YRRSDS 2024

The 20th Annual Meeting of the Young Researchers' Roundtable on Spoken Dialogue Systems

Proceedings of the Workshop

September 16 - 17, 2024
Kyoto, Japan

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 979-8-89176-162-9

Sponsor

The Association for Natural Language Processing

In Collaboration With

Preface

We are thrilled to present the opening remarks for the 20th Young Researchers Roundtable on Spoken Dialogue Systems (YRRSDS) 2024, a workshop dedicated to PhD candidates, PostDocs, and emerging researchers in the field of Spoken Dialogue Systems. YRRSDS 2024 was held in conjunction with the Special Interest Group on Discourse and Dialogue (SIGDIAL) 2024. The workshop took place on September 16-17, 2024, at Kyoto University in Kyoto, Japan. This year's YRRSDS was conducted in an in-person format.

Young researchers submitted a 2-page position paper detailing their current research topics, interests, and the key points they hoped to discuss during the workshop's roundtable sessions. Each submission was carefully reviewed by two senior researchers from our Advisory Committee. We extend our deep gratitude to the Advisory Committee members for their exceptional and insightful reviews. Their contributions have been invaluable in offering critical feedback to the workshop participants at this pivotal stage in their careers.

Participants accepted into the program were required to deliver a brief oral presentation based on their submissions. This year, YRRSDS accepted all 32 submissions received. The roundtable discussions covered topics such as LLMs, multimodality, explainability, evaluation, trustworthiness, ethics, safety, interdisciplinarity, human cognition, and the future of SDSs. Alongside the oral sessions and roundtables, the program featured two outstanding keynote presentations. We would like to express our gratitude and acknowledge our keynote speakers: Koichiro Yoshino (Associate Professor, Tokyo Institute of Technology) and Yoichi Matsuyama (Associate Research Professor, Waseda University and CEO of Equmenopolis, Inc.) for their inspiring talks.

We extend our gratitude to the organizers for making sure the conference ran seamlessly and was enjoyed by all attendees. Finally, we sincerely appreciate the support provided by our sponsor, The Association for Natural Language Processing (ANLP).

Organizing Committee, YRRSDS 2024

Organizing Committee

Organizers:

Koji Inoue, *Kyoto University*
Yahui Fu, *Kyoto University*
Agnes Axelsson, *Delft University of Technology*
Atsumoto Ohashi, *Nagoya University*
Brielen Madureira, *University of Potsdam*
Yuki Zenimoto, *Nagoya University*
Biswesh Mohapatra, *INRIA*
Armand Stricker, *Université Paris-Saclay*
Sopan Khosla, *Amazon Web Services AI Lab*

Advisory Committee:

Timo Baumann
Ryuichiro Higashinaka
Mikio Nakano
Marilyn Walker
Srinivas Bangalore
Ronald Cumbal
Luis Fernando D'Haro
Nina Dethlefs
Mikey Elmers
Tatsuya Kawahara
James Kennedy
Kazunori Komatani
Udo Kruschwitz
Marek Kubis
Divesh Lala
Pierre Lison
Alexandros Papangelis
Giuseppe Riccardi
Pawel Skorzewski
David Traum
Stefan Ultes
Nigel Ward
Hendrik Buschmeier
Kallirroi Georgila
Julia Hirschberg
Michimasa Inaba

Table of Contents

Conference Program

Monday September 16

13:00–13:10 Opening

13:15–14:00 Keynote 1: Koichiro Yoshino

14:20–15:20 Position Talks (Oral) 1

Conversational XAI and Explanation Dialogues
Nils Feldhus

Enhancing Emotion Recognition in Spoken Dialogue Systems through Multimodal Integration and Personalization
Takumasa Kaneko

Towards Personalisation of User Support Systems
Tomoya Higuchi

Social agents for positively influencing human psychological states
Muhammad Yeza Baihaqi

Personalized Topic Transition for Dialogue System
Kai Yoshida

Elucidation of psychotherapy and development of new treatment methods using AI
Shio Maeda

