Abstract

The turn-taking module of a conversational agent is a fundamental part of it, since if it works properly it gives to the user the feeling of an interactive and realistic conversation, otherwise it puts some barriers on the communication between the user and the conversational agent limiting and degrading the conversation experience. Such module should be reactive, to simulate the communication with a real person, and robust to noisy scenarios in order to perform correctly in different environment and use cases.

In this thesis we developed a remote and noise-robust turn-taking management system. At the core of this system there is the Silero VAD model a neural voice activity detection (VAD) model. We developed this module to be integrated in the Abel android, a hyperrealistic humanoid robot, used as a research platform in various AI application in the E. Piaggio research lab of the University of Pisa.

Furthermore, we validated our system creating an ad hoc noisy dataset, called Libri-Demand, that contains different noisy environments with different Signal to Noise Ratios (SNR), that is the level of noise with respect to the level of speech in the audio, comparing also the results of the system using the Silero VAD and the WebRTC VAD.