



The 2nd Workshop on

Social Robot Navigation: Advances and Evaluation.

In conjunction with: IEEE International Conference on Intelligent Robots and Systems

October 1-5th, 2023 - Detroit, MI, USA

Effortless Polite Telepresence

Jani Even



京都大学
KYOTO UNIVERSITY

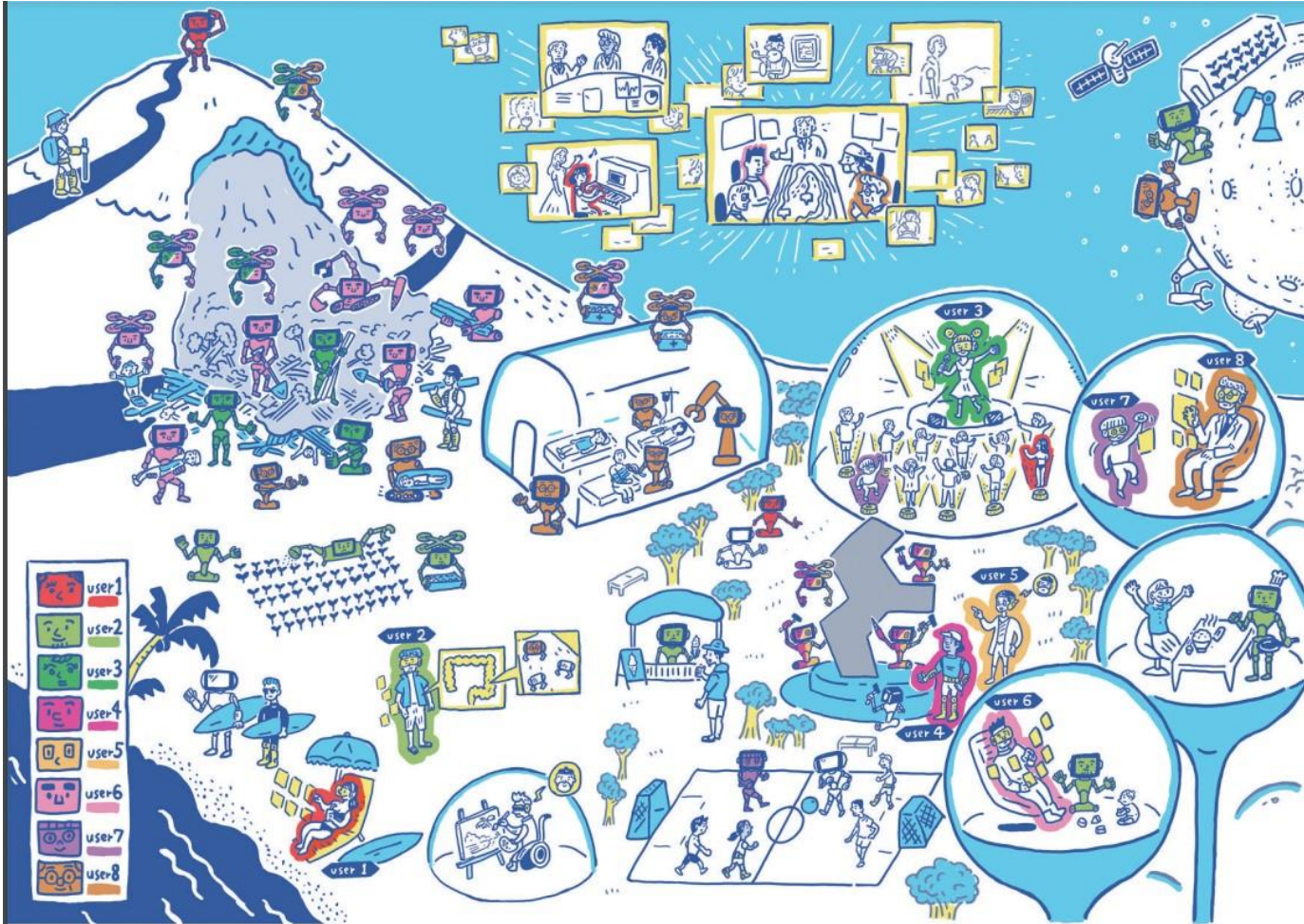
HRI Lab. at Kyoto University



京都大学
KYOTO UNIVERSITY

JST's Moonshot project

'By 2050, a society in which people are free from the constraints of body, brain, space and time'



(JST illustration of Moonshot program vision)