Assessing Interactional Competence with Multimodal Dialog Systems
Mao Saeki

Faithfulness of Natural Language Generation
Patricia Schmidtova

Knowledge-Grounded Dialogue Systems for Generating Interesting and Engaging Responses
Hiroki Onozeki

Towards a Dialogue System that Can Take Interlocutors' Values into Account
Yuki Zenimoto

15:45–16:15 Roundtable 1

Topic 1: Multimodality
Chair: Yuki Zenimoto, Koji Inoue

Topic 2: Data and Techniques
Chair: Yahui Fu, Armand Stricker

Topic 3: Evaluation
Chair: Brielen Madureira, Atsumoto Ohashi

16:45–17:45 Position Talks (Oral) 2

Multimodal Spoken Dialogue System with Biosignals
Shun Katada

Timing Sensitive Turn-Taking in Spoken Dialogue Systems Based on User Satisfaction
Sadahiro Yoshikawa

Towards Robust and Multilingual Task-Oriented Dialogue Systems
Atsumoto Ohashi

Toward Faithful Dialogs: Evaluating and Improving the Faithfulness of Dialog Systems
Sicong Huang

Cognitive model of listener response generation and its application to dialogue systems
Taiga Mori

Topological Deep Learning for Term Extraction
Benjamin Matthias Ruppik

Dialogue Management with Graph-structured Knowledge
Nicholas Thomas Walker

Towards a co-creation dialogue system
Xulin Zhou

Enhancing Decision-Making with AI Assistance
Yoshiki Tanaka

Ontology Construction for Task-oriented Dialogue
Renato Vukovic

18:00–19:00 Casual Dinner

Tuesday September 17

09:30–10:15 Keynote 2: Yoichi Matsuyama

10:30–11:45 Position Talks (Oral) 3

Generalized Visual-Language Grounding with Complex Language Context
Bhathiya Hemanthage

Towards a Real-Time Multimodal Emotion Estimation Model for Dialogue Systems
Jingjing Jiang

Exploring Explainability and Interpretability in Generative AI
Shiyuan Huang

Innovative Approaches to Enhancing Safety and Ethical AI Interactions in Digital Environments
Zachary Yang

Leveraging Linguistic Structural Information for Improving the Model's Semantic Understanding Ability
Sangmyeong Lee

Multi-User Dialogue Systems and Controllable Language Generation
Nicolas Wagner

Enhancing Role-Playing Capabilities in Persona Dialogue Systems through Corpus Construction and Evaluation Methods
Ryuichi Uehara

Character Expression and User Adaptation for Spoken Dialogue Systems
Kenta Yamamoto

Interactive Explanations Through Dialogue Systems
Isabel Feustel

Towards Emotion-aware Task-oriented Dialogue Systems in the Era of Large Language Models
Shutong Feng

Utilizing Large Language Models for Customized Dialogue Data Augmentation and Psychological Counseling
Zhiyang Qi

Toward More Human-like SDSs: Advancing Emotional and Social Engagement in Embodied Conversational Agents
Zi Haur Pang

13:30–14:00 Roundtable 2

Topic 4: Explainability and Trustworthy
Chair: Atsumoto Ohashi, Yuki Zenimoto

Topic 5: Taking Inspiration from Human Cognition
Chair: Brielen Madureira, Armand Stricker

Topic 6: Interdisciplinarity
Chair: Koji Inoue, Yahui Fu

14:30–15:00 Roundtable 3

Topic 7: Present and Future of SDSs
Chair: Atsumoto Ohashi, Yahui Fu

Topic 8: Possibilities and limits of LLMs
Chair: Armand Stricker, Yuki Zenimoto

Topic 9: Ethics and Safety
Chair: Koji Inoue, Brielen Madureira

15:30–15:45 Wrap up

15:45–16:00 Photo Session

16:00–18:00 Social Activity

Keynotes

Keynote 1: Multimodal Dialogue System Research and Careers, the Past 10 Years, the Future 10 Years

Koichiro Yoshino (Associate Professor, Tokyo Institute of Technology, Japan)

Abstract:

Over the past decade, advances in deep learning have made it easier for dialogue systems to handle various modalities in the real world, and research on multimodal dialogue systems has advanced enormously. Voice input devices such as Amazon Alexa and Google Home have entered our daily lives in the past ten years. Agent robots that handle various modalities will be realized as real-world services in another 10 years. What should young researchers consider in such an environment as they develop their careers? I participated as a PhD student at YRRSDS2014 10 years ago. Based on my experiences in academia over the past 10 years, I would like to discuss what career path you should follow in the next 10 years.

