CEUR Workshop Proceedings

http://ceur-ws.org

ISSN 1613-0073

Vol-2287

urn:nbn:de:0074-2287-2

Copyright © 2018 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

TOCAIS 2019
Towards Conscious AI Systems

Papers of the 2019 Towards Conscious AI Systems Symposium co-located with the Association for the Advancement of Artificial Intelligence 2019 Spring Symposium Series (AAAI SSS-19)

Stanford, CA, March 25-27, 2019.

Edited by

Antonio Chella *
David Gamez **
Patrick Lincoln ***
Riccardo Manzotti ****
Jonathan Pfautz *****

* University of Palermo, Department of Engineering, Palermo, Italy, & ICAR-CNR, Palermo, Italy
** Middlesex University, Computer Science Department, London, UK
*** SRI International, Computer Science Laboratory, Menlo Park, CA, USA
**** IULM University, Department of Business, Law, Economics and Consumer Behaviour, Milan, Italy
***** DARPA, USA

Table of Contents

Short and Position Papers

Regular Papers

2018-12-20: submitted by Antonio Chella, metadata incl. bibliographic data published under Creative Commons CC0

2018-12-22: published on CEUR-WS.org |valid HTML5|

Inner Speech for a Self-Conscious Robot

A. Pipitone[1][0000−0003−2388−5887], F. Lanza[1][0000−0003−4382−6366],
V. Seidita[1,2][0000−0002−0601−6914], and A. Chella[1,2][0000−0002−8625−708X]

[1] Dept. of Industrial and Digital Innovation (DIID), University of Palermo, Italy
[2] C.N.R., Institute for High-Performance Computing and Networking (ICAR),
Palermo, Italy

Abstract. The experience self-conscious thinking in the verbose form of inner speech is a common one. Such a covert dialogue accompanies the introspection of mental life and fulfills important roles in our cognition, such as self-regulation, self-restructuring, and re-focusing on attentional resources. Although the functional underpinning and the phenomenology of inner speech are largely investigated in psychological and philosophical fields, robotic research generally does not address such a form of self-conscious behavior. Existing models of inner speech inspire computational tools to provide the robot with a form of self-consciousness. Here, the most widespread psychological models of inner speech are reviewed, and a robot architecture implementing such a capability is outlined.

Keywords: Inner Speech · Cognitive Architecture · Robot Self-Consciousness · Robot Thought.

CEUR-WS.org/Vol-2287/paper14.pdf

1 Introduction

Inner speech plays a central role in daily life. A person thinks over her mental states, perspectives, emotions and external events by generating thoughts in the form of linguistics sentences. Talking to herself enables the person to pay attention to internal and external resources, to control and regulate her behavior, to retrieve memorized facts, to learn and store new information and, in general, to simplify otherwise demanding cognitive processes [1].

Moreover, inner speech allows restructuring the perception of the external world and the perception of self by enabling high-level cognition, including self-control, self-attention, and self-regulation.

Even if second-order thoughts may not need language but, for example, images or sensations, Bermudez [3], Jackendoff [5], among others, argue that genuine conscious thoughts need language. In the light of the above considerations, inner speech is an essential ingredient in the design of a self-conscious robot.

We model such a necessary capability within a cognitive architecture for robot self-consciousness by considering the underlying cognitive processes and components of inner speech.

It should be remarked that in the present paper such processes are taken into account independently from the origin of the linguistics abilities which are supposed acquired by the robot.

In Section 2 we show a brief overview of the cognitive models underlying the proposed robot architecture, which is detailed in Section 3. Conclusions and future works about the proposed robot architecture are discussed in Section 4.

2 Models of inner speech

Inner speech cannot be directly observed, thus reducing the scope for empirical studies. However, theoretical perspectives were developed during the last decades, and some of them are recognized in different research communities.

Vygotsky [10] conceives inner speech as the outcome of a developmental process during which the linguistics interactions, such as between a child and a caregiver, are *internalized*. The linguistically mediated explanation for solving a task thus becomes an internalized conversation with the self, when the learner is engaged in the same o similar cognitive tasks.

Morin [7][8] claims that inner speech is intrinsically linked to self-awareness. Self-focusing on an internal resource triggers the inner speech, and then it generates self-awareness about such a resource. Typical sources for the self-focus process are social interactions or mirror reflections by physical objects.