Bring new people into the workforce



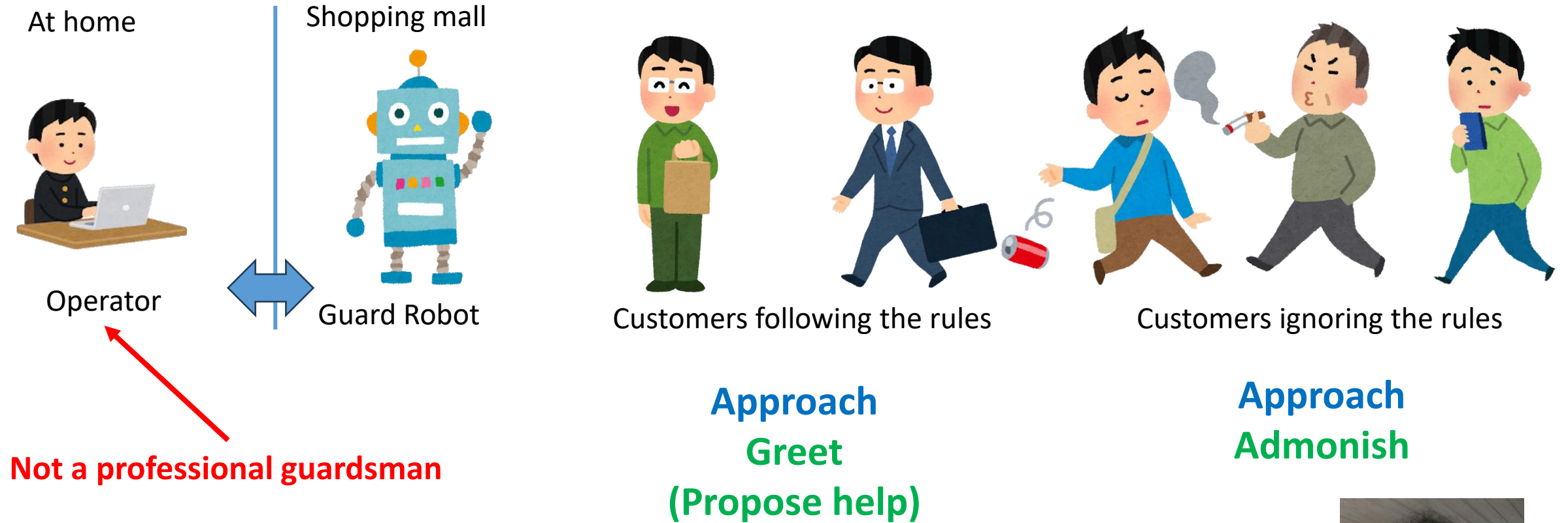
Avatars

- agents
- robots

Provide many kind of services

(JST = Japan science and technology agency)

Today's topic: Teleoperated guard robot

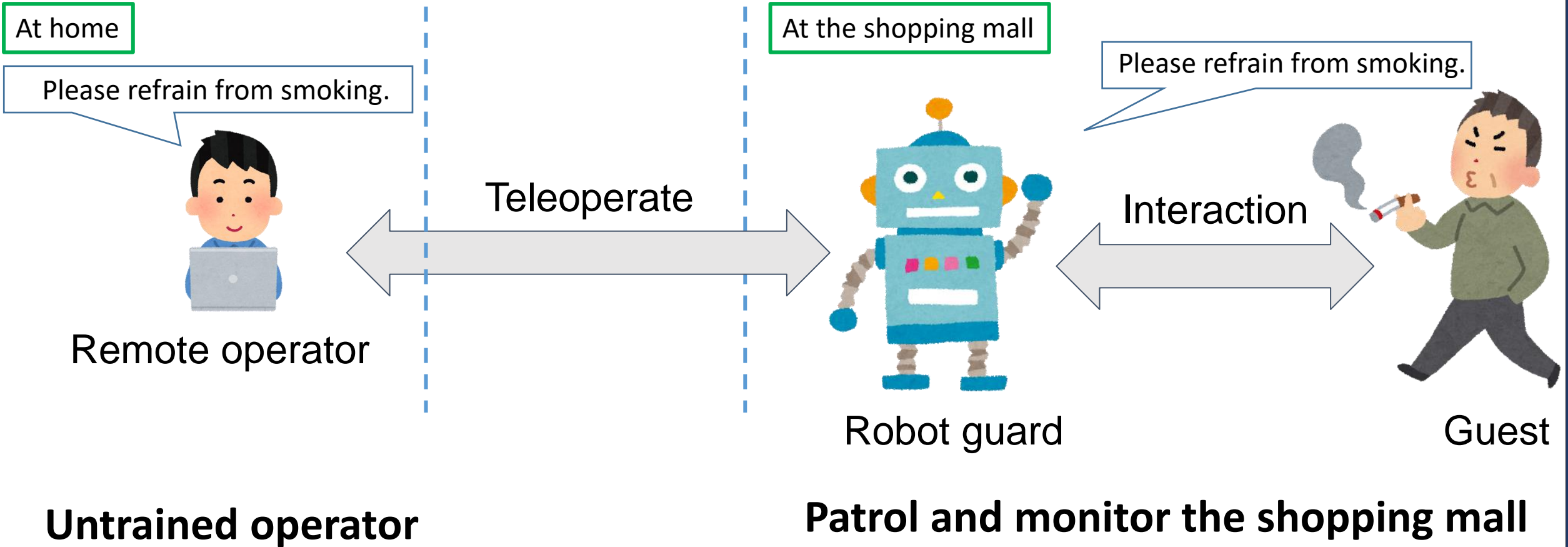


We want to create a support system to help the operator
We want the robot to provide polite service

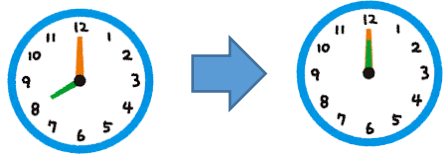


Morteza Daneshmand
(now at NIBIO in Norway)

Context: Guard Robot in a Shopping Mall



Context: Operator's fatigue



At home

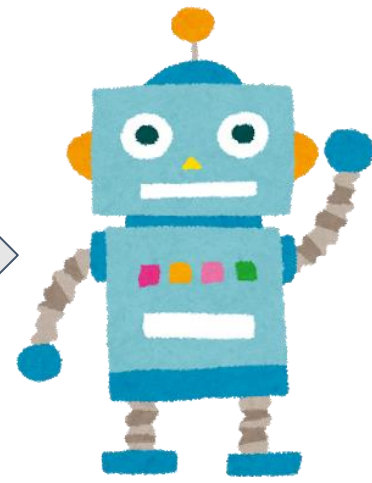
No smoking @##**#!!!



