Handling Uncertain Input in multi-user human-robot interaction

Presenter: Maham Tanveer
9th November, 2015

Fig. 1 [1]

Structure of Presentation

- Focus
- Background: Handling Uncertainty in HRI
- “Handling uncertain input in multi-user human-robot interaction”, JAMES Project
- Architecture
- Experimental Design and Results
- “Experiences with Mobile Robotic Guide for the Elderly”
- Conclusion
- Future Work

Focus of Presentation

- How to handle uncertainty in Human Robot Interaction by using POMDP in two scenarios, bartending robot and a robot assisting the elderly.
- How can human robot interactions be improved by catering uncertainty at all levels of robot control.

1

Background: Handling uncertainty in HRI

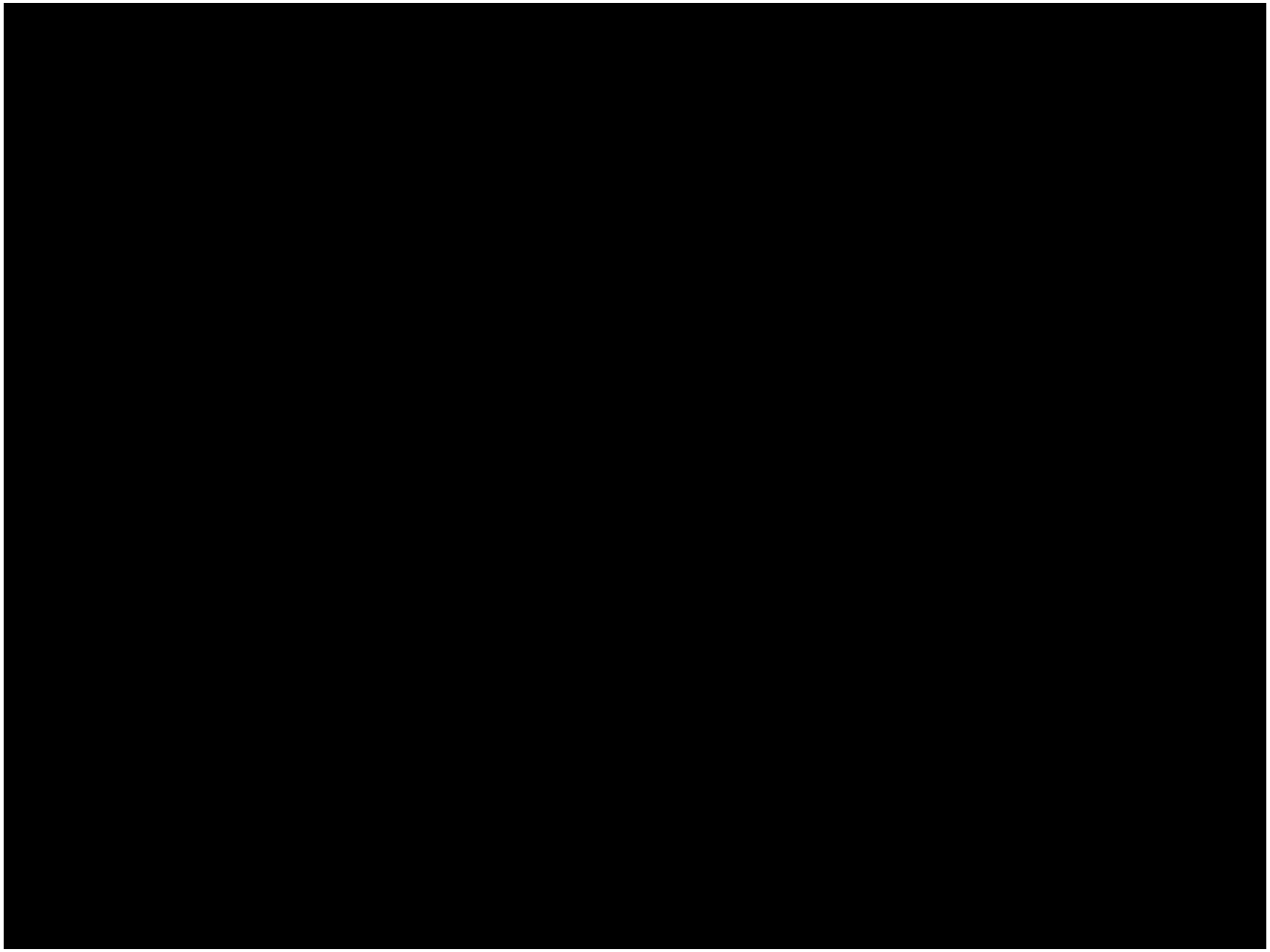
- What is Uncertainty in Human Robot Interaction?
- At which levels of robot control should uncertainty be tackled?

