Multiparty Dialogs

John Riebold

Participant Roles

- **Conversational Roles**
  - 2 participants:
    - Speaker
    - Addressee
  - 3+ participants:
    - Speaker
    - Addressee
    - Auditor (known, ratified)
    - Overhearer (known, non-ratified)
    - Eavesdropper (unknown, non-ratified)

    (Bell, 1984)

Participant Roles

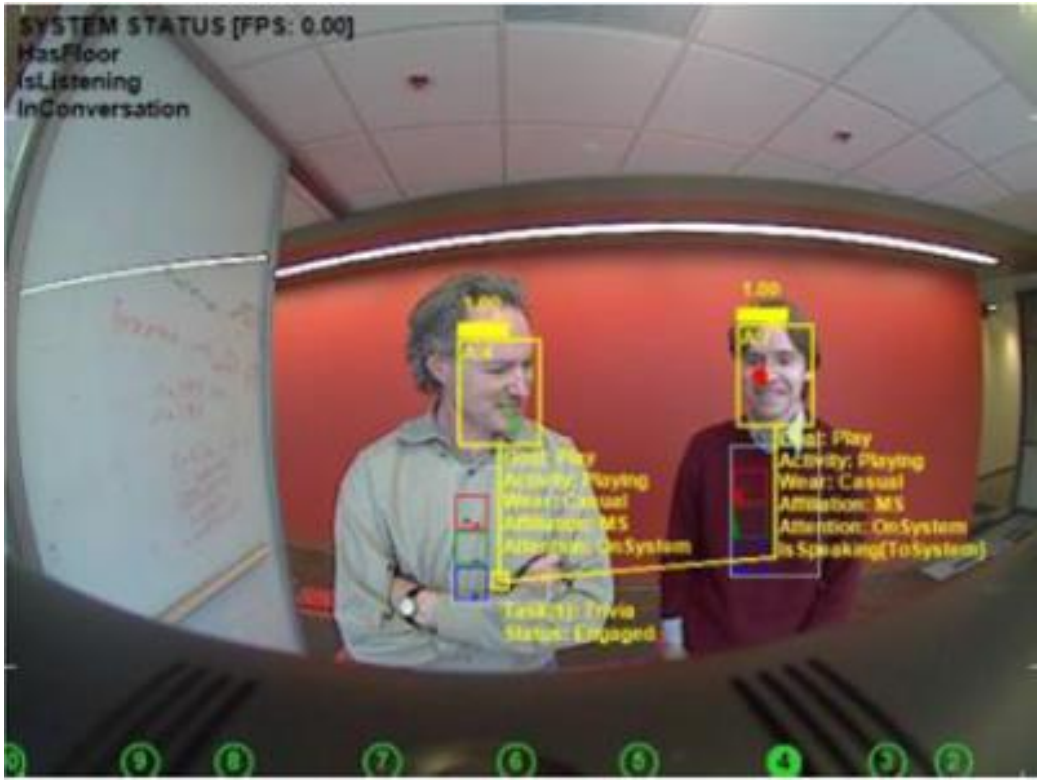
- **Speaker Identification**
  - Difficult in multiparty dialogs
  - Can be done acoustically, with a microphone array, or visually
- **Addressee Recognition**
  - Multiparty dialogs present many more possibilities
  - Addressee can be inferred from content (e.g. name, position/rank, etc.)
  - Can also be done with positional audio or video

Participant Roles

- **Addressee Recognition**
  - Jovanovic & op den Akker (2004) presents a set of features that could be used to perform addressee recognition:
    - Speech
      - Linguistic markers (e.g. to infer person, number)
      - Names
      - Rank/title?
      - Dialog acts (specifically, relation to previous conversation and effect on subsequent conversation
    - Gaze
    - Gesture
    - Context (e.g. user/conversation history, spatial organization)

Participant Roles

- **Speaker & Addressee Identification**
  - Bohus & Horvitz (2009) used video to identify speakers and addressees
  - Part of a more sophisticated engagement system



Interaction Management

- **Turn Management**
  - Turn-taking in multiparty dialog can be complex
  - More agents available to take a turn
  - Humans may drop some turn-taking expectations in conversation with a machine, but won't with other people
  - Depending on the system, crucial evidence may not be available (e.g. video, audio)

Interaction Management

- **Turn Management**
  - Bohus & Horvitz (2011)
    - Used Decision Theory to model turn-taking and allow the system to take the floor at relevant junctures
    - Leveraged audio/video info, previous turn info, time since previous turn, processing delays, and cost
    - Compared heuristic vs. learned (MaxEnt) models of floor release, and heuristic vs. Decision-theoretic models of turn-taking policy