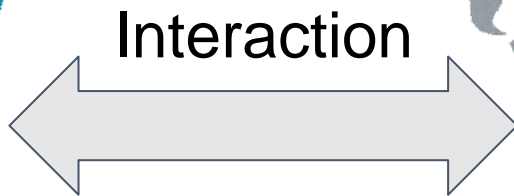
Remote operator



At the shopping mall



Robot guard



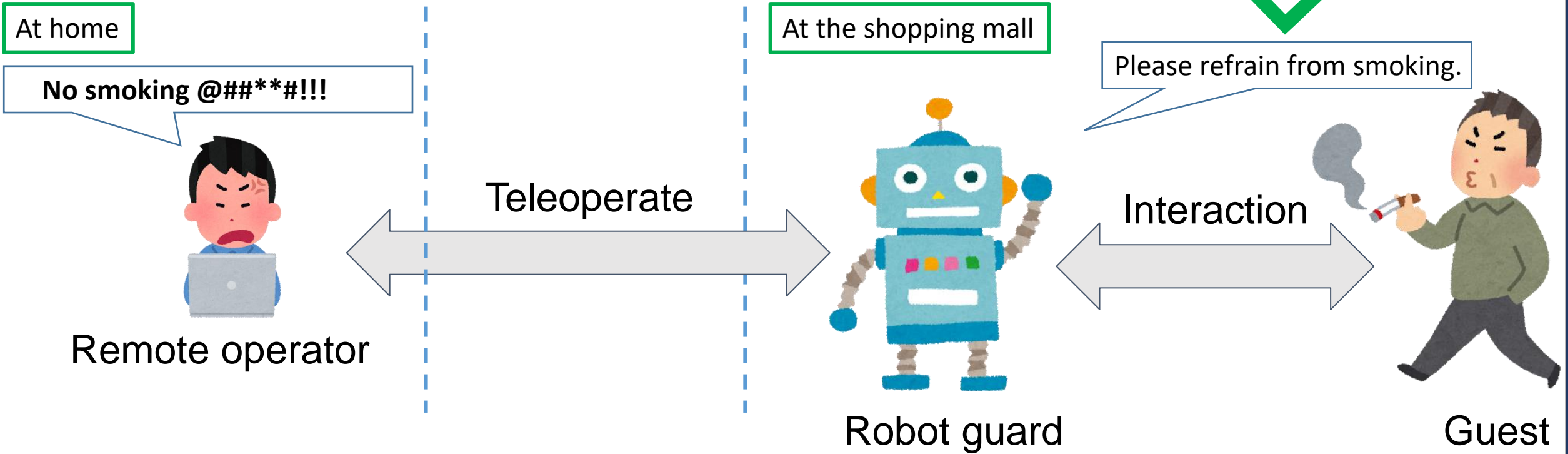
No smoking @##**#!!!



Guest

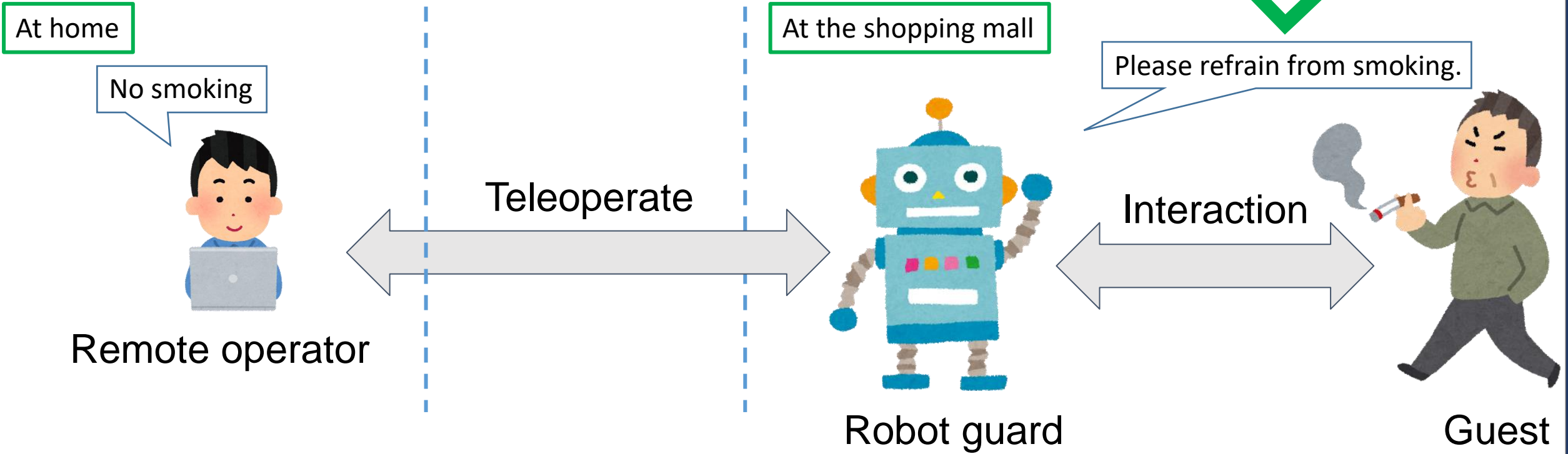
It is hard to teleoperate a robot for a long time keeping service quality