- Approaches to handle uncertainty:-
  - ➢ Kalman Filter Strategy:

    educated guess based on previous best estimate and correction of known external influences, stochastic state estimation from noisy sensor measurements, running estimate of robot's spatial uncertainty as a normal distribution
  - ➢ Partially observable Markov decision process (POMDP)

    Markov's Decision process: solving complex partially observable problems as a model of state synchronously interacting with the world, where uncertainty might be in actions but never in current state. (S,A, T, R)
    POMDP: MDP unable to compute its current state (S,A,T,R, Ω (finite set of obs.), O (SxA, prob. Dist. Over possible obs.)

Speech Recognition & Language Processing



Animation coutesy of : http://www.match-project.org.uk/resources/tutorial/Speech_Language/Speech_Recognition/Rec_4.html

"Handling uncertain input in multi-user human-robot interaction", JAMES Project

- Title: "Handling uncertain input in multi-user human-robot interaction"

Simon Keizer, Mary Ellen Foster, Andre Gaschler, Manuel Giuliani, Amy Isard, and Oliver Lemon, The 23rd IEEE International Symposium on Robot and Human Interactive Communication, August 25-29, 2014. Edinburgh, Scotland

- Topic:

User Evaluation of Bartender robot with two approaches:-

  - Handling uncertainty using threshold levels
  - Handling uncertainty using multiple input hypothesis and confidence levels.

1

Meet Bartender Robot JAMES!

- **JAMES:** Joint Action for Multimodal Embodied Social Systems (james-project.eu)
- 3.5 years project (2011-2014)
- Focus on socially appropriate, multi-party, multimodal interactions in a Robot bartending scenario.
- Interaction incorporate both ***task-based aspects*** & ***social aspects***
- Social modeling, learning, implementation & evaluation

Fig. 2 [1]

