The Roberta IRONSIDE project

A dialog capable humanoid personal assistant in a wheelchair for dependent persons.

Hugues Sansen, SHANKAA, Paris, France, hugues@shankaa.com

Maria Inés Torres, Speech Interactive RG, Universidad del País Vasco UPV/EHU, Bilbao, Spain
manes.torres@ehu.eus

Kristiina Jokinen, University of Helsinki, Finland and University of Tartu, Estonia
Kristiina.Jokinen@helsinki.fi

Gérard Chollet, Intelligent Voice, London, UK, gerard.chollet@intelligentvoice.com

Dijana Petrovska-Delacretaz, Télécom Sud Paris, Paris France, dijana.petrovska@telecom-sudparis.eu

Atta Badii, School of Systems Engineering, University of Reading, Reading, UK
atta.badii@reading.ac.uk

Stephan Schlögl, Department of Management, Communication & IT, MCI Management Center Innsbruck, Austria, stephan.schloegl@mci.edu

Nick Campbell, SFI Stokes Professor of Speech & Communication Technology, Trinity College Dublin, nick@tcd.ie

Human-robot interaction provides a new and promising area for the study of a dialogue partners' engagement in conversational interactions, and for experimenting with various presentation strategies, their cultural differences and the impact on the user's perception, understanding, and evaluation of the interaction. Robots have also opened up novel possibilities in practical applications, e.g. in scenarios which focus on providing tools for the aging population, such as maintaining companionship, detecting health problems, and assisting dependent people. In addition, adaptation to the user can be learned. Roberta Ironside is a personal assistant humanoid robot project that emphasizes speech capabilities for interacting with dependent persons. The project provides a unique opportunity for students to work on a real application and study verbal and non-verbal interaction possibilities (technical, dialogue, and social) between a human user and a human-sized robot.

2. **Project objectives (max. 1 page):**

Roberta Ironside is an Open Source indoor/outdoor humanoid assistant robot in an electric wheelchair designed to help dependent people. The project covers: speech processing and dialog management, affective computing, human behavior analysis, and human robot interaction and assistance. In this proposal we focus on the development of Roberta's **LifeLine** module, demonstrating its conversational skills.

The students will work on dialogue management. The task is to produce clear, natural, engaging, and intuitively easy-to-follow mixed-initiative interactions where both the user and the system are engaged in satisfying and smooth conversational-type situations. While the overall goal of Roberta Ironside is that both the human as well as the robot can initiate topics and ask questions, for this workshop we will focus on the initial encounter and presentation dialogs (i.e. Roberta provides general information e.g. about Calais, Paris, Twente, etc. and then asks the user to talk about related memories e.g. their summer in Calais, their life in Paris, etc.). In addition to spoken input Roberta will be able to obtain information from the environment and from the system through a dialog manager that integrates this information within its dialog tasks and strategies. It will be designed to help people tell stories about their lives in order to help them exercise their speech and memory capabilities. It further aims at acquiring a good level of knowledge about the person and his/her context and thus is expected to feature an open-domain conversational system, presenting useful and interesting information to the user. Roberta will observe user behavior, based on which it should infer a user's emotion and interest levels and tailor its presentation accordingly. The initial conversational subjects will deal with personal data, the user's life, and picture analysis. Additionally, Roberta will have a broad set of other discussion topics covering various domains students may use (if the timeframe allows for it) to develop their own interactive systems. In such a case, the system may obtain information from a suitable digital repository.

The novel contributions of this project can thus be summarized in providing a real world application:

- To build a robot-driven spoken dialog system that integrates multimodal observations and improved face recognition technologies into its dialog models,
- to find ways of using these technologies to identify basic user emotions, interactions goals and context,
- to utilize this data (e.g. context, dialog history, emotions, user goals, etc.) so as to establish a closer conversational relationship with the user,
- and based on this to eventually developed dialogue management techniques, especially presentation techniques, that take into account not only a user's spoken input but also his/her non-verbal signals

3. **Background information (max. 1 page):**

The system to be developed consists of an input perception layer, an understanding module, an interaction module that includes a dialog manager, a dialog history and context aware blackboard and an open domain conversational agent that will get information from a digital repository and a final reaction layer. Although all these components will eventually be integrated into the body of a humanoid robot, for the eNTERFACE work group this type of physical system representation can be neglected. That is, while we will try to bring an initial robot prototype the development can be based robot mockups and simulations.