Teleoperation system that guarantees politeness



Inappropriate language does not impact the target user

Teleoperation system that reduces workload

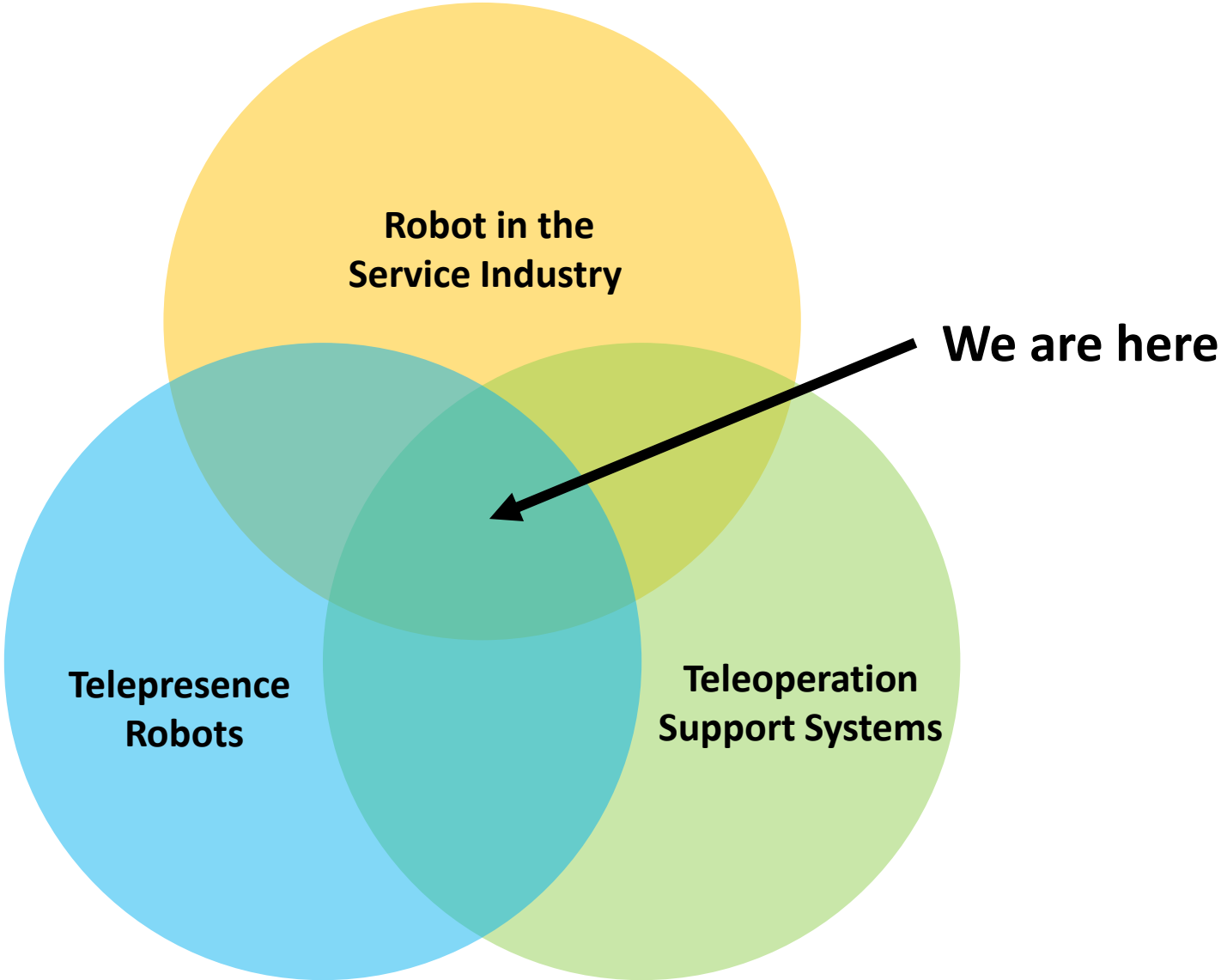


Operator may talk casually.

