Report of the visit under the European Community's Research Infrastructure Action - grant agreement VISIONAIR 262044 - under the 7th Framework Programme (FP7/2007-2013).

Number and acronym of project: 65 - Constructive Conversations - Modelling Interaction Space, Gesturing, and Gazing in natural interactive encounters with humans and robots

Access provider: MTA SZTAKI – 3DICC Laboratory

User: Kristiina Jokinen, University of Helsinki

Start date of the project: 28 January, 2013

End date of the project: 09 February, 2013


Abstract

The project focuses on the interaction between humans and robots within the VirCA virtual environment and the Nao WikiTalk robot system, and especially on natural ways to exchange information. In the context of automatic services, delivery of reliable and relevant information is an important design goal, and the effective transfer of information is often taken as the main criterion for the success of interaction. Recently, however, one of the challenges for designing interactive systems has been identified as related to social aspects of interactions: how to engage the partner in the interaction and keep their interest up, so that the speaker can deliver the message they intend to deliver, or provide rapport and affection so as to create a mutual bond and an understanding relationship. Our interest in studying conversational engagement derives from intelligent systems and interaction technology, where engagement is used to describe the user's willingness and involvement in the interaction with the automatic system. One of the ways to measure engagement is to study the interlocutors' multimodal signals concerning their interest, level of understanding, and focus of attention. If the human user's engagement level can be measured, it will be easier to adjust the robot's conversational strategies accordingly.

## 1. The motivation and goal of the visit

Natural language is used to exchange information, and the effective transfer of information is often taken as the main criterion for the success of interaction. Especially in the context of automatic services, the delivery of reliable and relevant information is an important goal for the design of such systems. Recently, however, one of the challenges for designing interactive systems has been identified as being related to social aspects of interactions: how to engage the partner in the interaction and keep their interest up so that the speaker can either deliver the message they intend to deliver, or can provide rapport and affection so as to create a mutual bond and an understanding relationship. Our interest in studying conversational engagement goes back to intelligent systems and interaction technology, where engagement is used to describe the user's willingness and involvement in the interaction with the automatic interactive system. If it is possible to measure the interlocutors' engagement level, it is easier to adjust the system's conversational strategies accordingly.

One of the ways to measure engagement is to study the interlocutors' paralinguistic signalling concerning their interest, level of understanding, and focus of attention. For instance, with motion tracker technology such issues as the participants' distance from each other can be measured, and combined with the research questions of the participants' engagement in the interaction, the novel technology can provided important and useful objective information of the interaction space and the participants' control over the space around them. This research will not enhance only our understanding of the comfortable communication between humans or between humans and virtual agents, but will also allow us to build models for the automatic recognition and management of issues related to appropriate and smooth communication.

Research questions to address were

a) to study multimodal signals, especially gesturing and participants' distance, i.e. interaction space in social conversational communication and interaction management

b) to investigate the interlocutors' engagement and synchrony in communicative activity through measuring their paralinguistic activity (gesturing, body posture)

c) to build (computational) models for the coordination and controlling of interaction, and constructing shared understanding,

d) to investigate techniques and means for automatic recognition of multimodal signals,

e) to compare human-human interactions with human-robot interactions and human interactions in the virtual world.

The current state of the art in interaction research has focussed on the function and correlation of such multimodal communicative means as overlapping speech, gaze, facial displays, hand gestures, head movement, and body posture. For instance, Campbell and Scherer (2010) and Jokinen (2011) describe utterance density as a measurement for engagement. On the other hand, research with Embodied Conversational Agents (ECAs) has especially brought forward several types of behaviors that are important when conducting natural conversations between humans, and which are also necessary when supplying natural intuitive communication models for interactions between humans and ECAs (Andre and Pelachaud, 2010; Misu et al., 2011).

Jokinen and Wilcock (2011) describe emergent verbal behaviour that arises when speech components are added to a robotics simulator. The robotics framework supports different behavioural paradigms, including finite state machines, reinforcement learning, fuzzy decisions, neural networks and evolutionary algorithms. For instance, by combining finite state machines with the speech interface, spoken dialogue systems based on state transitions can be implemented for classical closed-domain form-filling dialogues such as flight reservations. These closed-domain dialogue systems exemplify emergent verbal behaviour that is robot-initiated: the robot asks appropriate questions in order to achieve the dialogue goal. A demo of these different levels of emergent verbal behaviour is described by Wilcock and Jokinen (2011).

