AMSE JOURNALS-AMSE IFRATH Publication –2016- Modelling C; Vol. 77; N° 2; pp 169-181
Submitted Sept. 2016; Accepted Oct. 31, 2016

The Roberta IRONSIDE Project
A Cognitive and Physical Robot Coach for Dependent Persons

*H Sansen, **G Chollet, C Glackin, ***K Jokinen,****A Badii,
*****M Torres, ******D Petrovska-Delacretaz, J Boudy, *******S Schlögl

*Shankaa, France, Thiverval, (hugues@shankaa.com)
**Intelligent Voice, UK, London,
(gerard.chollet@intelligentvoice.com, neil.glackin@intelligentvoice.com)
***University of Helsinki; Finland, (kristiina.Jokinen@helsinki.fi)
**** School of Systems Engineering, University of Reading
UK, Reading, (atta.badii@reading.ac.uk)
*****Speech Interactive RG, Universidad del País Vasco UPV/EHU
Spain, Bilbao, (manes.torres@ehu.eus)
******Télécom Sud Paris
France, Paris, (dijana.petrovska@telecom-sudparis.eu, jerome.boudy@telecom-sudparis.eu)
*******Department of Management, Communication & IT MCI Management Center
Austria, Innsbruck, (stephan.schloegl@mci.edu)

Abstract

Many elderly and dependent people living at home suffer from a lack of social contact. With their strength and physical condition decreasing, they are also reluctant to walk outside. Even in sheltered accommodation or hospitals, they may have phases of loneliness when they are left alone or when the personnel cannot continuously take care of them. With an aging population and the financial difficulties of having a full time caregiver for every dependent person living at home, the proliferation of advanced assistant robots seems to be a viable future solution. However, as most of what can be done with a robot is also possible without it, it is sometimes difficult to quantify the real value this technology can add to the current situation. Hence we believe that such a robot should be a reliable assistant, capable of helping a person indoors as

well as outdoors. Furthermore, it should be a companion for dialoging, as well as a system capable of detecting health problems. The Roberta Ironside project proposes to build such an advanced robot. That is, we want to build a cognitive coach that on the one hand fulfills the lack of social contact, and on the other hand may transform into a physical coach when going outside. In this paper we propose an affordable and simplified design for this human-sized humanoid. It starts with an overall description of the robot, followed by a justification for the choice of putting it into an electric wheel-chair. Next, it emphasizes the technology that is used for the head and the face, and then ends with a discussion of the verbal and non-verbal communication capabilities of the robot, highlighting its distinct characteristics as an Embodied Conversational Agent.

Key words

Humanoid Personal Assistant Robot; Assisted Living; Natural Language Processing; Spoken Dialogue; Embodied Conversational Agent, Cognitive Coach, Robot Assisted Physical Coaching

1. Introduction

Life assistant robots for dependent people are beginning to arrive on the market (cf. Pepper from Aldebaran, Buddy from Blue Frog, etc.). To keep a low price tag, most robotic solutions advocate a simplistic approach, essentially comprising a 'tablet on wheels'. Furthermore, they are typically designed for indoor tasks. At the other extreme, robots such as the Robear from RIKEN and systems developed by the SRK Collaboration Center for Human-Interactive Robot Research [6] have the capability of lifting a patient from a bed into a wheelchair. In parallel, there is now a focus on the development of more humanoid-like robots with human robot interaction capability [33].

The Roberta Ironside project proposes a human-sized humanoid robot that sits on an electric wheelchair to serve as a companion, a cognitive as well as physical coach for dependent people.

Putting the robot into a wheelchair allows for both indoor and outdoor activities. In addition, the project emphasizes the need for human-robot interaction with advanced speech and non-verbal communication channels. Besides having a simpler architecture than the one used with bipedal robots, a humanoid robot on a wheelchair has many benefits, not least of which is a battery capacity that may significantly exceed traditional bipedal solutions, allowing for the embedding of greater computational power. The second benefit is that, by using a platform such as an electric wheelchair, the robot is de facto adapted for indoor and outdoor evolution, with the additional benefit of being able to go almost unnoticed in public places.The Positioning and Restrictions of the Concept

Taking the overall system architecture discussed above, the proposed system has to meet a number of additional restrictions. For example, it has to be cost efficient, which means if one wants to compare the system to a human assistant, its 3-year cost should not exceed 15k€, which suggests a limit for material costs of around 5k€. Next, the system should help a person go outdoors almost unnoticed and provide support for walking and shopping. It has to be capable of dialoging with a person, and consequently make up for a great part of the social interaction most elderlies are missing. This implies not only a need for advanced speech capabilities but also for associated deep learning strategies. Furthermore, the system requires an expressive face and an articulated neck so as to allow for non-verbal communication. It needs to have a reduced dependence on a network connection. That is, it has to be operational even if no Internet access is available, which implies a need for significant embedded computing power. It also needs arms to fetch objects or to apply medical devices such as an ultrasound scanner onto its user. Last but not least, it requires large batteries.

