The MOBOT Platform – Showcasing Multimodality in Human-Assistive Robot Interaction

Eleni Efthimiou[1(✉)], Stavroula-Evita Fotinea[1], Theodore Goulas[1],
Athanasia-Lida Dimou[1], Maria Koutsombogera[1], Vassilis Pitsikalis[2],
Petros Maragos[2], and Costas Tzafestas[2]

[1] Institute for Language and Speech Processing/ATHENA RC, Athens, Greece
{eleni_e,evita,tgoulas,ndimou,mkouts}@ilsp.gr
[2] Institute of Communication and Computer Systems–NTUA, Athens, Greece
{vpitsik,petros.maragos,ktzaf}@cs.ntua.gr

Abstract. Acquisition and annotation of a multimodal-multisensory data set of human-passive rollator-carer interactions have enabled the analysis of related human behavioural patterns and the definition of the MOBOT human-robot communication model. The MOBOT project has envisioned the development of cognitive robotic assistant prototypes that act proactively, adaptively and interactively with respect to elderly humans with slight walking and cognitive difficulties. To meet the project's goals, a multimodal action recognition system is being developed to monitor, analyse and predict user actions with a high level of accuracy and detail. In the same framework, the analysis of human behaviour data that have become available through the project's multimodal-multisensory corpus, have led to the modelling of Human-Robot Communication in order to achieve an effective, natural interaction between users and the assistive robotic platform. Here, we discuss how the project's communication model has been integrated in the robotic platform in order to support a natural multimodal human-robot interaction.

Keywords: Multisensory data · Multimodal semantics · Multimodal annotation scheme · Multimodal HRI model · Multimodal human-robot communication · Natural HRI

1 Introduction

The need to support mobility and vitality in our ageing society, as well as enhance independent living of elderly people and their quality of life [1] has inspired technological solutions towards developing intelligent active mobility assistance robots for indoor environments, providing user-centred, context-adaptive and natural support [2–5]. The MOBOT project[1] addresses this need envisioning cognitive robotic assistants that act (a) proactively by realizing an autonomous and context-specific monitoring of human activities and by subsequently reasoning on meaningful user behavioural

---

[1] www.mobot-project.eu/.

patterns, as well as (b) adaptively and interactively, by analyzing multi-sensory and physiological signals related to gait and postural stability, and by performing adaptive compliance control for optimal physical support and active fall prevention.

To address these targets, a multimodal action recognition system is being developed to monitor, analyse and predict user actions with a high level of accuracy and detail. Parallel to the enhancement of computer vision techniques with modalities such as range sensor images, haptic information as well as command-level speech and gesture recognition, data-driven multimodal human behaviour analysis has been conducted in order to extract behavioural patterns of elderly people. The aim here has been to import the basic elements of these behavioural patterns into a multimodal human-robot communication system [6], involving both verbal and nonverbal communication conceptually and systemically synthesised into mobility assistance models taking into consideration safety critical requirements.

By the end of the project, the different modules will be incorporated in a behaviour-based and context-aware robot control framework aiming at providing situation-adapted optimal assistance to users [7]. Direct involvement of end-user groups in various stages of the prototypes development has ensured that actual user needs are addressed by the functionalities and communication capabilities of the platform's prototypes. Thus, user trials have been conducted to evaluate and benchmark the overall system.

The next sections report on the technologies which have been integrated in the robotic platform's prototypes, the functionalities the latter provide, the HRI communication model adopted, as well as the various end user evaluation and usability studies conducted to ensure that the developed platform addresses actual user needs.

2   Platform Integrated Technologies and Functionalities

The development of the MOBOT platform has proven to be a rather ambitious experiment, since it envisioned integration and synergies of a wide range of technologies, which needed to reach a significant level of enhancement as a result of research work within the project, in order to reach the state of maturity required to meet the set targets in respect to functionalities and safety controls envisioned to be offered to the platform's end users.