Scientific challenges deal with the interlocutors reacting to their partner's actions and coordinating their turns in a manner that allows both to present their message in a cooperative manner. Engagement is an important sign of this kind of cooperation, and it is related to the interlocutors' experience of the interaction in general: more engaged the interlocutors are in the conversation, the more positively they may experience the interaction. This may be seen from the participants' body posture and gesturing in general. The hypothesis is that there are significant differences in the interaction patterns between interlocutors that experience their interaction either smooth or tedious and these can be modelled with the help of motion trackers and virtual agents. Besides the theoretical and interaction challenges, also technological challenges concerning the automatic methods and techniques will be met by

the experiments, and thus the proposed experiments will support interdisciplinary research.

Scientific merits concern better understanding of the participants' engagement is a complex process that involves various multimodal cues and signals, and the participants cooperation. Technological advancements (e.g. motion trackers) have allowed us to automatically recognize interaction issues that manual annotation and analysis may have found but due to their rare occurrence are difficult to observe to the sufficient extent to draw statistical conclusions. Technology can also provide more objective information, and more systematic studies on issues that are commonly established as important aspects of human communication. Finally virtual agents allow us to build, test, compare, evaluate, and sophisticate our models of interaction, based on empirical observations of the users.

The operational work description of the task is as follows.

1) Planning of the experiments focussing on engagement in interaction

2) Data collection

3) Manual and automatic data analyiss

4) Evaluation of the interaction models with virtual agents

5) Paper writing

We notice that the task 1 can start before the actual visit, while the tasks 2-4 require the actual setup devices. The task 5 can be mostly done after the experiments have been conducted, although the discussion and planning are best to conduct during the visit.

The research will be done in cooperation with Prof. Baranyi at MTA SZTAKI. The experimental setup needs motion tracker and virtual agent environment that are available at MTA SZTAKI. On 3I side, the visit and experiments will include both Jokinen and Wilcock.

In regard to ethics and other confidential issues, the investigators are committed to following strict ethical codes in all experiments that may be provided, and privacy issues as defined in the EU and national research policy.

We expect there will conference presentation and/or publication, and indeed paper writing occurred later on. Plans for papers in CogInfoCom conference, Sigdial, Interspeech, IJCAI etc. were envisaged and also a journal article. It was also expected that models and improved software for the analysis and

recognition would start  We also expect the project will bring demonstration given the robot interaction works. natural interaction models

Expected benefits of the project would be beneficial for both MTA SZTAKI and 3I research group, since the similarities and complementarities of our research expertise provide fruitful basis for collaboration. It would allow 3I to expand and consolidate previous work with empirically based communication studies, while 3I expertise in dialogue systems, theories of dialogue management, and machine learning techniques as well as experience in the use of eye-gaze tracking would build up new expertise in the MTA SZTAKI's growing research group. Moreover, the opportunity to work physically at the same place for a period fof time would be most rewarding, and it will also allow the respective groups to work more tightly in the CogInfoComm framework.

## 2. Report of the activities

The first week of my two weeks stay was devoted to research planning. The first days focussed on the writing of a joint project application, and we successfully managed to submit a project proposal for the EU FET-Open Xtrack.

The end of the first week included inspiring discussions on the research related to interaction and CogInfoCom, and planning of the writing of a joint article to be submitted to ACM TiiS Journal on Multimodal Interaction. The article concerns evaluation of the Nao WikiTalk robot system that constitutes the User's starting point for the current collaboration. The robot system was developed at the eNTERFACE workshop, led by the User, in Metz in summer 2012, and one of the students at the Access provider was a key member at the workshop and also takes part in the article writing.

The second week of my visit was devoted to demonstration and future planning of the collaborative work. The goal is to import the above mentioned Nao WikiTalk system into a virtual environment using the VirCA virtual collaborative arena of the Access Provider. The User's expertise in designing interactive systems could then be used to study natural communication capabilities and social aspects of interactions between humans and between humans and robots within the framework. For this goal, I saw a demonstration of the VirCA platform with an integration of WikiTalk-type access to Wikipedia articles, and was given the website for downloading the necessary software so as to be able to run the system at home. We discussed about the possibility to integrate the Nao WikiTalk in the VirCA framework so that it would be possible to use a physically dislocated robot through VirCA. We agreed on working towards such a system so that it would be possible to demonstrate it in the next CogInfoCom conference in Budapest in December 2013, and also at ICMI

(The 15th International Conference of Multimodal Interaction) in Sydney in December 2013. I was also invited to give a talk at the next CogInfoCom conference, which I accepted given that the time would not conflict with the ICMI conference which I chair.

3. Other visits

I also visited the Budapest University of Technology and Economics and discussed about research and collaboration using robots.

