Spoken Language Interaction with Robots: Discussion Session

Nigel G. Ward\textsuperscript{1}, Matthew Marge\textsuperscript{2}

\textsuperscript{1}University of Texas at El Paso  
\textsuperscript{2}Army Research Laboratory  
nigelward@acm.org, matthew.r.marge.civ@mail.mil

This session will be a discussion based on the recent report: Spoken Language Interaction with Robots: Research Issues and Recommendations [Marge \textit{et al.}, 2020], focusing on the 31 recommendations, as follows:

\textbf{User Experience Design}

Focus on language not only as a way to achieve human-like behaviors, but also as a way to support limited but highly usable communications abilities. (U1)

Deliberately engineer user perceptions and expectations. (U2)

Work to better characterize the list of communicative competencies most needed for robots in various scenarios. (U3)

Design for use in multi-party and team situations. (U4)

\textbf{Robustness and Adaptability}

Include partially redundant functionality. (R1)

Make components robust to uncertainty. (R2)

Focus not only on improving better core components, but also on cross-cutting issues. (R3)

Make every component able to support realtime responsiveness. (R4)

Make systems and components adaptable to users. (R5)

Develop new ways to make components more reusable across tasks and domains. (R6)

\textbf{Infrastructure}

Create and distribute one or more minimal speech and dialogue-capable robot systems. (I1)

Update the representation of a robot’s physical surroundings continuously. (I2)

\textbf{Audio and Speech Processing, Speech Recognition, and Behavior Signal Analysis}

Develop the following shared resources: general toolkits for front-end audio processing, a database of robot-directed speech, and a challenge task on speech recognition for robot-directed speech. (A1)

Better exploit context and expectations in speech recognition. (A2)

Consider creating a speech recognition system focused on the issues encountered in robotics (e.g., speech segregation and speech in noise where the noise may come from the robot and/or the surrounding environment, possibly while the robot is moving). (A3)

Better exploit prosody, emotion, and mental state. (A4)

Use audio scene and event analysis to better understand the environment. (A5)

\textbf{Language Understanding}

Develop language understanding models for robots that resolve referential and other ambiguities in spoken dialogue. (L1)

Develop methods to infer and represent more information about humans from robot sensors. (L2)

\textbf{Dialogue and Human Communication Dynamics}

Focus on highly interactive dialogue. (D1)

Explore the broad space of recovery strategies in spoken language interaction with robots, including when and how. (D2)

Work to elucidate the fundamental questions in realtime social interaction, both scientific and engineering. (D3)

\textbf{Language Generation and Speech Synthesis}

Extend the pragmatic repertoire of speech synthesizers. (S1)

Develop speech generators that support multimodal interaction. (S2)

Create synthesizers that support realtime control of the voice. (S3)

Develop the ability to tune speech generators to convey a desired tone, personality, and identity. (S4)

\textbf{Policy}

Create funding opportunities specifically for spoken language interaction with robots. (P1)

Prefer evaluation based on use cases. (P2)

Support many kinds of research and development activities. (P3)

Work to overcome the barriers to data sharing. (P4)

Explore novel public-private partnerships for open source software. (P5)

\textbf{References}