A Cognitive Architecture for Inner Speech

Antonio Chella[a,b], Arianna Pipitone[a,*]

[a]*Dipartimento di Ingegneria - Universitá degli Studi di Palermo*
[b]*C.N.R., Istituto di Calcolo e Reti ad Alte Prestazioni, Palermo, Italy*

Abstract

A cognitive architecture for inner speech is presented. It is based on the Standard Model of Mind, integrated with modules for self-talking processes. Briefly, the working memory of the proposed architecture includes the phonological loop as a component which manages the exchanging information between the phonological store and the articulatory control system. The inner dialogue is modeled as a loop where the phonological store hears the inner voice produced by the hidden articulator process. A central executive module drives the whole system, and contributes to the generation of conscious thoughts by retrieving information from long-term memory. The surface form of thoughts thus emerges by the phonological loop. Once a conscious thought is elicited by inner speech, the perception of new context takes place and then repeating the cognitive loop. A preliminary formalization by event calculus of some of the described processes, and early results of their implementation on the humanoid robot Pepper by SoftBank Robotics are discussed.

Keywords: Inner speech, Cognitive Architecture

1. Introduction

Daily, human beings are engaged in a form of inner dialogue, which enables them to high-level cognition, including self-control, self-attention and self-regulation. By inner dialogue, a person plans tasks, finds problem's solution, self-reflects, critical thinks,

*Corresponding author
Email addresses: antonio.chella@unipa.it (Antonio Chella), arianna.pipitone@unipa.it (Arianna Pipitone)

feels emotions, and restructures the perception of the world and of himself. Obviously, the inner dialogue cannot be directly observed, thus making empirical studies difficult. However, psychological and philosophical perspectives were developed during the last decades, and are recognized in research communities.

Alderson-Day and Fernyhough [1] states that to talk to oneself makes a person able to retrieve memorized facts, learn new knowledge and, in general, to simplify otherwise demanding cognitive processes. According , the self-dialogue is closely related to thought, and therefore is an essential component in the dynamics of information thinking. In fact, Carruthers [2], Jackendoff [3], among many others, claim that genuine conscious thoughts need language. Vygotsky [4] considers inner language as the result of an internalization process during which linguistic explanations by a caregiver to a children become an inner conversation of the children with the self when he is engaged in similar task. Morin [5] states that inner dialogue can be linked to self-consciousness. Self-concentration on internal resources triggers inner speech and generates self-awareness on these resources.

Baddeley [9] described the inner speech phenomena by a working memory architecture, which is suitable for the automation of the process into artificial agent. In particular, Baddeley claimed that the inner voice is the re-entrance of a sentences to an inner ear covertly produced by an articulatory system. The inner ear and the articulatory system form the phonological loop. A central executive module oversees the whole processes; the phonological cycle deals with spoken and written data, and the visuospatial sketchpad deals with information in visual or spatial form. The linguistic information are deal by the phonological loop, where the phonological cycle tests and stores verbal information from the phonological store, which is a kind of a short term memory.

The Baddeley's model is the base of the cognitive architecture for inner speech developed at the RoboticsLab of the University of Palermo [7], that integrates the Standard Model of Mind proposed by Laird et al. [8] with the working memory by Baddeley and with the perception loop introduced in Chella and Macaluso [10].

The goal of this paper is to show such an architecture and an early formalization by event calculus of the underlying processes. The paper is organized as follow: a

brief overview of inner speech for artificial agents is presented at 2. The cognitive architecture with first automation is described at 3. A case of study on inner speech in robot is shown at 4. Finally, conclusions and future works are discussed at 5.

2. Inner speech for artificial agents

In literature, few works investigate the role of inner speech for artificial agents. Steels [11] argues that language re-entrance allows refining the syntax of a grammar emerging during oral interactions within a population of agents. The syntax becomes more complex and complete by the parsing of previously produced utterances by the same agent. In the same line, Clowes et al. [12] discusses the effect of words back-propagation in a recurrent neural network. The output nodes are words interpreted as possible actions to take. When such words are re-entrant by back-propagating them to a specific input node, the selection of the plausible action for a task is more correct than the case without back-propagation.

The cited works demonstrated the positive effects for the artificial agents of re-parsing the produced linguistic utterances, but they do not investigate the underlying motivations. Why words back-propagation produces better results is unknow.