4. Summary of the project plan for CoCoRo -project

Constructive Conversations - Modelling Interaction Space, Gesturing, and Gazing in Natural Interactive Encounters with Humans and Robots

(Robotics, Cognitive Systems & Smart Spaces, Symbiotic Interaction)

The research targets systems that can operate autonomously in the real world through e.g. scene and context understanding, anticipation and reaction / adaptation to changes, manipulation and navigation, as well as symbiotic human-machine relations. The project concerns human-robot conversations and focuses on the various aspects of conversational engagement: how to measure the level of engagement, how to ensure and maintain the user's engagement and interest level in long-term conversational interactions. Application scenarios include smart public spaces such as airports, department stores, hospitals, or museums, where people spend significant amounts of time and would be glad to be told interesting and useful information by a robot guide. A specific application scenario is e.g. engaging the interest and raising the motivation of critically ill children in isolation units in hospitals by enjoyable and encouraging conversations with a Nao robot companion.

The goal of the project is to deploy the existing prototype Nao WikiTalk system in dynamic real-life environments, and to develop the system capabilities to reach new levels of autonomy and adaptability in human-robot interactions. Using Wikipedia means that WikiTalk is open-domain, which is very unusual compared with existing dialogue systems. However, Wikitalk can also be used with closed-domain information sources in web-page format. These can be existing websites (e.g. museum guides, department store websites) or new texts aimed specially at use by Wikitalk (e.g. children's stories for the hospital scenario). Smart public spaces in department stores and museums can have Nao robot guides greeting visitors/customers and telling what there is to see, where to go, special offers, etc. Most of the information exists in store/museum web pages already, and can be readily used by Nao WikiTalk. WikiTalk can readily be multilingual by using existing

French/Hungarian/other Wikipedia articles or other web pages, combined with existing TTS and ASR components for each language. Aldebaran support speech on Nao in gradually more languages and the hospital prototype has speech components in Hungarian that work on Nao.

The focus of the work-packages is on technology, methodology, and architecture of designing and developing intelligent interactive situated agents. There are such technical issues in robotics as being able to move independently and to have sensors for detecting and analyzing the environment. Besides these technical capabilities it is also important that robots should participate in interaction with human participants. The successful applications of robots in medical and health domains require the understanding and handling of the cognitive processes related to the interaction, and the presentation of information in public spaces requires understanding of human presentation techniques including multimodal aspects like gesturing, nodding, recognizing the user's interest level and presence in the interaction space. This way the artificial systems operating in dynamic real life can reach new levels of autonomy and interact in a symbiotic way with humans, extend past achievements in scientific research and introduce robotics technology in European service sectors.

Human-human face-to-face conversational interactions involve not just exchange of *verbal* feedback, but also that of *non-verbal* expressions. Often verbal expressions are accompanied by non-verbal expressions, such as *gestures* (e.g., hand, head and facial movements) and *eye-gaze*. They are not mere artefacts in a conversation, but intentionally used by the speaker to draw *attention* to certain pieces of information present in the verbal expression. For artificial agents, such as Nao WikiTalk, it is crucial to be able to understand and exhibit verbal and non-verbal behaviour in the application scenarios. Exhibiting non-verbal expressions would not only add to their ability to draw attention of the users to useful pieces of information, but also make them appear more expressive and intelligible which will help them build social rapport with their users. This is pertinent in real life situations.

Existing background technology contains:

(1) WikiTalk (UH) is a spoken dialogue system for open domain knowledge access using Wikipedia. The Nao version of WikiTalk greatly extends the robot's interaction capabilities by enabling open domain conversations with Nao. Figure 1 shows the overall architecture of the system and Figure 2 depicts users interacting with the Nao WikiTalk in the ENTERFACE Summer School in 2012.