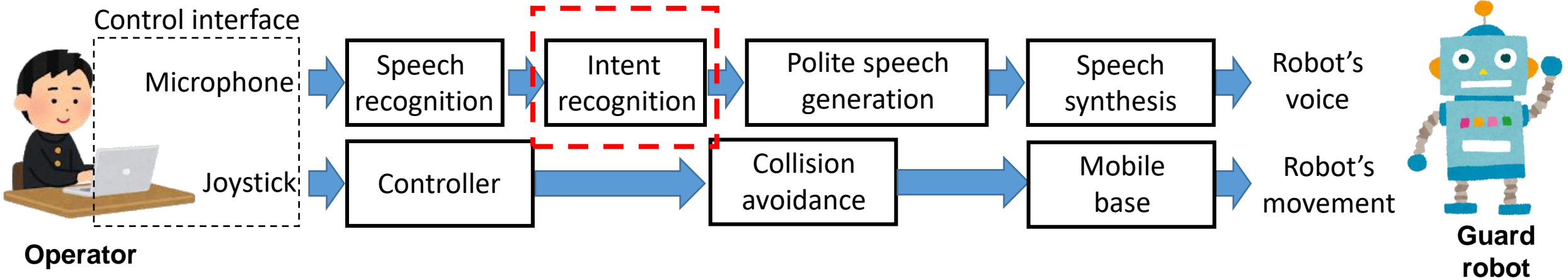
Hospitable Guardsman Robot



Related Work



How do we do that: Proposed System



➔ **The focus is on the conversation part**

Intent recognition

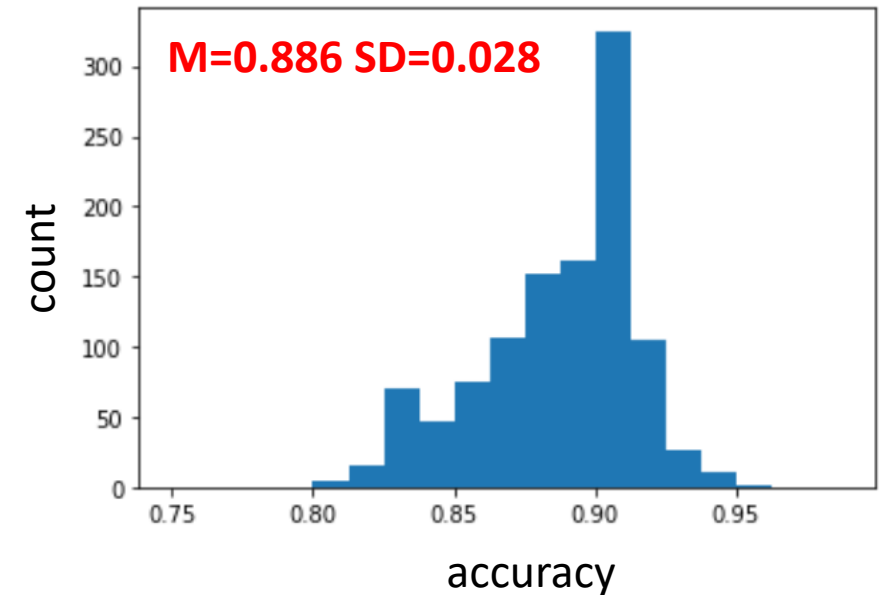
Tohoku University Natural Language Processing Group on the Japanese version of Wikipedia

Fine tune with our data a pre-trained BERT model

Data collection: 9 participants (40 utterances/intent) + testing

Intent categories	utterances
Greeting	34
Thanking	41
Asking if a visitor requires help	41
Asking to refrain from smoking	54
Asking not to litter	73
Asking not to use a phone while walking	59
Asking to wear a mask	47
Asking to be quiet	57
Asking not to run	46
Total	452

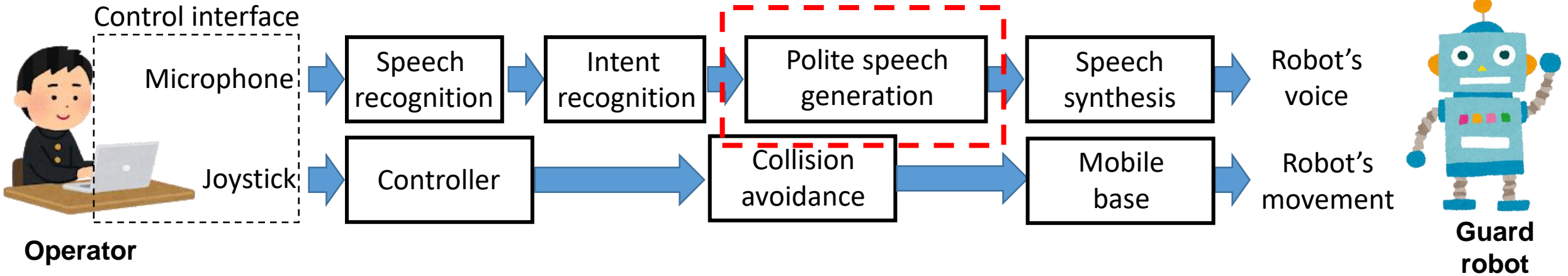
11 folds cross validation 100 runs
Predict test 40/intent
train and test from different participants



We can predict the intents

(0.95 out of intent/intent)