A preliminary study about the proposed cognitive architecture for inner speech was discussed at [7], where the modules of the Baddeley's phonological loop are integrated into the Standard Model of Mind by Laird et al.[8], and principles by Morin related to the inner speech triggering and self-consciousness are modeled too. Some of the authors suggested to integrate such an architecture into the IDyOT system [13].

By formalizing the processes of the architecture through event calculus, we attempt to identify the functions at the basis of inner speech phenomena. Thus we can implement them into the artificial agents, while observing and motivating the obtained results.

3. The proposed architecture

Figure 1 shows the proposed cognitive architecture for inner speech. The structure and processes of the Standard Model of Mind are further decomposed with the aims

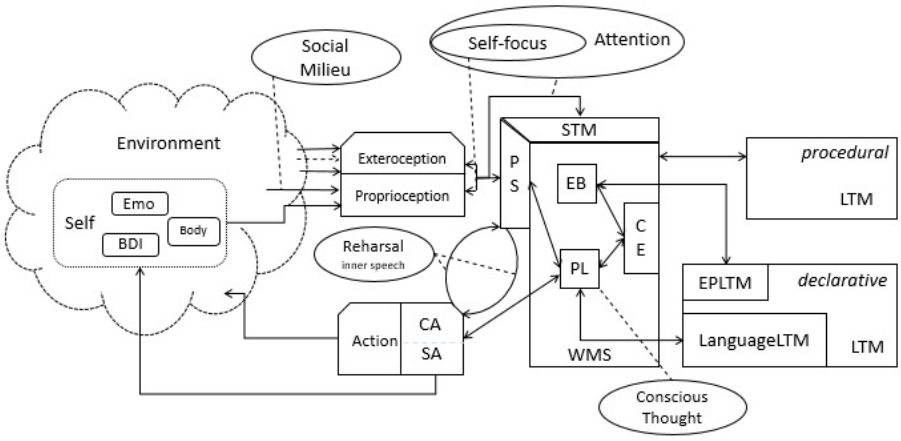

Figure 1: The proposed cognitive architecture for inner speech.

to integrate the components and the processes defined by the inner speech theories previously presented.

3.1. Perception and Motor Modules

The perception module of the proposed architecture includes two sub-modules, that are:

- the *proprioception* module which is related to the perception of the self, with emotions (Emo), belief, desires, intentions (BDI), and the physical body (Body) including all physical components of the agent;
- the *exteroception* module related to the perception of the entities into the environment, not including the self.

According to Morin [5], the proprioception is triggered by the *social milieu* (what the other tell about me), by self-reflection stimuli (mirrors, videocamera, etc...), and by the social interactions, as face-to-face interaction that foster self-world differentiation.

The motor module includes three sub-components:

- the *Action* module, which acts on the outside world producing modifications to the environment (not including the self) and the working memory;

- *Self Action* module (SA), which represents the actions that the agent takes on itself, i.e., self-regulation, self-focusing, and self-analysis;
- the *Covert Articulator* module (CA), which emulates the articulatory system by Baddeley, but in this case it is related to the silent articulation of sentences, i.e. it produces the inner voice, then rehearsed by the phonological store, thus modeling the phonological loop.

3.1.1. Automating perception and motor modules

To automate the perception of a fact from the whole environment (including the self), we define a set of modal operators from event calculus [14].

A fact is represented by a proposition ϕ. As consequence, we consider propositions that model facts related to the self (ϕ_{self}) and propositions that model facts of the domain, not including the self (ϕ_d). For example, the proposition $\phi_{self} = holds(battery(low), t)$ means that the battery of the robot is low at t. The typical $holds$ function of the event calculus returns the boolean value specifying such a condition, thus ϕ_{self} is a fact regarding the self. The $\phi_d = holds(on(apple, table), t)$ states that an apple is on a table, and regards a fact domain.

The modal operator **P** models the perception of a fact, that is $\mathbf{P}(\phi)$ means that the robot is perceiving the fact ϕ, that could be about the self (and in this case $\phi = \phi_{self}$) or about the domain (and in this case $\phi = \phi_d$).

The modal operator **A** defines the execution of a particular action whose effect on the environment is the proposition ϕ. As consequence, $\mathbf{A}(\phi)$ means that the robot is taking an action whose effect is the true condition of ϕ (which can be $\phi = \phi_{self}$ or $\phi = \phi_d$).

