

Abstract

Even as progress in speech technologies and task and dialog modeling has allowed the development of advanced spoken dialog systems, the low-level interaction behavior of those systems remains often rigid and inefficient.

The goal of this thesis is to provide a framework and models to endow spoken dialog systems with robust and flexible turn-taking abilities. To this end, we designed a new dialog system architecture that combines a high-level Dialog Manager (DM) with a low-level Interaction Manager (IM). While the DM operates on user and system turns, the IM operates at the sub-turn level, acting as the interface between the real time information of sensors and actuators, and the symbolic information of the DM. In addition, the IM controls reactive behavior, such as interrupting a system prompt when the user barges in. We propose two approaches to control turn-taking in the IM.

First, we designed an optimization method to dynamically set the pause duration threshold used to detect the end of user turns. Using a wide range of dialog features, this algorithm allowed us to reduce average system latency by as much as 22% over a fixed-threshold baseline, while keeping the detection error rate constant.

Second, we proposed a general, flexible model to control the turn-taking behavior of conversational agents. This model, the Finite-State Turn-Taking Machine (FSTTM), builds on previous work on 6-state representations of the conversational floor and extends them in two ways. First, it incorporates the notion of turn-taking action (such as grabbing or releasing the floor) and of state-dependent action cost. Second, it models the uncertainty that comes from imperfect recognition of user's turn-taking intentions. Experimental results show that this approach performs significantly better than the threshold optimization method for end-of-turn detection, with latencies up to 40% shorter than a fixed-threshold baseline. We also applied the FSTTM model to the problem of interruption detection, which reduced detection latency by 11% over a strong heuristic baseline.

The architecture as well as all the models proposed in this thesis were evaluated on the CMU Let's Go bus information system, a publicly available telephone-based dialog system that provides bus schedule information to the Pittsburgh population.