Architecture

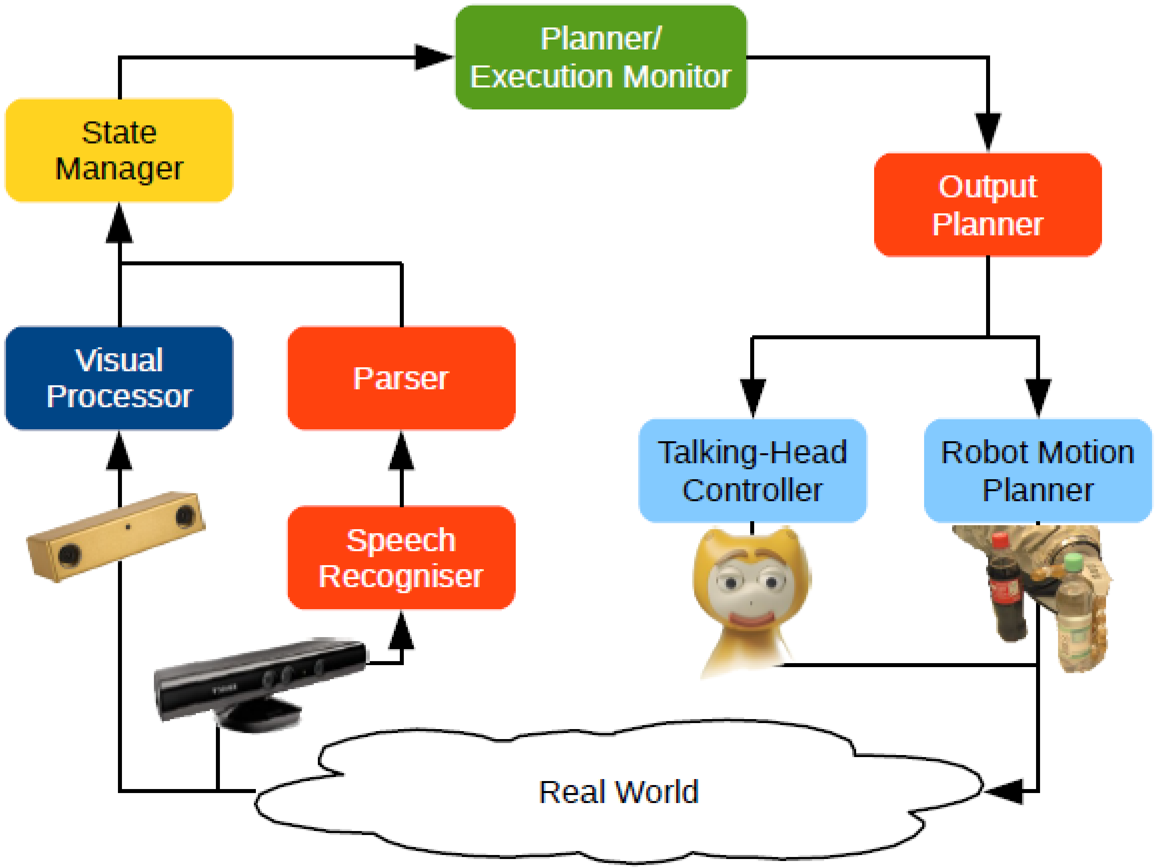


Fig. 3 [2]

Component	Hardware Used	Functionality
Visual processing component	• 2 Calibrated Stereo Cameras • Kinect Depth Sensor	• Location & Body orientation of multiple customers • Confidence values
Speech processing component	• Kinect ASR System • Open CCG	• Speech Recognition • Semantic Parsing
State Manager		• Fuses audiovisual input stream • Model of social state
Social Skills Executor		Selects response actions
Output Planner		• Performs actions • Talking Head Controller: looking at customer, nodding & speaking • Robot Motion Planner: Serving drinks, picking drinks & idle states

- Speech Application Processing Interface has two types: Text to Speech and Speech Recognizers.

Speech Recogniser

* N-best list of hypothesis
* Estimate of source sound angle
* Confidence Scores

(Range: 0-1, float)

* Low confidence signal is discarded
* Microsoft Speech API interfaces (Audio Interface, Grammar Compiler Interface & Speech Recognition Interface)

Semantic Parsing

* User defined grammar
* Dynamically loaded & unloaded for parsing
* Parse each hypothesis with Grammar defined
* Remove duplicate parses
* Convert parse > Communicative Act

Fig. 4 [3]