The standard intensional operators for belief **B**, desire **D**, intention **I** are included too, and have the same semantics of the previous ones.

All the modal operators are true conditions, and they are in turn propositions.

3.2. The Memory Structure

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 from the environment that were suitable coded by perception module. In the proposed architecture, it includes the *phonological store* (PS), that emulates the inner ear for inner voice. Information flow from perception to STM allows storing these coded signals. In particular, information from perception to the PS is related to *conscious* thoughts when they come from exteroception, and to *self-conscious* thoughts when they come from proprioception.

The reverse information flow from STM to perception provides expectations or possible hypotheses that are employed for influencing the *attention* process. In particular, the flow from the PS to proprioception enables the *self-focus* modality.

The long-term memory stores learned behaviors, knowledge, and experience. In our model, beyond these typical contents of the Standard Model of Mind, it includes:

- in the *declarative* LTM:
  - the *LanguageLTM* memory which contains the **linguistics data** including lexicon and grammatical structures;
  - the *Episodic Long-Term Memory* (EPLTM), which is the declarative long-term memory component which communicates to the *Episodic Buffer* (EB) within the working memory system, and acts as a 'backup' store of long-term memory data;
- in the *procedural* LTM, the composition rules according to which the linguistic structures are arranged for producing sentences at different levels of completeness and complexity.

Finally, the working memory system includes the *Central Executive* (CE) sub component which manages and controls the linguistic information of the rehearsal loop by the integrating (i.e., combining) data from the phonological store and also drawing on data held in the long-term memory. The working memory system deals with cognitive tasks such as mental arithmetic and problem-solving.

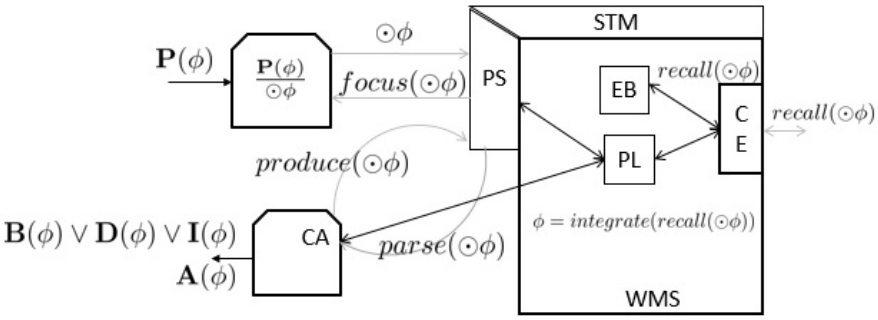

Figure 2: An excerpt of the whole architecture with details about the rehearsing process.

3.2.1. Automating the cognitive cycle

Figure 2 shows a detail of the architecture; it highlights the cognitive cycle of inner speech, and the corresponding functions we define for implementing the rehearsing process.

A cognitive cycle starts with the perception and the conversion of external signals in linguistics data, which are stored (as heard) into the phonological store. The perception of the fact ϕ is formalized by the modal operator $\mathbf{P}(\phi)$ inputted into the perception module. The inference schemata:

$$\frac{\mathbf{P}(\phi)}{\odot\phi}$$

automates the signal codification into linguistics form. In particular, the $\odot$ operator, applied to a proposition, returns the linguistic representation of that proposition, which is the set of words (that are not native of the event calculus) the proposition contains. For example, for the proposition $\phi_{self} = holds(battery(low), t)$, the set of not native words is $\odot\phi_{self} = \{battery, low\}$, begin $holds$ and t typical event calculus expressions. The inference schemata represented into the perception module formalizes the association of this linguistic form to each perceived fact.

The information flow from perception module to the phonological store is formalized by inputting $\odot\phi$ to PS, and it represents the information storing. The reverse flow is automated by the function $focus(\odot\phi)$, and allows to retrieve correlated fact to ϕ by perception.

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.

Formally, four steps implement the rehearsing process:

1. the working memory system receives $\odot\phi$;
2. the central executive recalls data from long and short term memories (including the episodic buffer) by the function $recall(\odot\phi)$;
3. the phonological loop integrates $\odot\phi$ with new retrieved information by the function $integrate(recall(\odot\phi))$;
4. the phonological loop triggers inner speech by producing and parsing the new extended form of $\odot\phi$; during production by function $produce(\odot\phi)$ the covert articulator inputs the new form to the phonological store, which perceives this stimuli by the function $parse(\odot\phi)$. Corresponding to this perception, the central executive activates new focus for further perceptions and evaluations, and the cycle restarts.