Biography: Koichiro Yoshino is an associate professor at Tokyo Institute of Technology, and is cross-appointed with RIKEN as a team leader. He received his B.A. from Keio University in 2009, and M.E. and Ph.D. in informatics from Kyoto University in 2014. He worked at Kyoto University and NAIST. From 2019 to 2020, he was a visiting research of Heinrich-Heine-Universität Düsseldorf, Germany. He is working on areas of spoken and natural language processing, especially robot dialogue systems. Dr. Koichiro Yoshino received several honors, including the best paper award of IWSDS2020, IWSDS2024, and the best paper award of the 1st NLP4ConvAI workshop. He is a member of IEEE Speech and Language Processing Technical Committee (SLTC), a member of Dialogue System Technology Challenge (DSTC) Steering Committee, an action editor of ACL rolling review (ARR), a board member of SIGdial and a board member of association for The Association for Natural Language Processing.

Keynote 2: From Dialogue System Research to Social Innovation

Yoichi Matsuyama (Co-Founder and CEO, Equmenopolis, Inc.)

Abstract: In the rapidly evolving field of dialogue systems, researchers hold a unique position, equipped to drive advancements in dialogue processing technologies while addressing emerging social needs. This talk will trace the journey from academic research to founding a university spin-out startup, showing how insights from cutting-edge technologies and user experiences can tackle real-world challenges and generate a social impact. Drawing from my experience as a dialogue systems researcher turned entrepreneur, I'll discuss the role of dialogue technologies in shaping the future of human-computer interaction and their broader implications for social innovation, encouraging young researchers to think creatively beyond academic boundaries.

Biography: Yoichi Matsuyama is the Co-Founder and CEO of Equmenopolis, Inc. and an Associate Research Professor at Waseda University in Tokyo. The mission of Equmenopolis is "Towards a Human-AI Co-Evolving Society," where we dispatch conversational AI agents to schools and workplaces to improve creativity and productivity. He specializes in developing computational models of human conversation, integrating AI, linguistics, social science, and human-agent interaction. Before his current role, he was a Postdoctoral Fellow at the ArticuLab, School of Computer Science, Carnegie Mellon University. His work has garnered attention from major media outlets, including MIT Technology Review, The Washington Post, CNBC, BBC, CNET, Popular Science, Nikkei, and NHK. He holds a B.A. in cognitive psychology and media studies, as well as an M.E. and Ph.D. in computer science from Waseda University, earned in 2005, 2008, and 2015, respectively.

Organizers' Notes of the Roundtable Discussions

Roundtable 1: Multimodality (Chair: Yuki Zenimoto, Koji Inoue)

Goal: Discuss the diverse aspects of multimodality in SDSs, including the utilization of visual information, environmental context, gestures, emotions, and personalization. Explore how these aspects can be effectively combined and what effect can be achieved.

Summary: In this discussion, we shared our individual works, focusing on multimodal dialogue systems as well as the challenges of sensing and annotation. We began by exploring the concept of ideal multimodal communication, emphasizing the need for dialogue to be smooth and duplex. In this context, we highlighted the significance of multimodal processing. We then addressed the challenges of annotating subjective phenomena like emotion labels. Additionally, we examined various measurable multimodal behaviors such as gestures, respiration, eye-gaze, and heart rate. Lastly, we delved into the interface of multimodal dialogue systems, comparing robots with CG agents/avatars and referencing the uncanny valley theory.