We present next the various types of integrated technologies and the respective functionalities they support, in order to illustrate how the adopted multimodal HRI communication model makes optimal use of the available technological solutions.

At this point, it is also important to notice that the MOBOT multimodal-multisensorial dataset [8] has been exploited towards the enhancement of all technologies explored in order to be integrated in the MOBOT robotic platform (Fig. 1).
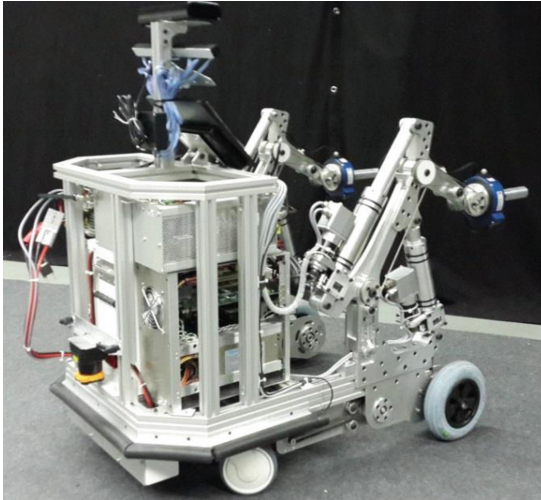
Fig. 1. The active rollator used at the first evaluation of the MOBOT rollator-type mobility assistant

2.1 The MOBOT Platform Technologies

Work on *visual action recognition* in continuous RGB-D video streams, captured by visual sensors on the MOBOT robotic platform [9], robust experimental results on *object detection* and advances in *human body pose estimation* have supported, in combination with other technologies, the detection capacity by the platform of human activity that denotes various user intentions such as to activate the robot [10, 11].

Exploitation of the MOBOT dataset in relation to research work in *action/gesture recognition*, provided the chance to apply the action/gesture recognition algorithms developed in the project to relevant data, where advancements in the field include (i) the development of an improved gesture recognition method that exploits specific articulatory points such as the arms and the hands of the subject, and (ii) the application and experimentation on actual MOBOT data, following research work with other datasets. The experimental framework concentrated mainly on HMM-type classifiers, using two visual cues for feature extraction: handshape (provided by the RGB stream) and 3D movement-position (provided by Kinect's depth stream and skeleton tracking) [12, 13].

For the processing of spoken commands, a *spoken command recognition system* is utilized, which in a first step uses a voice activity detector to detect in the audio stream the time segments with spoken commands and then in a second step trains a set of HMM models on these segments. Experimentation has taken place on the audio data from the benchmark dataset of the ACM 2013 Multimodal Gesture Challenge and also during work on developing a complete spoken command recognition system trained on MOBOT data and on integrating the corresponding software on the ROS platform with a MEMS microphone array to be used on the MOBOT active rollator prototype [14].

The adopted approach to *multimodal sensor fusion* for audio-visual gesture recognition exploits the color, depth and audio information captured by a Kinect sensor. Recognition of a time sequence of audio-visual gesture commands is based on an optimized fusion of all different cues and modalities (audio, movement-position, handshape). The methodology incorporates a generalized activity detection component, while extended experimentation and comparisons with several competing approaches have provided results which greatly outperform all other competing published approaches on the ACM 2013 benchmark dataset and achieve a 93 % accuracy, which corresponds to a 47 % error reduction over the best competing approach.

Processing of haptic data has been possible via the two force/torque sensors mounted on the two handles of the rollator type prototype. These sensors are used to detect and quantify haptic interactions between the robot and the user. Typical interaction patterns while standing up, sitting down and walking have been identified in this context [15].

Furthermore, *processing of physiological data* focuses on user fatigue, since this is considered an important physiological state that can strongly affect the human performance. Fatigue estimation is based on two specific features of the human heart rate and the total performed work. Moreover, available fatigue indicators suitable for elderly fatigue estimation [16] are extended in order to fit for their usage in the context of mobility assistive robots.