Perception layer:

Automatic Speech Recognition will be distributed between a local and a remote module. Locally, an HMM / DNN decoder running on an NVIDIA GPGPU platform will be available. Remote access to Intelligent Voice, Google, Microsoft and other servers will be used to manage richer linguistic data and knowledge bases. A number of languages are available but English will be the preferred one for the workshop.

Audio-visual identification will be based on open-source reference software developed and maintained by Telecom-SudParis. These include 2D face recognition, speaker verification and synchronization of lip movements with the speech signal.

Understanding module: The proposal includes the Phoenix Semantic decoder (Ward, 1994) for Natural Language Understanding purposes, a Semantic parser that uses statistical classifiers outperform semantic decoders that use manually constructed semantic grammars (Henderson et al, 1992). However they need to be trained by some amount of sentences representing the task.

Dialog Manager: The DM will be based on the Ravenclaw Dialog manager (Bohus and Rudnicky, 2009), which is a task independent agenda based DM. In research DMs are based on Statistical approaches, where POMDP (Young et al, 2013) is considered the state-of-the art. However, agenda-based dialog managers are used in many commercial systems as well as in very specific tasks. Moreover they have been extensively used as initial systems aimed at collecting real user dialogs that would further train a statistical DM. The DM is also extended with components taking care of multimodal information, and with components focusing on an appropriate presentation.

Open-domain conversational agent: One of the main problems in the acceptance of spoken dialogue systems is their boring presentation strategies and their lack of sensitivity to the user's needs, interests, and emotional state. In the workshop we will draw inspiration from concrete open domain information providing an application called WikiTalk (Jokinen & Wilcock 2014). This is a robot interface that

allows the users to navigate digital information repositories such as Wikipedia. It uses the robot, which enables interactions with the user via speech and/or gestures. The concept of affordance is often mentioned in interface design, and this will also be studied with regard to presentation strategies and modalities in dialogue contexts.

Output layer:

Text to speech uses OpenMary. Gesture and facial animation are currently under development.

Wizard of Oz: In order to design and shape interactions with complex technical solutions such as the ones envisaged by the Roberta Ironside project, researchers require tools and methods that allow for the early stage prototyping and evaluation of potential development directions. One approach often used for such low-fidelity evaluations is Wizard of Oz (WOZ) (Dahlbäck et al.). WOZ constitutes a prototyping method employed by researchers and designers to obtain initial feedback on features that usually require significant resources to be implemented. In a so called 'WOZ experiment' a human 'wizard' mimics the functionality of a system, either entirely or in part, which allows for the evaluation of potential user experiences and interaction strategies without the need for building the actual product. Language based interaction with systems has been one of the main fields of application for this type of prototyping. Similar to products whose interaction is based on Graphical User Interfaces (GUIs), also technology that uses speech and other forms of natural language (e.g. gestures) as an interaction channel requires thorough testing and evaluation during its design and development phase. While low-fidelity prototypes assessing GUI applications (e.g. sketches, storyboards, wireframes, etc.) can be built quickly and relatively inexpensively. The development of prototypes for natural language-based interaction tends to be rather costly and time intensive. For example, in order to be able to obtain a basic understanding of an envisioned spoken interaction, systems may require several hours of recorded speech, valid transcriptions, and an implemented framework of rules, as well as a solid error-recovery strategy. Obtaining these types of resources usually requires several person months of work. Here WOZ experimentation can help collect data as well as test overall design directions.

4. **Detailed technical description (max. 3 pages):**

   a. **Technical description:**

Roberta's problem solving is based on a blackboard like architecture with dedicated knowledge source processing units and a global controller. For the eNTERFACE project, the architecture is limited to speech, face recognition, and emotion as described on the right hand side of the following figure.