Not all apartments or houses allow the free movement of a wheel chair, and this significantly restricts the potential market for such a robot. However, Roberta Ironside is not targeting the mass market, but rather aims at installing a few hundred units a year, with a deployment limited to apartments that can receive wheelchairs, adapted homes, sheltered accommodation and hospitals. Future versions of the robot may be adaptable to other types of accommodation but this is currently outside the scope of this project.

2. Overall Description

Roberta does not target the DARPA robotics challenge [9]. Material costs should not significantly surpass 5k€, which implies simple and pragmatic solutions and probably different technological approaches. Basically, Roberta Ironside can be seen as an articulated mannequin on an electric wheelchair - the wheelchair being the traveling platform. The electric wheelchair is capable of two positions: sitting and standing. The different positions are particularly convenient for analyzing the person's vital signs while they are abed, or fetching an object from a shelf respectively. By choosing an electric wheel-chair we benefit from a high battery capacity that guaranties continuous usage throughout an entire day without having to recharge, as well as a space autonomy of about 40km. The computing architecture (Figure 1) is composed of several minicomputers that are dedicated to specific tasks, and are organized according to a tree structure similar to the blackboard architecture used in classical expert systems [26]. Some of these minicomputers are massively parallel computer systems based on the Nvidia Jetson TX1 kit (e.g.

for 3D vision, speech, deep learning, etc.), others are simple Android PCs or phones. Arduino kits and shields are used as stepper motor controllers.

