

Towards Effective Communication with Robotic Assistants for the Elderly: Integrating Speech, Vision and Haptics

B. Di Eugenio, M. Žefran, J. Ben-Arie, M. Foreman (*),
L. Chen, S. Franzini, S. Jagadeesan, M. Javaid, K. Ma
University of Illinois at Chicago / (*) Rush University
Chicago, IL, 60607, USA

Introduction and Motivation

Our goal is to develop an interface for older people to effectively communicate with a robotic assistant so that they can safely remain living in their home. We are devising a multimodal interface since people communicate with one another using a variety of verbal and non-verbal signals, including haptics, i.e., physical interactions. We view haptics as an integral component of communication, which in some cases drives the interaction between the user and the robot, and we study its relation to speech and gestures. The interaction of language, gestures and gaze has been well studied (Kawarazaki, Yoshidome, and Nishihara 2004; Morency et al. 2007; Mutlu et al. 2009). However, the role of haptics has hardly been explored in the context of interpersonal communication, in particular with robots (Gockley and Matarić 2006).

We will illustrate our points via excerpts from our ongoing data collection. Our data is being collected in a fully functional studio apartment in the College of Nursing at Rush University. It has a complete kitchen, bathroom, and living space (with bed, sofa, table with 4 chairs, dresser). We have equipped the apartment with an 8 camera video system. We are focusing on *Activities of Daily Living (ADLs)* that are relevant to the independent living of the elderly (Krapp 2002). We have settled on three ADLs: getting up from a bed or chair, ambulating in the apartment, and preparing a meal. So far, we have run extensive pilot studies, and one real subject. A student in gerontological nursing plays the role of the helper (HEL), both in pilot studies and with real subjects; she wears a data glove. In the pilot dialogues, one faculty member was playing the role of the elderly person (ELD). Both HEL and ELD are wearing a microphone.

In the first excerpt, our subject, a 96-year-old woman, is first getting up from a chair, and then walking towards the bathroom. In lines 2 through 5, and again in lines 7 through 8, haptics and vision are the primary modalities for exchange of information, with language providing only high level descriptions of goals. For example, in line 2, there is a haptic exchange between HEL pulling and ELD resisting the pull. Again in line 8, ELD's turning around is haptically sensed by HEL who is supporting ELD and follows suit.

1. HEL: "Ready?"
2. → HEL gently grabs and pulls on ELD's hands
3. → ELD retracts her hands
4. → ELD: "I don't know, I think I can do it this way"
5. ELD pushes herself up on the chair's armrests, while HEL moves her arm to support ELD at the waist
6. ... HEL and ELD are walking towards the bathroom; HEL has her arm around ELD's waist to support her
7. ELD: "Oh, I forgot my book"
8. → ELD turns around to go back to the table where the book is lying, HEL follows

This excerpt also illustrates an orthogonal point: the elderly we are targeting should not be seen as passive participants. This subject takes the initiative and finds the best way for herself to get up. The robotic assistant must be able to adapt to the elderly and to react appropriately.

In the next excerpt, taken from one of our pilot dialogs, in line 4 haptics is the only modality at work besides vision: HEL follows ELD on the cue she senses when ELD physically slightly tugs on the pot. Moving the pot from the sink to the stove is performed collaboratively and without uttering a single word.

1. ELD and HEL hold a pot under the faucet to fill it with water, each holding a handle
2. ELD: "OK alright"
3. HEL turns off the water
4. → ELD and HEL together move towards the stove holding the pot and put it down on the front right burner

In the final excerpt our real subject is setting the table. The interaction consists of a complex interplay between modalities, with haptics playing a subordinate role. Language provides ELD's high level intention; after HEL satisfies the request as best she can, in a silent exchange lasting 11 seconds, HEL rearranges the napkins and hands them to ELD one by one (steps 3 through 5 are repeated twice).

1. ELD: "could you please bring some paper towel"
2. HEL: "this is the closest I can find"
3. → HEL starts folding the paper napkins she found
4. → ELD extends her arm and opens her hand in front of HEL
5. → HEL hands one napkin to ELD, who lays it on the table

Research Plan and Interface Architecture

We are working on a systematic exploration of haptic interactions, and on the integration of technologies we have independently developed as the backbone of the communi-

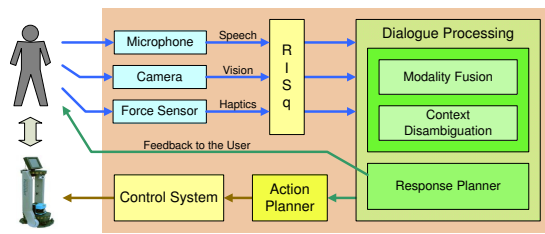


Figure 1: The architecture of the multimodal user interface

ation interface, whose architecture is shown in Figure 1. The speech, visual, and force data streams are processed by a novel highly adaptive recognition algorithm (Recognition by Indexing and Sequencing, RISq) (Ben-Arie 2008), that treats them as temporal vector sequences. For each mode, the vectors in the sequence are indexed into a hash table that stores all the models of that mode. E.g. in speech, the temporal vector sequences are the user’s voice patterns. Every sequence of such vectors represents one word. For each mode, we are constructing a RISq program with a separate hash table, which will match the incoming vector sequence to the most similar stored model. The symbolic output produced by RISq for each data stream will be fused into a coherent representation of what we call an *interaction event* and integrated with context via techniques adapted from those used in multimodal dialogue processing.