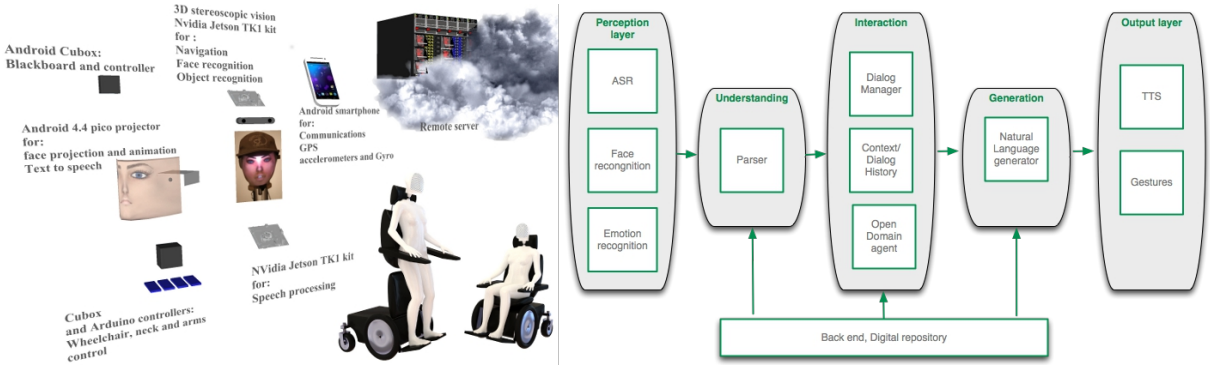

Figure 1: Left hand side: General architecture of the Roberta Ironside project.Right hand side: Functional architecture diagram of the speech system proposed for the project.

A dialog starts with visual face recognition and is adapted to the context as illustrated in figure 2 and the following use cases.

<u>Use case example 1: First meeting, person identification and start of a LifeLine dialog</u>

Roberta: "Good morning, my name is Roberta. I do not think we ever met."

Veronica: "Good morning Roberta. I am Veronica Doe."

Roberta: "Nice meeting you Veronica. Could you please spell your last name for me?"

Veronica: "Doe, D. O. E., Veronica Doe."

Roberta: " Veronica, do you mind if I ask you some questions to know more about you? I would like to hear about your life. All what you say will stay between you and me."

Veronica: "Nice meeting you Roberta. Please do."

Roberta: "Veronica, from what I understand you are a women. Can you confirm?"

Veronica: "Yes, I am a woman."

Roberta: "Veronica, can you tell me when you were born?"

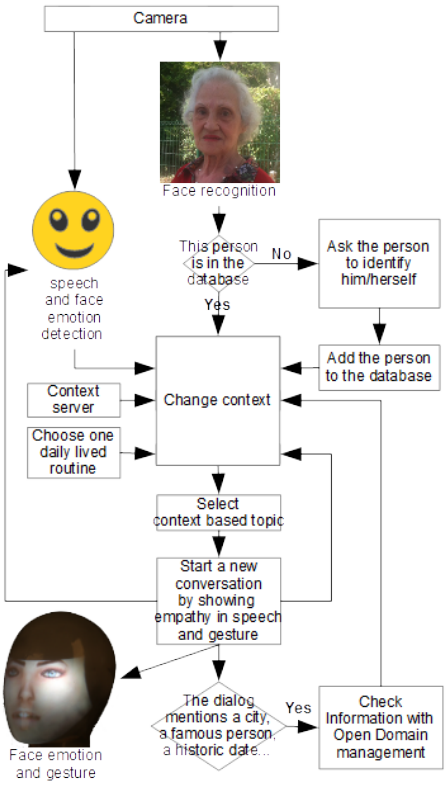

Figure 2: Simplified dialog flow. Elderly person image credit: Estelle Mansau

etc.

Use case example 2: Dialog enrichment with OpenDomain management

Roberta: " You said you used to live in Ensraday until the 70's before emigrating to the US. Could you please spell it for me"

Veronica: " E.N.S.C.H.E.D.E"

Roberta: "Can you confirm it is Enschede in the Netherlands?"

Veronica: " Yes, in the Netherlands"

Roberta: " Enschede had an important textile industry. Did you work in textile?"

etc.

Non exhaustive specification and use cases will be provided to the team leaving space for creativity.

Wizard of OZ: We plan to use WOZ as a means to design and evaluate different human-robot interaction scenarios before actually building and optimizing them. To do so we will use the WebWOZ prototyping platform (Schlögl et al.) and expand its features to better support the given setting of robots assisting people. WebWOZ is an open source WOZ tool, which currently focuses on supporting the integration of various language based interaction components via web services (e.g. Automatic Speech Recognition, Machine Translation, Text-to-Speech Synthesis, etc.). It already supports the use of pre-defined/pre-recorded audio and video files as well as on-the-fly generated responses, so that restricted as well as open dialogues may be tested. Going from here we plan to expand the number of available interaction modalities so that not only speech based interaction but also interactions that require additional modalities or sensor readings may be simulated (e.g. gestures, location-based services, etc.). This will help us test and evaluate multimodal human-robot interactions and consequently support the identification and optimization of scenarios where this type of interaction is both helpful and socially accepted.

