# Situated Intelligent Interactive Systems

**Zhou (Jo) Yu**
Computer Science Department
University of California, Davis

# Vision, Samantha

# Situation Intelligence

# Current Dialog Systems, Siri

# User Complaints

1. Bad at understanding me
2. Bad at remembering things
3. Bad at being coherent
4. Bad at being interesting
5. Bad at providing variety
6. Bad at being natural
7. Didn’t do anything!!!

# Human-Human Social Conversations

# Active Participation Strategy

# Grounding Strategy

# A Solution: Situation Intelligence Framework


**Situation Awareness**: Awareness of situation, such as people, time, environment, etc.

*e.g. User engagement in social conversation*

**Conversation Strategy**: System actions that react to situation.

*e.g. Active participation strategy to improve user engagement*


**Statistical Policy**: Policies that choose among strategies considering situation (especially history) to optimize towards a long-term natural interaction.

*e.g. Reinforcement learning policy*


# Situation Awareness

Definition: Awareness of situation, such as people, time, environment, etc

Examples:  Engagement in social conversation

Challenges: What to select, sense, track and reason? Various information from different channels.

Methods: Leverage task domain knowledge to select critical aspects of the situation that benefit the interaction most. Then use multimodal information to approximate these aspects automatically.

# Conversation Strategy

Definition: System actions that react to the situation.

Examples: Active participation strategies to improve user engagement

Grounding strategies to improve language understanding

Challenges: How to design natural and effective actions to facilitate communication?

Method: Leverage conversation theories to guide the use of knowledge bases and NLP methods to design actions


# Statistical Policy

Definition: Policies that choose among strategies considering situation (especially history) to optimize towards a long-term natural interaction.

Examples: Reinforcement learning policy that considers user sentiment

Challenges: Conversation actions hinge on history. A slight change will lead to different sequences. What aspects of the interaction history to consider?

Methods: Use reinforcement learning to optimize the sequential decision process and leverage conversation theories to design learning parameters.

# Engaging Social Conversation Systems

# Wide Applications

## Applicable in various areas, such as and entertainment, education and health care.

Entertainment: Create targeted advertisement (Yu et al., IJCAI 2017, SLT 2016)

Education: Provide training (Yu et al., IWSDS 2016) and facilitate discussion on MOOCs

Health care: Support therapy for depression (Yu et al., SEMDIAL 2013), aphasia and dementia

## Applicable in various platforms: virtual agents and robotics

Virtual: Build characters for games

Robots: Service robots, e.g. direction giving (Yu et al., SIGDIAL 2015), nursing and rescuing

# Outline

Situation Intelligence Framework

**Engagement Coordination**

**Situation Awareness**

Conversation Strategy

Statistical Policy

Social Conversations


Other Applications: Movie Promotion and Interview Training

Attention Coordination

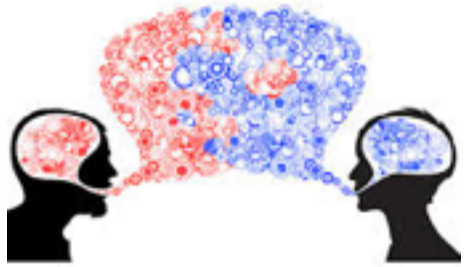
Direction Giving

Future Work

# Situation Awareness: Engagement

**Engagement:** Interest to continue, to contribute to the conversation
*(Peters et al., 2005)*
**Social Conversation:** Everyday social chatting.

# Supervised Multimodal Engagement Prediction



Verbal :

*Cloud ASR*

word count
turn length
time to respond

Visual:

*OpenFace (Baltrusaitis et al., 2015)*

head pose
action units
gaze direction

Acoustic:

*SphinxBase (Huggins-Daines et al., 2006)*

power
pitch

# Experiment

Conversation data: 23 (14 male) interactions (North American), 5+ minutes

Annotation scale: 1-5 Likert scale

Annotation unit: Per conversational exchange

Inter-annotator agreement: 0.93 in kappa (after collapsed into two-point scale, between two expert annotators)

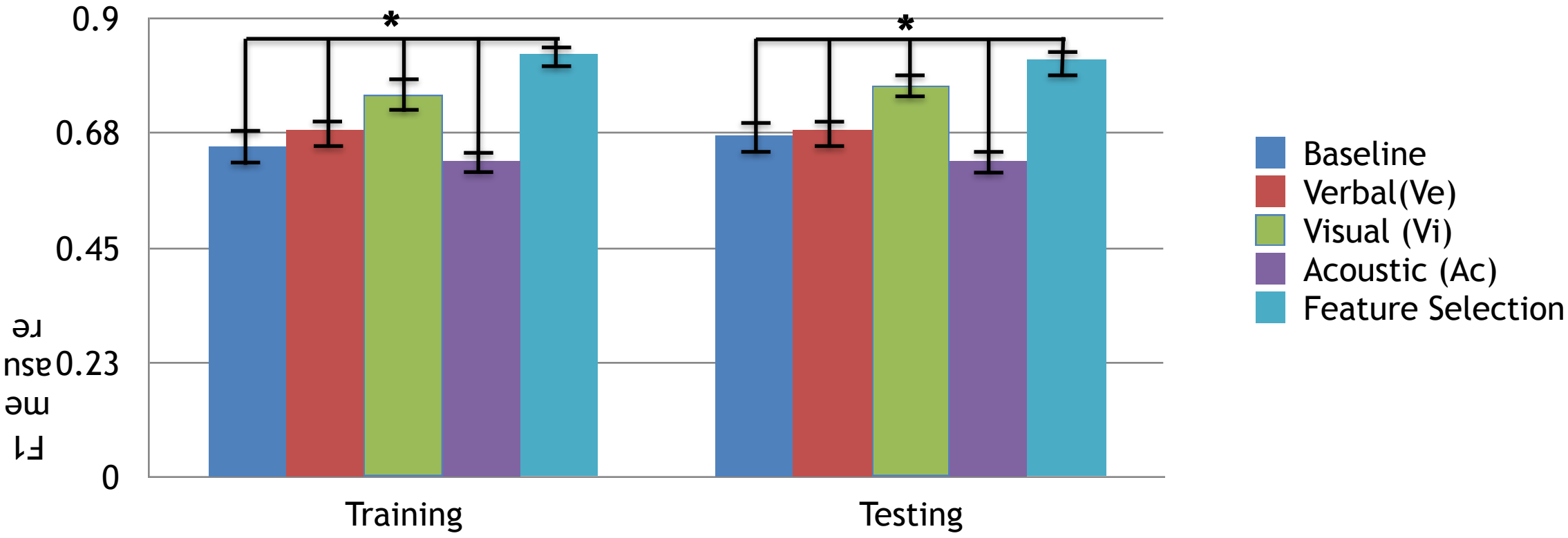
Machine learning setting: Early fusion for feature combination, leave one interaction out cross-validation and SVM with RBF Kernel

# Multimodal Engagement Prediction Results

Yu et al., Thesis, 2016

Combining Multimodal info achieves best result: 0.81 in F1- measure

Late Fusion in Prediction



# Outline

Situation Intelligence Framework

Engagement Coordination

Situation Awareness

**Conversation Strategy**

1. Active Participation Strategies
2. Grounding Strategies
3. Personalized Strategies

Statistical Policy

Other Applications: Movie Promotion and Interview Training

Attention Coordination

Direction Giving

Future Work

# Other Response Generation Methods

## Keyword Retrieval

Method: Search keywords in the database -> Return the corresponding response of the sentence with heights weighted score

Database: CNN Interview transcripts and Mturk collected human response

## Skip-Thought Neural Network Model

Method: Auto-encoder-> Decoder, with one turn context

Database: OpenSubtitle2016

# 1. Active Participation for Engagement Coordination

Definition: Actively participate in conversations to attract partner's involvement (Daniel Wendler, 2014).