Roundtable 2: Data and Techniques (Chair: Yahui Fu, Armand Stricker)

Goal: Discuss the challenges and innovative approaches related to data creation, collection, and learning techniques for advanced SDSs, such as adaptation for unseen data, controllability, effective use of LLMs, and efficiency.

Summary: In this discussion, we first examined the challenges of prompting robustness and parameter-efficient fine-tuning techniques like prefix-tuning and LoRA, highlighting that language models are sensitive to prompt variations and benefit from training on diverse prompts to adapt to unseen data. Then, we compared synthetic data generation with human data labeling: synthetic data offers scalability but risks model degeneration and lacks diversity, while human-labeled data is richer but costly and prone to subjective interpretations and low inter-annotator agreement, especially in emotion recognition tasks. Lastly, we discussed large language models' effectiveness in processing speech data, noting they handle ASR noise well in tasks like emotion recognition but face difficulties in slot filling and with non-English accents. These insights emphasize the need for innovative data creation and learning techniques to improve adaptability, controllability, and efficiency in advanced SDSs.

Roundtable 3: Evaluation (Chair: Brielen Madureira, Atsumoto Ohashi)

Goal: Critically examine the current evaluation practices for SDSs and their limitations. Explore innovative automated evaluation metrics and methodologies for various domains, such as in non-task-oriented dialogues.

Summary: In this discussion, we focused on the challenges of evaluating dialogue systems, particularly LLMs, and emphasized the limitations of existing evaluation metrics such as BLEU. These metrics are especially problematic for open-domain dialogue systems, where human-like qualities are difficult to measure objectively. We debated the need for a more holistic approach that considers aspects like coherence and common sense. Benchmarking was another key topic, with concerns raised about models becoming over-specialized and "gaming" the system to perform well on specific tests rather than improving general performance. The balance between human and automatic evaluations was discussed. Participants concluded by stressing the importance of real-world testing and aligning evaluations with user needs, rather than purely focusing on making systems human-like.

Roundtable 4: Explainability and Trustworthy (Chair: Atsumoto Ohashi, Yuki Zenimoto)

Goal: Discuss the importance and methods of making SDSs more explainable and trustworthy, including the development of conversational explainable AI (XAI), the evaluation of reliability, the controllability of language generation, and dealing with closed proprietary models.

Summary: In this discussion, we discussed various aspects of explainability in dialogue systems, focusing on the challenges of making ML models interpretable and understandable for both scientists and general users. We identified a gap in tools that enable interactive explainability for general users and discussed the ethical implications of trusting explanations generated by models. We also touched on methods for evaluating the quality of explanations, with simulatability being one approach where users predict the model's output based on its explanation. Additionally, we raised the challenges of working with large black-box models (e.g., those accessed via APIs), where researchers lack insight into their inner workings.

Roundtable 5: Taking Inspiration from Human Cognition (Chair: Brielen Madureira, Armand Stricker)

Goal: Explore how insights from human cognitive processes, language acquisition, and social interaction can inform the development of more advanced SDSs, including the integration of physiological signals and the development of collaborative and creative systems.

Summary: The session centered on examining how insights from human cognitive processes can inform the development of more advanced spoken dialogue systems (SDSs). By integrating elements such as theory of mind, physiological signals, multimodal perception, and visual cues, dialogue systems can become more adaptive, capable of keeping information up to date, and better at integrating new facts in real-time. A major focus was on how these systems can emulate human-like understanding and illocutionary intent, as well as the implications of such advancements for user expectations, where more human-like behaviors tend to amplify the impact of errors when they occur. Additionally, the discussion explored how long-term interactions with SDSs could be improved by employing strategies like smoother turn-taking, meta-learning, and curriculum learning, ensuring that systems adapt to communication pace over time.

Roundtable 6: Interdisciplinarity (Chair: Koji Inoue, Yahui Fu)

Goal: Discuss how we can foster collaboration between the fields of SDSs and other disciplines, such as linguistics, psychology, robotics, and social sciences. Explore the benefits and ways to incorporate insights from other fields into practical SGSs development.

