

Collaborating on Utterances with a Spoken Dialogue System Using an ISU-based Approach to Incremental Dialogue Management

Okko Buß, Timo Baumann, David Schlangen
Department of Linguistics, University of Potsdam

I. Abstract

- Silence-end-pointed SDS: user must finish speaking before system planning and generation can begin.
- Human-human dialogue: actions occur during or immediately at the end of speech input, e.g. back-channels, interruptions, delivery in instalments. This is lost in end-pointed systems.
- We explore using trial intonation in collaborative utterance building. Present and evaluate incremental dialogue manager capable of modelling such behaviour.
- Highlights of system:
 - reacts to reference resolution problems, and (un-)certainty of user, as signalled by prosody
 - can execute requests concurrently; raises implicit "correct?" question
 - judged more human-like and more reactive than non-incremental counterpart

II. Collaboration on Utterances

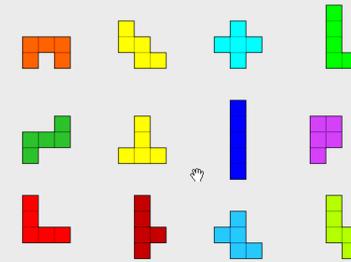
- Shape of an utterance determined not only by its speaker, but also by her addressee.
- Speakers monitor addressee responses online, to determine whether successfully grounded.
- Utterance itself may contain cues for grounding responses. E.g. (from Clark 1996)

A: A man called Annegra?–
B: yeah, Allegra
A: Allegra, uh, replied and, uh,...

A uses *try marker* (Schegloff 1979): "questioning upward intonational contour, followed by a brief pause", cueing response from B, which ultimately determines shape of the overall utterance.

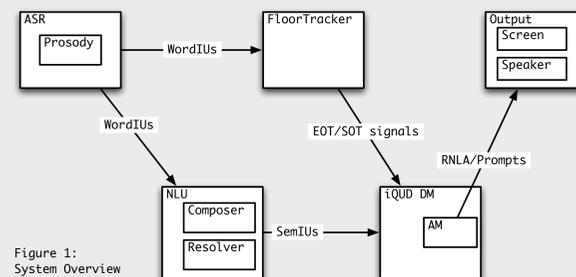
III. System, Domain

- Task-oriented, execute action (take, delete, etc.) on puzzle tiles (Pentomino):



IV. System, Overview

- Built with InproToolkit (cf. poster yesterday), middleware software package for building incremental spoken dialogue systems.
- Module network consists of speech recognition (ASR), with provisions for incrementality (Baumann et al., 2009), a FloorTracker, a unification-based natural language understanding component (NLU) and the iQUD DM (Section V.)

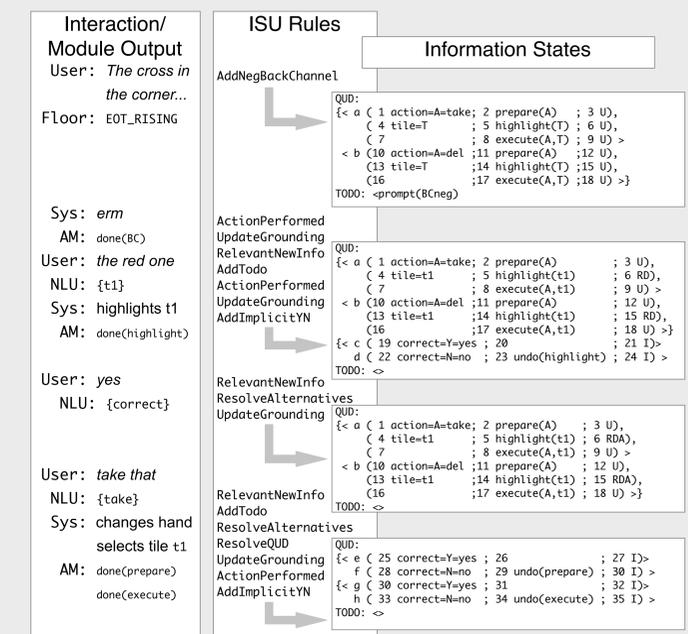


VI. Evaluation

- Overhearer evaluation against 800ms end-pointed system w/o concurrent actions.
- 30 minutes of interactions from both setups recorded (discarded 10% of "outlier" interactions.)
- Interactions from the incremental setting significantly shorter (t-test, $p < 0.005$).

V. System, Dialogue Management

- IS: question stack (iQUD) and a todo list. Update rules triggered by events from NLU, AM, FloorTracker, and changes to the IS.
- iQUD (variant of Ginzburg's (1996) QUD): collects relevant sub-questions, consisting of a semantic slot; a relevant non-linguistic action (RNLA); and grounding state.
- NLU events downgrade sub-questions, triggering rule sending RNLA to AM via todo list. Upon completion, AM rules update grounding state. FloorTracker rules cue system utterances.



- Judgments on helpfulness (1), human-likeness (2) and reactivity (3), Likert-scale (8 subj./34 randomly selected recordings). Incr. sys rated higher on 2 and 3 (Wilcoxon rank-sum test; $p=0.04$, $p>0.005$).

VII. References

Please see paper in proceedings.