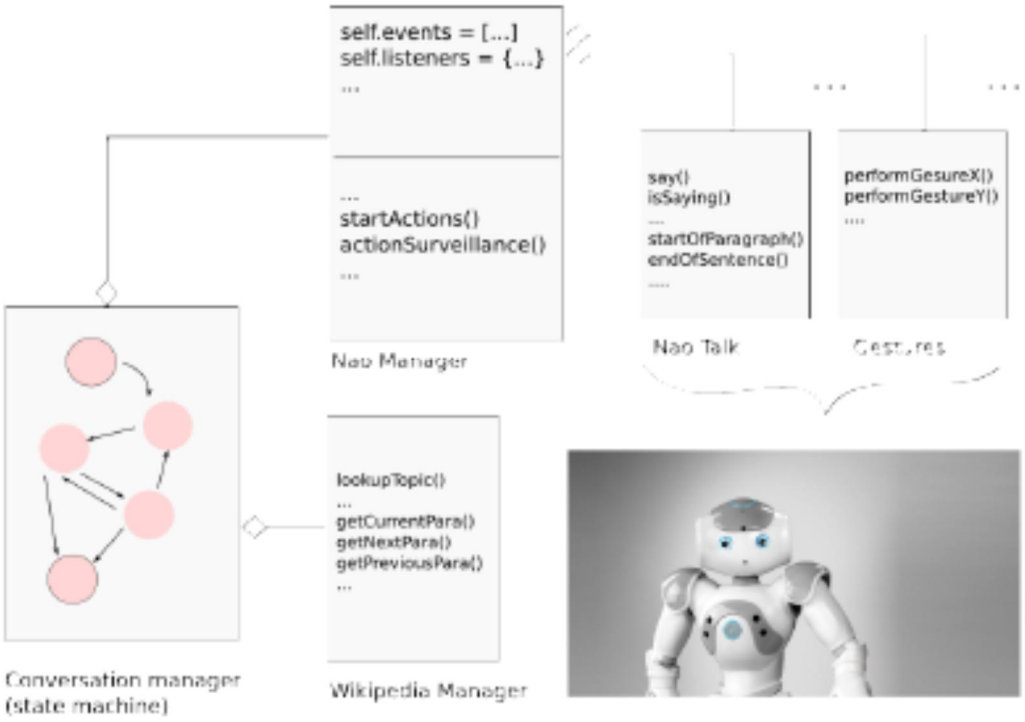


Figure 1. Nao WikiTalk architecture. From Csabo et al. (2012).

Figure 2. Users interacting with the Nao robot during the ENTERFACE-2012 Summer School.

(2) CDM (UH) is a communication framework starting from the viewpoint that speakers are rational agents who react to the situation according their beliefs, intentions, and understanding of the current situation, and who aim to achieve a goal (task based or just a general social chatting goal) for which they construct shared knowledge and build rapport. Nao WikiTalk is based on CDM, which is itself based on four enablements: contact, perception, understanding and reaction. These enablements directly map to the four terms in the call: sensing, perception, understanding and action. The CDM framework is shown in Figure 3, and the mapping between the terms used in the call and the enablement levels of CDM are as follows:

| FP7 text | CDM framework |
|---|---|
| sensing: | contact |
| perception: | perception |
| understanding: | understanding |
| action: | reaction |

