

# A User-Centric Design of Service Robots Speech Interface for the Elderly

Ning Wang, Frank Broz, Alessandro Di Nuovo, Tony Belpaeme  
and Angelo Cangelosi

**Abstract** The elderly population in the Europe have quickly increased and will keep growing in the coming years. In facing the elder care challenges posed by the amount of seniors staying alone in their own homes, great efforts have been made to develop advanced robotic systems that can operate in intelligent environments, and to enable the robot to ultimately work in real conditions and cooperate with elderly end-users favoring independent living. In this paper, we describe the design and implementation of a user-centric speech interface tailored for the elderly. The speech user interface incorporating the state of the art speech technologies, is fully integrated into application contexts and facilitates the actualization of the robotic services in different scenarios. Contextual information is taken into account in the speech recognition to reduce system complexity and to improve recognition success rate. Under the framework of the EU FP7 Robot-Era Project, the usability of the speech user interface on a multi-robots service platform has been evaluated by elderly users recruited in Italy and Sweden through questionnaire interview. The quantitative analysis results show that the majority of end-users strongly agree that the speech interaction experienced during the Robot-Era services is acceptable.

**Keywords** Human-robot interaction · Service robots · Elderly users · Speech interface

---

N. Wang (✉) · F. Broz · A. Di Nuovo · T. Belpaeme · A. Cangelosi  
Centre for Robotics and Neural Systems, Plymouth University, Plymouth, UK  
e-mail: ning.wang@plymouth.ac.uk

F. Broz  
e-mail: frank.broz@plymouth.ac.uk

A. Di Nuovo  
e-mail: alessandro.dinuovo@plymouth.ac.uk

T. Belpaeme  
e-mail: tony.belpaeme@plymouth.ac.uk

A. Cangelosi  
e-mail: a.cangelosi@plymouth.ac.uk

A. Di Nuovo  
Faculty of Engineering and Architecture, University of Enna Kore, Enna, Italy

# 1 Introduction

## 1.1 Robots for the Elderly

The increasing ageing population leads to a growing burden of elder care worldwide, which is especially true for the developed countries like those in the Europe. According to an European Union report, the number of people elder than 65 years old was 87 million in the Europe in 2010 [2]. It is estimated by the World Health Organization (WHO) that the elder population over the age of 60 is expected to be around 2 billion in 2050 [20]. One of the problems caused by the growing older population is that most of them still want to live in their own homes and to lead independent living as long as possible. With their gradually decaying physical and cognitive abilities, smart environments and assisting facilities, such as housekeeping, mobility support, social communication, and reminding systems are needed in their living places. In this circumstance, socially assistive robotic (SAR) platforms designed for improving independent living and caring for the elderly users, which can provide various services in both at-home and outdoor environments, are very much desired. For example, Robot *Robear*, which is developed by the Japanese robotics company RIKEN has human-like limbs to help move and carry objects. Others like the Aldebaran Robot *Pepper*, with emotion intelligence equipped, are designed to offer therapeutic care to the user. Meanwhile, elder adults living independently have also expressed their willingness to have robots live with them at home [17]. In the elder care domain, SAR platforms are usually integrated in an Ambient Assisted Living (AAL) environment [13], in which smart assistance systems and personal robots are designed and developed for a safer quality life at home.

## 1.2 Robotic User Interface

In order to achieve certain social, cognitive, and task outcomes goals in human-robot interaction (HRI), robots are needed to display appropriate social behaviours [6]. In realizing a SAR platform, one of the most challenging aspects regarding HRI is social communication. User interfaces in SAR can be keyboard, touch screen, gesture, and natural language, etc. Natural language based technologies such as automatic speech recognition (ASR), text-to-speech (TTS) synthesis, and language understanding have been evidently advanced in the past years, which can be seen from the great success gained by speech and language based technology products in consumer electronic markets, such as Siri<sup>1</sup> by Apple and Now<sup>2</sup> by Google. Speech technology is viewed as a major interaction modality in many application domains, for examples, customer

---

<sup>1</sup><http://www.apple.com/ios/siri/>.

