

R&D Theme

Research and development on unconstrained spoken dialogue

Progress until FY2022

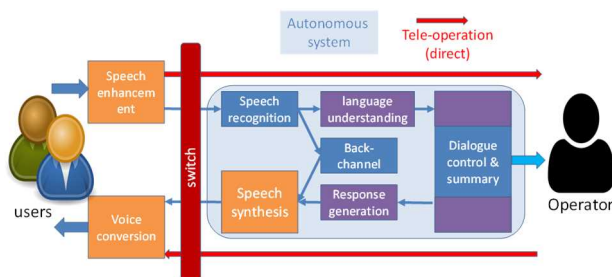
1. Outline of the project

We investigate automatic speech recognition and dialogue technologies to realize an autonomous spoken dialogue system with human-like hospitality, and develop flexible framework that allows the avatar to seamlessly switch between remote control dialogue and autonomous dialogue according to the operator's intentions and situation.

Thus, this group is responsible for the spoken language dialogue processing in this project. Spoken dialogue systems have been put to practical use in smart speakers and chatbots, but they are limited to uniform knowledge-level exchanges. To achieve human-like long and deep dialogues, it is essential to understand the user's situation (including inside and outside the dialogues) as well as to generate natural backchannels and empathetic responses.

2. Outcome so far

- Speech processing.



- (1) Real-time speech separation and recognition under real noisy environments are realized.
- (2) Low-latency speech synthesis with emotional expression

and high naturalness is realized. **The research team won the first prize in the Voice MOS Challenge in Interspeech 2022.**

- (3) We investigate a robot that generates backchannels and laughter in sync with the user. **This work has been covered by major world-wide media including BBC and Guardian, and selected as one of the best innovations in 2022 by the major French media.**

Scientists try to teach robot to laugh at the right time

Research team hopes system could improve natural conversations between humans and AI systems



- Natural language and dialogue processing.
- (1) We constructed the largest-scale dialogue data with persona information and the task corpus in Japanese.
 - (2) We designed highly naturalistic CG avatars and built a software environment that can operate both autonomously and remotely.
- Integrated system

- (1) We have developed a system that conducts attentive listening for three people in parallel and simultaneously.
- (2) We have built a system that provides explanations and guidance to three people simultaneously and in parallel. It was used for a field trial at an aquarium for one month.



3. Future plans

The individual modules of speech processing and dialogue processing have been developed to be deployed in the integrated system. We will test the system with subject experiments and field trials to feedback to the individual modules. We will also design more complex application scenario such as multi-party conversations, and develop the individual modules and the system.