Summary: This discussion emphasized the importance of interdisciplinary collaboration, particularly integrating insights from psychology, linguistics, robotics, and social sciences. Participants highlighted the need for collaboration to create more human-like and efficient systems. Understanding users' emotions and adapting to text-based and spoken communication styles was noted as crucial, with experiments requiring careful design and sufficient participant numbers, potentially more than 100 in diverse real-world settings. While large language models (LLMs) provide a cost-effective way to test systems, they cannot replace the need for real human input. Psychological insights can improve LLM performance, but human evaluations are essential for quality. In robotics, transferring knowledge between systems like CommU and ERICA presents challenges. Ultimately, interdisciplinary collaboration and real human interaction are key to advancing SDSs.

Roundtable 7: Present and Future of SDSs (Chair: Atsumoto Ohashi, Yahui Fu)

Goal: Critically think about the current directions in the SDSs field and the reasons and necessity for doing so. Discuss future research directions, including to what extent human-like SDSs are desirable and the ideal relationship between humans and SDSs.

Summary: In this discussion, we shared our individual motivations for SDSs, such as the need for AI systems that fulfill communication needs and offer companionship. One key point of debate was the necessity of a physical body in AI companions, where some argued that emotional bonds could be formed through voice alone, while others maintained that physical interaction was essential in certain contexts, such as companionship or therapeutic relationships. There was also a discussion about whether modular or integrated approaches like LLMs would be more effective for future AI systems. Some highlighted the advantages of modular systems (e.g., better control and faithfulness), while others pointed out that multimodal models could simplify interaction. Finally, the discussion concluded with thoughts on the future of AI, with some expressing optimism about integrating advanced models with robots for even more sophisticated interactions.

Roundtable 8: Possibilities and limits of LLMs (Chair: Armand Stricker, Yuki Zenimoto)

Goal: Discuss the capabilities and limitations of LLMs in the context of SDSs. Explore how to effectively incorporate LLMs into SDSs, such as architecture design, controllability, and handling of multimodal dialogues.

Summary: In this discussion, we discussed the capabilities and limitations of Large Language Models (LLMs) in spoken dialogue systems (SDSs). We began by questioning the necessity of incorporating LLMs into SDSs and shared both successful and challenging experiences. A key topic was the constraints of autoregressive models, which generate responses token-by-token, potentially limiting their planning capabilities and output fidelity. We debated whether vision-language models or LLMs trained on multimodal data suffice for capturing complex meanings or if symbolic representations are necessary. We also explored the need to look beyond model responses to understand internal behaviors. The discussion then turned to the architecture of multimodal dialogue systems, weighing end-to-end against modular designs, and considering the limitations of prompt optimization in ensuring controllable and adaptable systems. Participants suggested combining prompt and internal modifications, such as chain-of-thought decoding and control modules, to refine the generation process. Ultimately, we emphasized the importance of balancing general model capabilities with task-specific requirements for optimal performance in SDS applications.

Roundtable 9: Ethics and Safety (Chair: Koji Inoue, Brielen Madureira)

Goal: Raise awareness about the ethical considerations and potential risks associated with the development and deployment of SDSs, such as when working with powerful but opaque models and creating human-like SDSs. Address concerns related to privacy, data rights, toxicity, and misinformation.

Summary: This discussion emphasized the importance of addressing ethical concerns in the development and use of spoken dialogue systems. It calls for ethical oversight during human trials and data collection, especially in sensitive areas like emotions or mental health. Ethical responsibility should not be entirely shifted to external bodies, with individual accountability being crucial. The lack of transparency in newer technologies like large language models (LLMs) poses challenges in explaining errors. There is a need to manage harmful behaviors, such as toxicity and inappropriate personalization, with reversible actions and consideration of cultural diversity. It also warned about the risks of over-reliance on commercial LLMs and advocated for transparency and open models for development and debugging. Governments may need to regulate the field, and developers should work to raise awareness of these limitations and responsibilities.