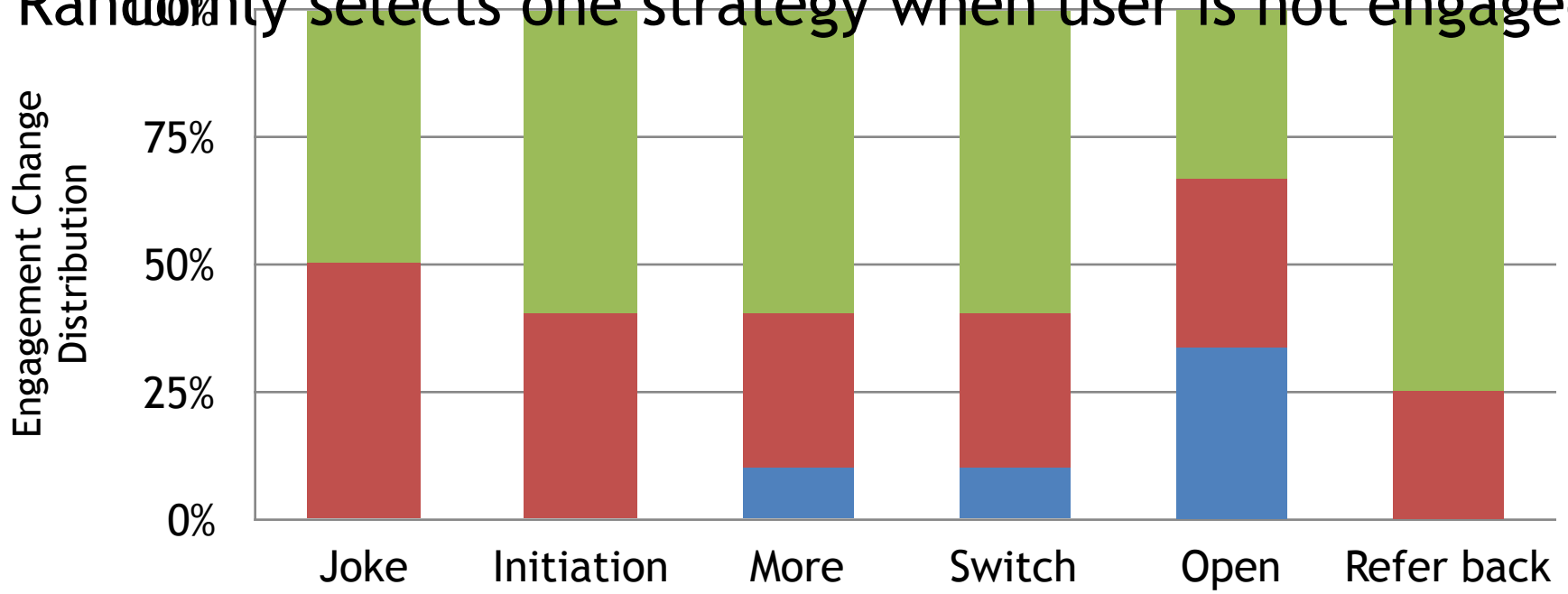
| Stay On Topic | Change Topic |
| --- | --- |
| Tell a joke (Joke)<br>e.g. Do you know that people usually spend far more time watching sports than playing any. | Switch topics (Switch)<br>e.g. How about we talk about movies? |
| Initiate activities (Initiate)<br>e.g. Do you want to see a game together some time? | End topics with an open question (Open)<br>e.g. That's interesting, could you share with me some interesting news on the Internet? |
| Talk more (More)<br>e.g. Let's talk more about sports. | Refer back to engaged topic (Refer back)<br>e.g. You said you really liked movies, do you |

# Results: Active Participation Strategies

Yu et al., SIGDIAL, 2016a

Experiment Setting: 30 multimodal conversations in a Wizard-of-Oz setting.

Policy: Randomly selects one strategy when user is not engaged



# 2. Grounding Strategies for Understanding

## Utilize static general knowledge base: Google Knowledge Graph, Wikipedia

### Ground on named entity (GroundEntity)

*e.g.* *User: Do you like Clinton?*

*TickTock: Are you talking about Bill Clinton, the 43rd president of United States or Hillary Clinton, the secretary of the state?*

Procedure: Detect named entity -> knowledge base (frequent referred entities) -> statistical template

### Ground on out of vocabulary words (GroundOOV)

*e.g.* *User: You are very confrontational TickTock.*

*TickTock: What do you mean by confrontational?*

Procedure: Not in knowledge base -> statistical template -> update knowledge base

### Ground on Single Word (GroundSingle)

*e.g.* *User: 1+2=?*

*TickTock: Can you say something in a complete sentence?*

Procedure: Single word -> not in knowledge base

# 3. Personalized Strategy for Adaptation

Utilize dynamic personal knowledge base: Personal opinions automatically extracted from each user's interactions.

Personal Suggestion Strategy (PersonalSuggest)

*e.g. User: Do you have any suggestions for movies?*

*TickTock: How about Captain America 2? (Knowing user said he liked superhero movies before.)*

Procedure: Topic detection -> personalized knowledge base -> statistical template

Don't Repeat Strategy (NoRepeat)

*e.g. User: What is your favorite movie?*

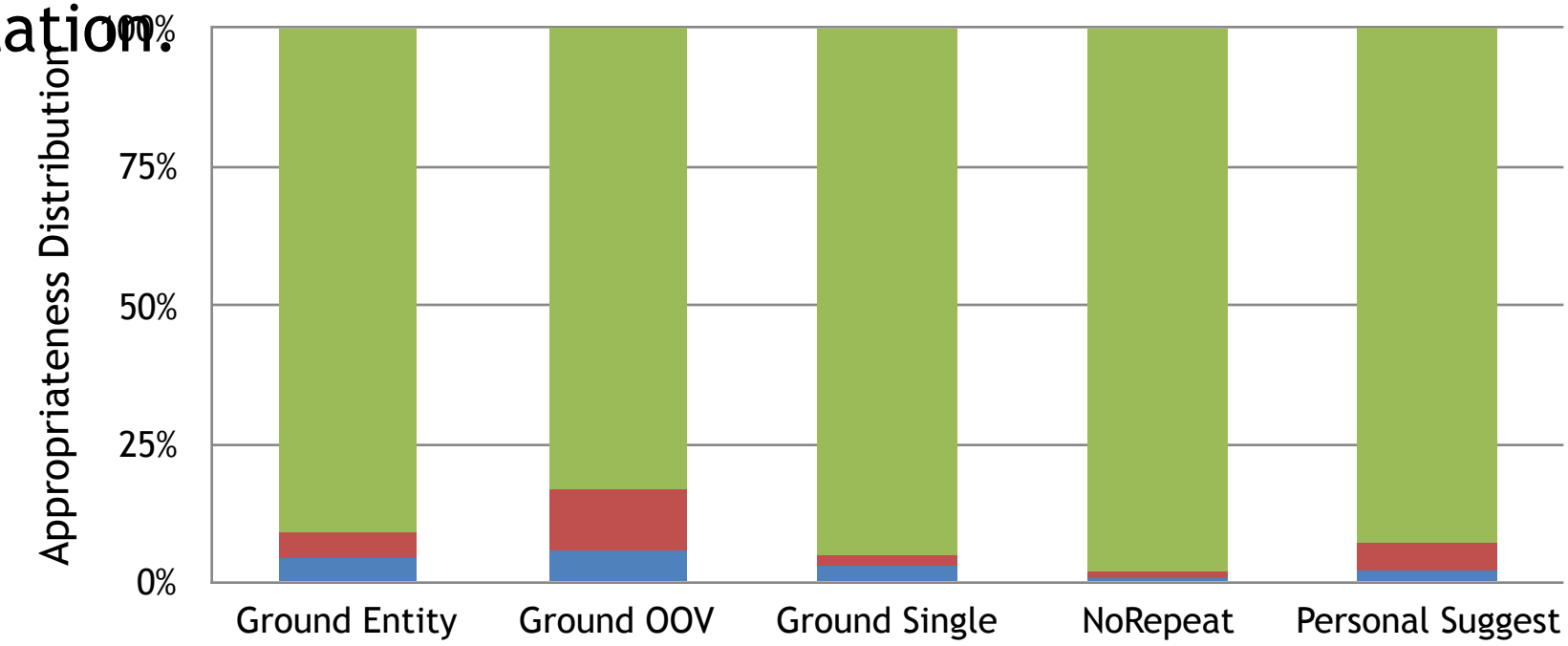
*TickTock: You already asked that.*

Procedure: Repeat utterance detection -> statistical template

# Results: Knowledge Base Strategies

Yu et al., SIGDIAL, 2016b

Appropriateness: The local coherence of system utterance
Experiment Setting: 50 typing conversations with self-reported annotation.



