



**HAL**  
open science

## A Multi-modal Perception based Architecture for a Non-intrusive Domestic Assistant Robot

Christophe Mollaret, Alhayat Ali Mekonnen, Julien Pinquier, Frédéric Lerasle, Isabelle Ferrané

► **To cite this version:**

Christophe Mollaret, Alhayat Ali Mekonnen, Julien Pinquier, Frédéric Lerasle, Isabelle Ferrané. A Multi-modal Perception based Architecture for a Non-intrusive Domestic Assistant Robot. 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2016), ACM; IEEE, Mar 2016, Christchurch, New Zealand. pp.481-482, 10.1109/HRI.2016.7451816 . hal-01296633

**HAL Id: hal-01296633**

**<https://hal.science/hal-01296633>**

Submitted on 1 Apr 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A Multi-modal Perception based Architecture for a Non-intrusive Domestic Assistant Robot\*

Christophe Mollaret<sup>1,2,3</sup>, Alhayat Ali Mekonnen<sup>2</sup>, Julien Piquier<sup>1</sup>, Frédéric Lerasle<sup>2,3</sup>, Isabelle Ferrané<sup>1</sup>

<sup>1</sup>IRIT, Univ de Toulouse, 118 route de Narbonne, 31062 Toulouse Cedex, France

<sup>2</sup>CNRS, LAAS, 7 avenue du Colonel Roche, F-31400 Toulouse, France

<sup>3</sup>Univ de Toulouse, UPS, LAAS, F-31400 Toulouse, France

{cmollare,aamekonn,lerasle}@laas.fr, {pinquier,ferrane}@irit.fr

**Abstract**—We present a multi-modal perception based architecture to realize a non-intrusive domestic assistant robot. The realized robot is non-intrusive in that it only starts interaction with a user when it detects the user’s intention to do so automatically. All the robot’s actions are based on multi-modal perceptions, which include: user detection based on RGB-D data, user’s intention-for-interaction detection with RGB-D and audio data, and communication via speech recognition. The utilization of multi-modal cues in different parts of the robotic activity paves the way to successful robotic runs.

## I. INTRODUCTION

In recent years, autonomous robot usage in hospitals, laboratories, and domestic homes has been increasing. In fact, robots serving various tasks and purposes in the social care and medical/health sectors beyond the traditional scope of surgical and rehabilitation robots are poised to become one of the most important technological innovations of the 21<sup>st</sup> century [2]. In this context, the main issue of Human-Robot Interaction (HRI) research is to endow the robot with suitable behaviors. A reactive behavior, which cycles through monitor and interaction phases has been widely accepted. The interaction generally starts when the user explicitly asks something to the robot or shows an interest to do so. In line with this, we focused on a user intention detection mechanism to transition from monitoring to interaction states. We refer this as *user’s intention-for-interaction* detection which is close to the concept of engagement.

In this paper we present an overview of our detection system designed to be non-intrusive. While most research works focus on interaction, the phase before it is also taken into account in our investigation. Following a use-case scenario, experiments involving elderly people have been carried out, giving preliminary evaluations that highlight the soundness of the proposed framework.

## II. COMPLETE SYSTEM OVERVIEW

The overall envisaged framework is illustrated in Fig. 1. This framework is very generic and can be applied to a whole range of scenarios involving HRI. It can be summarized as: detect the person in the environment; go to a garage position and wait until detecting the *user’s intention for interaction* based on the fusion of multimodal percepts (e.g., name calling,

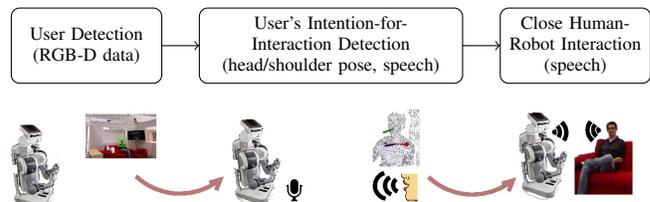


Fig. 1: Framework adopted to realize the complete non-intrusive autonomous assistive robotic system.

directing gaze, etc.); then approach this person and begin envisaged close interaction.

*a) User Detection:* Primarily relying on an embedded RGB-D sensor, we use a state-of-the-art upper body detector (head and shoulder), recently proposed by Jafari *et al.* [5], coupled with an optimized HOG based full body detector. The combined detector can detect users that are sitting or standing.

*b) Intention-for-Interaction Detection:* The user’s intention is estimated based on three important cues: user’s line of sight – inferred from head pose [3]; user’s anterior body direction – determined from shoulder orientation computed from openNI skeleton fitting [4]; and speech used to draw attention – captured via Voice Activity Detection (VAD). The head pose detection and shoulder orientation detection modules rely on depth image. Detection outputs are further filtered with a particle swarm optimization inspired tracker developed for this framework. Both, the VAD and tracker output are used as observation inputs and are fused to provide a probabilistic intention estimate using an HMM.

*c) Close Human-Robot Interaction:* During this phase, the human and robot engage in a verbal interaction. The person asks assistance, and the robot answers by providing a useful response or assistance. It is implemented via a static state machine dialogue module that monitors the interaction. Each specific question/request coming from the user can trigger transitions to different states leading to robotic service provision. It uses the CMU PocketSphinx and Google Speech APIs for speech recognition, and the Google TTS API for synthesis. Sentences the robot could say are defined *a priori*.

*d) Task-level Coordination:* Each component of the framework is conceptualized as an individual robotic task that can be coordinated to create and launch a complete working robotic demo. This complex system coordination is done using SMACH [1] – a task-level architecture for rapidly creating complex robot behavior.

\*This work was supported by a grant from the French National Research Agency (ANR) under RIDDLE project, grant number ANR-12-CORD-0003.