State Manager: Monitoring with Uncertain Input

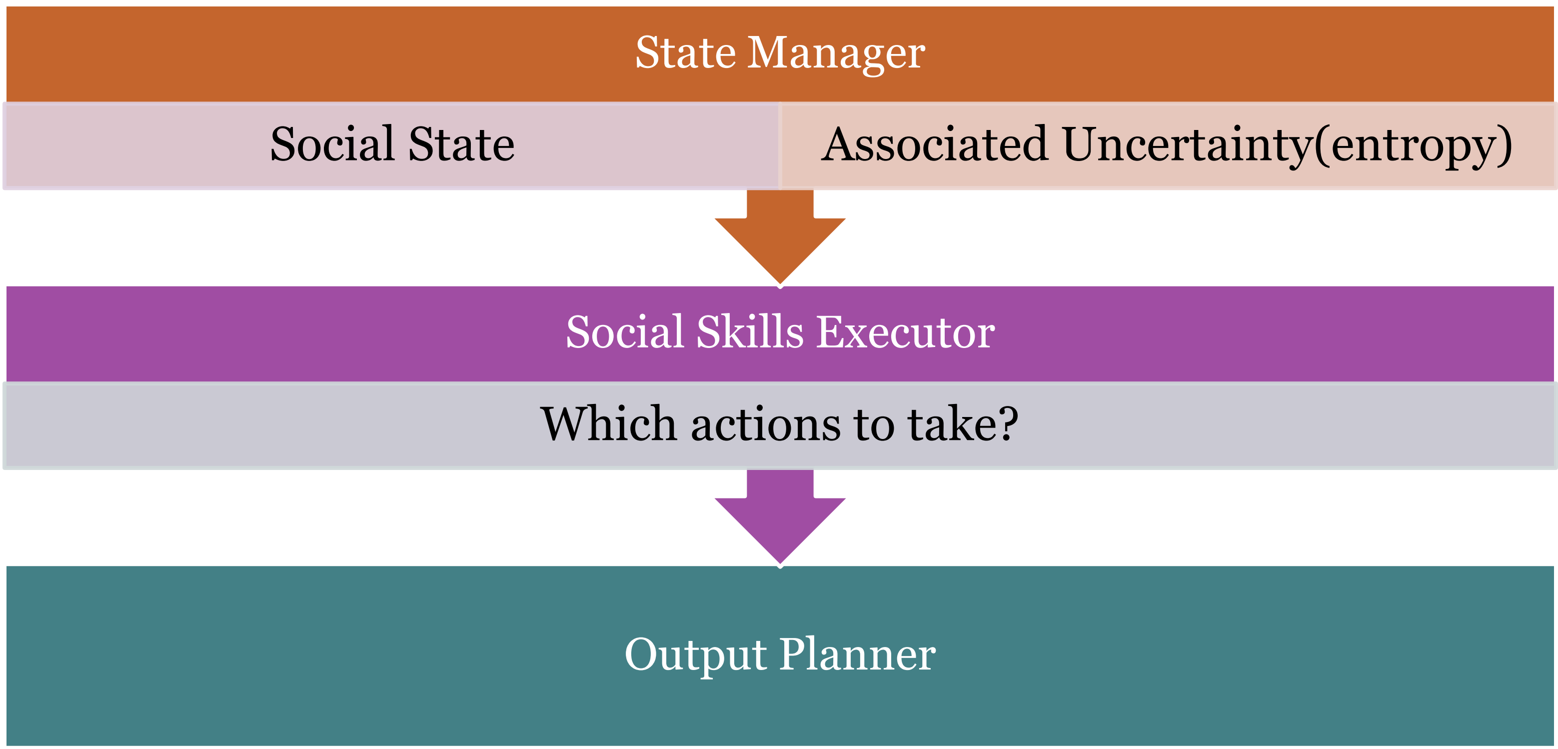
- Input is continuous stream of information from audio and visual components. Performs ***Fusion of audio visual input*** to assign a speech hypothesis and to estimate attention-seeking state of specific customer

- Information from audio visual components to associate ***Communicative Acts*** with customer

- Uses ***generic belief tracking*** procedure which maintains beliefs over user goals based on small number of domain independent rules using basic probabilistic operations.

- Maintains a ***dynamically updated list of possible drink orders*** made by each customer and associated ***confidence value*** for each order (social state).

Social Skills Executor: Action selection under uncertainty



Social Skills Executor (SSE)

- Action Selection Strategy
- Clarifications to exploit uncertainty

Stage 1 (Which customer to focus on its next action)

- Engage with customer seeking attention
- Ask them to wait
- Continue on-going interaction

Stage 2 (If interaction to be continued.)