# Outline

Situation Intelligence Framework

Engagement Coordination


Situation Awareness

Conversation Strategy

**Statistical Policy**

Other Applications: Movie Promotion and Interview Training

Attention Coordination

Direction Giving

Future Work

# A Strategy Failure

*TickTock:* Hello, I really like politics. Let's talk about politics.

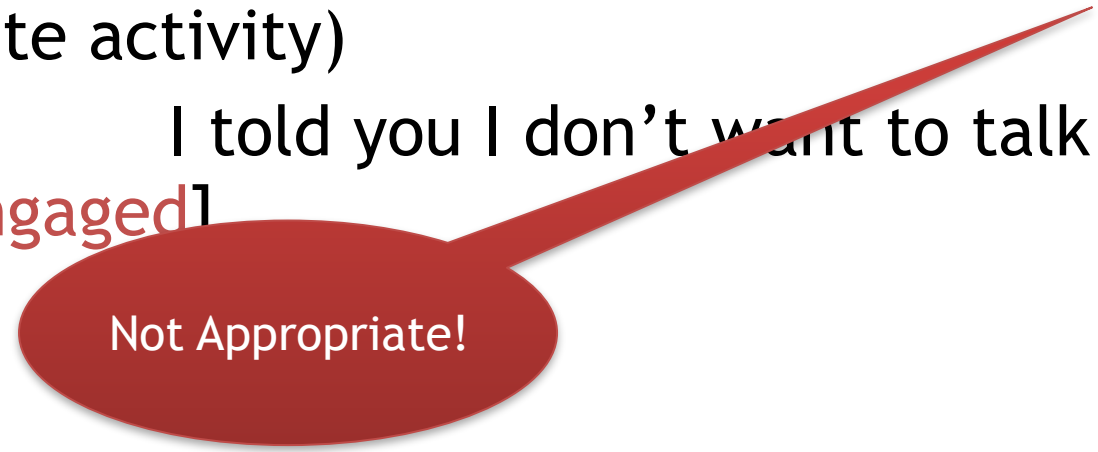
*User:* No, I don't like politics.

*TickTock:* Why is that?

*User:* I just don't like politics. [Disengaged]

*TickTock:* How about we watch a debate together some day? (initiate activity)

*User:* I told you I don't want to talk about politics! [Disengaged]



# Contextual User Sentiment

*TickTock:* Hello, I really like politics. Let's talk about politics. neutral

*User:* No, I don't like politics. negative

*TickTock:* Why is that? neutral

*User:* I just don't like politics. [Disengaged] negative

*TickTock:* OK, how about we talk about movies? (switch topics)

*User:* Sure. I do watch movies a lot. [Engaged] neutral


Take-home message: Consider conversation history in action planning!

# Statistical Policy

Goal: Long term effectiveness and naturalness considering **situation (history).**

Method: Reinf$_{Q: S \times A \rightarrow R}$rning$^{(S, A, R, \gamma, \alpha)}$

Q Learning :

State Variable (S):

- User engagement confidence
- System-appropriateness confidence
- All previous utterance-sentiment confidence
- Time of each strategy executed
- Turn position
- Most recently used strategy

Actions (A) : Conversation strategies and generated utterances

# Statistical Policy

Reward function(R): Weighted combination of accumulated appropriateness, conversation depth, information gain and overall user engagement

Pretrain-predictors

Appropriateness : Current response's coherence with the user utterance.

Conversation depth: Maximum number of consecutive utterances on the same topic.

Information gain: Number of unique tokens

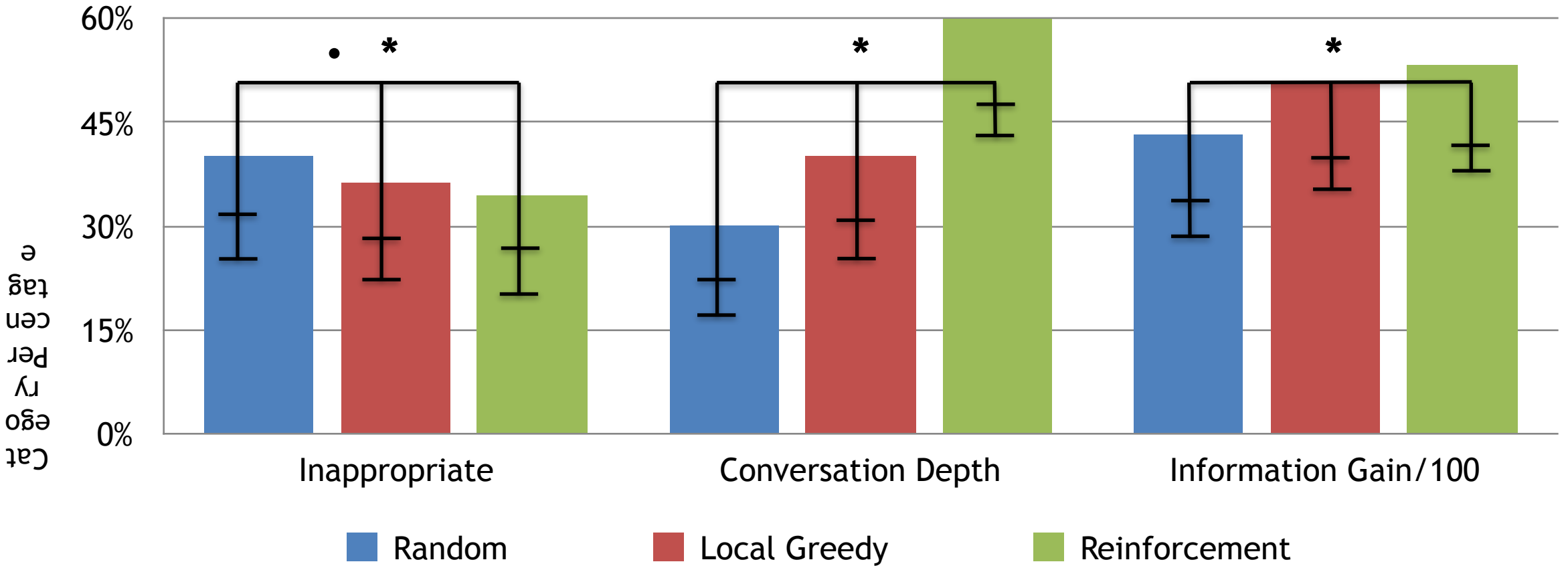
Overall user engagement: Overall assessment of users' engagement of the entire interaction.

Simulator: A.L.I.C.E. chatbot.

# Results: Policy

Yu et al., SIGDIAL, 2016b

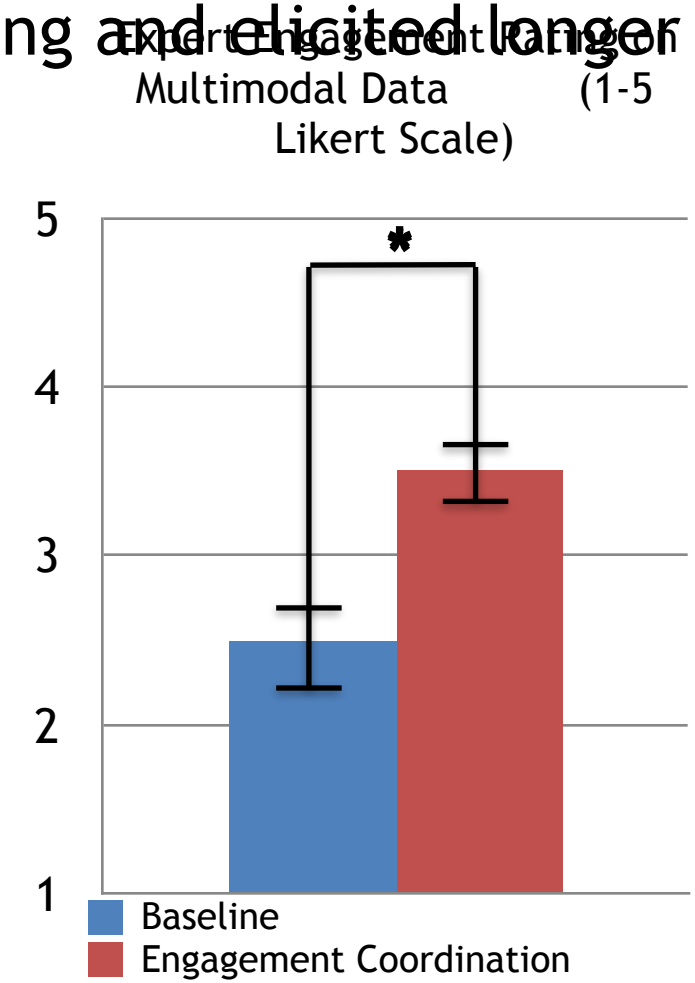
Two Baselines: Random selection and local greedy policy
Experiment Setting: 20 multimodal conversations per policy