<sup>2</sup><http://www.google.com/landing/now/>.

schedule information query and booking systems over telephone have employed voice-based interface for flights [8], trains [10] and restaurants and hotel booking [12]. At the same time, speech interface has also been widely engaged in multi-modal user interactive systems, such as those in smart homes [14] and AAL [7]. It has been found that among all human-machine interaction media, speech interaction is the one most accepted by users, especially for elderly people [14]. However, challenges exist in developing user-centric and high performance speech interaction tools. It is known that large vocabulary continuous speech recognition (LVCSR) is always time-consuming, substantial efforts such as data collection, model training, user adaptation, and parameter tuning are needed before actual deployment. LVCSR with state of the art recognition protocols and algorithms still report word error rates of around 20% on average [15, 16], or of 10% if trained for a more specific domain [19]. Accents, dialects and the mixed usage of multiple languages cause other failures in recognition. In domain-specific interaction tasks, contextual cues could be useful for enhancing speech recognition performance in either HRI or AAL cases. The central idea of context-sensitive speech recognition is to associate different contexts or dialogue states with individual language models or more specifically, grammars. In this case, grammar switching is indicated based on dialogue movement. This method has been shown led to more robust speech recognition [11]. Lemon showed in [11] the efforts on context-sensitive speech recognition in a dialogue system with more flexible and effective grammar switching strategy. Contextual information was also used to analyze the humans engagement towards the robot while using the dialogue system [9].

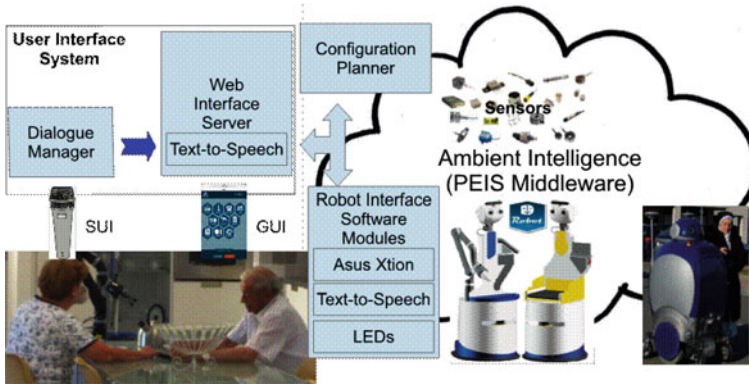
In this paper, we describe the speech interface employed in our elder-robot interaction investigation. At first, the elder service robotic platform developed in the EU FP7 Robot-Era Project is introduced. Secondly, an overview of the speech-based user interface deployed is given, in which the context-aware grammar-switching mechanism for speech recognition efficacy is highlighted. After that, preliminary results of HRI experiments on a series of the Robot-Era tasks aiming at evaluating the speech-based interface are shown. Finally, conclusive remarks of this investigation are gained.

## 2 Elder Service Robotic Platform in Robot-Era

The EU FP7 Project Robot-Era<sup>3</sup> [1] aims at integration and implementation of advanced robotic systems including SAR and AAL architectures, with an ultimate goal to provide intelligent environments and facilities in real scenarios for the ageing population. To this end, the general feasibility, scientific/technical effectiveness and social/legal acceptability of the package of robotic services offered as well as the smart environments where the robot is operating in, is assessed by real users in actual

---

<sup>3</sup><http://www.robot-era.eu/>.



**Fig. 1** Multi-robots platform and ambient intelligence architecture of the Robot-Era Project [1]

scenarios. The end users of this system are elderly people leading independent living in their daily life. In this project, several already available and commercial robotic components are adapted and integrated in both indoor (e.g., domestic, condominium) and outdoor environments to ensure independent, comfortable and safe living quality for the ageing population.

The Robot-Era services are specially designed to meet the needs of independently living elderly people in various scenarios ranging from indoor house keeping to outdoor walking support. Studies indicate that elderly people favor multi-modal user-robot interaction [18]. Among them, speech and gestures have been found most preferable by them [4]. Considering the fact that touch screen based mobile phone interface has gradually been accepted by more and more elderly users [3], in Robot-Era Project, two interfaces: graphic user interface (GUI) and speech user interface (SUI) are provided. Each of them can play as a sole-modal user interface, or they can work together as a dual-modal HRI platform. Figure 1 gives a whole picture of the multi-robots and ambient intelligent system architecture developed in the Robot-Era Project.