- Which Communicative Action to take?
- Whether drink will be served to customer or not

1

Algorithm 1 Selecting clarification actions (*conf* refers to the confidence score of the top drink order hypothesis, *entr* refers to the entropy of the drink order belief distribution, and the thresholds used in the experiment are listed in Table I.)

if (*conf* ≥ CONF_THR1) **or**
(*conf* ≥ CONF_THR2 **and** *entr* < ENTR_THR) **then**
select action based on top hypothesis;
(e.g., "Okay, a coke")
else if there is only one drink order hypothesis **then**
confirm the drink order with the user;
(e.g., "Did you say 'coke'?")
else
let user choose between top 2 hypotheses;
(e.g., "Did you say 'green' or 'blue' lemonade?")
end if

TABLE I

THRESHOLDS USED IN SELECTING CLARIFICATIONS

Description	**Threshold**	**Value**
Upper confidence threshold	CONF_THR1	0.65
Lower confidence threshold	CONF_THR2	0.40
Entropy threshold	ENTR_THR	0.25
parsing confidence threshold (baseline)	SCONF_THR	0.30
parsing confidence threshold (uncertainty)	SCONF_THR_UNC	0.10

Fig. 5 and Fig 6 [2]

1) *Customer (A1) enters the scene, seeking attention*
2) **System (looking at A1): "Hello"**
3) *A1 orders*
4) parser: *drink-order(green-lemonade) [0.02]*
5) state: speech input rejected
6) *Customer (A2) enters the scene, not seeking attention*
7) **System (to A1): "What can I get you?"**
8) *A1 answers*
9) parser: *drink-order(blue-lemonade) [0.05]*
10) state: speech input rejected
11) **System (to A1): "What can I get you?"**
12) *A1 answers*
13) parser: *drink-order(lemonade) [0.75]*
14) state: *drink-order(A1)=(blue-lemonade [0.37], green-lemonade [0.37])*
15) **System (to A1): "Did you say 'blue lemonade' or 'green lemonade'?"**
16) *A1 answers*
17) parser: *drink-order(lemonade) [0.53], drink-order(blue-lemonade) [0.53]*
18) state: *drink-order(A1)=(blue-lemonade [0.57], green-lemonade [0.43])*
19) **System (to A1): "Did you say 'blue lemonade' or 'green lemonade'?"**
20) *A1 answers*
21) parser: *drink-order(blue-lemonade) [0.69]*
22) state: *drink-order(A1)=(blue-lemonade [0.66], green-lemonade [0.34])*
23) **System (to A1): "Okay, a blue lemonade"**
24) *System serves blue lemonade to A1*

Fig. 7 [2] Interaction fragment where a user successfully orders blue lemonade

Uncertainty-Aware

Fig. 9 [1]

1) *Two customers enter the scene one of them (A1) approaches the bar*
2) **System (to A1): "Hello"**
3) *A1 returns greeting*
4) **System (to A1): "What can I get you?"**
5) *A1 tries to order a blue lemonade*
6) parser: *drink-order(coke) [0.332]*
7) state: *drink-order(A1)=(coke [1.0])*
8) **System (to A1): "Okay, a coke"**
9) *system serves A1 a coke*
10) **System (to A1): "Here you go"**
11) *A1 tries to order again but ASR fails*
12) **System (to A1): "Anything else?"**
13) *A1 tries to order again but speaker identification fails*
14) **System (to A1): "Will that be all?"**
15) *A1 responds negatively*
16) **System (to A1): "What can I get you?"**
17) *A1 once again orders a blue lemonade*
18) parser: *drink-order(blue-lemonade) [0.414]*
19) state: *drink-order(A1)=(blue-lemonade [1.0])*
20) **System: "Okay, a blue lemonade"**
21) *system serves A1 a blue lemonade*
22) *A1 thanks the system and leaves*

Fig. 8 [2] Interaction in which the system serves the wrong drink

Uncertainty-Unaware

User Evaluation

- Total participants: 24 (Male) (7 already took part in previous bartender robot evaluation), all native Germans
- Four drink ordering sessions
- Half of the sessions uncertainty-aware, other half uncertainty-unaware
- Half the times participant ordered for himself, in other half for his confederate
- Mean participant age: 27.5 (Range: 21-49)
- Mean of self-rating experience with robot (scale:1-7): 3.3
- Physical form of robot shown & not its interactive form before experiment start.
- All participants filled out computer based questionnaire after sessions.