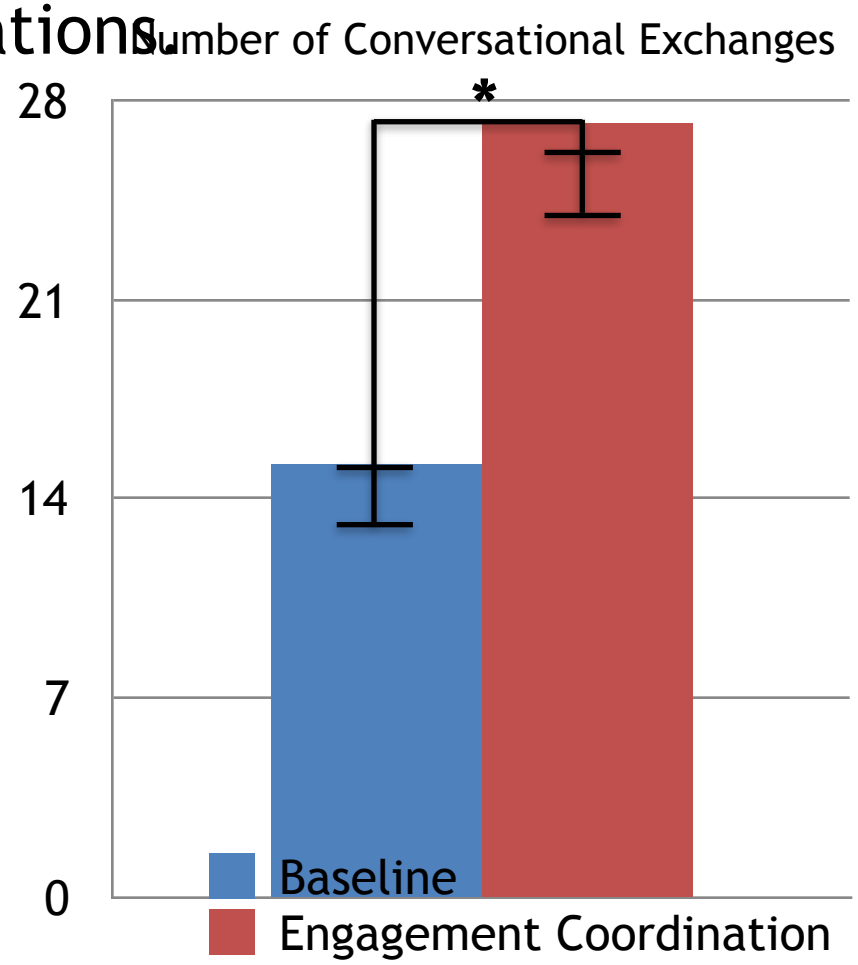


# Overall Results

Yu et al., Thesis, 2016

The system with engagement coordination was rated more engaging and elicited longer conversations



# Take-Home Message

**Active participation strategies** improve user engagement
(Deal with "*Bad at being interesting*")

**Grounding strategies** help open-domain language understanding
(Deal with "*Bad at understanding me*")

**Personalized strategies** help to adapt to different users
(Deal with "*Bad at remembering me*")

….

**Statistical policy** enables long-term optimal outcomes.
*(Deal with "Bad at being coherent and providing variety")*

# Outline

Situation Intelligence Framework

- Engagement Coordination
  - Situation Awareness
  - Conversation Strategy
  - Statistical Policy
  - **Other Applications: Movie Promotion and Interview Training**
- Attention Coordination
  - Direction Giving

Future Work

# Domain Generalization

Engagement predictor transfer:

Cold Start: Model from non-goal directed system

Later: Retrain model with adapted features

Conversation Strategy transfer:

Grounding strategies

Personalized strategies

# Entertainment Application: Targeted Movie Promotion

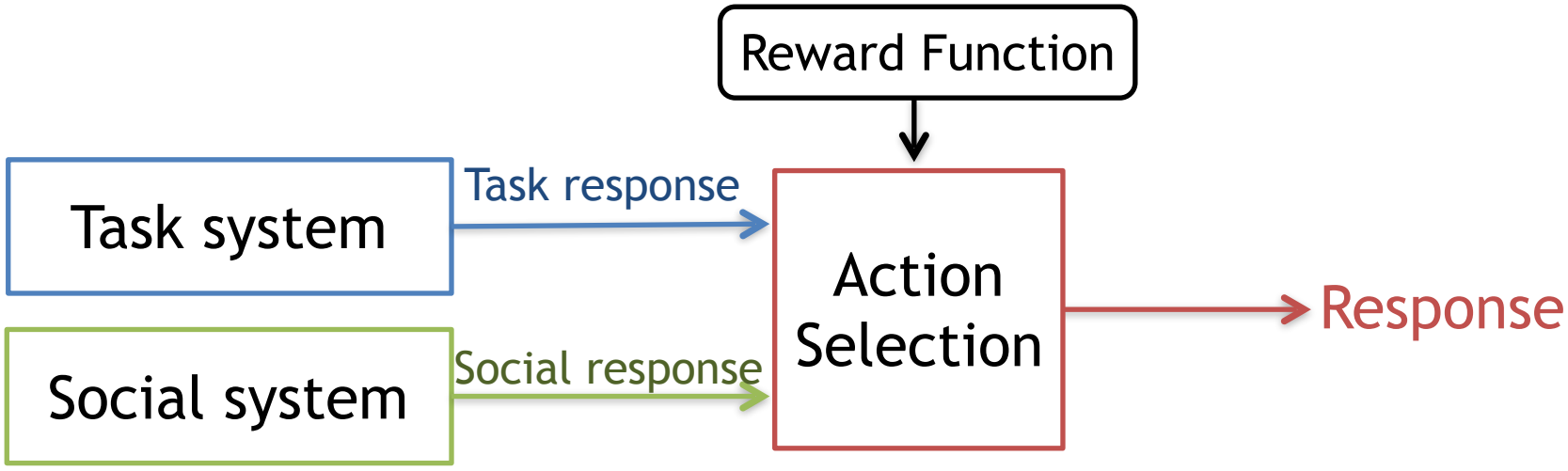
Yu et al., IJCAI, 2017

Goal: Elicit users' opinion about movies and then promote a movie based on their preferences.

Method: Interleave social content with task contents.

Task contents: A set of slots, such as: preferred movie type.

Reinforcement learning policy: Transition between social conversation and task conversation.