Baddeley [2] discussed the roles of rehearsal and working memory, where the different modules in the working memory are responsible for inner speech rehearsal. In particular, the *central executive* oversees the process; the *phonological loop* deals with spoken and written data, and the *visuospatial sketchpad* deals with information in a visual or spatial form. The phonological loop is composed of the *phonological store* for speech perception, which keeps information in a speech-based form for a very short time (1-2 seconds), and of the *articulatory control process* for speech production, that rehearses and stores verbal information from the phonological store.

Inner speech is usually conceived as the back-propagation of produced sentences to an inner ear: thus, a person rehears the internal voice she delivers. Steels [9] argued that the language re-entrance allows refining the syntax emerging during linguistic interactions within a population of agents. The syntax thus becomes more complex and complete by parsing previously produced utterances by the same agent.

In the same line, Clowes [4] discussed an artificial agent implemented by a recurrent neural network whose output nodes are words interpreted as possible actions (for example ‘up,’ ‘left,’ ‘right,’ ‘grab’). When such words are re-entrant by back-propagating the output to the input nodes, then the agent achieved the task in far fewer generations than in the control condition where words are not re-entrant.

3 The cognitive architecture for inner speech

Figure 1 shows the proposed robot cognitive architecture for inner speech. Such a representation refers to the Standard Model of Mind proposed by Laird et al.[6]. Here, the structure and processing of the Standard Model are decomposed with the aims to integrate the components and the processes defined by the inner speech theories previously discussed.

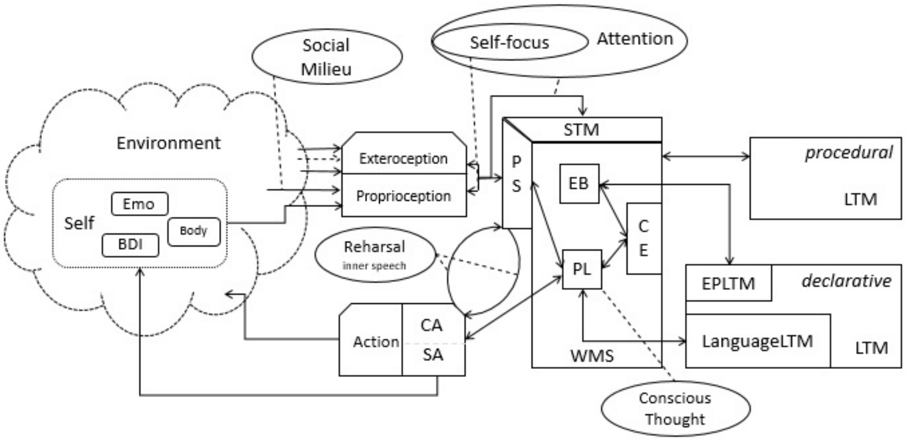

Fig. 1. The proposed cognitive architecture for inner speech.

3.1 Perception and Action

The perception of the proposed architecture includes the *proprioception* module related to the self-perception of the emotions (Emo), the belief, desires and intentions (BDI) and the robot body (Body), and the *exteroception* module related to the perception of the outside environment.

The proprioception module, according to Morin [7], is also stimulated by the *social milieu* which, in the considered perspective, includes the social interactions of the robot with the others entities in the environment, as physical objects like the mirrors and the cameras and others robots or humans, by means face-to-face interaction that foster self-world differentiation.

The motor module is decomposed in three sub-components: the *Action* module, the *Covert Articulation* module (CA) and the *Self Action* module (SA). In particular:

- The *Action* module represents the actions the agent performs on the outside world producing modifications to the external environment (not including the self) and the working memory.
- The *Covert Articulation* (CA) module rehearses information from the *Phonological Store* (PS), i.e., the perceptual buffer for speech-based data considered as a sub-component of the short-term memory (see below). Such a module

acts as the inner voice heard by the phonological store by rounding information in a loop. In this way, the inner speech links the covert articulation to the phonological store in a round loop.
- The *Self Action* (SA) module represents the actions that the agent performs on itself, i.e., self-regulation, self-focusing, and self-analysis.

3.2 The Memory System

The memory structure, inspired by the Standard Model of the Mind is divided into three types of memories: the short-term memory (STM), the *procedural* and the *declarative* long-term memory (LTM), and the working memory system (WMS).

The short-term memory holds sensory information on the environment in which the robot is plunged that were previously coded and integrated with information coming by perception. As previously mentioned, the short-term memory includes the phonological store.

Information flow from perception to STM allows storing the aforementioned coded signals. In particular, information from perception to the phonological store is related to *conscious* thoughts from exteroception, and to *self-conscious* thoughts from proprioception.