### 3 Speech Interaction

The SUI designed for the elderly people manipulating service robots is user-centric and domain-specific. We make efforts to provide a multilingual spoken dialogue system in various real-world scenarios. The spoken dialogue system employed is mainly composed of the following components: commercialized Nuance<sup>4</sup> speech

<sup>4</sup><http://www.nuance.com/>.

recognizer and parser, open-source Olympus<sup>5</sup> dialogue manager, and Acapela<sup>6</sup> voice-as-a-service (VAAS) speech synthesizer. The Ravenclaw-based dialogue manager simplifies the authoring of complex dialogs, having general support to handle speech recognition errors, and is extendable to multi-modal input/output design [5]. The speech recognition and parsing is based on service-tailored grammars. To operate the SUI in real-world scenarios, one of the most important issues is to provide satisfactory speech recognition accuracy as well as to manage the complexity and variety of speech-based interaction. Considering service-specific spoken contexts might apply to the SUI, we propose to handle this issue from the following aspects: (1) developing context-aware language models (i.e., grammars in this study) to help reduce speech recognition complexity [11]; (2) achieving continuous and flexible dialogue flow by switching among different tasks without restarting the speech interaction module. Another significant issue in SUI is error handling. These errors usually result from speech recognition failures or misunderstandings due to false positives. The Olympus dialogue manager uses an error handling policy based on repeated prompts for recognition failures and explicit confirmation to ground recognized concepts in order to manage these sources of error. To employ context-aware speech recognition in a dialogue system, we load grammars dynamically according to the context change of verbal interaction. At the beginning of a speech interaction, the dialogue could only be initiated by the user via saying the wake-up word, which is defined to be the name of the robot in this work. A grammar containing the full list of all available services will be loaded immediately. After a specific service is chosen by the user through speech, which can either be a short command or a complete sentence, the dialogue manager will indicate the engine to switch to the according service-specific grammar. In manipulating with the contextual information, real-time SUI operation is ensured by avoiding complex large vocabulary language models, meanwhile, more robust speech recognition performance is achieved.

Our SUI is fully employable on the Robot-Era multi-robots platform. The speech interaction mode is available for all Robot-Era services, which include *communication* (via mobile phone or skype), *shopping*, *cleaning*, *food delivery*, *indoor escort*, *object manipulation*, *garbage collection*, *laundry*, *reminding*, *surveillance*, and *mobility support*. These indoor or outdoor services are fully supported in three languages: English, Italian, and Swedish. Figure 2 shows the dialogue flow path in a real SUI operation. In this example, the Robot-Era services *food delivery*, *shopping*, *laundry*, and *garbage collection* are asked by the user sequentially. In each dialogue stage, a specific grammar is enabled. Transitions from one grammar to another are indicated by the dialogue manager. It is noted that before each dialogue movement, a confirmation with the user is made to avoid wrong action caused by speech recognition failures.

---

<sup>5</sup><http://wiki.speech.cs.cmu.edu/olympus/index.php/Olympus/>.

<sup>6</sup><http://www.acapela-group.com/>.



Fig. 2 Robot-Era services manipulated by speech

## 4 Preliminary Experiments on Usability

To evaluate the usability of our SUI by elderly users, we designed and conducted a series of tests on selected Robot-Era services. For the services *shopping*, *reminding*, *indoor escort*, and *garbage collection*, the experiments were carried out on our Italian test site in Peccioli, Italy. The services *laundry* and *food delivery* were tested on the Swedish test site in Angen, Sweden. For the last service *communication*, tests were conducted on both sites. At this stage, we recruited 35 Italian subjects (22 females and 13 males), of an average age of  $73.80 \pm 5.81$ , and 12 Swedish subjects (4 females and 8 males) aged averagely at  $70.67 \pm 5.37$  to participate in the HRI experiments.