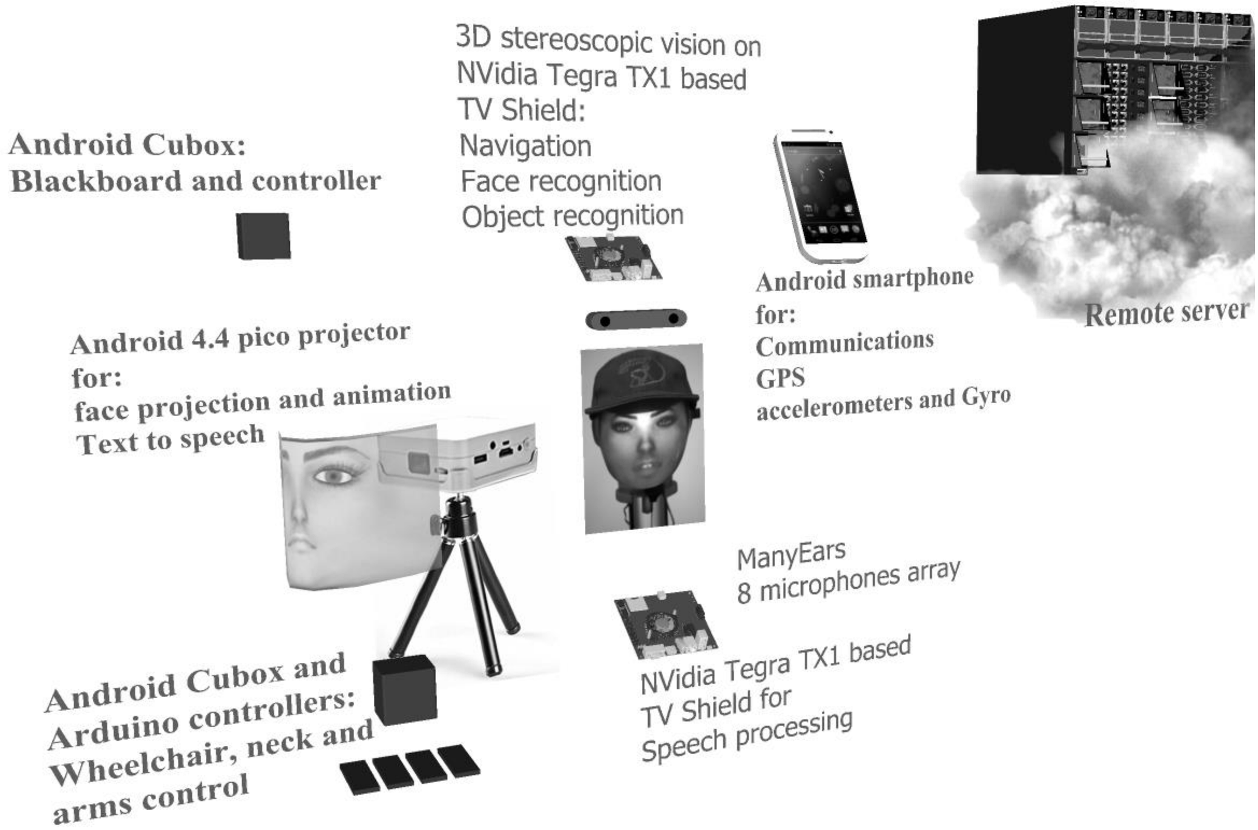

Figure 1: Global architecture of Roberta Ironside.

Roberta Ironside is fitted with a stereoscopic camera that is used as a normal camera for facial recognition [22], navigation and for visual contexts, to be used in the dialog. Text to speech uses an 8 microphone array with beam forming technique. Roberta is also fitted with an expressive face and it has two arms which are designed to (1) be used for non-verbal communication in the same way they are used with avatar animation and (2) to grab small objects (around 1kg) on shelves for which they must follow certain security rules for robots operating close to people.

All the mechatronics are designed to keep material costs low while chocks and conventional gimbals are replaced by rigid iron cables composing tripods. This architecture gives us a three axis rotation capability for the neck and the shoulder. Whenever possible, we use stepper motors that are widely used in 3D printers and shoulder rotation uses a specially designed worm/wheel to perform auto blocking when the power is off. Finger actions that need miniature motors present a real issue in pricing and size. Hence, in order to meet our pricing target, we are currently designing a new type of motor specifically dedicated for this task.

3. Human Like Expressive Face and Conversational Agent

Expressive faces are required for conveying emotions and for non-verbal communication. Realistic faces such as Nadine's [33] are impressive but necessitate very sophisticated and complex mechatronics in order to avoid the "uncanny valley" [19], if it really exists [13]. When a robot is designed for the market, the complexity of the mechatronics also means more maintenance and accompanying operational costs, which stands in great conflict with the business plan for this project. Besides, the fact that Roberta is handicapped, would already put it near the bottom of the valley [17]. Hence, to simplify the face while still preserving its capability to convey emotions during the dialog, we have opted for a projected face similar to the "Maskbot" [15, 23] as is illustrated by Figure 2.

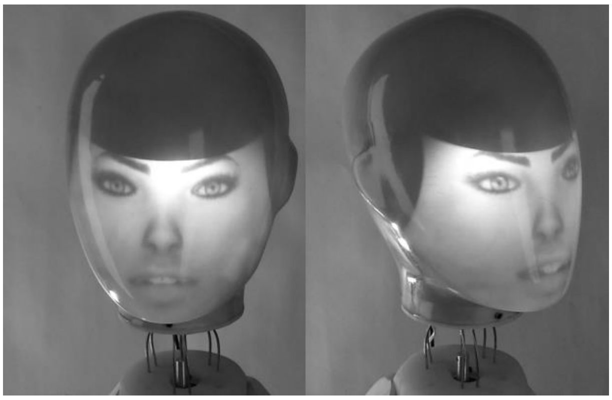

Figure 2 : Front and side view of Roberta's face showing the neck actuators.

4. Speech and Dialog Based Cognitive Coaching

The lack of human contact and loneliness is probably among the most painful experiences for the elderlies [24] and cognitive deficient people, even if a connection with an increased morbidity has not been demonstrated [32]. This lack of social contact and dialog impacts brain plasticity and the reorganization of memories. For this purpose, the Roberta Ironside project emphasizes speech and dialog.

With Roberta it is not planned to embed a screen as a Human Machine Interface (HMI). Rather speech will be the main interaction channel. As in the vAssist project, standard tablets, smartphones or Smart TVs will be later added as additional communication tool to interact with Roberta, e.g. to show a picture to the person as a basis for a dialog. However, they are not planned to be the standard HMI.

