Robotics and Autonomous Systems 75 (2016) 4–16

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

Starting engagement detection towards a companion robot using multimodal features[☆]

Dominique Vaufreydaz[a,b,*], Wafa Johal[b], Claudine Combe[a]

[a] Prima/Inria-LIG, CNRS - Zirst Montbonnot, 655 avenue de l'Europe, 38334 Saint Ismier cedex, France

[b] University of Grenoble - Alpes, LIG, CNRS - Maison Jean Kuntzmann, Domaine universitaire, 110 avenue de la Chimie, 38400 Saint-Martin-d'Hères, France

HIGHLIGHTS

- Multimodal approach for starting engagement detection using non-explicit cues.
- Results show that our approach performs better than spatial one in all conditions.
- MRMR strategy reduces the features space to 7 features without a performance loss.
- Validation of Schegloff (sociologist) meaningful features for engagement detection.
- A robot centered labeled corpus of 4 hours in a home-like environment.

ARTICLE INFO

Article history:
Available online 19 January 2015

Keywords:
Multimodal perception
Affective computing
Healthcare technologies
Companion robots

ABSTRACT

Recognition of intentions is a subconscious cognitive process vital to human communication. This skill enables anticipation and increases the quality of interactions between humans. Within the context of engagement, non-verbal signals are used to communicate the intention of starting the interaction with a partner. In this paper, we investigated methods to detect these signals in order to allow a robot to know when it is about to be addressed. Originality of our approach resides in taking inspiration from social and cognitive sciences to perform our perception task. We investigate meaningful features, i.e. human readable features, and elicit which of these are important for recognizing someone's intention of starting an interaction. Classically, spatial information like the human position and speed, the human–robot distance are used to detect the engagement. Our approach integrates multimodal features gathered using a companion robot equipped with a Kinect. The evaluation on our corpus collected in spontaneous conditions highlights its robustness and validates the use of such a technique in a real environment. Experimental validation shows that multimodal features set gives better precision and recall than using only spatial and speed features. We also demonstrate that 7 selected features are sufficient to provide a good starting engagement detection score. In our last investigation, we show that among our full 99 features set, the space reduction is not a solved task. This result opens new researches perspectives on multimodal engagement detection.

1. Introduction

Companion robots are entities that are intended to be used as assistants in everyday life, those being personal coach, desktop manager, etc. They could help to come up with tools that can potentially improve quality of life in the long run. Among usual embedded functions, one can find entertainment, video conference, objects grasping, activity monitoring, serious games and frailty evaluation [1–4]. Companion robots can assist therapy for autism [5]. This paper presents research on companion robots using the Kompai Robot (see Fig. 1).

As argued in [6,7], the primary challenge in building engaging companion robots is to provide social competency in perceiving, reasoning and expressing social and affective aspects of interactions with the human user. Companion robots are aimed to interact with humans in home environments. In order to stay credible, companion robots are expected to behave and react as per

[☆] This work is supported by the PAL (http://pal.inria.fr/) initiative from Inria and the FUI Pramad (http://pramad.inria.fr/) project from the French government.

[*] Corresponding author at: Prima/Inria-LIG, CNRS - Zirst Montbonnot, 655 avenue de l'Europe, 38334 Saint Ismier cedex, France.

E-mail address: Dominique.Vaufreydaz@inria.fr (D. Vaufreydaz).

http://dx.doi.org/10.1016/j.robot.2015.01.004
0921-8890/© 2015 Elsevier B.V. All rights reserved.

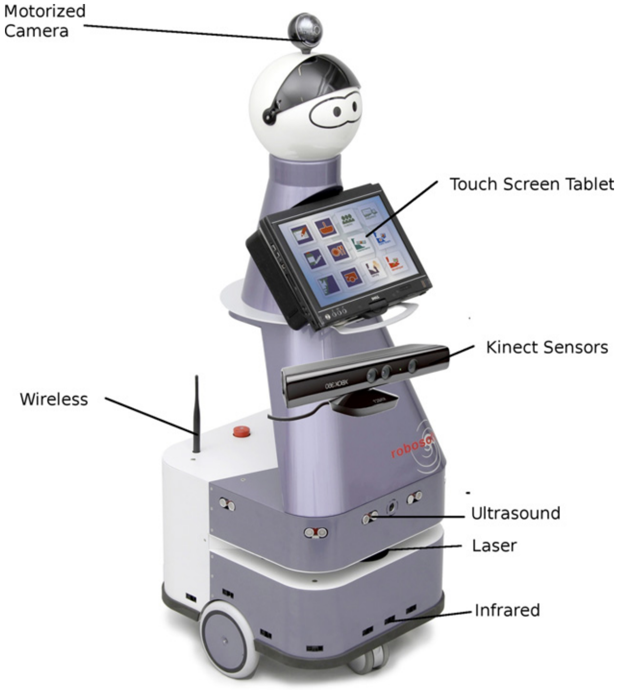

Fig. 1. The Kompaï Robot from our partner Robosoft is equipped with a laser range finder, ultrasound and infrared telemeter, a tablet PC and a webcam on top. We added a Kinect for our experiments.

predefined manners corresponding to instructions and social signals used by the user.