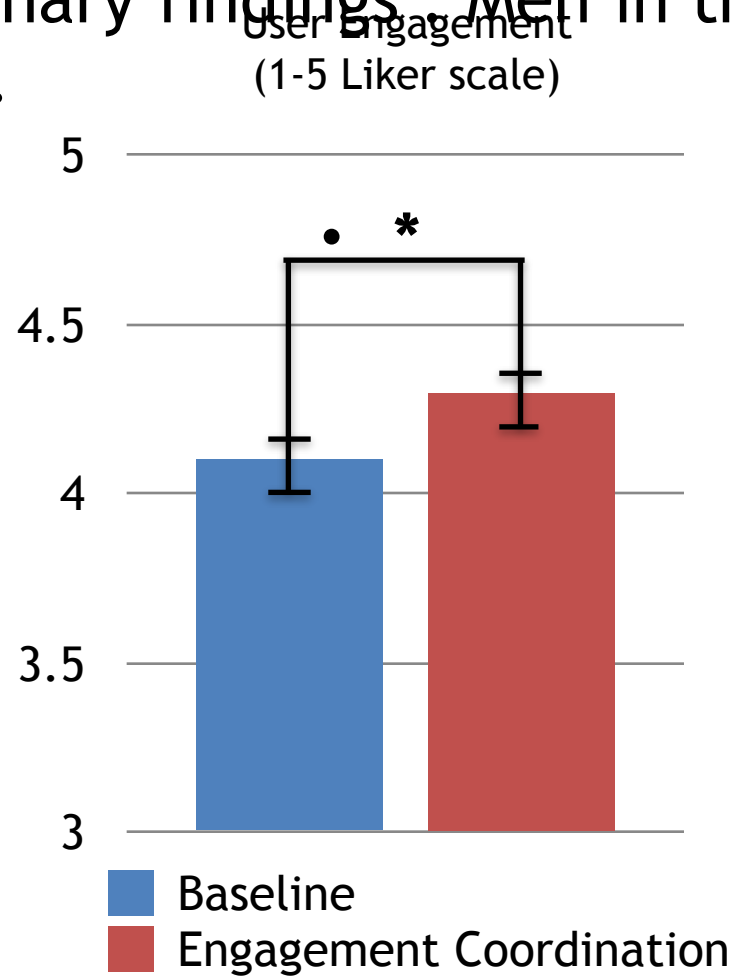
# An Example User Conversation

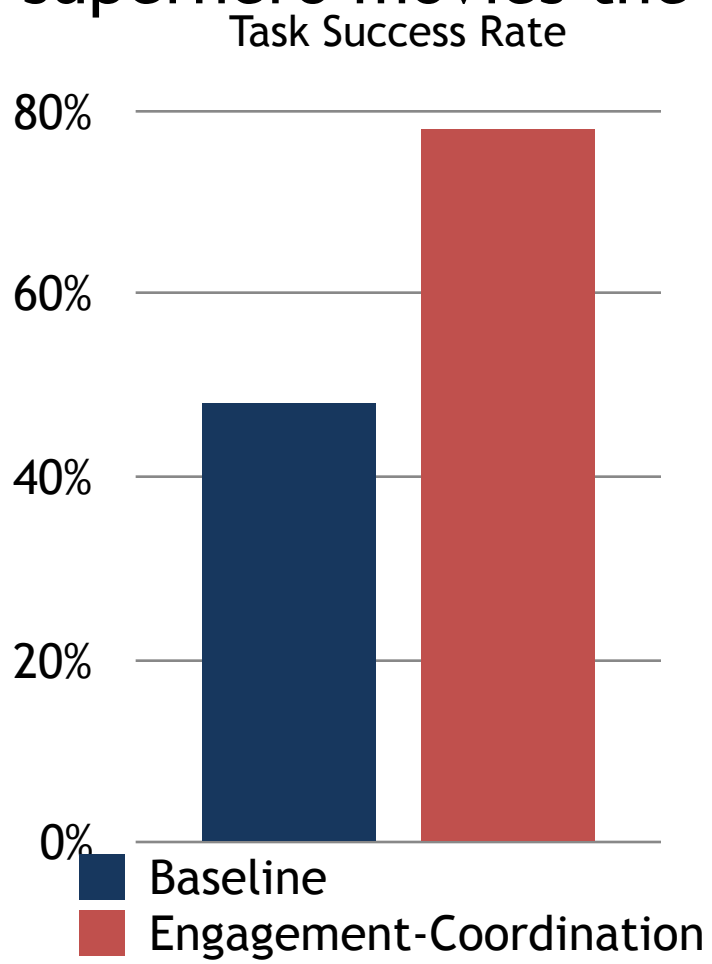
Yu et al., IJCAI, 2016

*User:* I like watching movies too.
*TickTock- task:* Do you like superhero movies or Disney movies?
*User:* I like superhero movies.
*TickTock- task:* My favorite superhero is Captain America.
*User:* I like spider-man.
*TickTock- social:* Do you watch them with your kids.
*User:* I don't have any children.
*TickTock- social:* What I meant to say was, what is that you hate?
*User:* I hated the last Fantastic Four movie.
*TickTock- social:* Are you talking about Fantastic Four, the 2005 film. (GroundEntity)
*User:* Yes. I am.
*TickTock- task:* I really like the first Avenger movie, have you seen it before.

# Overall Results

Experiment setting: 50 typing conversations for each system.
Preliminary findings : Men in their 20s like superhero movies the most in the US.



# Education Application: Adaptive Interview Training

Yu et al., IWSDS 2016, collaboration with ETS

Task: System interviews users for a job

Goal: Improve second-language learners' language ability

Experiment: 50 conversations per system(Baseline vs. Engagement Coordination)

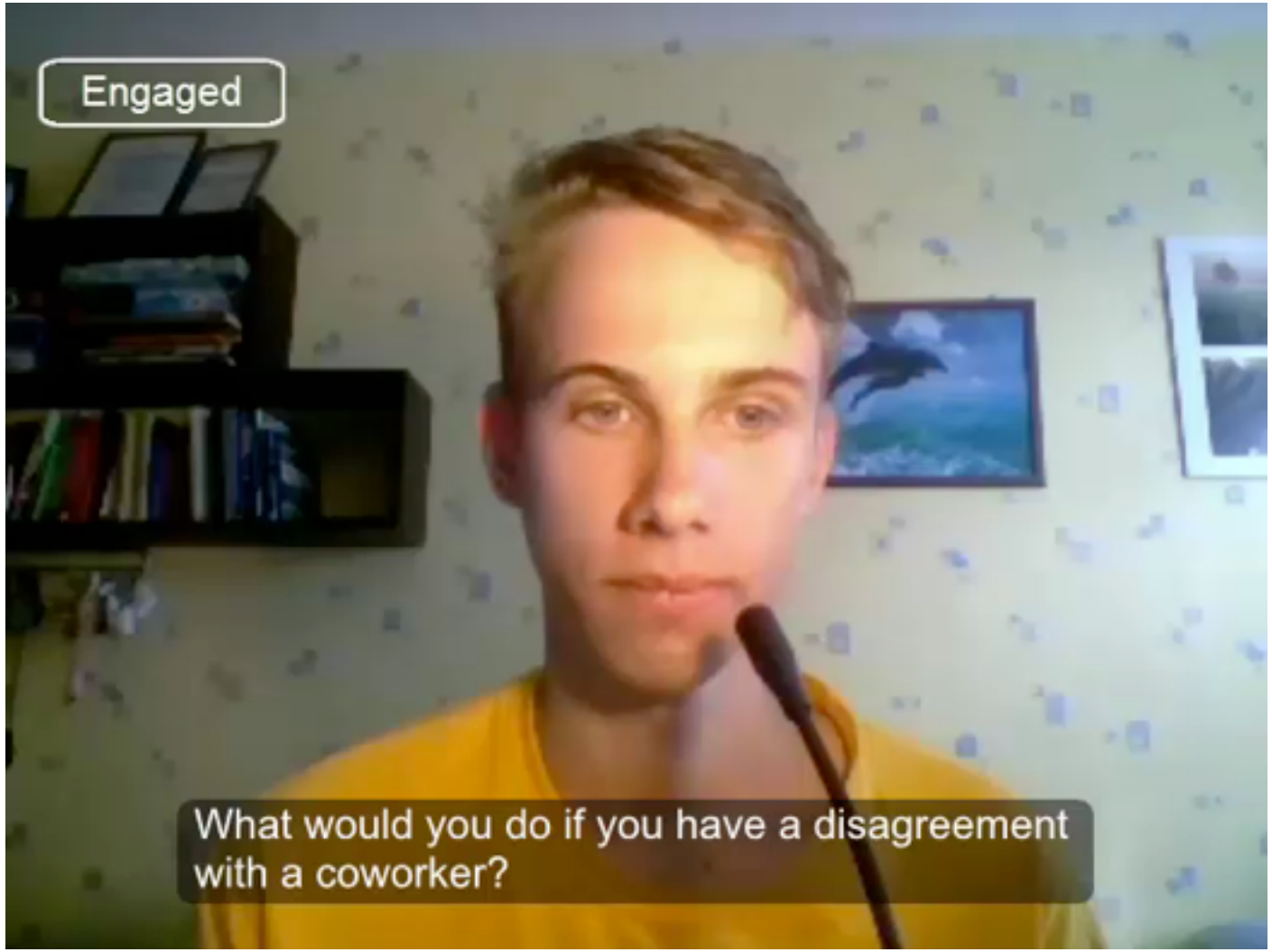
Result: The system with engagement coordination was rated as more engaging and elicited more user information