Roberta will be able to get information from the environment and from the system through a dialog manager that integrates this information in the 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 is also aiming at acquiring a good level of knowledge about the person. On the other hand, the system is expected to feature an open-domain conversational system that can present useful and interesting information to the user, following the examples of WikiTalk [11]. Roberta could observe the user's behavior, and on the basis of this, infer the user's emotion and interest levels so as to tailor its presentation accordingly. Conversational subjects can deal with personal data, a user's life, or picture analysis in the system initiative dialog management.

Additionally, Roberta will have a broad set of other discussion topics covering a number of potentially interesting domains. In order to ensure Roberta has a consistent situation-aware basis, it is important to have appropriate semantic representation of conversation contexts, care support contexts and context-specific privacy preferences of the care-recipient in place. To achieve this, we will deploy the methodologically-guided UI-REF Context-aware Framework for users' needs elicitation and formalization [2], see Figure 3 for the architecture. This will allow us to arrive at a consistent framing of the user's conversational use-contexts, including the various contexts within a care-recipient's lifestyle and activities of daily living, and the framing of the privacy protection requirements, comprising both default and user-specific privacy preferences. Essentially a topic-map-enabled semantic representation of the human-robot interaction frames (dialogue, and care–support) will provide a use-context (conversation-centric, care-support-centric), a hierarchy of care-

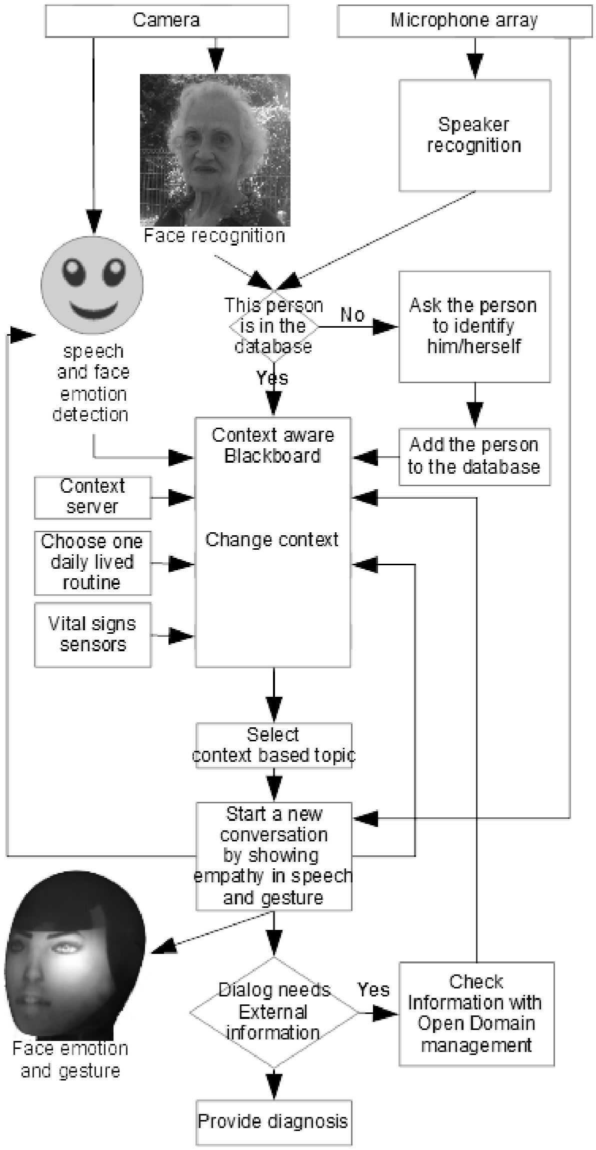

Figure 3: Context aware dialog architecture showing face recognition, speech processing, context aware topic selection and dialog processing. Picture of the elderly

support contexts semantically framed so as to underpin situation and context awareness.

A knowledge base of profiles of sensors, persons and use-contexts and relevant conversation frames will be established, as well as a black board structure to serve as a dynamic working memory supporting situation awareness. Historical models of user behavior and preferences will be used in light of data fusion and context-specific analytics to provide model-based and data-driven decision support to direct the pro-active context-aware dialoguing (via the Dialogue Manager), care support (via the Services Manager) and Privacy Protection (via the Privacy Controller).