Experiment Design : Independent Measures

- Variation in use of uncertainty
- Scenario where confederate orders for himself & then asks the participant to order on his behalf

Experiment Design: Dependent Measures

- Objective Measures
- Subjective Measures

Objective Measures

- The objective measures were based on the dimensions proposed by the PARADISE dialogue evaluation framework which provides predictive ***models for SLDS's*** as a function of ***task success*** and ***dialogue cost metrics*** measurable from system logs, without the need for extensive experiments with users to access user satisfaction.

- ***Task Success:*** No. of drinks served by system
- ***Dialogue quality:*** No. of user's attempted contributions below speech-recognition confidence threshold, no. of times the robot had to ask for order and no. of times clarification is asked in certainty aware systems
- ***Dialogue efficiency:*** time taken to serve the first drink in a trial, the time taken to serve all of the drinks, as well as the total duration of the trial as measured both in seconds and in system turns.

1

- **Objective Measures Results:**
  - Demographic features of participants did not affect the results
  - Only action-selection strategy affected the results
  - Mean result from each measure & significance level from paired Mann-Whitney Test

TABLE II

OBJECTIVE RESULTS

Measure	Baseline (sd)	Uncertainty (sd)	M-W
Drinks served	1.96 (0.14)	1.72 (0.39)	$p < 0.01$
Low ASR turns	3.2 (1.5)	2.0 (0.84)	$p < 0.001$
Order requests	5.7 (2.6)	5.5 (2.6)	n.s.
Choices	—	2.3 (2.3)	—
Confirmations	—	2.3 (2.0)	—
Time to first drink	49.6 (19.6)	71.3 (58.7)	$p < 0.05$
Time to last drink	94.2 (24.1)	107.7 (61.2)	n.s.
Duration	103.6 (25.3)	122.9 (61.2)	n.s.
System turns	14.1 (3.6)	17.6 (5.0)	$p < 0.05$

Fig. 10[2]

1

Baseline System	Uncertainty-aware System
SCONF_THR=0.30	SCONF_THR_UNC=0.10 (better process for dealing with low confidence utterances)
Served more drinks in a trial (out of max=2)	Served fewer drinks because of input processing issues, it sometimes never achieved sufficient confidence to serve all drinks
Never selected choices or asked for clarifications, hence reduced total trial time	Asked for clarifications several times within a trial increasing total time taken
Served 1st drink more quickly	Was slow in serving due to clarification

Subjective Measures:

- Used subjective ***GodSpeed Questionnaires*** before and after the trial and a short questionnaire to access overall impression and perceived success of experiment
  - **GodSpeed Questionnaires** are standardized measurement tool in HRI field, to measure user attitudes and as a performance criteria for service robots.
  - **Cronbach's Alpha** measures ***internal consistency reliability*** among a group of items that are combined to form a single state, ideal min value = ***0.7,*** high for both pre & post tests
  - **Linkert Scale**
  - **Anthromorphism** refers to ***human like*** form, human characteristics or behavior e.g. mechanical/humanlike
  - **Animacy** makes robots ***lifelike***, which involves users emotionally and can be used to affect users responses. E.g. Artificial/Lifelike & Inert/Inactive
  - **Likeability** is the ***positive first impression*** of robot on humans, e.g. factors like kind/unkind, friendly/unfriendly, pleasant/unpleasant and dislike/like,
  - **Perceived Intelligence** is ability of robot to ***act*** intelligently, hence factors like Incompetent/Competent and Unintelligent/Intelligent.
  - ***Responses decreased from pre to post tests, biggest decrease in Perceived Intelligence.***

1

TABLE III

Summary of responses to Godspeed questionnaire

Category	Pre-test		Post-test	
	α	Mean (sd)	α	Mean (sd)
Anthropomorphism	0.77	3.0 (1.1)	0.85	2.6 (1.3)
Animacy	–	3.6 (1.7)	–	3.2 (1.5)
Liking	0.82	5.3 (1.1)	0.91	4.8 (1.2)
Intelligence	0.90	4.5 (1.5)	0.85	3.7 (1.4)

