Robotics and Autonomous Systems 63 (2015) 22–35

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

# A review of verbal and non-verbal human–robot interactive communication

Nikolaos Mavridis *

*Interactive Robots and Media Lab, NCSR Demokritos, GR-15310, Agia Paraskevi, Athens, Greece*

## HIGHLIGHTS

- An extensive survey of human–robot interactive communication is provided.
- Verbal as well as non-verbal aspects of human–robot interaction are covered.
- The paper starts with a historical introduction.
- Ten special desiderata that human–robot systems should fulfil are proposed.
- The desiderata are examined in detail, culminating to a forward-looking discussion.

## ARTICLE INFO

*Article history:*
Received 9 February 2014
Received in revised form
14 September 2014
Accepted 26 September 2014
Available online 13 October 2014

*Keywords:*
Verbal
Non-verbal
Human–robot interaction
Human–robot communication
Survey

## ABSTRACT

In this paper, an overview of human–robot interactive communication is presented, covering verbal as well as non-verbal aspects. Following a historical introduction, and motivation towards fluid human–robot communication, ten desiderata are proposed, which provide an organizational axis both of recent as well as of future research on human–robot communication. Then, the ten desiderata are examined in detail, culminating in a unifying discussion, and a forward-looking conclusion.



## 1. Introduction: historical overview

While the first modern-day industrial robot, Unimate, began work on the General Motors assembly line in 1961, and was conceived in 1954 by George Devol [1,2], the concept of a robot has a very long history, starting in mythology and folklore, and the first mechanical predecessors (automata) having been constructed in Ancient Times. For example, in Greek mythology, the God Hephaestus is reputed to have made mechanical servants from gold ([3] in p. 114, and [4] verse 18.419). Furthermore, a rich tradition of designing and building mechanical, pneumatic or hydraulic automata also exists: from the automata of Ancient Egyptian temples, to the mechanical pigeon of the Pythagorean Archytas of Tarantum circa 400 BC [5], to the accounts of earlier automata found in the Lie Zi text in China in 300 BC [6], to the devices of Heron of Alexandria [7] in the 1st century. The Islamic world also plays an important role in the development of automata; Al-Jazari, an Arab inventor, designed and constructed numerous automatic machines, and is even reputed to have devised the first programmable humanoid robot in 1206 AD [8]. The word "robot", a Slavic word meaning servitude, was first used in this context by the Czech author Karel Capek in 1921 [9].

However, regarding robots with natural-language conversational abilities, it was not until the 1990s that the first pioneering systems started to appear. Despite the long history of mythology and automata, and the fact that even the mythological handmaidens of Hephaestus were reputed to have been given a voice [3], and despite the fact that the first general-purpose electronic speech synthesizer was developed by Noriko Omeda in Japan in 1968 [10], it was not until the early 1990s that conversational robots such as MAIA [11], RHINO [12], and AESOP [13] appeared. These robots cover a range of intended application domains; for example, MAIA was intended to carry objects and deliver them, while RHINO is a museum guide robot, and AESOP a surgical robot.

In more detail, the early systems include Polly, a robotic guide that could give tours in offices [14,15]. Polly had very simple interaction capacities; it could perceive human feet waving a "tour

* Tel.: +30 2310824077.
*E-mail address:* nmav@alum.mit.edu.

http://dx.doi.org/10.1016/j.robot.2014.09.031

wanted'' signal, and then it would just use pre-determined phrases during the tour itself. A slightly more advanced system was TJ [16]. TJ could verbally respond to simple commands, such as ''go left'', albeit through a keyboard. RHINO, on the other hand [12], could respond to tour-start commands, but then, again, just offered a pre-programmed tour with fixed programmer-defined verbal descriptions. Regarding mobile assistant robots with conversational capabilities in the 1990s, a classic system is MAIA [11,17], obeying simple commands, and carrying objects around places, as well as the mobile office assistant which could not only deliver parcels but also guide visitors described in [18], and the similar in functionality Japanese-language robot Jijo-2 [19–21]. Finally, an important book from the period is [22], which is characteristic of the traditional natural-language semantics-inspired theoretical approaches to the problem of human–robot communication, and also of the great gap between the theoretical proposals and the actual implemented systems of this early decade.