It is planned that some parts of the Assist project speech architecture will be reused. Text to speech relies on MaryTTS. Speech processing should rely on the embedded speech processing technology provided by Intelligent Voice on the NVidia Tegra based TX1 GPU platform, most probably the NVidia TV shield that is significantly cheaper than the TX1 developer kit.

To provide dialog capability to Roberta we will reuse the dialog manager developed by Telecom Paris Tech and the University of the Basque Country on top of the Disco LFF [18, 12, 31, 20].

Facial emotions and non-verbal interaction relies on the work undertaken with conversational agents at Telecom Paris [1]. Neck actuators as well as arms can be activated directly from Blender. Mechanical animation through Unity 3D is another option.

Dialog capabilities are organized into modules that manage different topics according to the context. For every new person, location or object, Roberta needs to put a name to him/her/it and ask the person for its own learning purpose: Who are you? What is it? What is it for? etc. to relate he/she/it to a semantic database, and establish the context for a dialog. We expect constant learning to create a dialog between Roberta and its user, again, reversing the role from assistant to dependent. The embedded system will also synchronize with servers in order to make the community of robots benefit from each other.

Roberta will be able to get information from the environment and from the system through a dialog manager that integrates this information in the dialog tasks and strategies.

5. The Example of the Life Line Dialog Module, Helping People Telling Their Life as a Support to Dialog

The LifeLine dialog module is designed to help people tell stories about their lives in order to help them exercise their speech, and long and short term memory capabilities. It is also aiming at acquiring a good level of knowledge about the person. The open-domain conversational system can enrich the dialog by proposing useful and interesting information to the person in relation to their current situation in life. Roberta can observe the user's behavior, and on the basis of this, infer the user's emotion and interest levels so as to tailor its presentation accordingly. Conversational subjects can deal with personal data, a user’s life, or picture analysis in the system initiative dialog management.

6. Learning and Wizard of Oz

A big part of our project relies on the capability of Roberta to learn. We are conscious that it will not be perfect from the beginning and even during its life. For that purpose, we introduce the same Wizard of Oz service as we have used in vAssist [30].

As mentioned for speech, a certain embedded learning capability will have to be included in Roberta. Potentially, it could benefit also from on line systems in particular for vision, e.g. by comparing what our system recognizes with the Google vision API results.

7. Roberta as a Physical coach

Physical activity influences the endogenous pharmacology of the brain to enhance cognitive and emotional functions in late adulthood [10]. For dependent people, an issue among others is their difficulties to walk outside and the risk of a fall. Because a wheelchair has a handle, Roberta Ironside can serve as a walker, making the daily exercise easier and safer while offering a less socially traumatic experience than with a classical walker. Sensors on the handles detect the pressure from which Roberta can adapt its speed. For cognitively deficient people it will be a constant reference with which they will never lose their way back home.

Vocal interaction as well as facial analysis with the camera can measure the physical state of the person. During a walk, Roberta will be able to quantify a person’s level of fitness. Potentially, the handle could be fitted with smart fabric used in Bebop Sensors' already mentioned above. If the person is too tired to walk back home, he/she can sit on the knees of the robot which then reverts to its most primitive function: i.e. help by being a wheelchair.

8. Privacy Protection

Default privacy rules and context-specific privacy preferences-aware control will enable personalization of the privacy protection of the three pillars of the care ecosystem namely the care-recipient's personal data management, the dialogue management, and the service provisioning by the care robot [2, 3, 4, 5].

In addition to data collected from the dialog, Roberta will collect a large amount of data going from health monitoring, location and habits as well as cartography and images. Although it is currently only a plan, we envision the introduction of blockchains to secure data access.

9. Conclusions

Many rational comments have questioned our choice to use a head with an animated face rather than a tablet, and to put our humanoid body in a wheelchair. Those comments make sense but the recent history of evermore sophisticated, larger screens and expensive cell phones shows that the rationale of the marketing people proved wrong, when this market was disrupted less than ten years ago. The evolution of 3D avatars has also proven that there is a certain need for us to see human-like creatures.

The initial specification of Roberta was a robot that can be a life companion, an early health alert that fills the gap of wearable objects that are not worn, a physical trainer that facilitates the life of a person when he/she leaves the home, and a platform that has enough energy storage for day long usage. By removing the screen, we emphasize the need for spoken dialog which is a serious bet. The choice of a humanoid on a wheelchair is our exploratory answer. It may not reach the market we targeted initially, but it may find other applications and above all it may evolve.