Polite speech generation



Dataset of curated utterances

3 utterances to politely express each intent as a guardsman would



Randomly select one of the three given the intent

User Study: Scenario

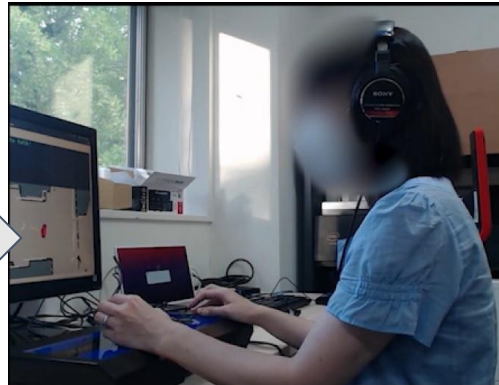
Two “Actors” roleplaying the customers



Corridor

Teleoperate

Participants are operators



Control room

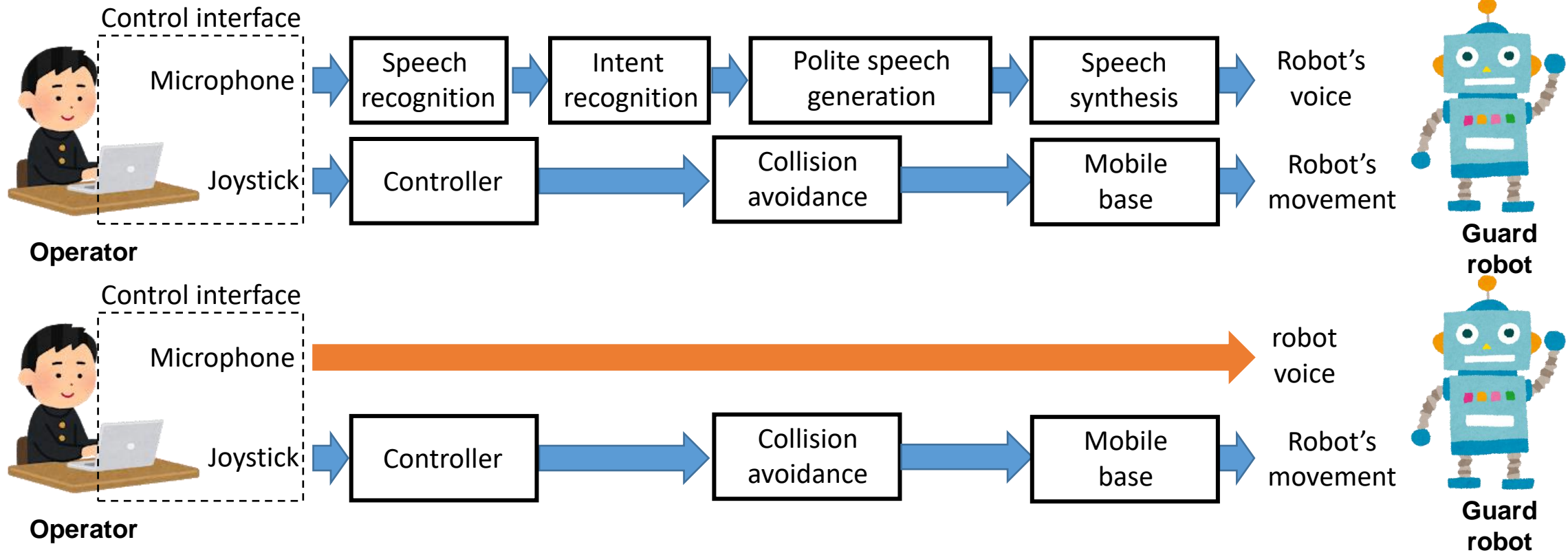
Each has a script with **four actions** with **timings**

One action is composed of four steps:

- Enter the corridor
- Perform a behavior
- Answer to robot first reaction
- Answer to robot second reaction
- Leave the corridor

(Answer = **ignore/respond/comply**)

User study: Hypothesis



Prediction 1 (“Workload”): An operator using the proposed teleoperation support system will have a lower workload than an operator using a baseline teleoperation system

Prediction 2 (“Politeness”): A robot controlled with the proposed teleoperation support system will speak more politely than a robot controlled with a baseline teleoperation system.

User Study: method

23 Participants (M=21.9 SD=7.3)

Within-subject design with counter-balanced Conditions

For each condition:

- Experimenter explains and demonstrates the system
- Participant tests the system
- Participant operates for 10 minutes
- Participant fills in a questionnaire ← **Workload: NASA-TLX**

+ Semi-structured interview at the end

Politeness:

- Three coders rated the transcripts of the robot's utterances
- 5 point Liker scale
- Take average of ratings

Results: Observation

Proposed

- Talked relatively politely
- Some adapted to the intent recognition
- Some repeatedly told utterances not recognized