User recognition In order to allow for a realistic dialogue with the human, it is important for Roberta to recognize what is the identity of the person with whom it is interacting. Télécom-SudParis will contribute an open source reference systems for recognizing persons based on their voice and 2D images. Their first version is described in *Petrovska-Delacrétaz, D., Chollet, G., & Dorizzi, B. (2009). Guide to biometric reference systems and performance evaluation. Springer*, and the software with guidelines for reproducible results can be found at / http://share.int-evry.fr/svnview-eph/

ASR: Multi-lingual Large Vocabulary Speech Recognition based on Deep Neural Nets will be provided by Intelligent Voice.

Emotion recognition: Basic feature emotion recognition from speech and face, provided by the University of Reading.

Vision: Although vision per se is not the purpose of this limited project we will use the Google vision API as a service to allow for a picture description and an understanding of the image analysis.

Understanding: The Phoenix Semantic decoder (Ward, 1994) is a robust semantic decoder. It is a hand crafted semantic decoder that has been extensively used in Spoken Dialog Systems. Thus there are many grammars that can be reused for this project (wiki.speech.cs.cmu.edu). Yet a set of task specific grammars have to be written manually.

Dialog Manager: The Ravenclaw DM is an agenda-based dialog manager that supports mixed initiative dialogues. The dialogue task is represented by a tree whose nodes represent the hierarchical structure of the dialogue, i.e. the pan for the interaction. This tree has to be manually designed using the Ravenclaw Task Specification Language (RTSL). To this end a set of dialog agents have to be defined. Each agent definition also includes a set of preconditions, triggers and succes/failure criteria to be managed at execution time, completing the embedded action definition to be taken by the DM.

Open domain management: We will build various presentation alternatives on the basis of the RavenClaw task tree (aka topic tree) and extend its presentation with special (software) agents that deal with more delicate presentation alternatives and multimodality. The course can draw inspiration from a concrete open domain information providing an adapted version of the Nao WikiTalk application. This is a robot interface that allows the users to navigate digital information repositories such as Wikipedia. It has been initially designed to use the Nao robot which enables interactions with the user via the help of speech and gesturing. Presentation of information in multilingual and multicultural contexts need to be taken into account and coordination of the modalities (e.g. speech and simultaneous gesturing) are to be studied carefully to avoid mismatches in timing and errors in information processing, which are confusing and affect the user's perception of the robot agent (Jokinen and Wilcock, 2014).

B. **Resources needed: facility, equipment, software, staff,…**

For the proposed workshop we will require

- a RavenClaw Dialog Manager (server based software provided by UPV),
- WikiTalk, Open-domain conversational agent (software provided by EFU),
- Context aware blackboard (software provided by University of Reading)
- a speech recognition/synthesis system (software provided by IMT/Intelligent Voice)
- a facial recognition system (software provided by IMT)
- a Wizard of Oz System (software provided by MCI)
- an interactive robot agent provided by Shankaa, integration mock-up consisting of:
  1. a torso,
  2. a head with active face, (with Android based projector).
  3. a 3 axis neck,
  4. a 3D stereoscopic camera,
  5. a 8 microphone array,
  6. NVIDIA Tegra TX1 based TV Shield,
  7. An independent tablet.
- computers (students have to bring their own devices)
- a good Internet connection (hopefully provided by U. Twente)

As for the students we need people who are able to write component software (good Unix programming skills; Android and Java skills appreciated) and/or be able to specify requirements clearly so that they can be implemented in the given framework. Ultimately, we are looking for dedicated PhD students who have keen interest in Human-Robot Interaction, designing and developing technology, and who like to conduct interesting experiments. The group as a whole should have various complementary skills, including people who can program, design interactions, design experiments and user studies, and conduct statistical studies. Overall, however, they should all be interested in HRI.

C. **Project management.**

Project management will use TRAC 1.0 or higher.
Trac is an enhanced wiki and issue tracking system for software development projects. Trac uses a minimalistic approach to web-based software project management. It helps developers write great software while staying out of the way. Trac should impose as little as possible on a team's established development process and policies.
It provides an interface to Subversion as well as other version control systems, an integrated Wiki and convenient reporting facilities.