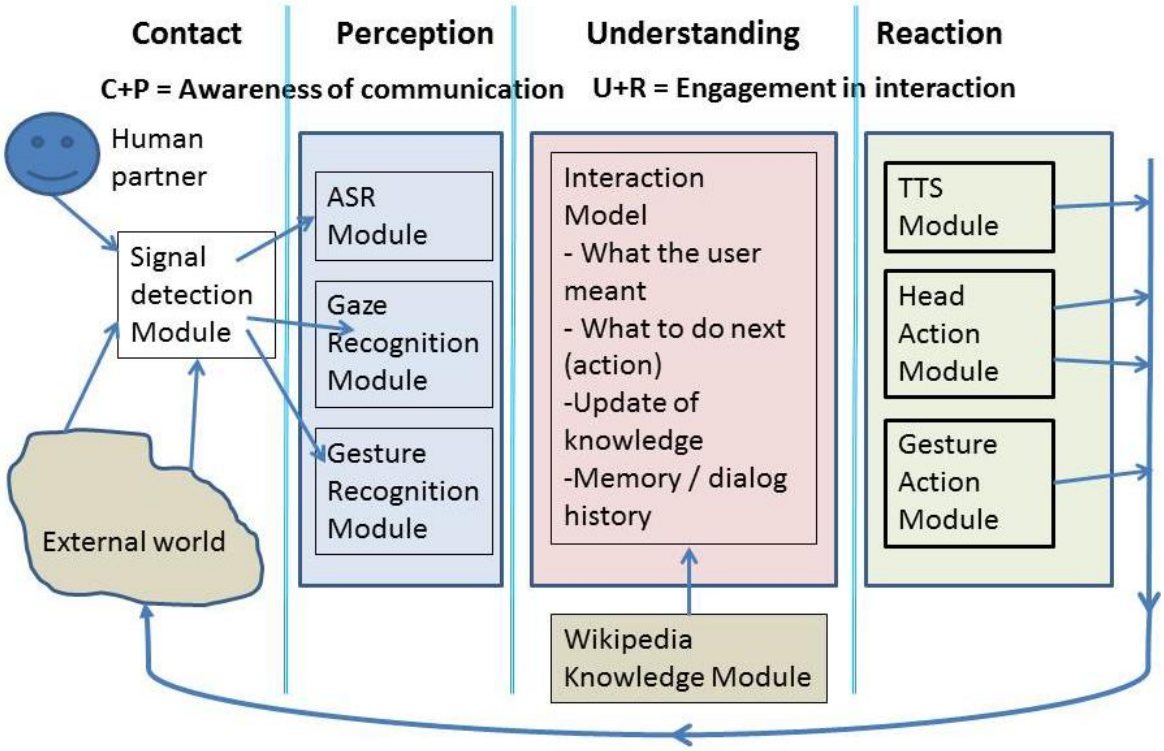


Figure 3. CDM architecture (From Jokinen & Wilcock, 2012).

(3) Reinforcement learning techniques will exploit the large database of interactive situations to optimize and adapt dialogue strategies with respect to different user groups (children, hospital staff, airport passengers, department store customers, museum visitors) and with respect to the user's perceived interest in the presented topics. They also allow modelling of adaptation over long stretches of time, and explorations on how user groups "co-evolve" with the robots. We can also study different roles of the robot with respect to the child (companion, guardian, assistant).

(4) The robot-child application concerning treatment of marrow-transplanted children is a realistic, already "prototyped" scenario with an existing collaboration. It differs from the Alize project (medical domain concerning diabetes) in that the situation is more critical, emotional, and long-lasting (thus social bonds can develop), interactions are longer (half year), and require more adaptation, motivation and sensitivity of the robot to the child's state and background.

(5) VirCA virtual collaboration arena supports smart spaces and virtual reality-based interaction, including human-robot interaction with Nao, using immersive 3D internet technologies. VirCA provides an integrated infrastructure for cognitive infocommunications (CogInfoCom), which goes beyond HCI or HRI by envisioning the robot in an infrastructure interacting with different types of users, for a long time, with interactions that can be interrupted for weeks or months (the interests of people might have changed etc.) and which is a public environment. It also allows cross-border interaction, e.g. interacting with a Nao in Finland while the partner is in the Budapest hospital.

A hospital can be seen as an infrastructure where patients with chronic diseases come recurrently (with their family also), visit different departments and doctors, have to wait a lot, and may have to stay for longer periods (like marrow-transplanted children). These recurrent visits provide the data for the adaptations by reinforcement learning, and the interactions between children, family, carers, and robots in different smart spaces will be enhanced by being treated as virtual collaborations in VirCa.

The design of interaction would be user-centered, i.e. having the user in the center of the design loop and measuring the system's usability directly through working with the end-users. This allows fast development and real interactions with the participants who want to work on the development. It would also allow us to collect a lot of data dynamically. Moreover, the child can learn and explore new things via WikiTalk, and the system thus can motivate the child to

keep up with their class and encouraging them to go back to their original class after the treatment. We may also see if a hospital in Finland would be interested in the project - we could then also plan some intercultural communication studies and even virtual connection between children in different countries. Another activity is to use the Virtual robot environment Virca to allow across-the boarders interaction, e.g. moving a Finnish Nao while the partner is in the Budapest hospital.

## 5. Conclusions

As a summary, scientifically the visit was successful, and resulted in possible concrete plans for further collaboration. However, the administration of the project especially concerning financial issues that are mentioned to be given to the project partner were appalling and beyond any reasonable expectations.

## 6. Biography, description of the visiting group, and some relevant publication

Kristiina Jokinen is Adjunct Professor at the University of Helsinki, Finland, and Visiting Professor at University of Tartu, Estonia. She obtained her first degree at University of Helsinki, and received her PhD at UMIST, Manchester in 1994. She lived four years in Japan. During 1994-1996 she was JSPS Research Fellow at NAIST (Nara Institute of Science and Technology) where her host was Prof. Yuji Matsumoto, and during 1997-1999 she was Invited Researcher at ATR-ITL (Advanced Telecommunications Research Laboratories), investigating the use of discourse information in a speech-to-speech translation system. In 1999 she set up CELE (Centre for Evolutionary Language Engineering) research department for the speech company Lernout and Hauspie in Flanders Language Valley. She moved to Finland in 2001, to take up Research Director post at University of Art and Design Helsinki. She was later invited to be Acting Professor at University of Helsinki.