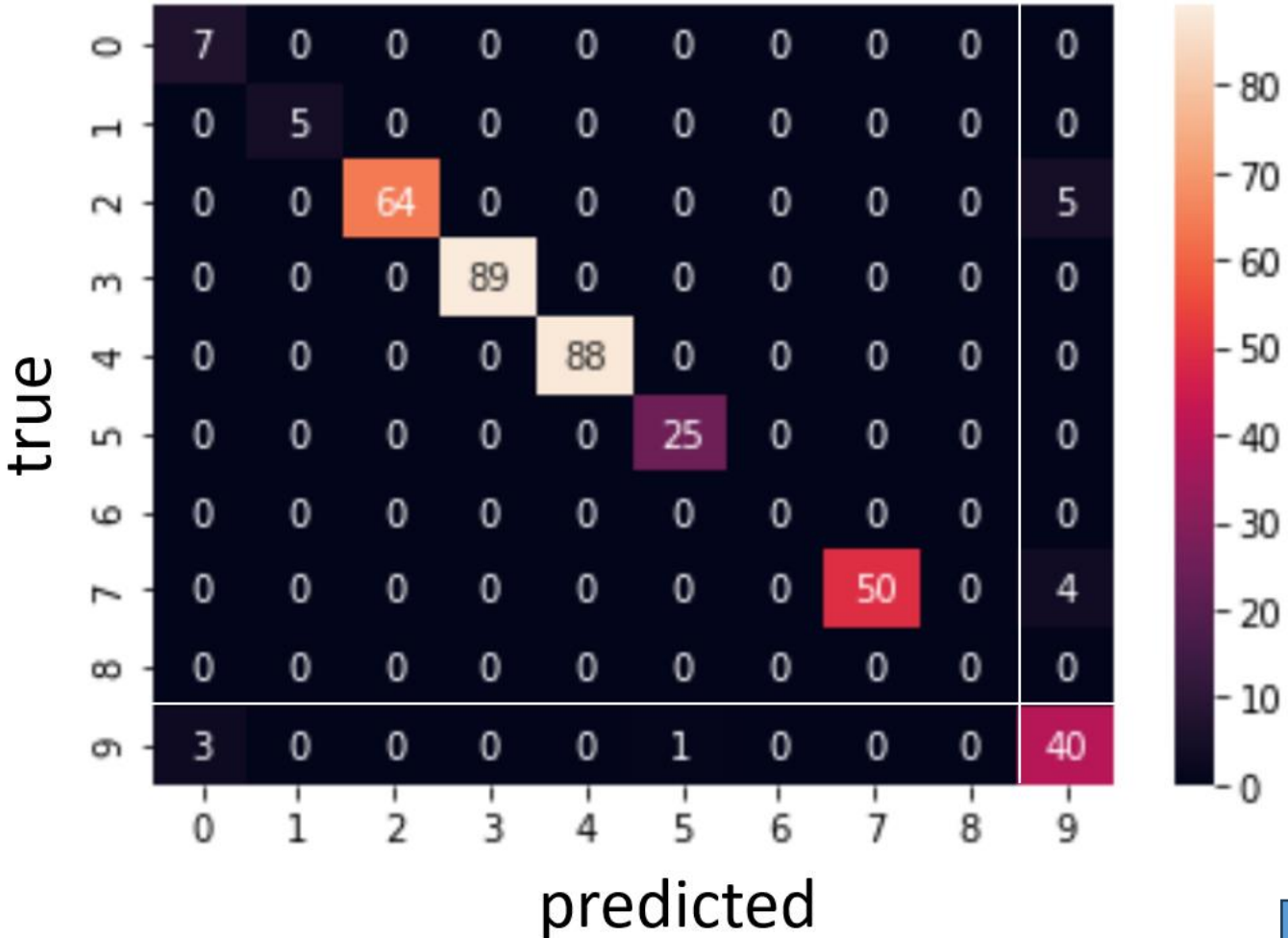
Baseline

- Took time to find the right words
- Talked politely

Both

- Participants had no problem to operate the system
- More hesitant to admonish “talking loud” or “walking with phone”
- No hesitation for “smoking” or “littering”
- Very few motion

Results: Intent recognition performance



Within intent set:

- 381 utterances
- Accuracy M=0.96 SD= 0.045

Out of intent set:

- 44 utterances
- 0.9 accuracy

➡ System performed well

Results: Workload NASA-TLX

	N	Mean	SD
baseline	23	62.739	12.693
proposed	23	46.072	14.365

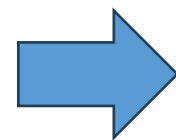
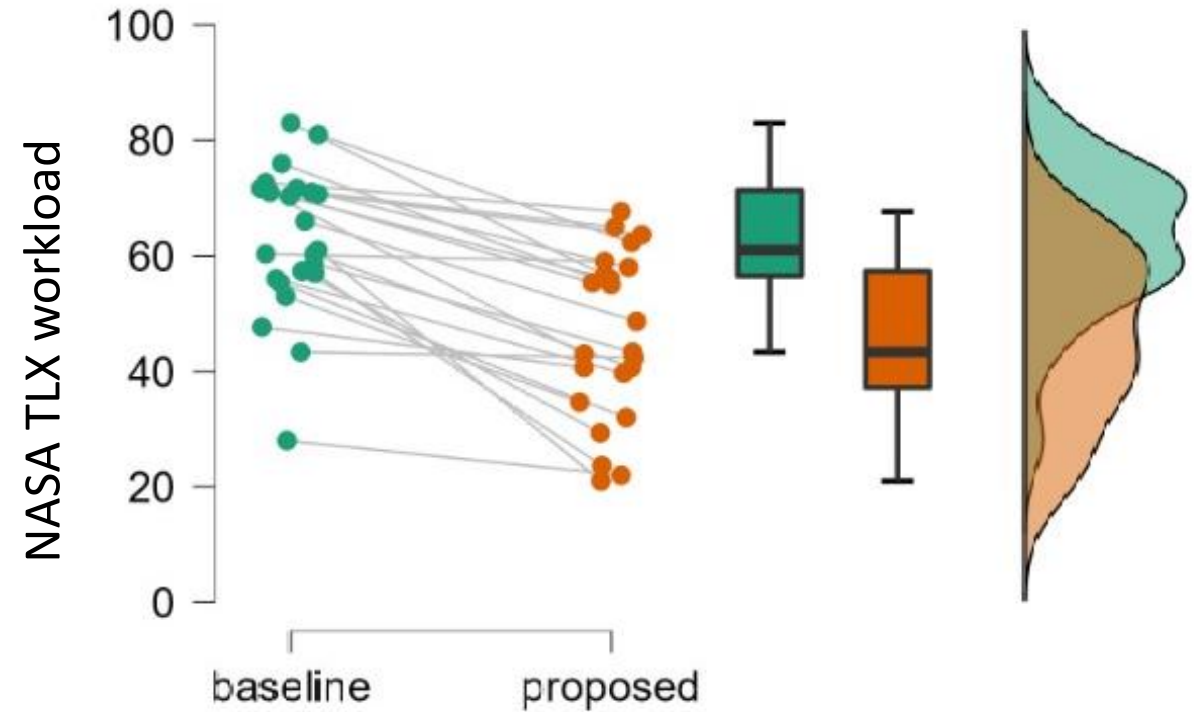
Test of Normality (Shapiro-Wilk)