Trac allows wiki markup in issue descriptions and commit messages, creating links and seamless references between bugs, tasks, changesets, files and wiki pages. A timeline shows all current and past project events in order, making the acquisition of an overview of the project and tracking progress very easy. The roadmap shows the road ahead, listing the upcoming milestones.

All the persons involved in the project: project team and project managers will have a Trac account to follow issues and project evolutions.
All the code and architecture should be commented in TRAC

.

5. Work plan and implementation schedule (max. 1 page):

Week 1:

Introduction, design, integration, and programming skills

- Woz-studies for data collection (i.e. interaction data to inform dialog management, acoustic data to support/train speech recognition, visual data to provide some basic features for emotion detection, etc.),
- Basic issues related to task requirements and specification
- Planning of interaction modeling
- Install Phoenix, run Phoenix (local and remotely), learn how to define grammars,
- Run Ravenclaw, learn the RavenClaw Task Specification Language (RTSL), etc.
- Check and assure connectivity between sites (UK, Netherlands, Spain,….) and servers.

Week 2 and 3:

- Speech processing integration via the Jetson TK1 kit
- Face recognition and emotion analysis
- Object recognition and physical context analysis
- Define grammars for Phoenix
- Plan the interaction task (and error recovering strategies) on RTSL.
- Insert non verbal information and context
- Task structure for the scenario, tasks structure modified to topic structure
- Extension with multimodal components and sophisticated presentation strategies
- Face animation and non verbal signals,
- NLG and TTS
- First interactive models with advanced presentation strategies and techniques
- Experiments related to user engagement using the developed model

Week 4:

Final evaluation, paper/report writing, wrapping up

- finalize code comments on the trac
- prepare instructions for download by future developers/users of the system
- Prepare the final demo
- Design a conference paper
- Design future work

6. Benefits of the research

The students should learn to
- design and develop conversational HRI systems,
- realize effective, interesting, and engaging response planning and presentation strategies for HRI
- use suitable multimodal components to study engagement in HRI
- evaluate HRI systems with users
- understand the effect of system responses on the user's perception of the system
- understand the complexity of human-robot interaction
- possible future research projects

Finally note that this project has been proposed by an international team with large experience in speech, language and interaction technologies and that it includes different profiles and approaches. Students will benefit from this diversity and scientific discussions among members.

7. **Profile team:**

b. Leader

Hugues Sansen

Hugues Sansen has started his career in the merchant marine with a Captain and Chief Engineer diploma before learning Artificial Intelligence and robotics in Marseilles.

In 2007, he has founded Shankaa, a French SME focused on telecom technologies for health. He has participated in the European Ambient Assisted Living project vAssist and is proposing the Roberta Ironside project as a continuation of vAssist benefiting from the strength and weakness analysis of the solution. He is also the Technical director of Assortiment Conseil for which he has developed the retail store performance analysis tool: Visumag. He regularly offers consulting services in heavy industry, health and finance when a broad knowledge is necessary.

Prior to Shankaa, he has developed a peer to peer VoIP solution which led to the creation of Peer2Phone in early 2003 and the GoSIP VoIP softphone for mobiles. He has proposed two navigation systems one based on cellular phone radio capacity, and one based on a hyperbolic usage of FM radio frequencies. He has worked as regional technical manager or regional sales manager for software houses such has Soprano Design (telco), Tibco Systems (telco, industry, and finance), Innovatron (smartcard inventor), Gemstone Systems (Object oriented database). He has worked 5 years on the conception of the French aircraft carrier as a member of the nuclear safety team. In the late 80's, he has created a blackboard based expert system engine in Smalltalk, based on logic 2+ and fuzzy logics that he has applied to ship-planning with ShipPlanner, the first expert system for the design of the loading plans of container ships.

He has a technical expertise in object oriented programming, artificial intelligence (expert systems and neural networks), telecoms, and radio navigation systems and has knowledge in speech processing and vision.

From his merchant marine education, he is a mechanical, electrical and electronics engineer.

c. Staff proposed by the leader .