During the academic year 2004-2005 she was Fellow at the Computer Laboratory at University of Cambridge, and in 2006 she was awarded Nokia Foundation Fellowship to conduct research at Stanford University. She received the Academy of Finland mobility grant for 2007-2008 to cooperate with Prof. Campbell on non-verbal communication in Japan, and she visited ATR/NiCT in 2007 and in 2008. She got the Japanese Government Fellowship (NICT-fellowship) to conduct research at Doshisha University for 2009-2010 and her host was Prof. Yamamoto).

Her research concerns human-computer interaction, adaptive spoken dialogue systems, natural multimodal and non-verbal communication, corpus analysis and annotation, modelling and system evaluation. She visited MTA SZTAKI in

2011 (host: Prof. Baranyi), and is involved in the conference series of Cognitive infocommunications (CogInfoCom) as a member of the Technical Program Committee.

She has published many papers and articles, and her book "Constructive Dialogue Modelling and Rational Speech Agents" (John Wiley and Sons, 2009) is a summary of the requirements and design principles for intelligent interactive systems. She published a popular textbook “SpokenDialogue Systems" together with Michael McTear, Morgan and Claypool, 2010), and edited a book "New Trends in Speech-based Interactive Systems" (Springer, 2010) together with Fang Cheng.

She has directed several academic and industrial research projects, e.g. she was the Scientific Coordinator of the EU FP5 project DUMAS, and the Finnish coordinator of the Nordic Multimodal Interfaces network MUMIN, and the collaboration project NOMCO. She is Workshop chair for EACL 2012. She is Secretary of SIGDial, the ACL/ISCA Special Interest Group for Discourse and Dialogue.

The organisation is University of Helsinki has regularly been ranked among Europe’s 10 to 15 best universities on worldwide ranking lists of research universities. Some 470 doctorates are completed annually and nearly 10,000 scientific articles or monographs are published yearly by the university’s researchers. The university, with almost 4,000 researchers and teachers, operates on four campuses in Helsinki and at 17 other locations. There are 35,000 students, and a further 30,000 participate in adult education. Founded in 1640, the University of Helsinki wants to strengthen its position among the world's leading multidisciplinary research universities and to actively promote the well-being of humanity and a fair society.

The organising research group is 3I (Intelligent Interactive Informatics) Group is a research group in the Institute of Behavioural Sciences at the University of Helsinki. Our research focuses on intelligent interaction and information systems, and our areas of expertise include multimodal interaction, interaction management, paralinguistic communication, corpus collection and annotation, machine learning techniques, clustering and classification of linguistic data, learning and interaction. We currently work on multimodal corpus analysis (top-down human annotations and observations as well as bottom-up signal analysis) on naturally flowing human-human conversations and first encounter interactions, especially focussing on eye-gaze, face, and hand gestures, and their use in signalling turn-taking and feedback, in order to develop models for interaction techniques and strategies.

The aim of the group is to pursue research on natural interaction, and to explore means and possibilities for its realization in intelligent interactive systems. The work draws especially on advances in dialogue modelling, human-computer interaction (Constructive Dialogue Management, speech interfaces, language technology) and research in human-human communication, and attempts to bridge the gap between recent advances in interface design, modelling and technology. Technological and social development affects our interaction with the environment: interactive systems are embedded in our environment, information flow increases, and interaction becomes more complex. In order to address challenges of the complex environment, to respond to needs of various users, and to provide possibilities to test innovative interactive systems, the group considers it important to investigate processes that underlie human-computer interaction, to provide models and concepts that enable us to experiment with various types of complex systems, and to design and build tools and prototypes that demonstrate the ideas and techniques in a working system. The main activities of the group focus on building computational models of natural interaction, prototyping and experimenting with various flexible interactive systems.

## 7. References


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.

Han, J., Campbell, N., Jokinen, K., Wilcock, G. (2012). Investigating the Use of Non-verbal Cues in Human-Robot Interaction with a Nao Robot. In *Proceedings of the 3rd IEEE International Conference on Cognitive Infocommunications (CogInfoCom) 2012*, Kocice, Slovakia, pp 679-683

Jokinen, K. (2007). User Interaction in Mobile Navigation Applications. In Meng, L., A. Zipf, and S. Winter (eds.) *Map-based mobile services - usage context, interaction and application*, Springer series on Geoinformatics and Cartography. Springer-Verlag Berlin/Heidelberg. pp. 168-197.

Jokinen, K. (2009). Constructive Dialogue Modelling – Speech Interaction and Rational Agents. John Wiley & Sons.