Today's technology makes the dreams of yesterday possible at an affordable price. It is no longer mandatory to be an electronic engineer to integrate electronics. All the parts needed to build a robot are available off the shelf. It is often even better when those parts are already in the consumer market, although their functions may be diverted from their initial purpose. Also when specific hardware parts are needed, 3D printing allows to easily re-produce them.

With a faster and simpler implementation process, it becomes therefore possible to build a robot that has a potential usage, coming directly from the expertise of previous projects and our understanding of the world. Is our vision appropriate and justifiable? This is hard to say. Business examples, like the smartphone market, show that a good concept may have existed long before someone brought in the one essential ingredient that makes the concept eventually successful.

A robot's value is limited by its intelligence. This is the reason why we emphasize the need for a real vocal dialog capability, mixed with other intelligence capabilities. If we want for

Roberta Ironside to be successful, we must trust the community to add functions to it which are adapted to specific needs and refine or completely modify some options that have been created by us. The best way to achieve this is through modularity to facilitate the developments of features by specialists. As a consequence, Roberta Ironside will be an entirely open design, and we do not plan to patent any innovation that we may introduce. This approach will lead to a cheaper platform and, we hope, to solutions from which dependent people will truly benefit.

10. References

1. Elisabeth André, Catherine Pelachaud, Interacting with Embodied conversational agents, 2010, Speech Technology: Theory and Applications, by Fang Chen, Kristiina Jokinen, pp.123-149, Springer.
2. Badii A, "User-Intimate Requirements Hierarchy Resolution Framework (UI-REF): Methodology for Capturing Ambient Assisted Living Needs", Proceedings of the Research Workshop, Int. Ambient Intelligence Systems Conference (AmI'08), Nuremberg, Germany November 2008
3. Badii et al Accessibility-by-Design: A Framework for Delivery-Context-Aware Personalised Media Content Re-purposing, proceedings of HCI and Usability for e-Inclusion: 5th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2009, Linz, Austria, November 9-10, 2009.
4. Badii et al, Visual context identification for privacy-respecting video analytics Multimedia Signal Processing (MMSP), 2012 IEEE 14th International Workshop on, Sept. 2012
5. Badii et al Holistic Privacy Impact Assessment Framework for Video Privacy Filtering Technologies, Signal & Image Processing: An International Journal, 4(6):13-32, 2013.
6. Changan Jiang, Ueno S, Hayakawa Y, Optimal control of non-prehensile manipulation control by two cooperative arms, Advanced Mechatronic Systems (ICAMechS), 2015.
7. Chollet G., Amehraye A., Razik J., Zouari L., Khemiri H., Mokbel C., 2010. Spoken Dialogue in Virtual Worlds. In Development of Multimodal Interfaces: Active Listening and Synchrony, Esposito A. (Ed.). Springer Verlag, LNCS-5967, pp. 423-443.
8. Chollet G., Černocký J., Gravier G., Hennebert J., Petrovska-Delacrétaz D., Yvon F., Towards Fully Automatic Speech Processing Techniques for Interactive Voice Servers, Speech Processing, Recognition and Artificial Neural Networks, 1999, Proceedings of the 3rd International School on Neural Nets "Eduardo R. Caianiello"

9. DARPA Robotics Challenge http://www.theroboticschallenge.org/ DARPA, Robots vs. Door at DARPA Robotics Challenge, IEEE Spectrum video, 2015 https://www.youtube.com/watch?v=UUOo8N9_iH0
10. Kirk I. Erickson, Ariel G. Gildengers, MD, Meryl A. Butters, PhD; Physical activity and brain plasticity in late adulthood, Dialogues Clin Neurosci. 2013 Mar; 15(1): 99–108.
11. 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
12. R. Justo, O. Saz, A. Miguel, M. I. Torres and E. Lleida (2013): "Improving language models in speech-based human machine interaction," International Journal of Advanced Robotic Systems, Vol 10, 87.
13. Katsyri, J., Forger, K., Makarainen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: Support for perceptual mismatch as one road to the valley of eeriness. Frontiers in Psychology, 6, 390.
14. M. Kumar, A. Veeraraghavan, and A. Sabharwal, "DistancePPG: Robust non-contact vital signs monitoring using a camera," Biomedical Optics Express, Vol. 6 Issue 4, pp.1407-1418 (2015)
15. Kuratate, T.; Matsusaka, Y. ; Pierce, B. ; Cheng, G. "Mask-bot": A life-size robot head using talking head animation for human-robot communication, 11th IEEE-RAS International Conference on Humanoid Robots, 2011
16. Mancini M., Pelachaud C.: Dynamic behavior qualifiers for conversational agents., Intelligent Virtual Agents, IVA'07, Paris, September 2007.
17. Maya B. Mathur, David B. Reichling, Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley, Elseviewer, Cognition Volume 146, January 2016, Pages 22–32
18. Milhorat P., Schlögl S., Boudy J, Chollet G., 2013a. What if Everybody could do it? A framework for easier Spoken Dialog System Design. In Symposium on Engineering Interactive Computing Systems. ACM SIGCHI, London, UK.
19. Mori, M., translation by MacDorman K. F. Norri Kageki N., (2012). The uncanny valley [Bukimi no tani]. IEEE Robotics and Automation, 19 (2), 98–100. (Original work published in 1970).

