

Concurrent Sub-turn Interaction Specification and Dialogue Management with an Application to Interactive Storytelling

Timo Baumann, Dorothee Geiser, Wolfgang Menzel,
Mario Mohr, Svenja Neef, Sören Nykamp, Nils Rokita*

*Universität Hamburg, FB Informatik
AB Natürlichsprachliche Systeme*

Conventional dialogue management centers around an interaction style that is best described as a ping-pong game, with full turns being the units at which speech is delivered (and expected) by the system, which greatly simplifies the interaction management, delivery and understanding components of the system. While the resulting mode of interaction works well for task-based systems, it is insufficient for more conversational interaction styles, where content is delivered and grounded in units finer than full turns (Poesio and Traum 1997) and where turns are delivered concurrently by both interlocutors and hence overlap more frequently. One domain with particularly frequent overlapping contributions is interactive storytelling: a storyteller that is responsive to listeners will integrate their feedback immediately while still speaking a current contribution, and listener's remarks or propositions regarding the story will typically be uttered immediately when the related content is delivered.

We present our work on a dialogue manager that leverages recent advances in *incremental* speech delivery and reception (Baumann 2013) to provide for an interactive and concurrent storytelling experience. Our system uses an *interaction graph*, which uses a word-by-word granularity and allows to specify for individual stretches of speech where (and with what content) users may interrupt/comment, how this is interpreted, and how it is integrated into the storytelling process, thus providing for in-utterance alternatives to be spoken even without audibly interrupting the system's ongoing utterance. The system's interaction graph is specified in an XML language and may be hand-crafted by a story designer but can also be automatically generated. In our current system, speech recognition results are interpreted only when the user utterance is finished; however, we plan to integrate incremental speech recognition and understanding capabilities, and to immediately react to the start of user contributions.

*Authors listed alphabetically.