2.2 The MOBOT Platform Functionalities

The MOBOT platform has two prototype demonstrators: a rollator type robot for walking and sit-to-stand assistance and a nurse type robot for sit-to-stand assistance. The MOBOT rollator is an assistive device comprising the main frame, the actuated handles, active wheels, user interface, an electronic control unit and a number of environment and user sensors.

Various functions are foreseen to be implemented on the MOBOT rollator by the end of the project which are both related to the mechanical design and the collected user requirements. The MOBOT rollator's functions are grouped next according to their main characteristics as:

(i) those which are dedicated to perceiving the user and involve the device's capacity
   a. to localize the user with respect to the rollator exploiting 3D coordinates and the state in which the user is with respect to the rollator as regards the "distant", "close" and "in contact" variables
   b. to track the articulated human body
   c. to detect walking patterns
   d. to recognize user gestures
   e. to recognize and interpret the user voice commands
   f. to monitor the human performance and postural stability and to detect unstable configuration and falls
   g. to recognize the human physiological state

    h. to recognize user actions

    i. to recognize user plans and user intentions

 (ii) those which are dedicated to detecting the environment, including the ability to detect obstacles, locomotion specific data (surface type, slip …), environment specific data (e.g. slopes) and creation of a map of the environment

(iii) the ability of the device to localize itself within the environment map

(iv) the ability of the device to approach the user from a distance

 (v) the ability of the device to assist the user by

    a. providing physical assistance during sit-to-stand and stand-to-sit transfers and

    b. assisting the user while walking in three ways which include following the user intention ("dock" to the user, accelerate, maneuver, decelerate, stop), balancing and stabilizing the user (fall prevention included), and assisting the user while passing through narrow passages and opening/closing doors

    c. assisting the user while standing

    d. assisting the user in proximity but in no contact mode

    e. assisting the user by following him/her

    f. assisting the user by providing sensorial assistance as when to avoid static and dynamic positive and negative obstacles, or assistance on slopes

    g. providing cognitive assistance

    h. assisting the user localization or guiding/navigating the user

(vi) the ability of the device to leave the user and go to parking position in autonomous mode, and finally

(vii) the ability of the device to perform autonomous charging

The features listed above, directly linked with the technological solutions in Sect. 2.1, have set the framework for the development of the platform's multimodal communication model.

3   The MOBOT Multimodal Communication Model

The technologies integrated in the MOBOT robotic platform enabled innovative synergies among modules towards the platform's target to provide walking and cognitive assistance to elderly users with slight walking and cognitive problems.

However, the potentials of the platform would remain unexploited, if there would be missing an adequate HRI communication model to support the most natural possible communication between the platform and its user, taking into account on the one hand the state of the integrated technologies, and on the other hand the ways the target user group communicate in their everyday activities.

Detection and interpretation of patterns of explicit interactional and behavioural cues may be a trivial task for humans, but it still remains a rather difficult -yet important to achieve- task for computer systems in view of the goal to realise a naturalistic, meaningful and engaging communication between humans and machines, including the performance of actions. Thus, building the HR communication model upon insights stemming from sets of acquired data allows for mining deeper into the semantics of

human actions, their sequence and correlation during the interaction, so that a more detailed representation of the human action (speech, audio-gestural) model in this specific environment can be drawn.

In this line, the MOBOT dataset was exploited also in respect to the information provided as regards human-to-human communications between elderly individuals and their cares while performing everyday activity tasks. The study of this information enabled the spotting of a number of natural multimodal interactions which are accompanying the core activities of the addressed audience.