Kristiina Jokinen is Adjunct Professor and Project Manager at University of Helsinki where she leads the 3I (Intelligent Interaction and Information Systems) Research Group. She is Visiting Professor at University of Tartu, Estonia. She has been a JSPS Fellow at NAIST (Nara, Japan) and Invited Researcher at ATR (Kyoto, Japan), and played a leading role in many academic and

industrial research projects. She is Secretary-Treasurer of SIGDial. She has served in several program and review committees, and organized international workshops and conferences, e.g. the recent IWSDS-2016 in January 2016 in Finland. She has also organized several research training courses, e.g. on Multimodal Communication in Tartu in 2014 and an eNTERFACE course on Nao robot in Metz in 2012. Her research focuses on interaction technology, multimodal communication (gesturing, eye-gaze), spoken dialogue systems, human-human and human-robot interaction, corpus collection, annotation, and analysis. Besides articles and papers, she has published three books: "Constructive Dialogue Modelling - Speech Interaction and Rational Agents" (John Wiley), "Spoken Dialogue Systems" (with M.McTear; Morgan & Claypool), and "New Trends in Speech-based Interactive Systems" (edited with F.Chen; Springer).

Prof. M. Inés Torres received her PhD in Physics from the UPV/EHU in 1990, including an internship at the Centre National d'Études des Télécommunications in Lanion (France) in 1988. She was also a visiting researcher at the Polytechnic University of Valencia (Spain) during the years 1991 and 1992. She was a member of the board of the Spanish Association of Pattern Recognition (IAPR), from 1995 to 2008. She is currently a Full Professor of Computer Science at the UPV/EHU where she has also held several academic management positions. She founded the Pattern Recognition and Speech Technology research group in 1990, which she has been leading ever since. Professor Inés Torres was also a visiting Faculty at the Language Technologies Institute in Carnegie Mellon University in 2012. She has published more than one hundred works in journals and international conferences, about thirty in national ones and edited three books. She has led many research projects as well as research under contract with technical centres and companies.

Dr. Gérard Chollet studied Linguistics, Electrical Engineering and Computer Science at the University of California, Santa Barbara where he was granted a PhD in Computer Science and Linguistics. He taught at Memphis State University and University of Florida before joining CNRS. In 1981, he was in charge of the speech research group of Alcatel. In 1983, he joined a newly created CNRS research unit at ENST (Telecom-ParisTech within Institut Mines-Telecom). In 1992, he joined IDIAP, a new research laboratory of the `Fondation Dalle Molle' in Martigny, Switzerland. IDIAP contributed to SpeechDat, M2VTS and other European projects. From 1996 to 2012, he was back full time at ENST, managing research projects and supervising doctoral work. Funding was secured from such projects as Eureka-Majordome and MajorCall, NoE-BioSecure, Strep-SecurePhone, IP-Companion@ble, AAL-vAssist, FET-ILHAIRE, etc. He supervised more than forty doctoral thesis. CNRS decided in July 2012 to grant him an emeritus status. He visited Boise State University in 2013 and the University of Eastern Finland in 2014. He is now VP of Research at Intelligent Voice. His

main research interests are in phonetics, automatic audio-visual speech processing, speech dialog systems, multimedia, pattern recognition, biometrics, privacy-preserving digital signal processing, speech pathology and speech training aids.

Dijana Petrovska-Delacrétaz (PhD EPFL 1990) obtained her degree in Physics and her PhD from the Swiss Federal Institute of Technology (EPFL) in Lausanne. She was working as a Consultant at AT&T Speech Research Laboratories and was as a Senior Scientist for four years at the Informatics Department of Fribourg University, Switzerland. Since 2004 she is an associate professor within the Mines-Télécom SudParis Intermedia group. She participated actively to the coordination of the FP6 NoE BioSecure (related to Multimodal Biometrics), and co-organized in 2005 the 1st BioSecure Residential Workshop of a one month duration with more than 100 participants.

Dr. Stephan Schlögl holds an MSc in Human-Computer Interaction from University College London and a PhD in Computer Science from Trinity College Dublin. Dr. Schlögl's main research interests lie in the design, exploration and adaptation of natural language based human-computer interfaces. In his doctoral research he investigated Wizard of Oz prototyping as a design and research instrument. This work was continued through a post-doctoral position at Institute Mines Télécom, Télécom ParisTech, where Dr. Schlögl joined the vAssist project team. In November 2013 Dr. Schlögl became an MCI faculty member. In addition to following his research interests he is also involved in teaching activities, holding courses in User Experience, Software Engineering, and Artificial Intelligence.