Impact: Deployment in China, Japan and Brazil Classroom

Integration with language learning application, ELSA (on going)

# An Example User Interaction

Collaboration with ETS




System advantages:  access via web-browser

# Outline

Situation Intelligence Framework

Engagement Coordination

Situation Awareness

Conversation Strategy

Statistical Policy

Other Applications: Movie Promotion and Interview Training

**Attention Coordination**

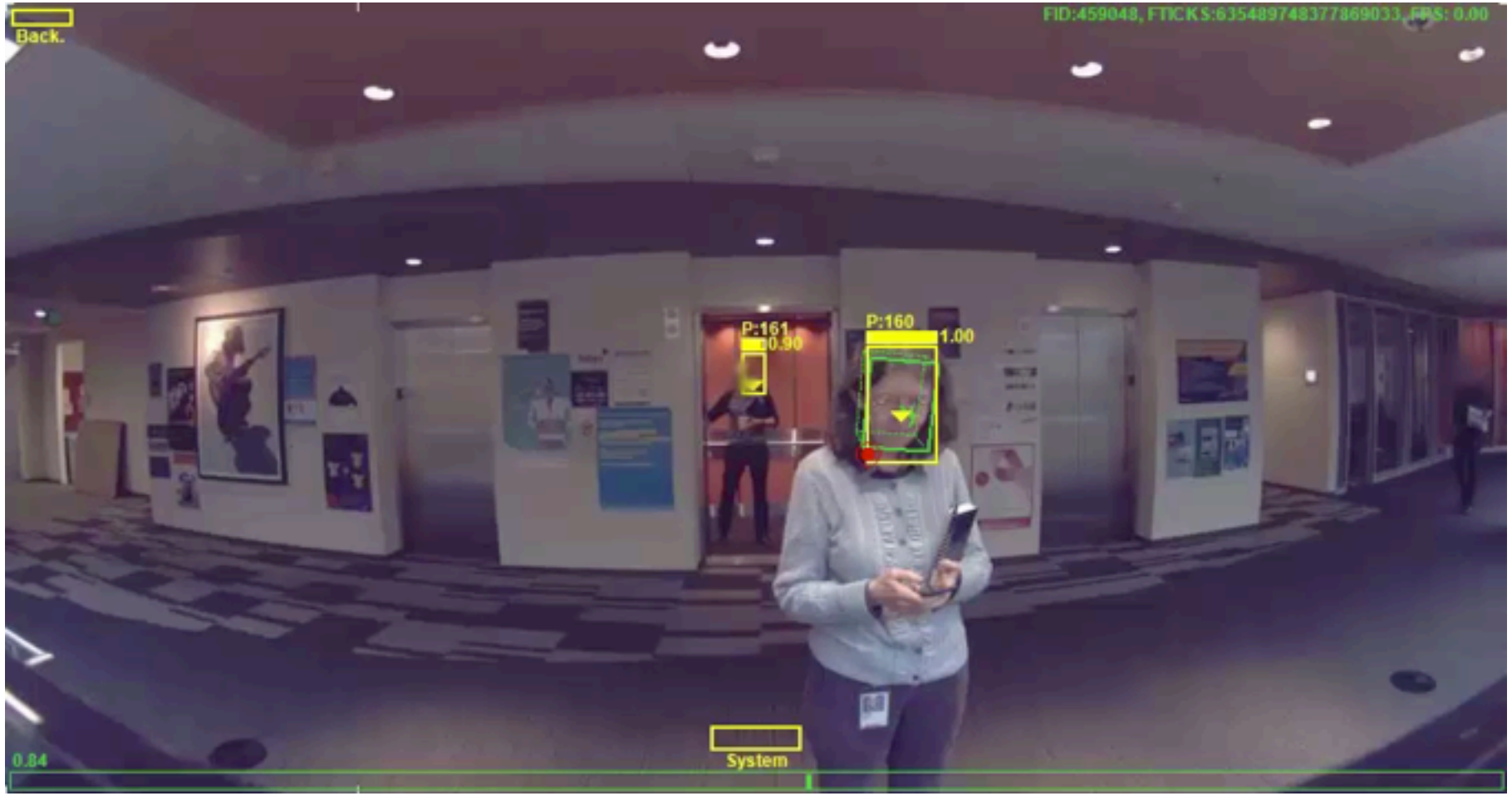
**Direction Giving**

Future Work

# Situation Awareness: Attention

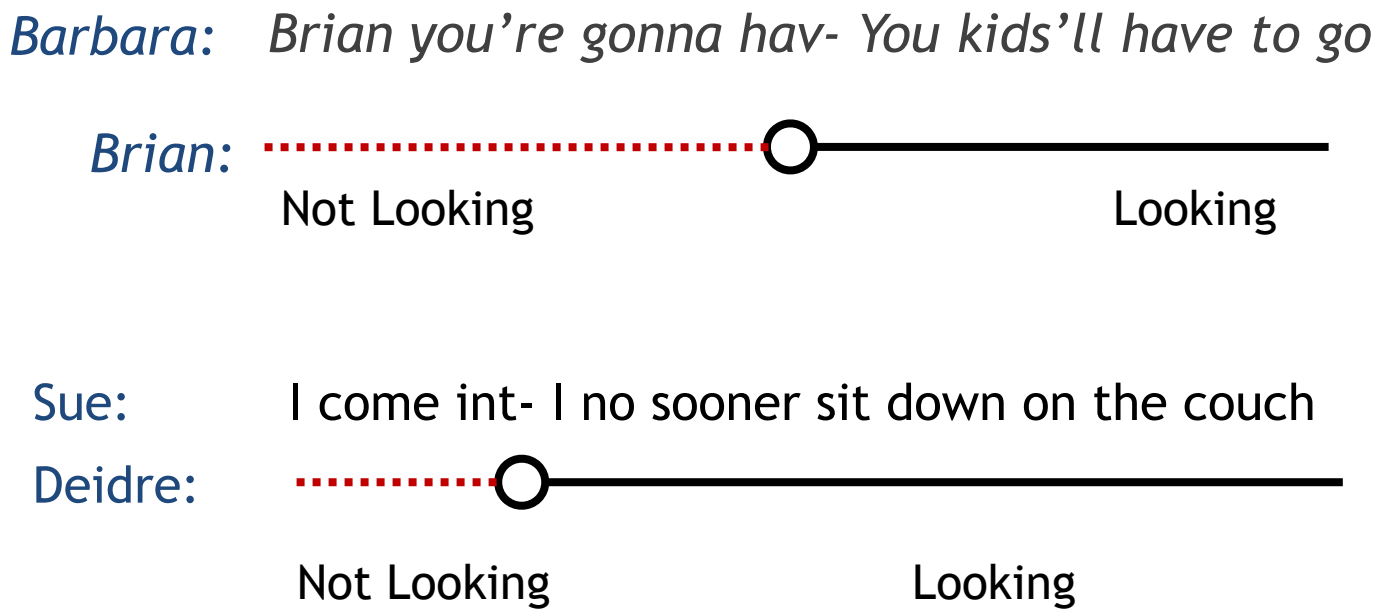
**Attention:** The visual focus of the users.