During the experiments, subjects weren't instructed step-by-step on using the SUI, instead, they were shown a demonstration for each task. For further investigation, their spontaneous speech interaction with the robot were recorded. A questionnaire was completed by each subject. The questionnaire consisted of several items, and two of them were about the usability of the SUI. Subjects were asked to use five-point Likert type scale (1. strongly disagree; 2. disagree; 3. no opinion; 4. agree; 5. strongly agree) to answer the questions. Figure 3 shows the median scores valued by all the subjects for the two questions, each in a radar plot. It is observed that most of the subjects were satisfied with the SUI performance in the five services: *shopping*, *indoor escort*, *food delivery*, *garbage collection*, and *reminding*. For the other two services: *communication* and *laundry*, they still accepted it. In Fig. 4, the service-wise acceptability scores are described by an error bar plot. It is found that among the subjects, the SUI in all services get average scores above four out of five. In specific, the elderly users were more satisfied with the SUI experience for the four services: *shopping*, *reminding*, *garbage collection*, and *indoor escort* during the experiments.

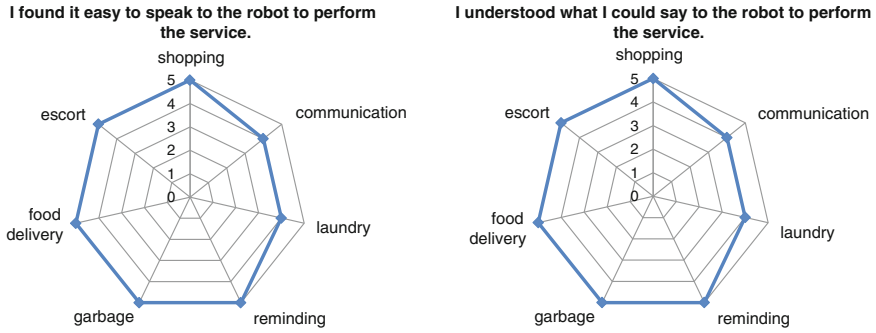


Fig. 3 User attitude towards the SUI under different application scenarios

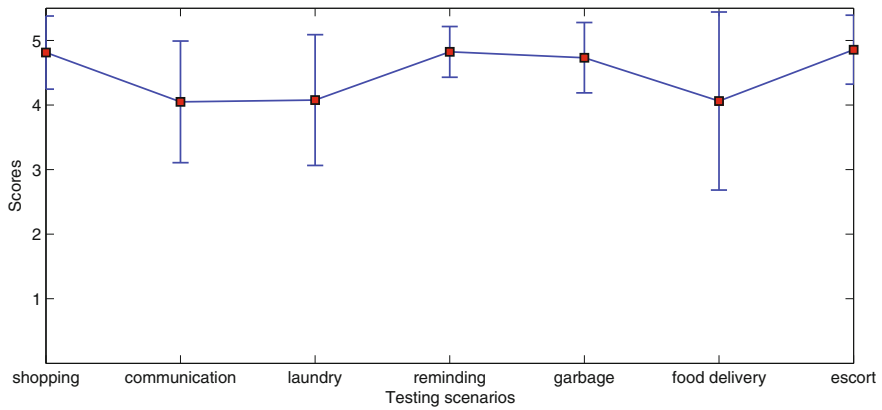


Fig. 4 Error bar plot of user acceptability scores in different testing scenarios by questionnaire interviews

## 5 Conclusions

The increasing ageing population in need of personal care and services causes huge pressure to the European countries, which turns out motivate the development of personal robots nowadays. One of the most challenging aspects for personal robots is social communication, which is especially difficult in the case of elder-robot interaction. In this paper, we introduce the design and implementation of a speech user interface specially tailored for elderly users. Although being developed under the framework of the EU FP7 Robot-Era Project, this interface can easily apply to other robotic platforms or used alone. The speech user interface is realized by a spoken dialogue system composed of the speech recognizer, parser, dialogue manager, and speech synthesizer, etc. To achieve low-complexity yet user-centric speech interaction system, we employ a context-sensitive speech recognition approach to enable flexible dialogue movement in real time. Evaluation made by elderly users in Italy

and Sweden shows that for the tested Robot-Era services, subjects totally agree that the speech interaction experience during operating the robot is acceptable. Further elder-robot interaction investigation is ongoing, it is therefore expected that more findings on elder-centric speech interface will be gained in future.

**Acknowledgments** This work is fully supported by the European Union Seventh Framework Programme (FP7/2007-2013) grant no. 288899.

