Multimodal Communication with Robots and Virtual Agents

a. Principal investigator: Mohamed CHETOUANI, Thierry DUTOIT, Jean-Claude MARTIN, Catherine PELACHAUD

b. Abstract

The purpose of this project is to develop a platform for human-robot and human-humanoid interface. One application of this platform is to study the impact of the feedbacks provided by the robot and the humanoid on the interaction. Several studies have shown the benefit of verbal and non-verbal analysis and synthesis for the improvement of Human-Agent Interaction and of Human-Robot Interaction (HRI). But very few studies have been conducted to compare both interaction types. In particular we aim to study how the differences in embodiment (physical vs. virtual, few expressive parameters vs. subtle expressiveness) can affect how humans interact with the device.

2. Project objective: providing the rationale for the proposed project

Significant advances on Human-Computer Interfaces (HCI) have been done on the processing of multi-modal signals resulting on efficient analysis, recognition and synthesis. Contrary to computers, robots have several advantages, some of them are shared with virtual agents. One of the most important is that human can touch the robots and create or enhance the social link. On the other hand, virtual agents can show a large panorama of expressions and body gesture. As for natural face-to-face interaction, providing a feedback to the partner showing your interest, your comprehension is fundamental. HCI and HRI can be clearly improved by the introduction of advanced techniques for the analysis and synthesis of behaviours (emotion/ expression). In this project, we propose to implement a common platform for agent and robot's behaviour allowing the regulation of the interaction. The robot/agent will analyze the verbal and non-verbal signals such as speech prosody, gesture, face but also tactile information. In parallel, the robot/agent should also be able to provide nonverbal feedback as a sign of its engagement in the conversation.

The foreseen application is to have the robot/agent be an active listener to humans telling stories. The robot/agent would be able to show its interest and engagement in the story, show emotional reactions to the story being said. They would be able to negotiate their speaking turn.

Careful evaluation will be carried on to study humans interacting with the devices. In particular corpus of human-robot/agent interaction would be gathered. Non-verbal behaviours (gaze direction, acoustic and visual feedback, emotional signs) will be annotated. Qualitative comparison would be done between both interaction devices. One measure would be to evaluate the impact of the physical presence of the robot over the virtual one of an agent.

3. Background information: a brief review of the related literature, so as to let potential participants prepare themselves for the workshop

Human-Robot Interaction:

Human-Robot Interaction (HRI) is a multi-disciplinary field involving research on robot control (planning, sensor,…), speech processing (recognition, synthesis), vision (human localization, environment characterization), artificial intelligence, cognitive science and other fields [DAUT07]. Several methods have been proposed for the improvement of the interaction. Among the proposed approaches, social robots are more dedicated to the interaction. The key idea of their design is to develop robots with different capabilities: establish/maintain interaction, expression/perception of emotions, dialog, natural gesture and gaze, exhibit distinctive personality or to learn/develop social capabilities [FONG03, BRE04, OUD07]. Theses social robots aim at naturally interact with humans by the exploitation of these capabilities.

In this project, we will investigate one aspect of this social interaction: the engagement in the conversation [SID05]. The engagement process makes it possible to regulate the interaction between the human and the agent. This process is obviously multi-modal (verbal and non-verbal) and requires an involvement of both the partners. Some mechanisms as motivation, curiosity can be useful for this purpose [OUD07]. In this project, we propose to improve the human-robot interaction by exploiting the advances done for human-agent studies.

The term Embodied Conversational Characters (ECAs) has been coined in Cassell et al. [CAS00] and refers to human-like virtual characters that typically engage in face-to-face communication with the human user. Examples are REA [CAS99], an early system that realizes the full action-reaction cycle of communication by interpreting multimodal user input and generating multimodal agent behaviour, the pedagogical agent Steve [RIC99] which functions as a tutor in training situations, MAX [KOP04] a virtual character geared towards simulating multimodal behaviour, Carmen [MAR03] a system that supports humans in emotionally critical situations such as advising parents of infant cancer patients. Other systems realize presentation agents [AND00], i.e. one or more virtual agents present some information to the user. They can adopt several roles, such as being a teacher [JOH05; BAY05], a museum guide [GUS99; MAR06; KOP04] or a companion [BIC05; HAL06].

4. Detailed technical description:
    a. Technical description

    The project needs the following workpackages:
    - Speech Feature Extraction: prosody analysis, feature extraction for other workpackages
    - Partner tracking: face localization and tracking, gesture analysis.
    - Robot/agent behaviour modelling.
    - Dialog management, emotion and feedback model
    - Multimodal (speech and behavior) synthesis
    - Application scenario
    - Evaluation, corpus and annotation

    b. Resources needed: facility, equipment, software, staff etc.

