Multimodal Interaction and Attentional Regulation in HRI

Francesco Cutugno  Alberto Finzi  Salvatore Iengo  Silvia Rossi  Mariacarla Staffa

I. INTRODUCTION

In order to work with humans, a robotic system should be able to understand the users’ commands and intentions, and to safely interact within a shared workspace. In this context, one of the main issues is to adapt the interaction to dynamic environments and to offer a set of rich HRI schemata. For this purpose we propose an integrated architecture, where multimodal HRI is regulated by an attentional system.

II. GENERAL FRAMEWORK

We present a general framework for HRI that integrates a multimodal interaction module, deployed to support the overall HRI, with an attentional module able to orient the robot perceptive and cognitive resources towards relevant objects, humans, and activities.

a) Multimodal Interaction Module: The multimodal interaction module is to recognize the multiple human commands and actions, such as utterances, gestures or body postures, and to provide an interpretation of user’s intentions according to a dialogical context. It is composed of three layers: the lower layer contains the classifiers of the single modalities (in our system speech, emotions, and gestures); the middle layer is the fusion engine, it provides a late fusion integration of the multiple inputs (see [2] for details) (here we use Support Vector Machines); the upper layer (dialogue manager), is modeled as a Partially Observable Markov Decision Process, and coordinates the dialogue between the human and the robot. It accomplishes the semantic interpretation of the observations according to the context and the inner knowledge. In the proposed HRI architecture, each layer provides a N-bets list of possible interpretations, which are disambiguated in cascade; the latter layer interprets the human actions and defines the interaction policy.

b) Attentional System Module: This module is to regulate the interaction exploiting bottom-up and top-down attentional mechanisms. It is composed of a behavior-based architecture (BBA) and an executive system. The first one is endowed with bottom-up attentional mechanisms, while the latter provides top-down regulations. We deploy frequency-based attentional mechanisms that tune the resolution at which each behavior is monitored and controlled (see [1] for details). This regulation can depend on salient stimuli (bottom-up) or task/context related activities (top-down). Depending on the dialogue context and the BBA attentional state, the executive system is to decide which action to execute while providing top-down modulations of the BBA sampling rates.

III. APPLICATIONS AND CONCLUSIONS

The proposed architecture offers many advantages in human-robot co-worker scenarios, where the ability to interact with a multitude of objects and humans in a safe, natural and effective way is strictly required. It is specifically designed to be easily modified and expanded, either by adding or modifying input channels, changing multimodal fusion strategies, or by adding new behavioral skills without impacting the rest of the system. In particular, we considered the following case studies:

c) Navigation: a mobile robot shares the workspace with several users interacting by voice or gestures. The robot has a list of user dialogue models describing possible patterns of multimodal interactions. In this context, the task is to pick and place objects according to the ambiguous human intentions. Our tests shows not only that the multimodal HRI system enhances the performance of individual classifiers, but also that the dialogue manager can further improve both classification and the quality of the interaction. Moreover, the attentional system can provide additional regulation and deliberative capabilities (e.g., contextual disambiguation and on-line velocity modulation).

d) Manipulation: a robotic arm interacts with humans in tasks of object handover. Here, we mainly tested the attentional system effectiveness in regulating the interaction according to the human movements. The collected results show that both attentional task-switching and velocity modulation allow a natural, readable, and safe interaction.

REFERENCES