Furthermore, the underlying notion behind the part of data acquisition which entails the closed set of combinations of the MOBOT audio-gestural commands is that any system attempting to model human actions in terms of interactional behaviour, needs to have access to knowledge on structures of human actions. Specifically, assistive systems dealing with human-machine interaction must be able to decompose human behaviour activities into measurable and machine detectable features, so that they are able to make decisions and plan support actions on the basis of heterogeneous sensory data [6, 17]. To this end, a set of recognizable actions that are associated with specific forms of human behaviour needs to be identified to deduce information on underlying human intentions and needs.

The MOBOT human-robot communication model was built as a structured tree of possible multimodal action-reaction interactions engaging both audio and gestural signals, enriched with a number of cognitive assistance assertions from the part of the platform, which assimilate human reinforcement to elderly individuals while performing a trivial task. The platform may interact in three modes: (i) in hands-on mode, (ii) in following mode, and (iii) in stand-by mode, while user input focuses on developing a multimodal dialogue strategy, which takes into account the options of (a) communication via body posture in silence and thus complete absence of any other speech or gesture signal, processing information that can be linked with the platform's action recognition module, and (b) communication via speech and/or gesture signals, information that can be linked with the audio and gestural signal recognition.

The system's ability to learn how to navigate using a map is linked with a number of cognitive support messages which are sent to the user in the form of orally uttered questions or reinforcement messages, similar to those received from the part of human cares during natural human interaction.

The question type messages are linked with the activation and approach of the robotic device, decision making in respect to the route selected each time or obstacle avoidance situations, and demand some verification from the part of the user.

Reinforcement type messages help users be encouraged to complete a task successfully. Both types of messages are provided in a manner that prevents message generation from being part of a routine that does not help or even gets annoying for the receiver.

The scenarios of use, for which action trees are constructed, are extracted from real user needs as defined during preparation of acquisition of the MOBOT dataset and depicted in the scenarios actually used in creating the MOBOT corpus, and also through end user evaluation of the intermediate stages of the platform's prototypes.

4   MOBOT Platform End User Evaluation

The MOBOT platform functionalities and devices are steadily tested and evaluated by end user groups, since one main aim of the project has been to perform and carry out an overall benchmark verification of the developed intelligent mobility assistants via intensive evaluation studies involving end-users.

The first evaluation of the MOBOT rollator-type mobility assistant was conducted at the BETHANIEN-Hospital/Geriatric Centre at the University of Heidelberg from end of October 2014 to beginning of December 2014.

The subject recruitment procedure, the different test scenarios developed to evaluate specific functionalities of the MOBOT device, adequate quantitative and qualitative performance measures and preliminary results of this evaluation study are reported in detail in [18].

A detailed description of validation studies on existing devices by a systematic review, as well as a detailed account of the use cases included in the evaluation process, have been reported in [19] and have been updated in [20].

Use cases represent typical scenarios for the development of technical devices to guide technical development and allow validation of devices. The MOBOT project defined such use cases for the specified setting of geriatric rehabilitation and long-term care.

The exploited use cases identify tasks or situations which differ with respect to duration, frequency of activity and clinical relevance. Some use cases take only seconds while other may take minutes, while some situations occur frequently during a day while others may occur only once a year. However, some of the tested situations occur rarely but are crucial for the task of the device to support the motor stability of the user, thus, preventing injurious falls.

Technical functionalities were assessed with respect to accuracy, validity and reliability of technical function. Clinical evaluation targeted at the interaction between a human and the device.

The evaluation study was carried out with 36 participants who met predefined inclusion criteria and who were recruited from geriatric rehab wards of the BETHANIEN-Hospital, hospital-associated nursing homes and from a rehab sports club. The aim of the evaluation study was to validate the functionalities mounted on the MOBOT active rollator, which comprised the sit-to-stand assistance, sensorial assistance tested via obstacle avoidance, as well as basic walking assistance. To fit these functionalities, specific test scenarios and "tailored" assessment methods were developed including qualitative as well as quantitative performance measures. Most of the assessment procedure proved to be feasible and adequate for the needs of the testers. Furthermore this initial trial proved a valuable experiment for planning all following trials and evaluation studies. Most crucially, the gained results proved promising with respect to the tested STS assistance, the basic walking assistance, and the subjective user perception of the MOBOT device and its assistance systems.

