ISSN 2186-7437

NII Shonan Meeting Report

No. 163

Spoken Multimodal Dialogue Systems Technology for Pervasive Computing

Wolfgang Minker
Keiichi Yasumoto
Zoraida Callejas

September 30– October 03, 2019

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-Ku, Tokyo, Japan

Spoken Multimodal Dialogue Systems Technology for Pervasive Computing

Organizers:
Wolfgang Minker (Ulm University, Germany)
Keiichi Yasumoto (Nara Institute of Science and Technology, Japan)
Zoraida Callejas (Universidad de Granada, Spain)

September 30– October 03, 2019

1 Description of the Meeting

The technological development over the past decades and the wide-spreading of IoT, smart devices and sensors has enriched the number and variety of intelligent systems embedded in our environment. These systems have the capability to monitor our daily life and provide personalized and context-aware services. In this setting, designing the boundary between human and machine/intelligent environment through natural and seamless human-computer interaction is a hot research topic and within this context spoken multimodal dialogue systems become indispensable to offer the overwhelmed user the possibility to operate the plethora of intelligent devices, sensors and appliances in an efficient and natural manner. Therefore, there is a prominent interdisciplinary potential of both fields to generate sophisticated models that take into account the requirements for complex interactions in pervasive systems.

Interdisciplinary work coupling dialogue systems and pervasive computing technologies could have a significant impact in the adoption of ambience intelligence and ubiquitous computing solutions. On the one hand, dialogue systems can provide a more natural and human-like communication with pervasive systems, making them more transparent and easy to use and also more inclusive. However, although they have become the most featured and cutting-edge consumer interface, most consumer systems are still in an early stage because they predominantly use spoken dialogue only at a shallow level (e.g. for controlling home appliances, playing music, or searching the web). The integration of multimodal dialogue and sensing technologies developed in the pervasive computing area might allow to realize more complex context-aware interfaces and services.

On the other hand, pervasive applications have a huge potential to gain information about the users and their daily routines, providing very valuable information to generate rich user and multi-user models. Such models can be employed to infer relevant information about the users' knowledge, emotions and intentions as well as the context in which the interaction takes place. The use of these types of models can be relevant for intelligent multimodal systems to react adequately and in a user-friendly manner. Providing spatially distributed information, pervasive computing technologies are realised either in a

close contact with their users or smoothly integrated into the environment. Being a part of the smart environment, each smart device serves as a mini personal assistant and requires an ability to interact with the user so as to be able to be more adaptive, user-friendly and effective. This allows for user-adaptiveness and proactiveness in an ongoing interaction and makes the conversation more natural and human-like, e.g. using argumentative dialogue models which take multimodality and user preferences into account and thus, extend the system's communicative repertoire to cater various new interface possibilities.

While both research fields have recently been of interest for the respective research groups, questions of integration between pervasive computing and spoken multimodal dialogue systems have not yet been sufficiently discussed within each community. In other words, despite the described potential, the interdisciplinarity of both fields have not yet been investigated in detail. An active cooperation between both communities may open the way for more powerful and user friendly pervasive computing systems. We aimed to create the first and unique venue for discussion and collaboration between experts from these disciplines. This meeting has helped to explore possible challenges and jointly develop a research agenda for main directions. Topics of this workshop included, but were not limited to:

Research challenges:

- Challenges of spoken and multimodal dialogue interaction for pervasive systems.
- Challenges for specific components of multimodal dialogue systems (natural language understanding, speech recognition, dialogue management, specific modalities...).
- Single and multiple user/speaker identification, modelling and tracking within smart environments.
- Context awareness for multimodal dialogue systems.
- Personalization and user-centered development of dialogue systems for pervasive systems.
- Evolution of the technology and ideas for future research and application scenarios.

Development, testing and evaluation:

- Experimental design, user studies and evaluation of pervasive multimodal dialogue systems.
- Engineering approaches to dialogue interaction in pervasive systems: lifecycle, requirement elicitation, robustness to change, standardization and simulation.
- Development models, tools and strategies: middleware, tools, languages...
- Description, development and sharing of resources: corpora compilation, annotation tools and approaches, crowdsourcing approaches...