Jokinen, K. (2010a). Gestures and Synchronous Communication Management. In: Esposito, A., Campbell, N., Vogel, C., Hussain, A., and Nijholt, A. (Eds.) Development of Multimodal Interfaces: Active Listening and Synchrony. Springer Publishers.

Jokinen, K. (2010b). Eye-gazing for turn-taking and feedback. The 5th Workshop on Disfluency in Spontaneous Speech and The 2nd International Symposium on Linguistic Patterns in Spontaneous Speech, Tokyo, Japan, September 2010.

Jokinen, K. and McTear, M. (2009). *Spoken Dialogue Systems*. Synthesis Lectures on Human Language Technologies. Morgan and Claypool.

Jokinen, K., M. Nishida and S. Yamamoto (2013). Modelling eye-gaze behaviour for interaction management. ACM TiiS Journal .

Jokinen, K. and S.Scherer (2012). Embodied Communicative Activity in Cooperative Conversational Interactions - studies in Visual Interaction Management. *Acta Polytechnica. Journal of Advanced Engineering*

Jokinen, K. and G.Wilcock (2011). Emergent Verbal Behaviour in Human-Robot Interaction. In Baranyi, P. et al (ed.): Proceedings of the International Conference on CogInfoCom (Cognitive Information Communication), Budapest.

Meena, R., Jokinen, K., Wilcock, G. (2012). Integration of Gestures and Speech in Human-Robot Interaction. In *Proceedings of the 3rd IEEE International Conference on Cognitive Infocommunications (CogInfoCom) 2012*, Kocice, Slovakia. pp 673-678.

Wilcock, G. and K. Jokinen (2011). Adding Speech to a Robotics Simulator. Proceedings of the Third International Conference on Spoken Dialogue Systems: Ambient Intelligence, Granda, Spain.

# VISIONAIR EVALUATION FORM

10. How did you learn about the VISIONAIR project? *
I talked to Professor Baranyi in a conference and visited MTA-SZTAKI two years ago, and learnt about the possibility to get funding for a short term visit within the Visionair project.

11. Why did you choose to apply? *
The Virca virtual environment seem like an interesting opportunity to apply my own Constructive Dialogue Model to a virtual environment and also to study multimodal dialogue management in human robot interaction

12. What laboratory/facility would you like to choose as first and second priority to visit (see lab list on VISIONAIR website)? * Please mark #1 for your first choice, # 2 for your second choice

#1 3DICC MTA SZTAKI
#2 CRVM UNIVMED France

13. Why have you chosen this specific laboratory/facility? *
My group has worked with a robot interaction system and one of the students at MTA-SZTAKI took part in a summer workshop that we led on multimdoal human-robot interaction based on the WikiTalk system.
It would be interesting to apply the same interactive WikiTalk system in a virtual environment, so that the human need not be physically in the same space as the robot.

15. What would you consider as the most relevant reason for your visit? *

f. Interested in sharing knowledge generated from the project with colleagues

My group already has the WikiTalk system running and it would be beneficial to develop the system further and explore possibilities for its application in the virtual computing contexts. We have complementary expertise which fit together well within the proposed project (my group's expertise in dialogue systems and human-computer interaction and the host's expertise on virtual reality), and we believe that the research will bring forward new expertise and fruitful results for both groups, as well as for European research in general.

17. What would you consider the most relevant expectation you might gain following your visit? *

c. Initiate professional collaboration at the academic level

It is important to strengthen collaboration on the shared interests concerning human interaction with intelligent agents and to realise the potential regarding virtual human-robot interaction envisaged by the partners. Some collaboration has already been conducted in related context, but it is important to consolidate the preliminary research and build a solid project around the possibilities for extending human-robot interaction capabilities with virtual environment.

18. What would you consider to be a successful and productive visit? * Please indicate the criteria for your judgment

To be able to work towards establishing a project within the EU-framework on the topic, and be able to collaborate on the shared interests in a concrete manner (e.g. data collection, evaluation, paper writing)

19. Please specify what topics you would like to see in the VISIONAIR newsletter?

Human communication possibilities in virtual environments, multimodal interaction (based on motion capturing, gesture interaction, eye-gaze, etc.)