20. Olaso J.M., Milhorat P., Himmelsbach J, Boudy J., Chollet G., Schlögl S, Torres M.I., A Multi-lingual Evaluation of the vAssist Spoken Dialog System. Comparing Disco and RavenClaw, 7th IWSDS, Finland, 2016.
21. F. Pecune, M. Mancini, B. Biancardi, G. Varni, Y. Ding, C. Pelachaud, G. Volpe, A. Camurri, "Laughing with a Virtual Agent", In Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems (pp. 1817-1818), 2015
22. Petrovska-Delacrétaz Dijana, Chollet Gérard, Dorizzi Bernadette, Guide to Biometric Reference Systems and Performance Evaluation, Springer, 2009
23. Pierce B, Kuratate T, Vogl C and Cheng G, “Mask-Bot 2i”: An active customisable Robotic Head with Interchangeable Face, 2012 12th IEEE-RAS International Conference on Humanoid Robots, Nov.29-Dec.1, 2012. Business Innovation Center Osaka, Japan
24. Kvaal K., Halding A.G., Kvigne K., Social provision and loneliness among older people suffering from chronic physical illness. A mixed-methods approach, Scandinavian Journal of Caring Sciences, Volume 28, Issue 1, pages 104–111, March 2014
25. Sansen H., 1989, "ShipPlanner: Un système d'aide au chargement de navires porte-conteneurs", Conference on Artificial Intelligence and Expert systems, Avignon
26. Sansen H., 1990, "ShipPanner: A conception support tool for the bay plan of container ships", Technology of Object Oriented Languages and Systems, Paris:
27. Sansen H., Morlier D., Chollet G., Crespel H., 2012. Visumag, un logiciel de présentation des résultats commerciaux des supermarchés ou la “tammetisation” des chiffres. International workshop IHM 2012 SETIT, Sousse, Tunisia (an IEEE sponsored conference)
28. Sansen H., Baldinger J-L., Boudy J., Chollet G.,Milhorat P., Schlögl S., vAssist: Building the personal assistant for dependent people, Healthinf 2014
29. Sansen H., Chollet G., Glackin C., Badii A., Torres M.I., Petrovska D., Schlögl S., Boudy J.. The Roberta IRONSIDE project: A Humanoid Personal Assistant in a Wheelchair for Dependent Persons, ATSIP 2016.
30. Schlögl S., Chollet G., Milhorat P., Deslis J., Feldmar J., Boudy J., Garschall M., Tscheligi M.. 2013a. Using Wizard of Oz to collect Interaction Data for Voice Controlled Home Care and Communication Services. In International Conference on Signal Processing, Pattern Recognition and Applications. SPPRA, Innsbruck, Austria.
31. Serras m., Pérez N., Torres M., del Pozo A, Justo R. (2015): "Topic Classifier for Customer Service Dialog Systems" 18th International Conference on Text. Speech and Dialooue, to ne published in LNCS-. Plzen, Czech Republic, September, 2015

32. Stessman J., Rottenberg Y., Shimshilashvili I., Ein-Mor E., Jacobs J., Loneliness, Health, and Longevity, Journal of Gerontology, 2013.
33. Xiao, Y., Zhang, Z., Beck, A., Yuan, J., & Thalmann, D.(2014). Human robot interaction by understanding upper body gestures. Presence: teleoperators and virtual environments, 23(2), 133-154.