- Pervasive sensing devices and frameworks, the Internet of Things and its relation to multimodal dialogue systems.
- Coordination of conversational artifacts within smart environments (e.g. service discovery, coordination and argumentation...).

Use cases and industrial applications:

- Specific application domains for multimodal dialogue systems in pervasive computing: smart cities, public space (e.g. interactive digital signage, guidance systems), assistive environments (e.g. elderly care, hospitals)...
- Success stories, functional systems and industrial challenges.

Ethics and societal impact:

- Social responsibility.
- Data protection and privacy by design and default.
- Legal issues.
- Social design and development of conversational pervasive systems.

2 Meeting Schedule

Spoken Multimodal Dialogue Systems Technology for Pervasive Computing

Time Table	Arrival Day	1st Day	2nd Day	3rd Day	Final Day
	September 29th	**September 30th**	**October 1st**	**October 2nd**	**October 3rd**
7:00 - 7:30					
7:30 - 8:00		Breakfast	Breakfast	Breakfast	Breakfast
8:00 - 8:30					
8:30 - 9:00					
9:00 - 9:30		Introduction	Keynotes: **Catherine Pelachaud &**	Keynotes: **Satoshi Nakamura &**	Presentation of
9:30 - 10:00		Pecha Kucha 1	**Kristiina Jokinen**	**Leo Wanner**	Final Results
10:00 - 10:30		Break	Break	Break	Break
10:30 - 11:00		Pecha Kucha 2	Working Groups 2	Working Groups 4	Wrap up
11:00 - 11:30					Setting up a schedule for the
11:30 - 12:00					book proposal
12:00 - 12:30	Early check-in (negotiable)	Lunch	Lunch	Lunch	Lunch
12:30 - 13:00					
13:00 - 13:30		Group Photo Shooting		Excursion to Kamakura:	
13:30 - 14:00		Keynotes: **Roger Moore &**	Keynotes: **Iryna Gurevych &**	Visiting Engakuji and Kenchoji	
14:00 - 14:30		**Jaqueline Urakami**	**Yutaka Arakawa**		
14:30 - 15:00		Working Group Allocation	Break		
15:00 - 15:30	Regular check-in	Break			
15:30 - 16:00		Working Groups 1	Working Groups 3		
16:00 - 16:30					
16:30 - 17:00					
17:00 - 17:30		Presentation of	Presentation of		
17:30 - 18:00		Preliminary Results	Preliminary Results		
18:00 -18:30	Pre-meeting with Shonan staff	Dinner	Dinner		
18:30 - 19:00					
19:00 - 19:30	Welcome Banquet			Banquett	
19:30 - 20:00					
20:00 - 20:30					

3 Working groups

The participants of this meeting were divided into the following four Working groups according to their interests and preferences.

- Human-centered design development (adaptivity and personalization) of dialogue systems for pervasive systems and its ethical and societal impacts.
- Development of models and tools for pervasive multiple user identification and multimodal dialogue systems.
- Recommendation and behavior change support through multimodal dialogue systems in pervasive computing.
- Situation-based use cases and (industrial) applications.

Overview of Talks

Optimizing communication interaction

Roger K. Moore, Chair of Spoken Language Processing Dept. Computer Science, University of Sheffield

Whilst there is considerable interest in the possibility of people interacting with 'intelligent' agents for a wide range of applications, as yet there is no underpinning science as to how create such artefacts to ensure that they are capable of providing an effective and sustainable interaction. This is especially true in the case of human-robot interaction where the look, sound and behaviour of a robot may not be consistent - a situation that can trigger feelings of eeriness and repulsion in the users. Prof. Moore has developed the first mathematical model of this so-called 'uncanny valley' affect, and he is currently investigating the implications for designing autonomous social agents (such as robots) whose visual, vocal, cognitive and behavioural 'affordances' are appropriate to the role for which they are designed.
Prof. Moore is also interested in the computational mechanisms that underpin interaction between agents, especially (a) vocal interaction and (b) interaction between mismatched agents (e.g. humans and machines). To support this work, he has initiated a series of international workshops on 'Vocal Interactivity in-and-between Human, Animals and Robots' (VIHAR). Also he has developed a series of system architectures for modelling communicative interaction between agents that intentionally blur the distinction between the core components of a traditional spoken language dialogue system.