The integration of new contents into ϕ will change the environment in the form of new beliefs, desires and intentions of the agent, and also in term of a reactive action to take. In particular:

- by function $produce(\odot\phi)$ new beliefs, desires or intentions could emerge, affecting the environment;
- by the function $parse(\odot\phi)$ new focus (or perception in general), and/or new beliefs, desires and intentions may emerge, generating new ϕ.

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

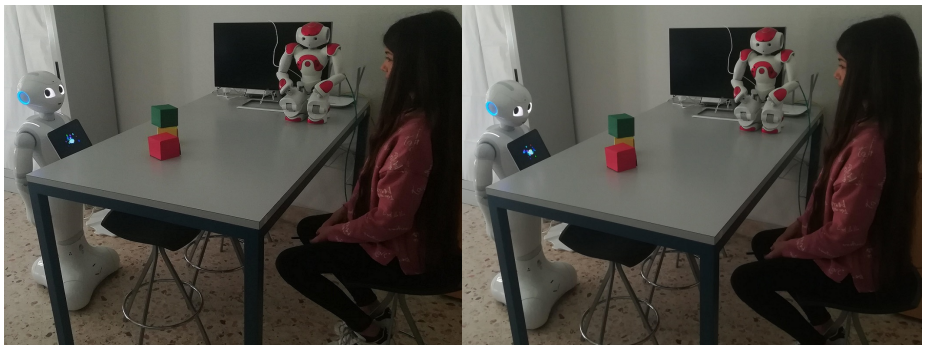

Figure 3: The robot perceives the query by user (i.e. *"Where is the green box"*) and then it activates the cognitive cycle for reasoning about the positions of the boxes.

4. Implementation and results

An early implementation of the proposed architecture on the Pepper Robot allows us to estimate preliminary results in a simple scenario.

Figure 3 shows the experimental session we conducted. A set of boxes with different colors are on the table in front of the robot. The user asks to the robot where is the green box. The robot is engaged in describing the box in respect to the other ones. By the inner speech the robot queries itself to retrieve useful information that allow it to answer to our request.

We model the linguistic knowledge of the robot by associating to the question words $where, who, when$ the typical adverbs and prepositions used for answering to these questions. For example, for the question word $where$, the typical adverbs and prepositions are: *to, from, on, left, right, up, down*. In the same way, for the question word $when$, typical adverbs and prepositions which are used for answering are: *since, at, from*. We manually annotate question words, thus build the LanguageLTM for this scenario.

The followed facts formalize the states of the boxes in the environment, that is:

- $(box \wedge red)$ means that there is a red box;
- $(box \wedge yellow)$ means that there is a yellow box;
- $(box \wedge green)$ means that there is a green box;

- $(on\ (box \wedge green)(box \wedge yellow))$ means that the green box is on the yellow one;
- $(right\ (box \wedge green)(box \wedge red))$ means that the green box is to the rigth of the red one;
- $(right\ (box \wedge yellow)(box \wedge red))$ means that the yellow box is to the left of the red one.

At start time the robot believes each fact of the environment, that is: $\mathbf{B}(box \wedge red)$, $\mathbf{B}(box \wedge yellow)$, $\mathbf{B}((box \wedge green))$, $\mathbf{B}(on\ (box \wedge green)(box \wedge yellow))$, $\mathbf{B}(right\ (box \wedge green)(box \wedge red))$, and $\mathbf{B}(right\ (box \wedge yellow)(box \wedge red))$. These beliefs are stored in the episodic buffer, because they represent short term memory related to the actual context (the episodical memory).

When the query specifying the task (i.e. *"Where is the green box?"*) is perceived by the robot by its speech recognition routines, it formalizes such a query by $\mathbf{P}(where \wedge box \wedge green)$. The linguistic form of ϕ is $\odot\phi = \{where, green, box\}$.

From the linguistic knowledge in the long term memory, the central executive retrieves the set of linguistic rules corresponding to that perception. At this time, the recall function performs a simple string matching, and returns from memories the information which match to one of the word in the linguistic form $\odot\phi$.