5   Conclusion

In this work we discussed the integration of a multimodal communication model into the MOBOT robotic platform to support a natural multimodal human-robot interaction. The communication model was built through the analysis and annotation of acquired multi-sensory data of users interacting with mobility aids and their emerging multi-modal human behavioural patterns. At the same time, the analysis of walking patterns and critical safety requirements in mobility assistance, as well as the detection of abnormalities were taken into account for the model development. Since the aim of the platform is to provide situation-adapted optimal assistance to users, the platform functionalities and communication capabilities have been initially tested by end user groups to evaluate and benchmark the overall system. This first trial was positively assessed in terms of feasibility and adequacy to the user needs (especially regarding the basic walking assistance, and the subjective user perception of the MOBOT device and its assistance systems), and it has served as a valuable experiment for planning all following trials and evaluation studies.

In parallel, a goal that has been addressed is the implementation of context-dependent instantiations and flows of the communication model following the different interaction states and user needs, successfully through the use of the audio and visual channels of communication with the device.

This Human-Robot Interaction model directly responds to the communication re-quirements as an integral part of an architecture towards monitoring, analysing and predicting user actions with a high level of accuracy and detail. The extension of the communication model with an enhanced set of audio-gestural commands as well as with device-initiated cognitive information regarding localization information and information of general interest (e.g. date, time, weather, patients' personal info, etc.) is currently work in progress. Moreover, the plans for the finalization of the communi-cation scheme include an algorithmic flow diagram of all possible interaction states within the communication model. This flow will also specify the states implemented in the device at every step of the integration process together with the progress of the individual components of the spoken (audio-gesture) dialogue system (i.e. speech/gesture recognition, language understanding, dialogue management, commu-nication with external system, response generation, speech output etc.).

An effective, natural interaction between users and the assistive robotic platform is a crucial step towards the development of safe hardware platforms for mobility aids which are tailored for indoor environments. Aspects of such platforms include software modules for robot navigation, human-adaptive and proactive robot motion control, and low-level active compliance control and shared control behaviours.

Intensive evaluation studies will keep being performed to include more end-users towards an overall benchmark verification of the developed intelligent mobility assistants.

References

1. Hirvensalo, M., Rantanen, T., Heikkinen, E.: Mobility difficulties and physical activity as predictors of mortality and loss of independence in the community-living older population. J. Am. Geriartric Soc. **48**, 493–498 (2005)

2. Chuy, O.J., Hirata, Y., Wang, Z., Kosuge, K.: Approach in assisting a sit-to-stand movement using robotic walking support system. In: IEE/RSJ International Conference on Intelligent Robots and Systems, China (2006)

3. Chugo, D., Asawa, T., Kitamura, T., Jia, S., Takase, K.: A rehabilitation walker with standing and walking assistance. In: IEE/RSJ International Conference on Intelligent Robots and Systems, Nice, France (2008)

4. Wakita, K., Huang, J., Di, P., Sekiyama, K., Fukuda, T.: Human walking-intention-based motion control of an omnidirectional-type cane robot. IEEE/ASME Trans. Mechatron. **18**(1), 285–296 (2013)

5. Hirata, Y., Komatsuda, S., Kosuge, K.: Fall prevention control of passive intelligent walker based on human model. In: IEE/RSJ International Conference on Intelligent Robots and Systems, IROS, Nice, France (2008)

6. Fotinea, S-E., Efthimiou, E., Koutsombogera, M., Dimou, A-L., Goulas, T., Maragos, P., Tzafestas, C.: The MOBOT human-robot communication model. In: Proceedings of 6th IEEE Conference on Cognitive Infocommunications (CogInfoCom 2015), Győr, Hungary, 19–21 October (2015)