The information flow from the working memory system to perception provides expectations or possible hypotheses that are employed for influencing the *attention* process. In particular, the flow from the phonological store to proprioception enables the *self-focus* modality.

The long-term memory holds learned behaviors, semantic knowledge, and experience. In the considered case, the *declarative* LTM contains the linguistics information in terms of lexicon and grammatical structures, i.e., the *LanguageLTM* memory. The declarative linguistics information is assumed acquired, as specified above, and represent the *grammar* of the robot. Moreover, the *Episodic Long-Term Memory* (EBLTM) is the declarative long-term memory component which communicates to the *Episodic Buffer* (EB) within the working memory system, that acts as a 'backup' store of long-term memory data.

The *procedural* LTM contains the composition rules according to which the linguistic structures are arranged for producing sentences at different levels of completeness and complexity. A procedure does not concern the grammatical plausibility of the structures only. Other rules concerning the regulation, the focusing and the restructuring of resources within the whole environment (including the self) are to be considered.

Finally, the working memory system holds task-specific information 'chunks' and streamlines them to the cognitive processes during the task execution, step by step according to the cognitive cycle of the Standard Model of the Mind. The working memory system deals with cognitive tasks such as mental arithmetic and problem-solving. The *Central Executive* (CE) sub-component manages and controls the linguistic information of the rehearsal loop by the integrating (i.e., combining) data from the phonological loop and also drawing on data held in the long-term memory.

3.3 The Cognitive Cycle

In brief, a cognitive cycle starts with the perception that converts external signals in linguistics data and holds them into the phonological store. The central executive manages the inner thinking process by enabling the working memory system to selectively attend to some stimuli or ignore others, according to the rules stored within the LTMs, and by orchestrating the phonological loop as a slave system.

At this stage, a conscious thought emerges as a result of a single round between the phonological store and the covert articulation triggered by the phonological loop, once the central executive has retrieved the data for the process. The phonological loop enables the covert articulation which acts as a motor for the internal production, and whose output stream is heard to the phonological store. The output stream also affects the self which is then regulated and restructured.

Once the conscious thought is elicited by inner speech, the perception of the new context could take place, repeating the cognitive cycle.

4 Conclusions

In this paper, a cognitive architecture for inner speech cognition is presented. It is based on the Standard Model of Mind which was decomposed for including some typical components of the inner speech's models for human beings.

The working memory system of the architecture includes the *phonological loop* considered by Baddeley as the main component for storing spoken and written information and for implementing the cognitive rehearsal process.

The covert dialogue is modeled as a loop in which the *phonological store* hears the inner voice produced by the *covert articulator* process. The *central executive* is the master system which drives the whole system.

By retrieving linguistic information from the long-term memory, the central executive contributes to creating the linguistic thought whose surface form emerges by the phonological loop.

Acknowledgments

This material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-17-1-0232.

References

1. Alderson-Day, B., Fernyhough, C.: Inner Speech: Development, Cognitive Functions, Phenomenology, and Neurobiology. Psychological Bulletin **141**(5), 931–965 (2015)
2. Baddeley, A.: Working Memory. Science **255**(5044), 556–559 (1992)
3. Bermudez, J.L.: The Paradox of Self-Consciousness. MIT Press, Cambridge, MA (1998).

4. Clowes, R.: A Self-Regulation Model of Inner Speech and its Role in the Organisation of Human Conscious Experience. Journal of Consciousness Studies **14**(7), 59–71 (2007)
5. Jackendoff, R.: How Language Helps Us Think. Pragmatics & Cognition **4**(1), 1–34 (1996)
6. Laird, J.E., Lebiere, C., Rosenbloom, P.S.: A Standard Model of the Mind: Toward a Common Computational Framework across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics. AI Magazine *Winter 2017*, 13–26 (2017)
7. Morin, A.: A Neurocognitive and Socioecological Model of Self-Awareness. Genetic, Social, and General Psychology Monographs **130**(3), 197–222 (2004)
8. Morin, A.: Possible Links Between Self-Awareness and Inner Speech. Journal of Consciousness Studies **12**(4-5), 115–134 (2005)
9. Steels, L.: Language Re-Entrance and the 'Inner Voice.' Journal of Consciousness Studies **10**(4-5), 173–185 (2003)
10. Vygotsky, L.: Thought and Language. Revised and expanded edition. MIT Press, Cambridge, MA (2012)