For the perceived query, the recall function returns from the LanguageLTM the set of annotated words corresponding to the question word $where$ (this word matches to the word $where$ in $\odot\phi$), that are: {*on, left, right, up, down*}. These information are integrated by phonological loop and then produced by the covert articulator in the form of $\odot\phi = \{\{on, left, right, up, down\}, green, box\}$, which becomes the new information flow inputted into the phonological store.

By parsing such a new proposition, the central executive retrieves by string matching from the episodic buffer the corresponding beliefs which allow to answer to the query. In particular, the results of this cycle are the retrieved beliefs $(on\ (box \wedge green)(box \wedge yellow))$ and $(right\ (box \wedge green)(box \wedge red))$ because they match to the words on and $right$ in the integrated form $\odot\phi$. These beliefs are in turn in-

tegrated in $\odot\phi$ generating a new integrated form of type $\{(on\ (box \wedge green)(box \wedge yellow)), (right\ (box \wedge green)(box \wedge red)), green, box\}$.

The corresponding propositions are re-produced and re-hearsed by phonological store. Considering that the central executive does not retrieve further new information, the phonological loop will not restart a new cycle, and the process ends with the overt articulation of the conjunction of the last propositions by the speech production routines. As result, the robot correctly answers to the query. The goal was reached by a form of inner dialogue enabled by the proposed architecture; the approach is general and produces same results for any kinds of objects with different properties (shape, dimension), by adding the corresponding facts and beliefs.

5. Conclusions

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

The working memory system of the architecture includes the *phonological loop* as component for storing spoken and written information, and for managing the rehearsal process.

The inner speech 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 these components that act as slave systems.

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. Also, the central executive retrieves related facts to the perception from the other memories, as the episodic buffer (that is a new component defining the episodical memory).

A preliminary event calculus allows to automatize some of the defined processes, and it enables us to implement an early inner dialogue into the Pepper Robot; the robot was engaged in the simple task to describe an object in respect to the others in the same context, and by a form of inner dialogue it correctly answers to the request.

Future works regard the extension of the architecture for enabling high-level cognition for robot, as planning, regulation, and consciousness.

Acknowledgments

This material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-19-1-7025.

References

[1] Alderson-Day B, Fernyhough C. Inner Speech: Development, Cognitive Functions, Phenomenology, and Neurobiology. Psychological Bulletin 2015;141(5):931-65.

[2] Carruthers P. Conscious Thinking: Language or Elimination? Mind & Language 1998;13(4):457-76.

[3] Jackendoff, R.: How Language Helps Us Think. Pragmatics & Cognition **4**(1), 1–34 (1996)

[4] Vygotsky, L.: Thought and Language. Revised and expanded edition. MIT Press, Cambridge, MA (2012)

[5] Morin, A.: Possible Links Between Self-Awareness and Inner Speech. Journal of Consciousness Studies **12**(4-5), 115–134 (2005)

[6] Baddeley, A.: Working Memory. Science **255**(5044), 556–559 (1992)

[7] Pipitone, A., Lanza, F., Seidita, V., Chella, A.: Inner Speech for a Self-Conscious Robot. Proc. of AAAI Spring Symposium on Towards Conscious AI Systems, CEUR-WS.org, (2019), online `http://ceur-ws.org/Vol-2287/paper14.pdf`.

[8] Laird, J.E., Lebiere, C., and 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, 13–26 (2017).

[9] Baddeley, A.: Working Memory. Science **255**(5044), 556–559 (1992)

[10] Chella, A., Macaluso, I.: The Perception Loop in CiceRobot, a Museum Guide Robot. *Neurocomputing*, 72, 760–766, (2009).

[11] Steels, L.: Language Re-Entrance and the 'Inner Voice.' Journal of Consciousness Studies **10**(4-5), 173–185 (2003)

[12] Clowes R, Morse AF. Scaffolding Cognition with Words. In: Berthouze L, Kaplan F, Kozima H, Yano H, Konczak J, Metta G, Nadel J, Sandini G, Stojanov G, Balkenius C, editors. Proceedings of the Fifth International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, vol. 123. Lund:Lund University Cognitive Studies; 2005.

[13] Chella A.; Pipitone A. (2019) The inner speech of the IDyOT: Comment on "Creativity, information, and consciousness: The information dynamics of thinking" by Geraint A. Wiggins. Phys Life Rev.

[14] Shanahan, M. (2000). The Event Calculus Explained. In Artificial Intelligence LNAI. 1600. 10.1007/3-540-48317-9_17