# A Problematic Real Interaction



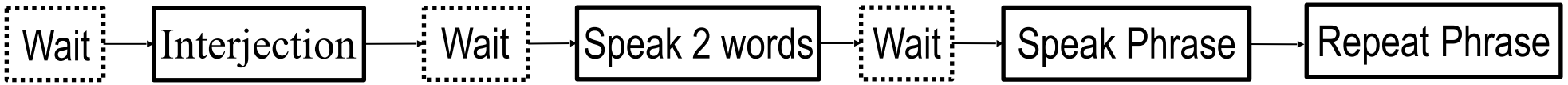
# Disfluencies as Conversation Strategies

Human-human conversations: Speaker's speech coordinates with listener's gaze, behavior level (*Goodwin 1981*)



# Attention Coordination Policy

Yu et al., SIGDIAL 2015, collaboration with MSR




… … … Excuse me! … … … To get … … … To get to 3800… To []

Trigger disfluency strategies only when system action demands user attention.

# Examples

Yu et al., SIGDIAL 2015, collaboration with MSR

# Overall Summary

**Multimodal** information is useful for interaction planning.

Using **computational methods** (e.g. NLP and Knowledge base) to encode conversational theories facilitates interaction.

**Reinforcement learning** leads to long-term optimum interaction.

Situation Intelligence framework is applicable in **various domains**, and **augmentable** on different existing systems.

# On Going Work: Efficient Multimodal Models

Challenge:

- High dimension and noise in raw multimodal features
- Inefficiency in training and testing

Method:

- Introduce sparsity (Group sparsity regularization for multimodal feature representation and pruning for neural network architecture)

Expected Results:

- Computational models, real systems and multimodal dataset.

# Movie Recommendation System

- https://github.com/kevinjesse/chatbox

# On Going Work: Learning Structures

Challenge:

Domain expert-designed dialog actions and flows for each domain.

Method:

Automatic extraction of domain knowledge and conversation structures from existing unlabeled conversations

Automatic encode them back to conversations

Expected Results:

Seq2Seq, hierarchical, reinforcement learning models with structures

Conversational systems driven by these models

# Future Work: Awareness

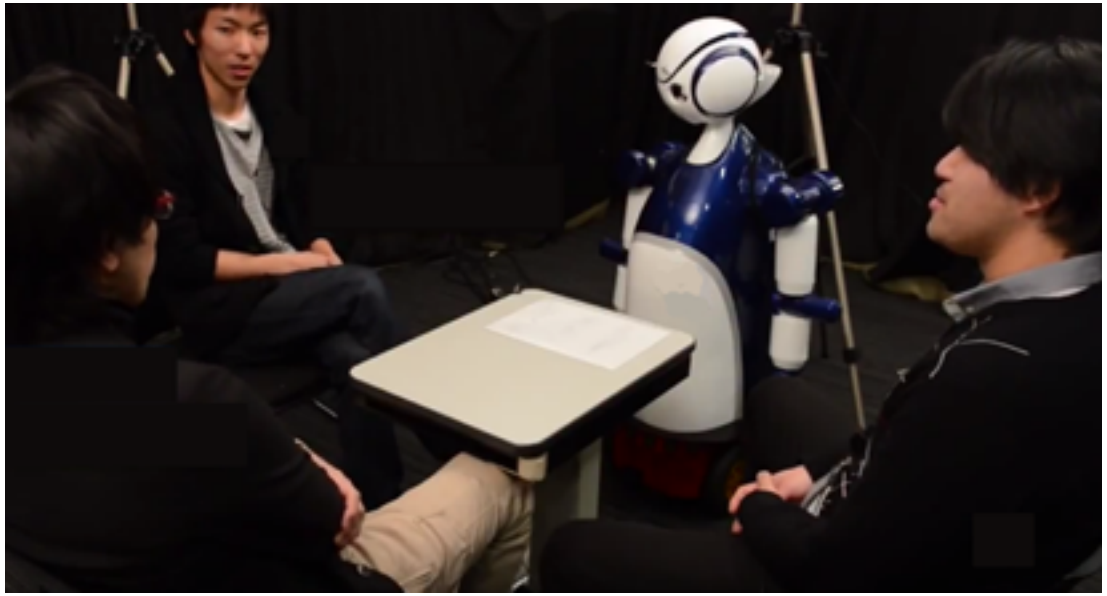
Individual-Aware

Interpersonal-Aware

**Social Cultural-Aware**

# Interpersonal Relationship

Preliminary work: Friendship prediction,
Yu et al., SIGDIAL 2013



Group collaboration,
such as tutoring and meetings

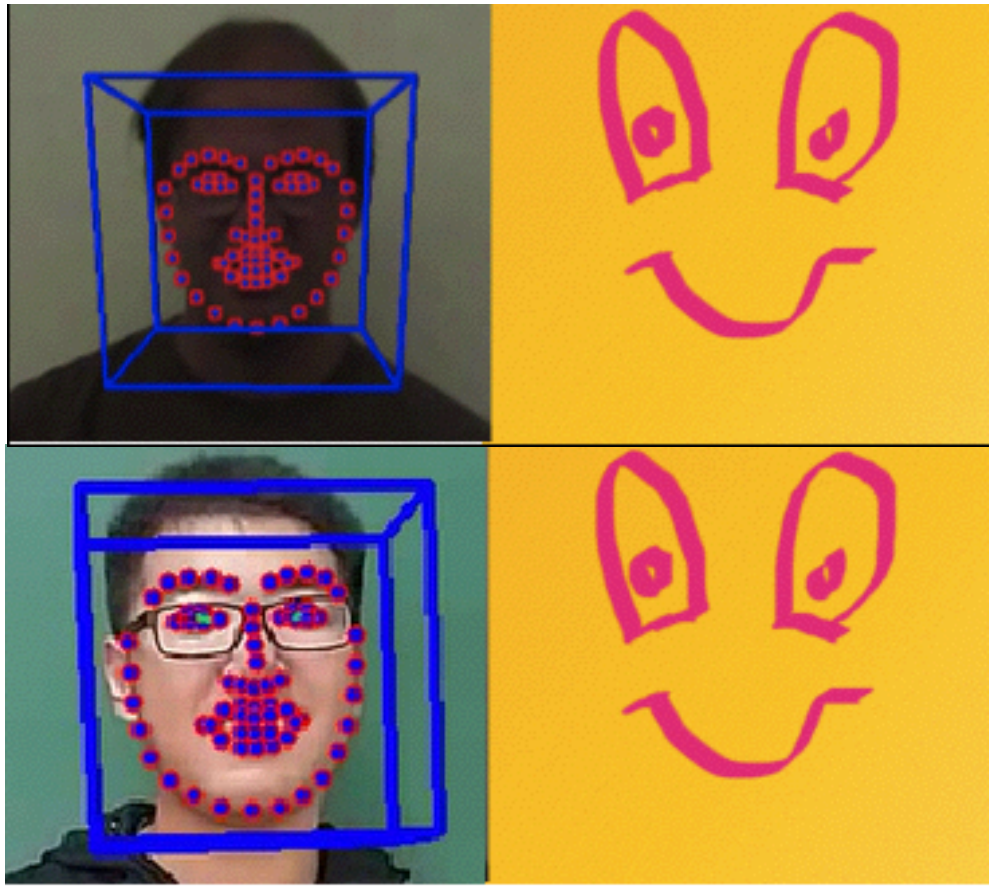
# Future Work: Awareness

Individual-Aware

Interpersonal-Aware

**Social Cultural-Aware**

# Social Cultural Awareness



Preliminary work: Engagement Model with Culture Adaptation (Chinese VS Americans), Yu et al., IVA 2016a

Race, age, gender, education level, etc.