As concerns haptic interaction, we are currently identifying *physical interaction acts* (Enriquez, MacLean, and Chita 2006), namely, force patterns that describe the intent of the participant such as *grab, let go, push, pull* etc. Additionally, analogously to much research on the structure of dialogue, we are exploring which sequences occur in haptic dialogue and how those sequences mesh with linguistic ones: e.g., can a haptic act take the place of linguistic acts, as in Excerpt 1 where there is no linguistic utterance following ELD’s assertions? From the language point of view, we are adapting techniques we have previously developed to build the semantic representation of instructions and to recognize dialogue acts via a combination of Latent Semantic Analysis and K-nn methods (Di Eugenio, Xie, and Serafin 2010).

As concerns vision, we developed a reliable method for 360 degrees of human head detection in video based on a combination of HOG-like filters (Histogram Of Gradients) and Laplacian histograms, followed by Support Vector Machines based pattern recognition. We are currently working on the much more difficult problem of hand detection in video, since the hand appearance depends not only on the hand’s pose but also on the individual poses of all 5 fingers.

The interpreted commands will be passed to a Response Planner module, and an Action Planner module. We will use simple techniques for the response planner module, which is not a focus of our project. The Action Planner determines the physical motion of the robot necessary to respond to the user intent. In order to avoid the full complexity of mapping the interaction events into sequences of very detailed commands to the robot, we will devise translation rules that map classes of interaction events into low-level robot behaviors

(but see e.g. (Pineau et al. 2003) for a more principled approach). However, we are addressing the control architecture that needs to guarantee that various low-level robot behaviors can be activated without compromising the safety of the user and stability of the robot (Wei and Žefran 2006).

Clearly some of the phenomena illustrated in the three excerpts presented earlier, especially in the third, are well beyond the state of the art, and will be beyond our interface as well. However, we believe we will manage to substantially contribute to the next generation of human-robot interfaces, and to shed light on the role haptics plays in human-human interactions. When the symposium takes place in November, we will present our progress, including findings from the data we will have collected from the subjects we have recruited from two assisted living facilities.

Acknowledgments

This work is supported by award IIS 0905593 from the National Science Foundation.

References

- Ben-Arie, J. 2008. Method for recognition of human motion, vector sequences and speech. US patent 7,366,645 B2.
- Di Eugenio, B.; Xie, Z.; and Serafin, R. 2010. Dialogue act classification, higher-order dialogue structure, and instance-based learning. *Journal of Discourse and Dialogue Research* 1(2):1–24.
- Enriquez, M.; MacLean, K.; and Chita, C. 2006. Haptic phonemes: Basic building blocks of haptic communication. In *the 8th International Conference on Multimodal Interfaces, ICMI’06*.
- Gockley, R., and Matarić, M. 2006. Encouraging physical therapy compliance with a hands-off mobile robot. In *1st ACM/IEEE Conference on Human-Robot Interaction*.
- Kawarazaki, N.; Yoshidome, T.; and Nishihara, K. 2004. Assistive robot systems using gesture and voice instructions. In Keates, S., ed., *Second Cambridge Workshop on Universal Access and Assistive Technology*, 65–70.
- Krapp, K., ed. 2002. *Encyclopedia of Nursing & Allied Health*. Gale Group, Inc. Chapter Activities of Daily Living Evaluation.
- Morency, L.-P.; Lee, C.; Sidner, C.; and Darrell, T. 2007. Head gestures for perceptual interfaces: The role of context in improving recognition. *Artificial Intelligence Journal* 171(8-9):568–585.
- Mutlu, B.; Shiwa, T.; Kanda, T.; Ishiguro, H.; and Hagita, N. 2009. Footing in Human-Robot Conversations: How Robots Might Shape Participant Roles Using Gaze Cues. In *4th ACM/IEEE Conference on Human-Robot Interaction*.
- Pineau, J.; Montemerlo, M.; Pollack, M.; Roy, N.; and Thrun, S. 2003. Towards robotic assistants in nursing homes: Challenges and results. *Journal of Robotics and Autonomous Systems* 42:271–281.
- Wei, S., and Žefran, M. 2006. Towards a formal design of behaviors for autonomous robots. In *IEEE Int. Conf. On Networking, Sensing, and Control*, 66–71.