

Introduction

Online crowd-sourcing

- A large number of conversations happen online.
- Deploying dialogue systems online gives access to a large number of participants.
- HTML5 technologies support developing dialogue systems better.
- Crowd-sourcing saves money and time.

Designing and development of interactive spoken dialogue systems online

- Human-Human (N=326) & Human-Agent (N=150) systems built and deployed.
- Data from crowd-sourced Human-Human interaction used for building the agent
- NLU, ASR, Dialogue Manager(DM) incremental in nature

Task

- RDG-Image game: Two player collaborative spoken dialogue game.
- Director : Describes the image highlighted
- Matcher : Guesses the image based on the clues from Director.
- Game is fast-paced with acknowledgements, overlapping speech, disfluencies, laughter etc.

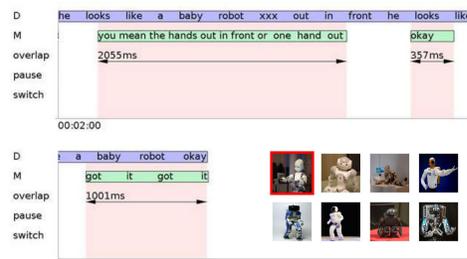


Fig 1 : Shows users conversing about the highlighted image.

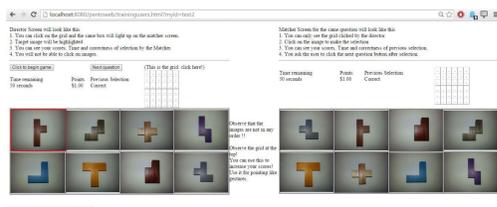
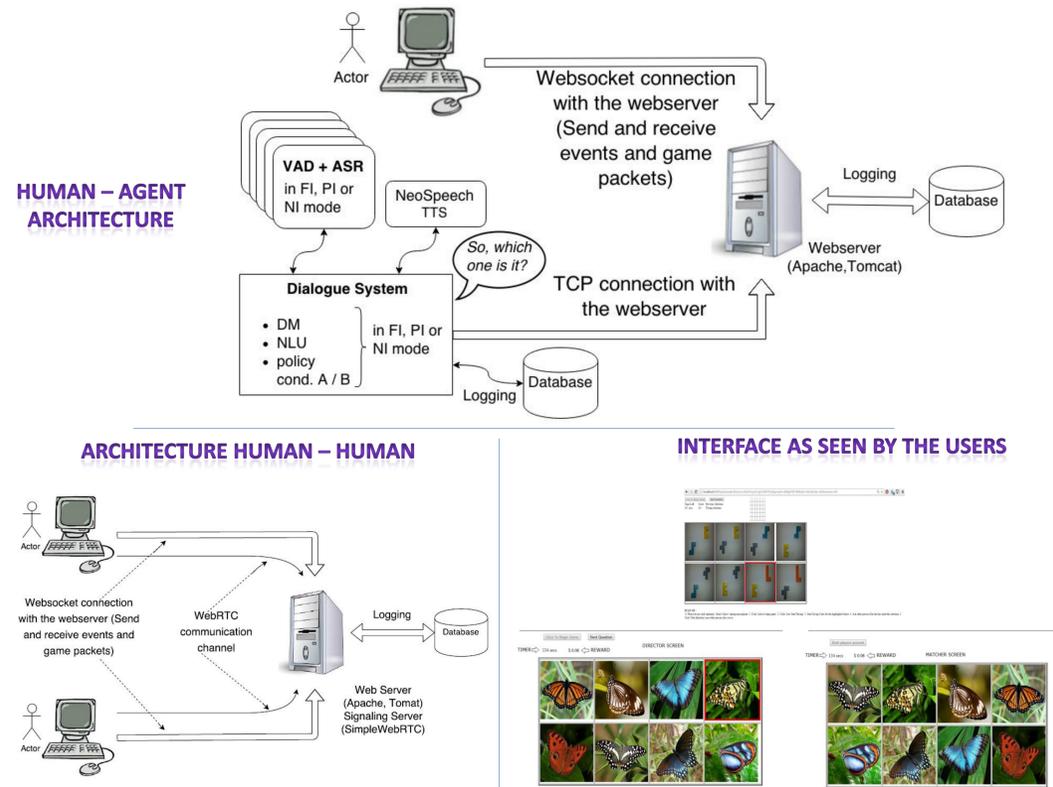


Fig 2 : Shows Game interface used for another version of the game used for English & German data collection

- Users recruited web on Amazon Mechanical Turk.
- No additional installations. Google Chrome dependent.
- Uses latest web paradigms (WebRTC, Websockets etc.)

Design & Interface



Data Collection (PMU Framework)

- Study deployed over Amazon Mechanical Turk.