Human-Centered Design for Pervasive Systems and Its Ethical and Societal Impacts

Jaqueline Urakami, Tokyo Institute of Technology

In the Affective Laboratory research on how human affects influences technology use, satisfaction, and overall user experience is done. In recent projects we focused on empathic technology. Empathy is the ability to understand the other's situation, feelings and thoughts. By incorporating empathy into Human Agent interaction it is believed that people can build up long term relationships to intelligent agents that increase interaction enjoyment and motivates to repeatedly use these agents. Furthermore, empathy might enhance people's understanding of the capabilities and limitations of these systems. Another line of research focuses on cross-cultural design. Peoples' interaction with technical devices is affected by cultural learned patterns of communication, decision making, and perception. Her previous research has focused on cultural differences in people's use of gestures for table top systems, differences in people's communication pattern of mobile messenger services, and cultural differences in peoples' perception and decision making of predictive forecasts.

Exploring boundary crossing robots

Kristiina Jokinen, AI Research Center AIST Tokyo Waterfront

Social robots are expected to make important contributions in society in the coming years. Robots can be used in everyday life as companions, instructors, or providing support for those who have age-related challenges. Consequently, spoken interactions with robot agents are expected to increase, so it is important to design dialogue models to support natural interactions that the human partners can understand. Design should be based on a human-centred approach which emphasises natural multimodal communication and social aspects of interaction. Moreover, the robot should provide truthful and reliable information, and deal with sensitive information in a way that is regarded as socially tactful and trustworthy. While social robots should address the user's emotional needs, also safe and ethical data management has become their crucial feature, especially in smart environments with interconnected devices.
In her talk, Prof. Jokinen discussed her research related to designing and developing models for social interacting robots that can cooperate with humans in pervasive environment, share knowledge and information, and attend the human user in a socially competent manner. She discussed challenges and also shared examples from her work on human-robot interaction and dialogue modelling, in order to achieve deeper understanding of how human socio-emotional stance manifests itself in various interactive settings with robot agents.

Council of coaches

Catherine Pelachaud, CNRS – ISIR, Sorbonne University

Within the EU H2020 project Council of Coaches she is developing an interactive platform where user (a patient) can dialogue with several virtual coaches. Each coach has their specialized knowledge. We have conducted a perceptual study where we examine the impact of multiple agents on user's persuasion for behavior change. She considers three conditions: 1 agent, 2 agents communicating directly with the user and 2 agents relying on vicarious persuasion technique that discussed together while the user was an observer. She is continuing her research on using group of virtual agents to induce behaviors change into user. Currently she is analyzing a corpus of 2 or 3 persons in medical domain interacting with a patient. The data is analyzed along two cohesion dimensions, task and social, and multimodal behaviors. Lately, an architecture has been developed that allows an ECA to adapt to user's nonverbal behaviors during a dyadic interaction. As a first step she has conducted three studies that look at different levels of adaptation: conversational strategy, nonverbal behaviors, signals. Each adaptation mechanism was implemented in the same architecture that includes multimodal analysis of user's behavior using the platform Eyesweb (Volpe et al, 2016), a dialog manager (Flipper, van Waterschoot et al, 2018), a virtual agent GRETA-VIB. The architecture was tailored for each study. A same scenario was used for the three studies that were conducted in the science museum Cité des sciences et de l'industrie of Paris.