    The ISIR laboratory will provide at least two Aibo robots for the experiments. The robots will be remotely driven by computers.

    The LINC lab will provide an ECA platform, Greta. The virtual agent is able to communicate using facial expression, gesture and gaze behaviors.

    Equipment: computers with IEEE 1394 interface, webcams, microphones, tripod, digital camera.

    Software: Windows, Linux, C/C++ , Java compilers

    Project management

    M. Chetouani, T. Dutoit, J.C. Martin, C. Pelachaud,

5. Work plan and implementation schedule: a tentative timetable detailing the work to be done during the workshop
    - Week 1 : scenario, presentation of a model of functions and non-verbal signs of engagement, illustration of annotation and representation of feedback, turn taking and related signs
    - Week 2 : introduction to virtual agent and robot software modules, implementation of the expression of non-verbal signs on the two platforms (robot + agent)
    - Week 3 : implementing common perception capabilities : speech feature extraction, partner tracking (face, body)
    - Week 4 : integration and evaluation, objective and subjective comparison of robot vs virtual agent interactions

6. Benefits of the research: expected outcomes of the project
NB : we insist on the fact the all the software components used for this project, and all the software built during the project, should be free for use, and available as such to all participants (including after the workshop).

Contrary to HCI and face-to-face interaction, HRI is less developed. The reproducibility of results and experiences is very limited due to the robot itself but also to the methodology used for the evaluation. One of the benefits of this project is the definition and the development of an open-source interface that can be exploited by different robots.

In addition, the two plateforms (agent robot) will be compared by annotation techniques.

7. Profile of team:

a. Leader (with a 1-page-max CV)

Mohamed Chetouani

He received the M.S. degree in Robotics and Intelligent Systems from the University Pierre and Marie Curie (UPMC), Paris, 2001. He received the PhD degree in Speech Signal Processing from the same university in 2004. In 2005, he was an invited Visiting Research Fellow at the Department of Computer Science and Mathematics of the University of Stirling (UK). Dr. Chetouani was also an invited researcher at the Signal Processing Group of Escola Universitaria Politecnica de Mataro, Barcelona (Spain). He is currently an Associate Professor in Signal Processing and Pattern Recognition at the University Pierre et Marie Curie. His research activities, carried out at the Institute of Intelligent Systems and Robotics, cover the areas of non-linear speech processing, feature extraction and pattern classification for speech, speaker and language recognition. He is a member of different scientific societies (ISCA, AFCP, ISIS). He has also served as chairman, reviewer and member of scientific committees of several journals, conferences and workshops.

b. Staff proposed by the leader (with 1-page-max CVs)

i. *You may propose some members of your future team. If possible, try to avoid having too many people from your group: part of the benefit of eNTERFACE is to let people meet and share experiences from different places, and possibly in different languages.*

CV Jean-Claude MARTIN

Jean-Claude Martin is Associate Professor at CNRS-LIMSI. He passed his habilitation to direct research in 2006 on « Multimodal Human-Computer Interfaces and Individual Differences. Annotation, perception, representation and generation of situated multimodal behaviors ». His research interests include the study of multimodal communication, both in human-human and human-computer contexts ; the study of individual differences wrt. the multimodal expression and perception of social behaviors and the evaluation of user's multimodal interaction with embodied conversational agents. He received the PhD degree in Computer Science in 1995 from the Ecole Nationale Supérieure des Télécommunications (ENST, Paris). He is the head of the « Conversational Agents » topic of research created in 2003 within the Architecture and Models for Interaction Group (AMI) at LIMSI-CNRS. He is and has been scientific leader at LIMSI for several European research projects related to multimodal communication such as: HUMAINE Network of Excellence (Human-Machine Interaction Network on Emotions) (2004-2008), IST-NICE (Natural Interactive Communication for Edutainment) which features multimodal input, Embodied Conversational Agents, Wizard of Oz, educational games (2002-2005), IST-ISLE (International Standards for Language Engineering) which featured multimodal corpora, coding schemes for multimodal behavior, surveys and guidelines (2001-2003). He co-organised a series of three international workshops on multimodal corpora at LREC 2002, 2004, 2006 and he

is a guest editor of a special issue of the international Journal on Language Resources and Evaluation to appear in 2007 on multimodal corpora.