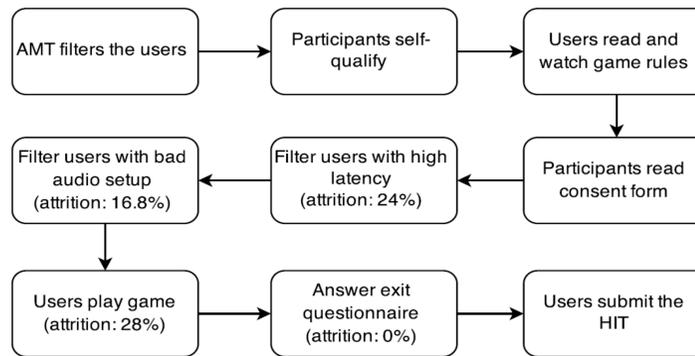


Fig 3 : Protocol for conducting the study on Amazon Mechanical Turk.

- Spoken dialogue data collection, transcription and evaluation performed in crowd-sourced manner.

- We use Pair Me Up (PMU) [1].

- Data synchronization important.

- Measuring latency helps get fairly accurate timing for the packets.

- High latencies can impact data and user performance [1]

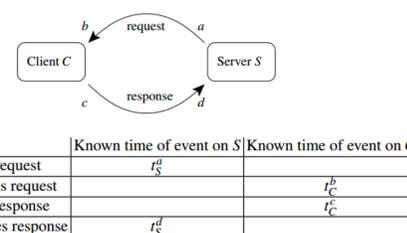


Fig 4: Latency measurement protocol

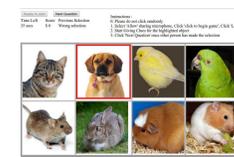
Results

- Study 1: Human-Human Data Collection 1 (N=196) [1]



- Study 2: Human-Agent Evaluation Study (N=200) [2][3]

- Developed 3 different versions of the agent operating on 2 different optimization policies.
- Money and time saved.
- Agent is very interactive, real-time and responds at word level.



	Web	Lab
Participant Fees	\$1.24	\$15
Staff time per participant	2.5 min	~35 min
Cost of Server Time	\$0.72/hour/machine	-
Participant Time	1193.1 sec	~1800 sec

Fig 5: Savings in time and cost in lab vs web.

- Agent: {1. The normal case} Human: It's a dog with pink ok Agent: Got it Agent: Got it Human: Correcting the description} Human: It's a cat Agent: Got it Human: Sorry, it's a dog with pink collar Agent: oh I see! I got it
- Agent: {3. If it doesn't make sense to keep listening} Human: oh my god! I think it's a a.. Agent: I don't think I can get that one. Lets move on. And many others....

- Study 3: Human-Human Data Collection 2 (N=80) (Ongoing)

- Ongoing study collects spoken interaction data in English and German.
- Understand the complex scene descriptions (colors, spatial language, plurals, relations) using [4]

Future Directions

- Develop multimodal rapid conversational agent .
- Modify the task into a tutoring task and develop a pedagogical agent.
- Develop complex policies for the agent.

ACKNOWLEDGMENTS:

Special thanks to David DeVault, David Schlangen & Maikae Paetzel. Also thanks to Dialogue Systems Group, Bielefeld University. We also like to thank the people for sharing the images. (Names mentioned in [1][2][3][4]) This work was supported by the National Science Foundation under Grant No. IIS-1219253 and by the U.S. Army. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views, position, or policy of the National Science Foundation or the United States Government, and no official endorsement should be inferred.

REFERENCES:

- [1] Ramesh Manuvinakurike and David DeVault. 2015. "Pair Me Up: A Web Framework for Crowd-Sourced Spoken Dialogue Collection." In International Workshop on Spoken Dialogue Systems (IWSOS)
- [2] Maikae Paetzel, Ramesh Manuvinakurike, and David DeVault. 2015. "So, which one is it?" The effect of alternative incremental architectures in a high performance game-playing agent. In The 16th Annual SIGDIAL Meeting on Discourse and Dialogue (SIGDIAL).
- [3] Ramesh Manuvinakurike, Maikae Paetzel and David DeVault. 2015. "Reducing the Cost of Dialogue System Training and Evaluation with Online, Crowd-Sourced Dialogue Data Collection." In 19th workshop on Semantics and Pragmatics of dialogue (SemDial)
- [4] Maikae Paetzel, David Nicolas Racca, and David DeVault. 2014. A multimodal corpus of rapid dialogue games. In Language Resources and Evaluation Conference (LREC).
- [5] Casey Kennington and David Schlangen. 2015. "Simple Learning and Compositional Application of Perceptually Grounded Word Meanings for Incremental Reference Resolution." In Proceedings of the Conference for the Association for Computational Linguistics (ACL).