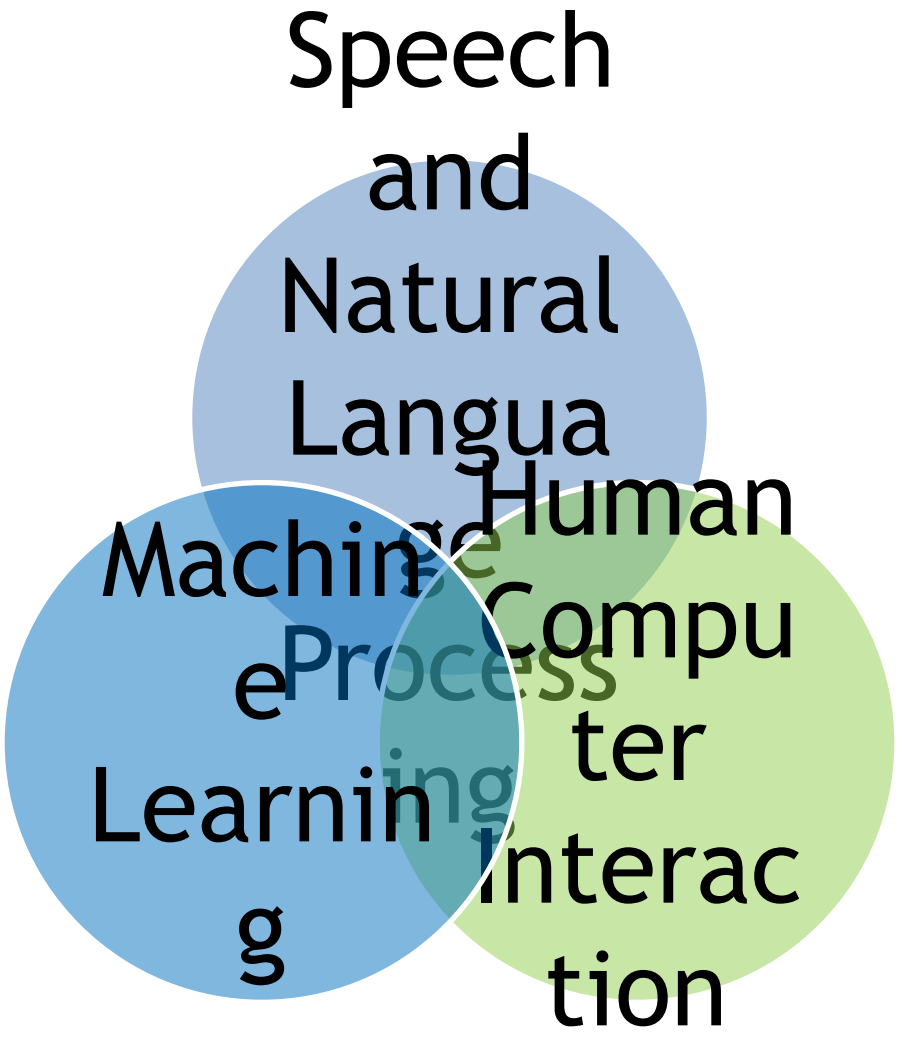
# Acknowledgement

Alan Black, Alex Rudnicy, Louis-Phillip Morency, David Suendermann-OeFT, Dan Bohus, Eric Horvitz, Justine Cassell, Alexandros Papangelis, Vikram Ramanarayanan, Shrimai Prabhumoye, Xinrui He and Leah Nicolich-Henkin

# Questions?

zhouyu@cs.cmu.edu

Thanks For Coming!



# References


- Timothy W. Bickmore, Laura Pfeifer Vardoulakis, Daniel Schulman: Tinker: a relational agent museum guide. Autonomous Agents and Multi-Agent Systems 27(2): 254-276 2013
- Zhou Yu, David Gerritsen, Amy Ogan, Alan Black and Justine Cassell, Automatic Prediction of Friendship via Multi-model Dyadic Features, SIGDIAL 2013
- Zhou Yu, Stefan Scherer, David Devault, Jonathan Gratch, Giota Stratou, Louis-Philippe Morency and Justine Cassell, Multimodal Prediction of Psychological Disorder: Learning Verbal and Nonverbal Commonality in Adjacency Pairs, SEMDIAL 2013
- Zhou Yu, Dan Bohus and Eric Horvitz, Incremental Coordination: Attention-Centric Speech Production in a Physically Situated Conversational Agent, SIGDIAL 2015
- Zhou Yu, Vikram Ramanarayanan, David Suendermann-Oeft, Xinhao Wang, Klaus Zechner, Lei Chen, Jidong Tao and Yao Qian, Using Bidirectional LSTM Recurrent Neural Networks to Learn High-Level Abstractions of Sequential Features for Automated Scoring of Non-Native Spontaneous Speech, to appear ASRU 2015.

# References


- Teruhisa Misu, Antoine Raux, Rakesh Gupta, and Ian Lane. 2014. Situated language understanding at 25 miles per hour. In. Proc. of the SIGDIAL - Annual Meeting on Discourse and Dialogue.
- Wafa Benkaouar and Vaufreydaz Dominique. Multi-sensors engagement detection with a robot companion in a home environment. In Workshop on Assistance and Service robotics in a human environment at IEEE International Conference on Intelligent Robots and Systems (IROS2012), pp. 45-52. 2012.
- Tomislav Pejsa, Sean Andrist, Michael Gleicher, Bilge Mutlu. Gaze and Attention Management for Embodied Conversational Agents. TiiS 5(1): 3:1-3:34,2015
- Candace L Sidner, Christopher Lee, Cory D Kidd, Neal Lesh, and Charles Rich. Explorations in engagement for humans and robots. Artificial Intelligence, 166(1):140–164, 2005
- Christopher Peters, Catherine Pelachaud, Elisabetta Bevacqua, Maurizio Mancini, and Isabella Poggi. A model of attention and interest using gaze behavior. In Intelligent Virtual Agents, 5th International Working Conference, 2005

# References


- Alex Papangelis, Ran Zhao, Justine Cassell Towards a computational architecture of dyadic rapport management for virtual agents, 2014
- Daniel Gatica-Perez, Iain A. McCowan, Dong Zhang, and Samy Bengio. Detecting group interest-level in meetings. In IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), no. EPFL-CONF-83257. 2005.
- Aasish Pappu, Ming Sun, Sridharan Seshadri, and Alex Rudnicky. Situated multiparty interaction between humans and agents. In Human-Computer Interaction. Interaction Modalities and Techniques, pages 107–116. Springer.2013
- Paul Boersma and David Weenick. Praat: doing phonetics by computer [computer program]. Version 5.3. 03, retrieved 21 November 2011 from http://www. praat. org. 2006
- Florian Eyben, Martin Wollmer and Bjorn Schuller. Opensmile: The munich versatile and fast open-source audio feature extractor. In Proceedings of the International Conference on Multimedia, MM ’10, pages 1459–1462, New York, NY, USA. ACM, 2010
- Tadas Baltrusaitis, Peter Robinson and Louis-Philippe Morency. 3D constrained local model for rigid and non-rigid facial tracking. In CVPR, 2012 IEEE Conference, 2010

# References


- Zhou Yu, Xinrui He, Alan W Black and Alexander Rudnicky, User Engagement Modeling in Virtual Agents Under Different Cultural Contexts, to appear IVA 2016.
- Zhou Yu, Ziyu Xu, Alan W Black and Alexander Rudnicky, Strategy and Policy Learning for Non-Task-Oriented Conversational Systems, to appear SIGDIAL 2016.
- Zhou Yu, Leah Nicolich-Henkin, Alan W Black and Alexander Rudnicky, A Wizard-of-Oz Study on A Non-Task-Oriented Dialog Systems that Reacts to User Engagement, to appear SIGDIAL 2016.
- Zhou Yu, Ziyu Xu, Alan W Black and Alexander Rudnicky, Chatbot evaluation and database expansion via crowdsourcing, In Proceedings of the RE-WOCHAT workshop of LREC,2016.
- Ryan Kiros, Yukun Zhu, Ruslan Salakhutdinov, Richard S. Zemel, Antonio Torralba, Raquel Urtasun, Sanja Fidler, Skip-Thought Vectors, arXiv, 2015
- Zhou Yu, Alexandros Papangelis, Alexander Rudnicky, TickTock: Engagement Awareness in a non-Goal-Oriented Multimodal Dialogue System, AAAI Spring Symposium on Turn-taking and Coordination in Human-Machine Interaction 2015
- Weizenbaum, Joseph (January 1966). "ELIZA – A Computer Program For the Study of Natural Language Communication Between Man and Machine" (PDF). *Communications of the ACM*. **9** (1). Retrieved September 16, 2016 – via Stanford University.
- Rafael E Banchs and Haizhou Li. 2012. Iris: a chat-oriented dialogue system based on the vector space model.In Proceedings of the ACL 2012 System Demonstrations, pages 37–42. Association for Computational Linguistics