ARTICLE IN PRESS

Biologically Inspired Cognitive Architectures xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Biologically Inspired Cognitive Architectures

journal homepage: www.elsevier.com/locate/bica

Research article

Robots on stage: A cognitive framework for socially interacting robots

Igor Rodriguez*, Aitzol Astigarraga, Elena Lazkano, José María Martínez-Otzeta, Inigo Mendialdua

Department of Computer Sciences and Artificial Intelligence, University of the Basque Country UPV/EHU, Donostia-San Sebastian 20018, Spain

ARTICLE INFO

Keywords:
Social robotics
Affective perception
Cognitive architecture

ABSTRACT

This article is an attempt to characterize the cognitive skills involved in the development of socially interacting robots. We argue that performative arts, such as oral improvised poetry, can serve as a useful testbed for the development and evaluation of robots that interact with humans. The paper presents a speech-based humanoid poet-performer that can (1) listen to human commands and generate poems on demand; (2) perceive audience's feedback and react displaying the corresponding emotional response; and (3) generate natural gesticulation movements enriched with social signals depending on sentiment processing. We discuss each of the involved abilities, present working implementations and show how they are combined in an embodied cognitive architecture to achieve the fluent coordination and joint-action timing needed in live events.

Introduction

Social robotics (Breazeal, 2004) aims to provide robots with artificial social intelligence to improve human–machine interaction and to introduce them in complex human contexts. The demand for robot's sophisticate behaviors requires to model and implement human-like capabilities to sense, to process, and to act/interact naturally by taking into account emotions, intentions, motivations, and other related cognitive functions. And, of course, the ability to communicate through natural language and non-verbal signs is in the front line of research.

Nowadays, the development of control architectures for robots while taking into account the complexity of social human-robot interaction is a real challenge. It requires various cognitive features to be present: emotions, attention allocation, creativity and reactive and deliberative levels of perception and action.

Ever since the pioneering research on cognitive architectures (Newell, 1994), several architectures can be found in literature: SOAR (Laird, Kinkade, Mohan, & Xu, 2012), ACT-R (Anderson, 2005), CLARION (Sun, 2006), iCub (Vernon, Metta, & Sandini, 2007a) and ICARUS (Choi & Langley, 2018), among others. A good review of the literature can be found in (Vernon, Metta, & Sandini, 2007b; Thórisson & Helgasson, 2012; Langley, Laird, & Rogers, 2009). But, in spite of the numerous contributions in the field of cognitive architectures, robots that can listen to human speech, understand it, interact according to the conveyed meaning and respond still represent major research and technological challenges. Therefore, a research on different approaches to build control architectures oriented for interaction able to deal with cognitive capabilities such as emotion and social aspects of human-robot interaction (HRI) is highly useful.

In the last years research in the field of social robotics with conversational capabilities has grown up, and several robots have been designed and developed in this area. Such applications, most of them not concerned about being a faithful model of cognition, comprise several cognitive abilities and provide robust adaptive behaviour for human-robot interaction. Relevant works include: industrial and manufacturing robots (Cherubini, Passama, Crosnier, Lasnier, & Fraisse, 2016; Heyer, 2010); assistive robots and robots focused on aiding users with special needs (Bemelmans, Gelderblom, Jonker, & De Witte, 2012; Fasola & Mataric, 2012; Gómez Esteban et al., 2016; Kachouie, Sedighadeli, Khosla, & Chu, 2014; Luria, Hoffman, Megidish, Zuckerman, & Park, 2016; Tapus, Tapus, & Mataric, 2009); interactive teachers and educational assistants (Fridin & Belokopytov, 2014; Kanda, Shimada, & Koizumi, 2012); lab or household robotic assistants (Dautenhahn et al., 2005; Wisspeintner, Van Der Zant, Iocchi, & Schiffer, 2009); shopping mall guides (Chen et al., 2015); persuasive robots (Chidambaram, Chiang, & Mutlu, 2012; Lee & Liang, 2016); museum robots (Kanda, Arai, Suzuki, Kobayashi, & Kuno, 2014; Rashed, Suzuki, Lam, Kobayashi, & Kuno, 2015) and tour guides (Kanda et al., 2014); companion robots (Moyle et al., 2013); and robots more oriented to the entertainment area, such as robotic theater actors (Fernandez & Bonarini, 2013; Hoffman, 2011), musicians, storyteller robots (Bruce, Knight, Listopad, Magerko, & Nourbakhsh, 2000; Bae et al., 2012; Costa, Brunete, Bae, & Mavridis, 2016; Wu, Wang, Tay, & Wong, 2017) and dancers (Kosuge, Hayashi, Hirata, & Tobiyama,

* Corresponding author.
E-mail address: igor.rodriguez@ehu.eus (I. Rodriguez).

https://doi.org/10.1016/j.bica.2018.07.014
Received 12 June 2018; Received in revised form 11 July 2018; Accepted 11 July 2018
2212-683X/

2003). We could denominate this last group as performance robots, that is, robots close to performative arts that execute their task on a stage. An approach that brings the gap between cognitive models and social robotics is presented in Augello, Infantino, Pilato, Rizzo, and Vella (2015) and Augello et al. (2016). In those works authors propose a cognitive architecture for computational creativity, making a humanoid robot able to dance.