Argument Mining: Applications in a multimodal Dialogue System

Iryna Gurevych, TU Darmstadt

Iryna Gurevych is a Professor in the Dpt. of Computer Science at the Technische Universität Darmstadt, Germany. She is the Director of the UKP Lab and the Research Training Group "Adaptive Preparation of Information from Heterogeneous Sources" (AIPHES), funded by the German Research Foundation. Her interests are in natural language processing (NLP), such as computational argumentation, text semantics, or question answering. She has recently served as the PC Co-Chair of ACL 2018 in Melbourne, Australia. She co-authored the book "Linked Lexical Knowledge Bases: Foundations and Applications" (2016).
UKP Lab is a high-profile NLP lab, which conducts both research in Conversational AI and foundational NLP techniques for advanced AI dialogue skills. Examples include our research on searching and generating of arguments for recommendation behavior change support, community-based Question Answering and dialogue systems, or models of quality such as dialogue coherence assessment. We consider situation-based applications of dialogue systems for customer support, spoken interfaces to real-life databases and chatbots for scientific communication.

Potential of collaboration between ubiquitous computing and Spoken Dialog System

Yutaka Arakawa, Kysuhu University / JST PRESTO

His research topic was mainly focused on human activity recognition by using various sensors and machine learning techniques. However, the interest is slightly shifting to the intervention part after sensing, e.g. how to change human behavior by information. As an experimental place, he started the collaboration with several companies and occupational health psychologist who cares about the mental health of employees. In Japan, companies with more than 50 employees are required to have a mental check once a year. However, the form is just a simple questionnaire. Therefore, he is trying to estimate a continuous and quantitative psychological state by using wearable devices such as Fitbit and sensors on smartphones. The Spoken Dialog System is very attractive for him because it can ask the participant directly about his/her state, and also it can use for the intervention. He hopes for the creation of a next-generation chatbot therapy by combining sensing technologies and spoken-dialog system.

Spoken language technologies for tourism information analytics and social skill training

Satoshi Nakamura, Nara Institute of Science and Technology and Center for Advanced Intelligence Project, Riken

Currently he is working on two projects, tourism information analytics, and social skills training system related to dialog systems research.

Tourism information analytics project: Japan's Inbound tourism is drastically growing. The demands from tourists have been becoming stronger and more diverse and thus it is necessary to provide appropriate and real-time information for the users. His research goal is to develop IoT-to-human technologies, which analyze, anonymize, and visualize big data obtained by IoT-based sensing and social media, and provide information in real time by language-centered human-comprehensive multimedia.
Social skills training system: Increase of people who are not good at social communication and social anxiety disorder in schools and workplace (3 to 13%). His project is trying to solve by verbal / non-verbal interactive training system by Embodied Conversational Agents (ECA) in everyday situation. Here target population are general population, social anxiety disorders (SAD), and autism spectrum disorders (ASD). Social skills training (SST) will be applied for behavioral training and cognitive behavioral therapy (CBT) for cognitive training.

Advanced personal conversational assistants and data privacy and ethics in basic and healthcare. A tightrope walk?

Leo Wanner, ICREA and TALN, DTIC, Pompeu Fabra University

His main research interests revolve around a number of multilingual natural language processing applications, including, e.g., natural language generation, parsing, concept extraction, computer-assisted language learning, and computational lexicology and lexicography. In the context of spoken language analysis and generation, he is also interested in problems related to personalized multimodal interaction between intelligent conversational agents and humans, and, in particular, in social, societal and cultural competence of conversational agents. He recently coordinated the EU-funded KRISTINA Project on the development of an intelligent multilingual conversational agent with social competence and human interaction capabilities in the basic and healthcare domain and is the coordinator of the just approved WELCOME Project on the development of teams of conversational agents for the reception and integration of migrants in the European Union.