Model		Cost
Floor Release Inference	Policy	
Heuristic	Heuristic	0.43
Learned	Heuristic	0.29
Learned	Decision-theoretic	0.21

Interaction Management

- **Channel Management**
  - Multiparty dialogs may have multiple channels (i.e. multiple conversations)
  - May share a single channel (i.e. single topic, one speaker at a time)

Interaction Management

- **Thread/Conversation Management**
  - Multiparty systems must manage a complex set of shifting (and often linked) topics
  - Side conversations can entail an entirely separate set of threads
  - Current thread bears on turn-taking, obligations, grounding, etc.

Interaction Management

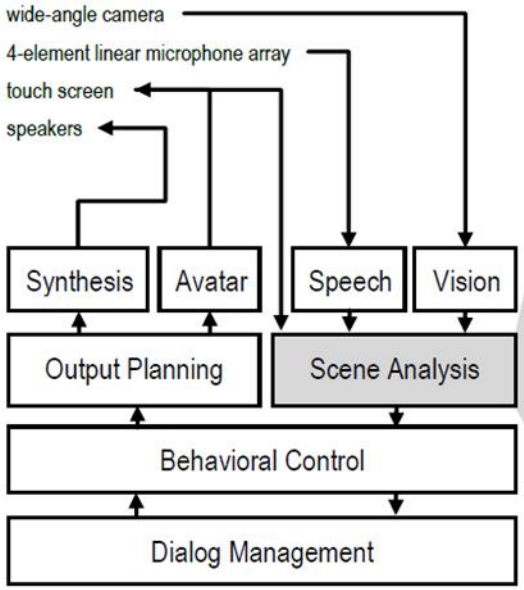
- **Thread/Conversation Management**
  - Purver, et al. (2007) look at the automatic detection of subdialogs
  - Detection of subdialogs is done with classifiers using various features:
    - ngrams
    - Utterance length
    - Prosody
    - Time expression tags
    - Dialog acts
    - Context
  - Classifiers outperform the baseline, but take a hit when using errorful ASR input

Interaction Management

- **Initiative Management**
  - Multiparty may have unevenly-distributed initiative
  - Speakers can defer to others
  - Interruptions are more likely

Interaction Management

- **Attention Management**
  - Managing multiple (possibly uninvolved) participants is necessary in multiparty systems
  - Bohus & Horvitz (2009) model multiparty engagement using acoustic, positional, visual, and tactile information



Grounding and Obligation

- Multiparty dialogs may have very complex grounding and obligations
- If information is presented in one conversation, must it be grounded in another?
- How should a system handle transfer of obligation?

Grounding and Obligation

- Purver, et al. (2007) also look at the automatic detection of ‘action items’ (obligations)
- They train a classifier to rank phrases based on various features:
  - Phrase length
  - Phrase probability
  - Parse probability
  - Syntactic features (class, theta roles, main verb, head noun, etc.)
  - Time expression tags
- Evaluated based on amount task descriptions covered by top-ranked fragment
- Results for timeframe phrases were above baseline, but still relatively low (f-score 0.51, precision 0.62). Results for description were worse, with no feature set outperforming the baseline.

Discussion

- What possible use cases are there for systems like MSR's Situated Interaction?
- Would it be worth implementing these systems in commercial applications?
- Are there other cues or types of information that aren't being used in these models?

References

Bell, A. (**1984**) Language Style as Audience Design. In Coupland, N. and A. Jaworski (eds.) *Sociolinguistics: a Reader and Coursebook*, pp. 240-50. New York: St Martin's Press Inc.

Bohus, D. & Horvitz, E. (**2009**) Models for Multiparty Engagement in Open-World Dialog. In *Proceedings of SIGdial 2009*.

Bohus, D. & Horvitz, E. (**2011**) Decisions about Turns in Multiparty Conversation: From Perception to Action. In *ICMI-2011*.

Jovanovic, N. & op den Akker, R. (**2004**) Towards automatic addressee identification in multi-party dialogues. In *Proceedings of Sigdial 2004*.

Purver, M., Dowding, J., Niekrasz, J., Ehlen, P., Noorbaloochi, S., & Peters, S. (**2007**) Detecting and Summarizing Action Items in Multi-Party Dialogue. In *Proceedings of SIGdial 2007*, p. 18-25.

Traum, D. (**2004**) Issues in multiparty dialogues. In F. Dignum (ed.), *Advances in Agent Communication*. Springer-Verlag LNAI 2922, p. 201-211.