As speech being multimodal in a face-to-face interaction, non-verbal communication also uses a variety of channels to convey messages. New areas explore techniques for the multimodal aspect of human communication in order to design robots able to read and express communicative signals in a social manner. Non-verbal cues of communications have been well studied for detection of emotions [8–11]. In this paper, we propose to use these cues for intention recognition, and in particular an intention of interaction.

Recognition of intention is a basic skill acquired by infants early in their development. According to Vernon [12], among other skills, the perception of others' attention is crucial for the infant to master social interactions. The perception of intentions and emotions, present in newborn infants, helps to set their "preparedness" for social interaction [12]. Intention recognition allows the interacting agent to take quick decisions and to respond better to the user's need or state of mind. Some of the non-verbal communication signals are cues to subdued goals and intentions of the humans, and therefore a good way to improve adaptability of the robots' behaviors is by predicting their intentions. A part of human cognition is anticipation, allowing reading intentions and guessing goals in order to react quickly to stimuli. This skill is also very important for turn-taking in interaction.

In neurocognition, Broca's area, responsible for language comprehension, action recognition & prediction and speech-associated gestures, would be the host of intention recognition in the human brain. According to Vernon, studies have shown that the activation of Broca's area is significantly higher when a subject observes goal-directed actions with intentional cues rather than meaningless gestures.

As humans instinctively detect the intention of someone who wants to ask for way in the street, we are interested in the opening engagement phase of the process during which humans subconsciously express their intentions to interact. Our goal is to investigate techniques to detect and recognize signals for non-verbal communication reflecting this intention and in our particular case, the intention of a user to engage an interaction with a robot.

Intention of engagement is a real question, especially when it comes to environments such as the work place or home, where people are not used to interact with robots [13]. Classically, the criterion for a user's intention of engagement is the spatial distance between the user and the communicant interface [14]. Some investigations have improved on this idea by also considering the speed of movement of the user [15]. These studies have chosen to use the relative spatial position of the concerned agents as criteria. The following assumption is made behind this choice: if the user is close to the robot, there is an intention to interact. Using distance and sometimes speed of the human provides with satisfactory results, but for a companion robot in real situations at home, close distance does not necessarily signal a desire for engagement. For instance, many times during the day, one can pass in front of the refrigerator without the wish to open it. Following the same logic, despite the physical distance of the user from the robot, a robot should be able to detect when it is about to be solicited, and anticipate the interaction in order to be more comfortable and socially acceptable.

In this study, we propose a multimodal approach for detecting a starting engagement using a RGB-D sensor mounted on a companion robot. Getting inspiration from social and cognitive sciences, our goal is to select features in order to improve the re-usability in other situations and/or with other sensors. In our approach, the idea here is to get rid of the usual way to do such experiment i.e. putting all available features together, combining them in a more optimized representation and let the training paradigm filter everything. Doing this, we might have good performances, but we may not learn anything about detecting intention of engagement. We will see that less than 10% of our features are crucial for starting engagement detection. In another context, one can make well-founded choices among sensors to reflect this knowledge. It will be more efficient to design a new device or robot knowing which particular features are of importance. This prospective research aim to build a set of meaningful features extracted from multimodal sensors useful for the description, recognition and discrimination of the intention of engagement.

This paper aims to contribute on the following statements:

- There exist subconscious social signals expressed by humans that characterize their will to interact with a robot and these signals are detectable.
- Some features from literature in the social and cognitive sciences are computable on a companion robot (notably Schegloff metrics [16]).
- Multi-modal features will perform better than spatial features to detect this starting of engagement in a home-like environment. A realistic dataset in a home-like environment can help us to validate this hypothesis.
- The set of relevant features for starting of interaction detection can be reduce without loss of performance using a feature space reduction process using the Minimum Redundancy Maximum Relevance (MRMR) method [17] never used in this context.

2. Multimodal social signal processing for non-verbal communication

2.1. Social signal processing

A communicative agent does not use only the verbal channel, but many channels to send and receive various messages while interacting [18]: human communication is intrinsically multimodal. To make human–robot communication fluent and acceptable, the