20. VISIONAIR published a periodic Newsletter. For publication in our Newsletter, please briefly describe (a) You and your team – WHO WE ARE? (b) focus of your study (c) your potential contribution to VISIONAIR *

(a) Kristiina Jokinen is Adjunct Professor at the University of Helsinki, Finland, and Visiting Professor at University of Tartu, Estonia. She obtained her first degree at University of Helsinki, and received her PhD at UMIST, Manchester. She was JSPS Research Fellow at NAIST (Nara Institute of Science and Technology) and Invited Researcher at ATR-ITL (Advanced Telecommunications Research Laboratories) in Kyoto, Japan. Her research concerns human-robot interaction, adaptive spoken dialogue systems, natural multimodal and non-verbal communication, corpus analysis and annotation, modelling and system evaluation. She is the author of the books "Constructive Dialogue Modelling and Rational Speech Agents" (John Wiley and Sons, 2009) and "SpokenDialogue Systems" (together with Michael McTear, Morgan and Claypool, 2010), and edited a book "New Trends in Speech-based Interactive Systems" (Springer, 2010). She has directed several academic and industrial research projects. She has chaired several conferences and workshops and is involved in the conference series of Cognitive infocommunications (CogInfoCom) as a member of the Technical Program Committee.

She directs the 3I (Intelligent Interactice Informatics) Group is a research group in the Institute of Behavioural Sciences at the University of Helsinki together with Dr. Graham Wilcock.

http://www.ling.helsinki.fi/~kjokinen/Web_3I/index-start

(b) The research focuses on intelligent interaction and information systems, and we currently work on multimodal corpus analysis (top-down human annotations and observations as well as bottom-up signal analysis) and the Nao WikiTalk system on natural interaction capabilities of human-robot interaction.

(c) The project would be beneficial for Visionair, since the existing cooperation has shown how similarities and complementarities of our research expertise provide a fruitful basis for further collaboration. It would allow 3I to expand and consolidate previous work with empirically-based communication studies, while 3I expertise in dialogue systems, theories of interaction management, and machine learning techniques would build up new expertise in Visionair context.

Other overall comments

The Visionair setup is very interesting and useful for short-term visits and for starting collaboration. It might be good to be able to present the work conducted within the Visionair framework to the consortium at the end of the project in a "science fair" or the like so as to learn what has been done within the project and what kind of synergy could be built forward.

# VISIONAIR POST-VISIT QUESTIONNAIRE

1. Number of the EC Grant Agreement that financed the user group's access to the research infrastructure:

VISIONAIR (262044)

2. User Project Acronym: ConCon

3. Person filling in the questionnaire *(normally the User Group Leader): Kristiina Jokinen*

4. Where did you first find out about the possibilities of access supported through the EC grant agreement?

- ☐ EC Research Infrastructures Action web-site
- ☐ CORDIS databases
- ☐ Grant Agreement web-site
- ☐ Infrastructure web-site
- ☐ Announcement in journal
- ☐ Announcement at conference
- ☑ Direct mailing from infrastructure
- ☑ Personal contact *(please specify)*

5. Without the support of this EC grant agreement would you still have been able to carry out your work at this research infrastructure?

○ Yes ◉ No

If no, please indicate the reason *(you may indicate more than one choice)*

- ☐ Not otherwise eligible to apply for access to the infrastructure(s)
- ☐ Too difficult to obtain access by applying directly
- ☐ Unable to pay the user fee
- ☑ Unable to pay travel & subsistence for one or more of the group members

Other *(please specify)*

6. assess the services provided by the grant agreement with respect to the following points rating them on a scale from 'very poor' to 'very good'.

| *(Please provide at least 4 ratings. Leave blank when the point is not applicable*) | |
|---|---|
| Publicity, made by the infrastructure, concerning the access financed by the EC | fair |
| Practical information provided on how to apply for access | fair |
| Advice to use the most appropriate installation or infrastructure note 2 | |
| Information provided, once your project was accepted, on how to use the facility | fair |
| Scientific support to set up your experiments and interpret the results | good |
| Technical support to make best use of the installation(s) | fair |
| Logistic support at the facility (*office space, computing, libraries, accommodation*) | very good |
| Administrative support (*including the reimbursement of travel & subsistence expenses*) | very poor |
| The intellectual environment | very good |
| Overall appreciation of the services provided (*mandatory*) | good |

Please indicate any comments you would like to make on the services provided (*here you can also differentiate your appreciation with respect to the different installations or infrastructures*

It would be good if the costs incurred during the visit could be refunded in a uniform manner that is is clearly explained on the website and acceptance form, so that it would be known to the visitor (User) right from the beginning of the project and not afterwards.

7. Please indicate any further comments or suggestions you would like to make concerning your access to the infrastructure

Some later state of Visionair, it would be interesting and potentially useful to know what other projects have been realised within the framework, and what kind of results have been produced or facilitated during the visits. Maybe a science fair or a similar event would be possible to organise, in order to learn about synergies and collaboration possibilities within the partners and users of the Visionair grant.