TABLE IV

Summary of results to session questionnaire

Measure	Baseline (sd)	Uncertainty (sd)	M-W
Perceived precision	0.92 (0.26)	0.97 (0.17)	n.s.
Perceived recall	0.90 (0.21)	0.81 (0.33)	n.s.
Overall impression	4.4 (1.0)	3.7 (1.2)	$p < 0.01$

Fig. 11 and Fig 12 [2]

1

Subjective Measures (Contd.)

- People’s expectations of a robot’s interactive capabilities tend to outstrip their actual experience of interacting with it, even when they have previous experience with the same robot.
- Results from additional subjective questionnaire shown in Table IV:

Systems	Perceived Precision	Perceived Recall	Overall impression
Baseline	Lower	Higher	Higher
Uncertainty-Aware	Higher	Lower	Lower

1

- Stepwise multiple linear regression analysis carried out to test what aspects of uncertainty-aware system effected the user's overall impression of interaction, with R2=0.235

$$\begin{aligned} \text{Overall} \quad = \quad & 4.04 - 3.1 \cdot \mathcal{N}(\text{LastDrinkTime}) \\ & + 3.04 \cdot \mathcal{N}(\text{Duration}) + 0.91 \cdot \mathcal{N}(\text{NumDrinks}) \\ & - 0.49 \cdot \mathcal{N}(\text{Choices}) - 0.36 \cdot \mathcal{N}(\text{AskOrder}) \end{aligned}$$

- Scores were higher when interaction with user was longer & Number of drinks served was higher as well
- Scores were lower when duration to serve drinks was longer, more queries were asked by robot and when robot repeatedly asked for an order.
- Main contributors to satisfaction were no. of drinks served, system response time and the number of turns discarded due to low ASR with similar R2 value.

1

Results

Baseline System	Uncertainty-aware System
Serving Time: ***Faster***, served drinks right away	Serving Time: ***Slower*** as it always asks for clarifications
No. of drinks served ***more***	No. of drinks served ***less***
Serves more, but ***served incorrect orders*** as well. E.g. if there were 2 hypothesis both with same values, it chooses randomly between the two, which could be incorrect order	***Never served an incorrect order*** as it takes care of uncertainty by asking ***clarifications*** and using ***confidence levels*** for input hypothesis, but sometimes did not serve any drink as it failed to accumulate enough confidence and user lost patience
In case the threshold is greater than coded for comparison, the system fails to recognize the error	Recovers from misunderstanding by asking for clarification

"Experiences with Mobile Robotic Guide for the Elderly"

- Introduction to paper

  Building on a robot navigation system , new software modules specifically aimed at interaction with elderly people were developed.

- Robustness of probabilistic techniques for real world tasks
- Feasibility of using mobile robots as an assistance to the elderly
- Handling safety concerns during robot-elderly interaction
- Uses POMDP in robot's high level control system
- Handles uncertainty in all levels of decision making

Conclusions:

- Since selection of confidence thresholds was arbitrary, Building on previous work on using ***reinforcement learning*** for optimizing action selection strategies for multi-user human-robot interaction, a learned strategy will have incorporated the optimal thresholds automatically.
- Taking into account safety measures during Human Robot Interaction

1

References

- [1]: "Planning for social interaction in a robot bartender domain" by Ronald P.A Patrick and Marry Ellen Foster (Proceedings of 23rd international conference on automated planning and scheduling)
- [2]: "Handling uncertain input in multi-user human robot interaction" by Simon Keizer, Marry Ellen Foster, Andre Gaschler, Manuel Giuliani, Amy Isard and Oliver Lemon (The 23rd International IEEE International Symposium on Robot and Human Interactive Communication, August 25-29, 2014. Edinburg, Scotland, UK)
- [3]: http://dailydotnettips.com/2014/01/23/accepting-kinect-speech-commands-after-a-specific-level-of-confidence/