Prof. Nick Campbell's background is in experimental psychology and linguistics, but most of his practical experience is in speech technology. He is an advocate of corpus-based approaches and he has pioneered advanced (and paradigm-shifting) methods of speech synthesis and natural conversational speech collection in a multimodal environment. His principal interest is in multi-modal and nonverbal speech processing, extending this research to social interaction to show how the voice is used in discourse to express personal relations alongside propositional content. Most of his previous work has used speech materials collected in Japan and, through his move to Ireland, he now works at TCD to confirm the universality of his previous findings and to produce a friendlier speech-based human-machine interface for web-based information, customer-services, games, and robotics, while trying to understand how humans perform such often perfect communication. His recent work includes biosensors in interactive dialogue to estimate user cognitive states from physical signals.

Prof. Atta Badii is a high ranking professor at the University of Reading where he is the Director of the Intelligent Systems Research Laboratory, at the School of Systems Engineering. He holds the Chair of Secure Pervasive Technologies (UoR) and the designation of Distinguished Professor of Systems Engineering and Digital Innovation (UCC) and is an International Privacy-by-Design Ambassador as designated by the Canadian Information and Privacy Commission. Atta is Director of the European Virtual Centre of Excellence for Ethically-guided and Privacy-respecting Video-Analytics (VideoSense), Director of the European Observatory for Crowd-Sourcing and Collective-Awareness Platforms for Transformative Government (Citizens' Say), Chair of the International Companion Robotics Institute (CRI), and, the Internet of People Things and Services (IoPT) Research Forum as well as the European Privacy by Co-Design Research Cluster. He is also a panel member of the Companion Robotics and Human-Robot-Interaction research roadmap groups of the European Robotics Research Forum and has previously served on other European and UK research roadmap panels e.g. as the Chair of the European Security Architectures and Virtualisation Taskforce of the SECURIST Research Roadmap Initiative and as a leading member of the TSB-RCUK Expert Group for Secure Internet of Things Research Strategy.

d. Other researchers needed (describing the required expertise for each, max. 1 page).

8. **References:**

Anastasiou, D., Jokinen, K., Wilcock, G. (2013). Evaluation of WikiTalk ñ User Studies of Human-Robot Proceedings of the Interaction HCI International Conference, July 2013. In C. Stephanidis (Ed.): Universal Access in Human-Computer Interaction, Part II, Lecture Notes in Computer Science (LNCS).

http://wiki.speech.cs.cmu.edu/olympus/index.php/RavenClaw
Bohus, D., Rudnicky, A. (2009) - The RavenClaw dialog management framework: architecture and systems, in Computer Speech and Language, Volume 23, Issue 3, pages 332–361

Csapo, A., Gilmartin, E., Grizou, J., Han, J., Meena, R., Anastasiou, D., Jokinen, K., Wilcock, G. (2012). Multimodal Conversational Interaction with a Humanoid Robot. In Proceedings of the 3rd IEEE International Conference on Cognitive Infocommunications (CogInfoCom) 2012, Kocice, Slovakia, pp 667-672.

Dahlbäck, N., Jönsson, A., and Ahrenberg, L. (1993). Wizard of Oz studies - Why and How. In Proceedings of the ACM International Workshop on Intelligent User Interfaces, Orlando, Florida, USA, pp. 193-200.

Henderson, M.; Gasic, M.; Thomson, B.; Tsiakoulis, P.; Kai Yu; Young, S., "Discriminative spoken language understanding using word confusion networks," in Spoken Language Technology Workshop (SLT), 2012 IEEE , vol., no., pp.176-181, 2-5 Dec. 2012

Jokinen, K. (2009). Constructive Dialogue Modelling ñ Speech Interaction and Rational Agents. John Wiley & Sons, Chichester, UK. http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470060263.html

Jokinen, K. and McTear, M. (2009). Spoken Dialogue Systems. Synthesis Lectures on Human Language Technologies. Morgan and Claypool. doi:10.2200/S00204ED1V01Y200910HLT005 http://www.morganclaypool.com/doi/abs/10.2200/S00204ED1V01Y200910HLT005?journalCode=hlt

Jokinen, K. and Cheng, F. (2010, eds.). New Trends in Speech-based Interactive Systems. Springer Publishers.

Jokinen, K. and Pelachaud, C. (2013). From Annotation to Multimodal Behaviour. In Rojc, M. and Campbell, N. (Eds.) Co-verbal Synchrony in Human-Machine Interaction. Chapter 8. CRC Press, Taylor & Francis Group, New York.