We believe that stage performance is valuable both as an implementation platform and as a testing ground for interaction-oriented cognitive architecture research. On the one hand, the event setting is constrained to some degree, limiting thus the perception and actuation possibilities of the robotic system. On the other hand, it provides a unique environment in which humans and robots collaborate incorporating dialog, sensory processing, action selection and behavior coordination.

Our robotic system, called Bertsobot, should be framed within performance robots: an autonomous robot that participates on live events, improvising poems under given constraints and performing them on stage. Thus, Bertsobot brings together capabilities and characteristics from many of the previously mentioned performance robots: theatrical staging, verbal and non-verbal communication, people detection and key stage elements perception, affect detection and emotional response, timing and coordination, etc.

In this article we present working implementations of the involved cognitive skills and show how they are combined to achieve the fluent coordination and joint-action timing needed in live events.

We do not claim to address here the issue as a whole. This article attempts however to organize it into a coherent challenge for social robotics, and to explain and illustrate some of the paths that we have investigated on our robots, which result in a robot architecture designed for human-robot interaction that implements cognitive skills.

Improvised poetry and *Bertsolaritza*

Writing poetry requires both creativity to construct a meaningful message and lyrical skills to produce rhyme patterns and follow metrical constraints. Furthermore, oral poetry, poetry constructed without the aid of writing (Lord, Mitchell, & Nagy, 2000), implies that a work has to be composed and performed at the moment, with no prior preparation. Nowadays many improvisational oral practices exists around the world, such as Serbo-Croatian guslars (Lord et al., 2000), freestyle rap (Pihel, 1996) and Basque *bertsolaritza* (Garzia, Sarasua, & Egaña, 2001).

Bertsolaritza, the art of improvising verses in *Euskara* (the language of the inhabitants of the Basque Country) is one of the manifestations of traditional Basque culture that is still very much alive. Events and competitions in which the verse-makers, *bertsolari*-s, have to produce impromptu compositions about topics or prompts are very common. A typical scenario involves an emcee suggesting a topic to the *bertsolari*, who must then, within the space of less than a minute, come up with a verse on that topic that must obey certain rules; in other words, it must fit in with a prescribed verse-form that also involves a rhyme scheme and a melody (chosen from among hundreds of tunes). And of course perform that verse, before an audience and without any musical accompaniment (see Fig. 1).

The Bertsobot cognitive architecture

The Bertsobot system endows the robots with some of the *bertsolari*-s' capabilities that allow to take part in public performances. Therefore, our Bertsobot system is able to perceive the feedback and emotions of the audience through their applause and react accordingly, as human oral improvisers do, modifying in real time the sentiment of the poem and its corporal expression accordingly. We focused on creating a practical cognitive architecture that follows the dynamics of real events, as verse-makers do:

Fig. 1. Typical scenario.

1. Wait sitting for its turn.
2. When demanded, place itself in front of the microphone and listen to the exercise proposed by the emcee.
3. Compose and sing the verse to the public.
4. Observe and receive audience's feedback and react accordingly.
5. Go back to its sitting place.

A robot capable of performing the aforementioned tasks involves the development of several cognitive capabilities. Although Bertsobot's main task is to compose verses, which requires a high cognition level, there are other important capabilities that at lower-level manage perceptions and representations of the environment. Specifically, it requires certain abilities to understand verbal instructions, move around the stage, recognize the different key elements of the scenario, interact with other agents and the audience, and show the same degree of expressiveness that *bertsolari*-s show on stage.

The cognitive architecture is the framework that facilitates us the development of cognitive functions, providing a structure within which to embed the mechanisms for perception and action, motivation and social interaction (Vernon, von Hofsten, & Fadiga, 2016). The cognitive description of our robotic system has been inspired by Augello et al. (2018). The framework is suitable to model aspects such as motivation and emotions that are integrated with perceptual and reasoning processes. Fig. 2 shows an overview of the proposed framework for Bertsobot.

In our framework, The Long Term Memory (LTM) stores all the knowledge required by the robot to accomplish the task. It contains the postures model with which the robot will be aware of its body configuration, the rules and corpora to generate extemporary poems, a set of melodies for singing composed verses, a gesture repertoire related to the expression capabilities of the robot, and a linguistic dictionary. On the other hand, the Short Term Memory (STM) or working memory stores temporal information about robot's and audience's emotional state. Drives comprises all the basic behaviors to interact with the environment and extract information from it. Finally, the Social Interaction module guides the human-robot interaction through verbal communication and body expression.

Our cognitive framework has been designed as the basis for the development of an adaptive robot for human-robot interaction, integrating a wide range of components in a scalable ROS[1] based control architecture. It is composed by different behaviours or modules that make the robot act in a consistent manner and resemble a real *bertsolari*.

In the following sections, we introduce in detail the main cognitive capabilities used in our robot performer.

[1] http://www.ros.org/.