What is common to all the above early systems is that they share a number of limitations. First, all of them only accept a fixed and small number of simple *canned commands*, and they respond with a set of *canned answers*. Second, the only *speech acts* (in the sense of Searle [23]) that they can handle are requests. Third, the dialogue they support is clearly not flexibly *mixed initiative*; in most cases it is just human-initiative. Four, they do not really support *situated language*, i.e. language about their physical situations and events that are happening around them; except for a fixed number of canned location names in a few cases. Five, they are not able to handle *affective speech*; i.e. emotion-carrying prosody is neither recognized nor generated. Six, their *non-verbal communication* [24] capabilities are almost non-existent; for example, gestures, gait, facial expressions, and head nods are neither recognized nor produced. And seventh, their dialogue systems are usually effectively stimulus–response or stimulus-state-response systems; i.e. no real *speech planning* or purposeful dialogue generation is taking place, and certainly not in conjunction with the motor planning subsystems of the robot. Last but quite importantly, no real *learning*, off-line or on-the-fly is taking place in these systems; verbal behaviours have to be prescribed.

All of these shortcomings of the early systems of the 1990s, effectively have become desiderata for the next two decades of research: the 2000s and 2010s, which we are in at the moment. Thus, in this paper, we will start by providing a discussion giving motivation to the need for existence of interactive robots with natural human–robot communication capabilities, and then we will enlist a number of desiderata for such systems, which have also effectively become areas of active research in the last decade. Then, we will examine these desiderata one by one, and discuss the research that has taken place towards their fulfilment. Special consideration will be given to the so-called ''symbol grounding problem'' [25], which is central to most endeavours towards natural language communication with physically embodied agents, such as robots. Finally, after a discussion of the most important open problems for the future, we will provide a concise conclusion.

## 2. Motivation: interactive robots with natural language capabilities but why?

There are at least two avenues towards answering this fundamental question, and both will be attempted here. The first avenue will attempt to start from first principles and derive a rationale towards equipping robots with natural language. The second, more traditional and safe avenue, will start from a concrete, yet partially transient, base: application domains existing or potential. In more detail:

Traditionally, there used to be a clear separation between design and deployment phases for robots. Application-specific robots (for example, manufacturing robots, such as [26]) were: (a) designed by expert designers, (b) possibly tailor-programmed and occasionally reprogrammed by specialist engineers at their installation site, and (c) interacted with their environment as well as with specialized operators during actual operation. However, not only the phenomenal simplicity but also the accompanying inflexibility and cost of this traditional setting is often changing nowadays. For example, one might want to have broader-domain and less application-specific robots, necessitating more generic designs, as well as less effort by the programmer-engineers on site, in order to cover the various contexts of operation. Even better, one might want to rely less on specialized operators, and to have robots interact and collaborate with non-expert humans with a little if any prior training. Ideally, even the actual traditional programming and re-programming might also be transferred over to non-expert humans; and instead of programming in a technical language, to be replaced by intuitive tuition by demonstration, imitation and explanation [27–29]. Learning by demonstration and imitation for robots already has quite some active research; but most examples only cover motor and aspects of learning, and language and communication is not involved deeply.

And this is exactly where natural language and other forms of fluid and natural human–robot communication enter the picture: Unspecialized non-expert humans are used to (and quite good at) teaching and interacting with other humans through a mixture of natural language as well as nonverbal signs. Thus, it makes sense to capitalize on this existing ability of non-expert humans by building robots that do not require humans to adapt to them in a special way, and which can fluidly collaborate with other humans, interacting with them and being taught by them in a natural manner, almost as if they were other humans themselves.

Thus, based on the above observations, the following is one classic line of motivation towards justifying efforts for equipping robots with natural language capabilities: why not build robots that can comprehend and generate human-like interactive behaviours, so that they can cooperate with and be taught by non-expert humans, so that they can be applied in a wide range of contexts with ease? And of course, as natural language plays a very important role within these behaviours, why not build robots that can fluidly converse with humans in natural language, also supporting crucial non-verbal communication aspects, in order to maximize communication effectiveness, and enable their quick and effective application?

Thus, having presented the classical line of reasoning arriving towards the utility of equipping robots with natural language capabilities, and having discussed a space of possibilities regarding role assignment between human and robot, let us now move to the second, more concrete, albeit less general avenue towards justifying conversational robots: namely, specific applications, existing or potential. Such applications, where natural human–robot interaction capabilities with verbal and non-verbal aspects would be desirable, include: flexible manufacturing robots; lab or household robotic assistants [30–33]; assistive robotics and companions for special groups of people [34]; persuasive robotics (for example, [35,36]); robotic receptionists [37], robotic educational assistants, shopping mall robots [38], museum robots [39,40], tour guides [41,42], environmental monitoring robots [43], robotic wheelchairs [44,45], companion robots [46], social drink-serving robots [47], all the way to more exotic domains, such as robotic theatre actors [48,49], musicians [50], and dancers [51].

In almost all of the above applications, although there is quite some variation regarding requirements, one aspect at least is shared: the desirability of natural fluid interaction with humans supporting natural language and non-verbal communication, possibly augmented with other means. Of course, although this might be desired, it is not always justified as the optimum choice, given