7. Papageorgiou, X.S., Tzafestas, C.S., Maragos, P., Pavlakos, G., Chalvatzaki, G., Moustris, G., Kokkinos, I., Peer, A., Stanczyk, B., Fotinea, E.-S., Efthimiou, E.: Advances in intelligent mobility assistance robot integrating multimodal sensory processing. In: Stephanidis, C., Antona, M. (eds.) UAHCI 2014/HCII 2014 Part III. LNCS, vol. 8515, pp. 694–703. Springer, Heidelberg (2014)

8. Fotinea, E.-S., Efthimiou, E., Dimou, A.-L., Goulas, T., Karioris, P., Peer, A., Maragos, P., Tzafestas, C., Kokkinos, I., Hauer, K., Mombaur, K., Koumpouros, I., Stanzyk, B.: Data acquisition towards defining a multimodal interaction model for human – assistive robot communication. In: Stephanidis, C., Antona, M. (eds.) UAHCI 2014, Part III. LNCS, vol. 8515, pp. 613–624. Springer, Heidelberg (2014)

9. Trulls, E., Kokkinos, I., Sanfeliu, A., Moreno, F.: Dense segmentation-aware descriptors. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2013)

10. Trulls, E., Kokkinos, I., Sanfeliu, A., Moreno, F.: Superpixel-grounded deformable part models. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014)

11. Boussaid, H., Kokkinos, I., Paragios, N.: Discriminative learning of de formable contour models. In: International Symposium on Biomedical Imaging (ISBI) (2014)

12. Pitsikalis, V., Katsamanis, A., Theodorakis, S., Maragos, P.: Multimodal gesture recognition via multiple hypotheses rescoring. J. Mach. Learn. Res. **16**(1), 255–284 (2015). http://www.jmlr.org/papers/volume16/pitsikalis15a/pitsikalis15a.pdf

13. Kardaris, R.N., Pitsikalis, V., Mavroudi, E., Katsamanis, A., Tsiami, A., Maragos, P.: Multimodal human action recognition in assistive human-robot interaction. In: Proceedings of 41st International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016), Shanghai, China (2016)

14. Skordilis, Z.I., Tsiami, A., Maragos, P., Potamianos, G., Spelgatti, L., Sannino, R.: Multichannel speech enhancement using MEMS microphones. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-2015), Brisbane, Australia, April 2015

15. Corredor, J., Sofrony, J., Peer, A.: Deciding on optimal assistance policies in haptic shared control tasks. In: IEEE International Conference on Robotics and Automation, pp. 2679–2684 (2014). http://dx.doi.org/10.1109/ICRA.2014.6907243

16. Ho Hoang, K.-L., Corradi, D., Mombaur, K.: Identification and classification of geriatric gait patterns - multi-contact capturability. In: French-German-Japanese Conference in Humanoid and Legged Robots, 12–14 May 2014, Heidelberg, Germany (2014)

17. Papageorgiou, X.S., Chalvatzaki, G., Tzafestas, C.S., Maragos, P.: Hidden Markov modeling of human normal gait using laser range finder for a mobility assistance robot. In: Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA-2014), Hong Kong, China, pp. 482–487, June 2014. http://dx.doi.org/10.1109/ICRA.2014.6906899

18. MOBOT Deliverable D5.3 - Report on performance metrics and first evaluation study. http://www.mobot-project.eu/userfiles/downloads/Deliverables/MOBOT_WP5_D5.3.pdf

19. MOBOT Deliverable D5.1: Preliminary report on use cases and user needs. http://www.mobot-project.eu/userfiles/downloads/Deliverables/MOBOT_WP5_D5.1_v1.5.pdf

20. MOBOT Deliverable D5.2: Report on use cases, performance metrics and user study preparations. http://www.mobot-project.eu/userfiles/downloads/Deliverables/MOBOT_WP5_D5.2_v1.5.pdf