## References

1. Robot-Era project: Implementation and integration of advanced robotic systems and intelligent environments in real scenarios for the ageing population. In: FP7-ICT-Challenge 5: ICT for Health, Ageing Well, Inclusion and Governance. Grant agreement number 288899
2. Active ageing and solidarity between generations. European Union (2011)
3. Al-Razgan, M., Al-Khalifa, H., Al-Shahrani, M., Al-Ajmi, H.: Touch-based mobile phone interface guidelines and design recommendations for elderly people: a survey of the literature. In: *Lecture Notes in Computer Science*, vol. 7666, pp. 568–574. Springer, Berlin Heidelberg (2012)
4. Arch, A., Abou-Zahra, S., Henry, S.: Older users online: WAI guidelines address the web experiences of older users. *User Experience Mag.* **8**(1), 18–19 (2009)
5. Bohus, D., Rudnicky, A.: The ravenclaw dialog management framework: architecture and systems. *Comput. Speech Lang.* **23**(3), 332–361 (2009)
6. Breazeal, C., Kidd, C., Thomaz, A., Hoffman, G., Berlin, M.: Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In: *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 708–713 (2005)
7. Di Nuovo, A., Broz, F., Belpaeme, T., Cangelosi, A., Cavallo, F., Esposito, R., Dario, P.: A web based multi-modal interface for elderly users of the Robot-Era multi-robot services. In: *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, pp. 2186–2191 (2014)
8. Karpov, A., Ronzhin, A., Leontyeva, A.: A semi-automatic wizard of oz technique for Let'sFly spoken dialogue system. In: *Lecture Notes in Computer Science: Text, Speech and Dialogue*, pp. 585–592. Springer (2008)
9. Kipp, A., Kummert, F.: Dynamic dialog system for human robot collaboration—Playing a game of pairs. In: *Proceedings of 2nd International Conference on Human-Agent Interaction*, pp. 225–228 (2014)
10. Lamel, L., Rosset, S., Gauvain, J., Bennacef, S.: The LIMSI arise system for train travel information. In: *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 501–504 (1999)
11. Lemon, O.: Context-sensitive speech recognition in information-state update dialogue systems: results for the grammar switching approach. In: *Proceedings of Eighth Workshop on the Semantics and Pragmatics of Dialogue*, pp. 49–55 (2004)
12. Lemon, O., Georgila, K., Henderson, J., Stuttle, M.: An ISU dialogue system exhibiting reinforcement learning of dialogue policies: generic slot-filling in the TALK in-car system. In: *Proceedings of Eleventh Conference of the European Chapter of the Association for Computational Linguistics*, pp. 119–122 (2006)
13. Mayer, P., Beck, C., Panek, P.: Examples of multimodal user interfaces for socially assistive robots in ambient assisted environments. In: *Proceedings of IEEE International Conference on Cognitive Infocommunications*, pp. 401–406 (2012)
14. Portet, F., Vacher, M., Golanski, C., Roux, C., Meillon, B.: Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects. *Pers. Ubiquit. Comput.* **17**(1), 127–144 (2013)



15. Sainath, T., Mohamed, A., Kingsbury, B., Ramabhadran, B.: Deep convolutional neural networks for LVCSR. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 8614–8618 (2013)
16. Sainath, T., Peddinti, V., Kingsbury, B., Fousek, P., Ramabhadran, B., Nahamoo, D.: Deep scattering spectra with deep neural networks for LVCSR tasks. In: Proceedings of INTERSPEECH, pp. 900–904 (2014)
17. Smarr, C., Prakash, A., Beer, J., Mitzner, T., Kemp, C., Rogers, W.: Older adults preferences for and acceptance of robot assistance for everyday living tasks. In: Proceedings of Human Factors and Ergonomics Society Annual Meeting, pp. 153–157 (2012)
18. Tang, D., Yusuf, B., Botzheim, J., Kubota, N., Chan, C.: A novel multimodal communication framework using robot partner for aging population. *Expert Syst. Appl.* **42**(9), 4540–4555 (2015)
19. Weninger, F., Schuller, B., Eyben, F., Wöllmer, M., Rigoll, G.: A broadcast news corpus for evaluation and tuning of German LVCSR systems (2014). eprint [arXiv:1412.4616](https://arxiv.org/abs/1412.4616)
20. World Health Organization: 10 facts on ageing and the life course. <http://www.who.int/features/factfiles/ageing/en/> (2014)