4 Summary of discussions

This section shall give a brief overview on the very fruitful and wide-ranging discussions during the Meeting. The first working group reformulated its topic as follows: Human-centered design of multi-modal interactive systems for ubiquitous systems and its ethical and societal impacts. Especially the following keywords were identified as very important in this context: ethical and societal impacts, human-centered design, dialogue systems for pervasive systems. It needs to be stressed that human-centered is not equivalent to user-centered design, even though in traditional literature it might be the same. However a human is not only the user of a system but also third party, which has to be considered. Thus the conclusion was derived that human-centered development should not only be technology-driven but should also consider people beyond the user-aspect.
The second working group focused on the questions why 'user identification' is needed and what the modalities are in multimodal systems? Furthermore, aspects of embodiment, such as "how do you embody a ubiquitous system?" and "physical embodiment is the interface?" were thoroughly discussed. The working group hereby focused on the related dimensions such as time (long term interactions), space (macro environments), agent (dyadic – multi-agent interaction), system (distributed), world (dynamic), modality (omnimodal) and communication (verbal and non-verbal signals).
During its discussions the third working group revised their topic to focus on technologies for dialog-based behavior change support. Behavior change/ adaptation to different situation need continuous support, which is hard to deliver by a human. The group developed as an exemplary scenario, which is concerned with the supervision of PhD students. This example was chosen as an illustration of the complexity of related areas, such as dialogue systems, persuasion techniques/strategy, data collection/acquisition, virtual coaches, learning for coaching (corrective actions), pervasive ubiquitous validation/control, behavior assessment. In various discussions the group tried to find answers to questions like "when to intervene?", "what is the best interface for a given context/person/topic/...?" and "How to collect (and share) the necessary information?".
The forth working group discussed various use cases, like smart cities/ campus/ shopping and corresponding dimensions like modalities, multi-lingual approaches, user expectations, pervasive computing and its limitations, current dialogue approaches and limitations, as well as new possibilities with IoT. Thereby research challenges and possible solutions were extensively explored.

5 Summary of new findings

There were multiple interesting findings derived from exchange within each working group. Maybe the most prominent one was establishing a clear distinctions between terms that a priori could seem synonymous. The multidisciplinary nature of the meeting helped to identify terms that had different conceptualizations in different academic communities, and the discussion lead to framing such concepts and bringing new shared meaning.

This was the case of user-centered vs. human-centered technology, personalization vs. adaptation, pervasive vs. ubiquitous, multimodal vs. omnimodal communication, dialogue vs. conversation and coaching vs. persuasion.

6 Summary of identified issues

An interesting result of the meeting was to identify challenges for future research, some of the open questions envisioned for IoT-based spoken dialogue applications are:

- When to speak?
- How to use in public area?
- How to target a person?
- How to guarantee privacy?
- How to abstract discourse?
- How to mix domains and use several applications?
- How to determine and get the right information?
- Can we predict?
- How to collect and share information?
- What information – session vs. core
- How to build a distributed context?
- How to share that information between devices/users?
- Everything is not application...

List of Participants

- Wolfgang Minker, Ulm University
- Zoraida Callejas, Universidad de Granada
- Keiichi Yasumoto, Nara Institute of Science and Technology
- Annalena Aicher, Ulm University
- Yutaka Arakawa, Kyushu University
- Martin Baumann, Ulm University
- David Griol, Universidad Carlos III de Madrid
- Iryna Gurevych, TU Darmstadt
- Naoya Isoyama, Nara Institute of Science and Technology
- Kristiina Jokinen, AIRC AIST Tokyo Waterfront
- Shogo Kawanaka, Nara Institute of Science and Technology
- Philippe Lalanda, Grenoble-Alpes University
- Bing Liu, University of Illinois at Chicago
- Guillaume López, Aoyama Gakuin University
- Takuya Maekawa, Osaka University
- Yuki Matsuda, Nara Institute of Science and Technology
- Roger Moore, University of Sheffield
- Satoshi Nakamura, Nara Institute of Science and Technology
- Catherine Pelachaud, CNRS-ISIR, Sorbonne University
- Francois Portet, Université Grenoble Alpes
- Niklas Rach, Ulm University
- Stephan Sigg, Aalto University
- Yasuyuki Sumi, Future University Hakodate
- Hirohiko Suwa, Nara Institute of Science and Technology / Riken
- Jacqueline Urakami, Tokyo Institute of Technology
- Leo Wanner, ICREA and Pompeu Fabra University
- Stefan Wesner, Ulm University
- Graham Wilcock, University in Helsinki
- Hirozumi Yamaguchi, Osaka University
- Takuya Yoshihiro, Wakawama University