow of external events. The EM cannot be implemented as a pure log/video, because these are bad data structures (including human level). Why are they bad? First, they produce too large data. Second, they are not well organised with respect to future possible queries: neither a log nor a video have appropriate indexes in terms of database systems. A better approach is needed.

2 Overview of Our Model

We have started to pursue a systematic research on episodic memory for IVAs, beginning with investigating a single episodic memory module, what we now call the *core*, and continuing with widening the core by adding new subsystems that modulate the core and/or other subsystems. The core has been already presented [3]. It is a hierarchically organised memory for complex events (e.g. cooking a dinner) that can happen in large environments (house), and the IVA with the core can live in its world over long intervals (weeks). The memory features a gradual forgetting mechanism. For example, the IVA can originally remember that it was cooking a goulash yesterday morning, including all subtasks, but later forget the subtasks, keeping only the high-level information about the cooking. The forgetting is based on the age of episodes and their emotional salience. Thanks to the core, the agent can express itself, i.e. to answer questions such as: "What did you do yesterday between t_1 and t_2?".

The core has been extended with several memory appendages recently. First, the IVA can now express itself in a given number of sentences: "What did you do yesterday between t_1 and t_2? Summarise in 3 sentences." Second, instead of using exact time for dating episodes, which is not plausible, the memory now can learn human-like notions of time such as "after breakfast," "late afternoon," "when tired after lunch" etc. based on the history of the agent's interaction [4]. This part of the memory also

Zs. Ruttkay et al. (Eds.): IVA 2009, LNAI 5773, pp. 484–485, 2009.

allows the agent to adapt to temporal time shifts like a change of time zone. Additionally, thanks to this component, similar episodes can be blended in some situations; e.g., an agent who was watering a garden every evening cannot recall details of any particular watering episode when cued by time, but it can recall both that it was watering evenings last week and some detail of each watering when cued with time *plus* other details of the episode. Third, the memory has now a new module for representing long-term information about possible objects' locations in a familiar environment. This module is presented elsewhere [1], suffice it to say here that the module solves the issue of plausible estimation of possible objects locations, of those objects that are passive but can be moved by external forces beyond the agent's capabilities (e.g. a pen is moved by a fellow agent and put elsewhere). Finally, we have been implementing a module for plausible subjective representations of proximal space; basically intermediate-term allocentric and egocentric representations of locations of objects in the IVA's immediate surrounding. Our preliminary results suggest that our model will be able to organise some psychological data [e.g. 7, 6].

While the memory core and the timing module were investigated in a 3D world, other modules have been tested only in a 2D world. The most notable future works include porting all the modules to the 3D world, development a mechanism for emergence of false memories, and development a component for reconstructing high-level goals of a human user from the flow of the user's atomic actions. Without it, the memory is not able to represent efficiently what human users do in virtual worlds.

Acknowledgement. This work was partially supported by the Program "Information Society" under project 1ET100300517, by the Ministry of Education of the Czech Republic (Res. Project MSM0021620838), and by GA UK 21809. We thank to students working on this project: T. Soukup, J. Vyhnánek, J. Kotrla, and R. Kadlec.

References

1. Brom, C., Korenko, T., Lukavský, J.: How Place and Objects Combine? "What-Where" Memory for Human-like Agents. In: Ruttkay, Z., et al. (eds.) IVA 2009. LNCS (LNAI), vol. 5773, Springer, Heidelberg (2009)
2. Brom, C., Lukavsky, J.: Towards Virtual Characters with a Full Episodic Memory II: The Episodic Memory Strikes Back. In: Proc. Empathic Agents, AAMAS workshop, pp. 1–9 (2009)
3. Brom, C., Pešková, K., Lukavský, J.: What does your actor remember? Towards characters with a full episodic memory. In: Cavazza, M., Donikian, S. (eds.) ICVS-VirtStory 2007. LNCS, vol. 4871, pp. 89–101. Springer, Heidelberg (2007)
4. Burkert, O.: Connectionist Model of Episodic Memory for Virtual Humans. Master thesis. Dept. Software & Comp. Sci. Education. Charles University in Prague (2009)
5. Castellano, G., Aylett, R., Dautenhahn, K., Paiva, A., McOwan, P.W., Ho, S.: Long-term affect sensitive and socially interactive companions. In: 4th Int. Workshop on Human-Computer Convers (2008)
6. Holmes, M.C., Sholl, M.J.: Allocentric coding of object-to-object relations in overlearned and novel environments. Jn. Exp. Psychol.: General 119, 1069–1087 (2005)
7. Waller, D., Hodgson, E.: Transient and Enduring Spatial Representations Under Disorientation and Self-Rotation. Jn. of Exp. Psychol.: Learning, Memory and Cognition 32(4), 867–882 (2006)