Catherine Pelachaud

Since 2002 Catherine Pelachaud is a professor at the University of Paris 8, IUT of Montreuil. She currently in delegation at INRIA. She leads the research Laboratory in INformatics and Communication (LINC). She received her PhD in Computer Graphics at the University of Pennsylvania, Philadelphia, USA in 1991. Her research interest includes representation language for agent, embodied conversational agent, nonverbal communication (face, gaze, and gesture), expressive behaviors and multimodal interfaces. She has been involved in several European and national projects related to multimodal communication (EAGLES, IST-ISLE), emotion (FP5 NoE Humaine, FP6 IP CALLAS, FP7 STREP SEMAINE), to believable embodied conversational agents (IST-MagiCster, FP5 PF-STAR, RIAM ACE, ANR My-Blog-3D), and to handicap (CNRS-Robea HuGeS, RIAM LABIAO).

c. Other researchers needed (describing the required expertise for each)

People interested in joining the team should present expertise in related fields: Virtual agents, Speech Processing, Gesture and Face processing. Robotic.

8. References

[DAUT07] K. Dautenhahn (2007) Methodology and Themes of Human-Robot Interaction: A Growing Research Field. International Journal of Advanced Robotic Systems 4(1) pp. 103-108.

[FONG03] T. Fong, I. Nourbakhsh and K. Dautenhahn (2003): A Survey of Socially Interactive Robots, Robotics and Autonomous Systems 42(3-4), 143-166.

[BRE04] C. Breazeal (2004). "Social Interactions in HRI: The Robot View," R. Murphy and E. Rogers (eds.), in IEEE SMC Transactions, Part C.

[OUD07] Oudeyer P-Y, Kaplan , F. and Hafner, V. (2007) Intrinsic Motivation Systems for Autonomous Mental Development, IEEE Transactions on Evolutionary Computation, 11(2), pp. 265--286.

[SID05] Sidner, C.L.; Lee, C.; Kidd, C.D.; Lesh, N.; Rich, C., "Explorations in Engagement for Humans and Robots", Artificial Intelligence, May 2005

[CAS99] Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsson, H. and Yan, H. (1999). "Embodiment in Conversational Interfaces: Rea." Proceedings of the CHI'99 Conference, pp. 520-527. Pittsburgh, PA.

[CAS00] Cassell, J., Sullivan, J., Prevost, S., and Churchill, E. (eds). Embodied Conversational Agents. MIT Press, 2000.

[KOP04] S. Kopp and I. Wachsmuth, Synthesizing Multimodal Utterances for Conversational Agents, The Journal Computer Animation and Virtual Worlds, 15(1), PP 39-52, 2004.

[MAR03] S. Marsella, Interactive Pedagogical Drama: Carmen's Bright IDEAS Assessed. Intelligent Virtual Agents, pp 1-4, 2003.

[BIC05] T. Bickmore, J. Cassell, Social Dialogue with Embodied Conversational Agents", in J. van Kuppevelt, L. Dybkjaer, N. Bernsen (Eds) Advances in Natural, Multimodal Dialogue Systems, Kluwer Academic, New-York, 2005.

[AND00] E. André, T. Rist, S. van Mulken, M. Klesen and S. Baldes, The automated design of believable dialogues for animated presentation teams, in J. Cassell, J. Sullivan, S. Prevost and E. Churchill (Eds) Embodied Conversational Characters, MITpress Cambridge, MA, 2000.

[BAY05] A.L. Baylor, S. Kim, C. Son and M. Lee, Designing effective nonverbal communication for pedagogical agents, Proceedings of AI-ED (Artificial Intelligence in Education), Amsterdam, July 2005.

[RIC99] J. Rickel and W.L. Johnson, Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control, Applied Artificial Intelligence, 13, pp. 343-382, 1999.

[HAL06] L. Hall, M. Vala, M. Hall, M. Webster, S. Woods, A. Gordon and R. Aylett, FearNot's appearance: Reflecting Children's Expectations and Perspectives, Intelligent Virtual Agents IVA, Marina del Rey, USA, pp. 407-419, 2006.

[JOH05] W.L. Johnson, H. Vilhjalmsson, S. Marsella, Serious Games for Language Learning: How Much Game, How Much AI?, 12th International Conference on Artificial Intelligence in Education, Amsterdam, The Netherlands, July 18-22, 2005.

[GUS99] J. Gustafson, N. Lindberg, M. Lundeberg, The August sopken dialog system, Proceedings of Eurospeech'99, Budapest, Hungary, 1999.

[MAR06] Martin, J.C., Buisine, S., Pitel, G., and Bernsen, N. O.: Fusion of Children's Speech and 2D Gestures when Conversing with 3D Characters. In T. Dutoit, L. Nigay and M. Schnaider (Eds.): Multimodal Human-Computer Interfaces. Special Issue of Signal Processing (Elsevier) Vol. 86, Issue 12, December 2006