Jokinen, K. and Wilcock, G. (2014). Multimodal Open-domain Conversations with the Nao Robot. In: Mariani, J., Rosset, S., Garnier-Rizet, M. and Devillers, L. (eds.) Natural Interaction

with Robots, Knowbots and Smartphones - Putting Spoken Dialog Systems into Practice. Springer Science+Business Media. pp. 213-224

Jokinen, K. (2010). Rational Communication and Affordable Natural Language Interaction for Ambient Environments. In: Nakamura, S., Geunbae Lee, G., Mariani, J., and Minker, W. (Eds.) The Second Workshop on Spoken Dialogue Systems Technology. Springer Lecture Notes in Computer Science. Vol. 6392 pp. 163-168.

Jokinen, K., H. Furukawa, M. Nishida and S. Yamamoto (2013). Gaze and Turn-taking behaviour in Casual Conversational Interactions. ACM Transactions on Interactive Intelligent Systems (TiiS) Journal, Special Section on Eye-gaze and Conversational Engagement, Guest Editors: Elisabeth AndrÈ and Joyce Chai. Vol 3, Issue 2.

Levitski, A., Radun, J. and K. Jokinen (2012). Visual Interaction and Conversational Activity. In Proceedings of The 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction: Eye Gaze and Multimodality, at the 14th ACM International Conference on Multimodal Interaction (ICMI-2012), October 26, 2012, Santa Monica, California, U.S.

Sansen H., Torres M-I., Chollet G., Petrovska-Delacretaz D., Glackin C., Schlögl S., Badii A., & Boudy, J. (2016). The Roberta-Ironside project. ATSIP, Monastir, March 2016.

Sansen H, Chollet G., Glackin C., Jokinen K, Badii A., Torres M-I., Petrovska-Delacretaz D., Boudy, J, Schlögl S., (2016). The Roberta IRONSIDE project, A cognitive and physical robot coach for dependent persons, Handicap 2016, Paris, June 2016.

Sansen H., Baldinger J-L., Boudy J.,Chollet G.,Milhorat P., Schlögl S., vAssist: Building the personal assistant for dependent people, Healthinf 2014

Schlögl, S., Doherty, G., & Luz, S. (2015). Wizard of Oz Experimentation for Language Technology Applications: Challenges and Tools. Interacting with Computers 27(6), 592-615.

Schlögl, S., Milhorat, P., Chollet, G., & Boudy, J. (2014). Designing Language Technology Applications: A Wizard of Oz Driven Prototyping Framework. In Proceedings of the EACL Conference of the European Chapter of the Association for Computer Linguistics, pp. 85-88. Gothenburg, Sweden: April 28-30.

Schlögl, S., Luz, S, & Doherty, G. (2013). WebWOZ: A Platform for Designing and Conducting Web-based Wizard of Oz Experiments. In Proceedings of the SIGdial Meeting on Discourse and Dialogue, pp. 160-163. Metz, France: August 22-24.

Schlögl, S., Chollet, G., Garschall, M., Tscheligi, M., & Legouverneur, G. (2013). Exploring Voice User Interfaces for Seniors. In Proceedings of the PETRA International Conference on Pervasive Technologies Related to Assistive Environments. Island of Rhodes, Greece: May 29-31.

Schlögl, S., Chollet, G., Milhorat, P., Deslis, J., Feldmar, J., Boudy, J., Garschall, M., & Tscheligi, M. (2013). Using Wizard of Oz to Collect Interaction Data for Voice Controlled Home Care and Communication Services. In Proceedings of the IASTED SPPRA International Conference on Signal Processing, Pattern Recognition and Applications, pp.511-518. Innsbruck, Austria: February 12-14.

Vels, M. and Jokinen, K. (2015). Detecting Body, Head, and Speech Signals for Conversational Engagement. In proceedings of the IVA Workshop "Engagement in Social Intelligent Virtual Agents" (ESIVA-2015), August 25, 2015, Delft, The Netherlands.

Ward, Wayne (1994): "Extracting information in spontaneous speech", In ICSLP-1994, 83-86.

Wilcock, G. and Jokinen, K. (2014). Advances in Wikipedia-based Interaction with Robots. The ICMI Workshop Multimodal, Multi-Party, Real-World Human-Robot Interaction. Istanbul, Turkey. pp.13-18.

Steve Young, Steve, Milica Gaˇsi´c, Blaise Thomson and Jason D. Williams, “POMDP-Based Statistical Spoken Dialog Systems: A Review”, Proceedings of the IEEE, Vol. 101, No. 5, May 2013

9. **Other information (optional):** please add other information you consider useful for the project proposal evaluation.