	W	p
baseline - proposed	0.964	0.550

Note. Significant results suggest a deviation from normality.

Paired Samples T-Test

Measure 1	Measure 2	Test	Statistic	z	df	p	Effect Size
baseline	proposed	Student	7.792		22	< .001	1.625



Data supports hypothesis 1
Operators perceive less workload

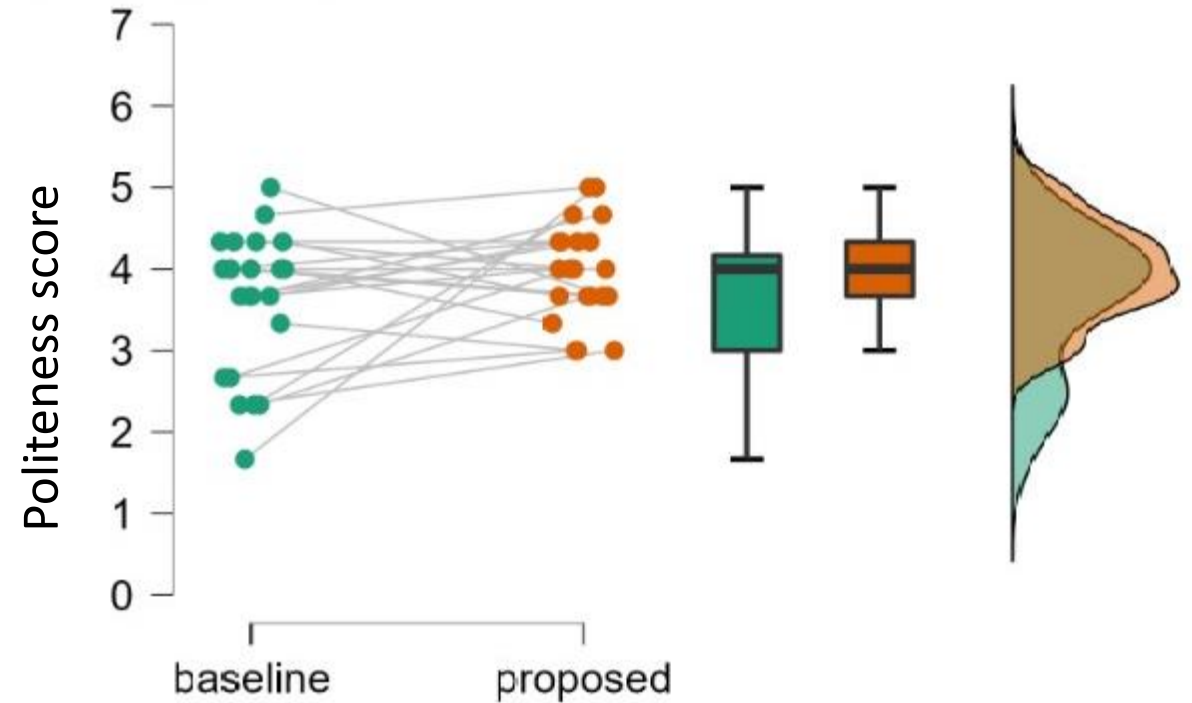
Results: Politeness

	N	Mean	SD
baseline	23	3.609	0.868
proposed	23	3.971	0.585

Test of Normality (Shapiro-Wilk)

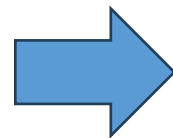
		W	p
baseline	-	0.906	0.033

Note. Significant results suggest a deviation from normality.



Measure 1	Measure 2	Test	Statistic	z	df	p	Effect Size
baseline	-	Student	-1.696		22	0.104	-0.354
		Wilcoxon	69.500	-1.325		0.189	-0.338

Note. For the Student t-test, effect size is given by Cohen's *d*. For the Wilcoxon test, effect size is given by the matched rank biserial correlation.



Data does support hypothesis 2

Results: Interview

Proposed:

- mentally less demanding (17/23).
- Not having to pay attention to the wording (7/23)
- Delay when out of intent set (9/23)
- Felt protected from customers' aggressiveness (11/23)

Baseline

- Hard to find the appropriate words (13/23)
- Felt under pressure to be polite (6/23)
- Better as could express more nuance (9/23)
- Felt customers' aggression (10/23)

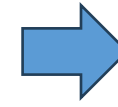
Discussion

Reduced workload

Questionnaires: Perceived less workload

Interviews: No pressure to find words or be polite

But still relatively polite!



Different levels of politeness?

Safeguard = Peace of mind ?

Cultural difference

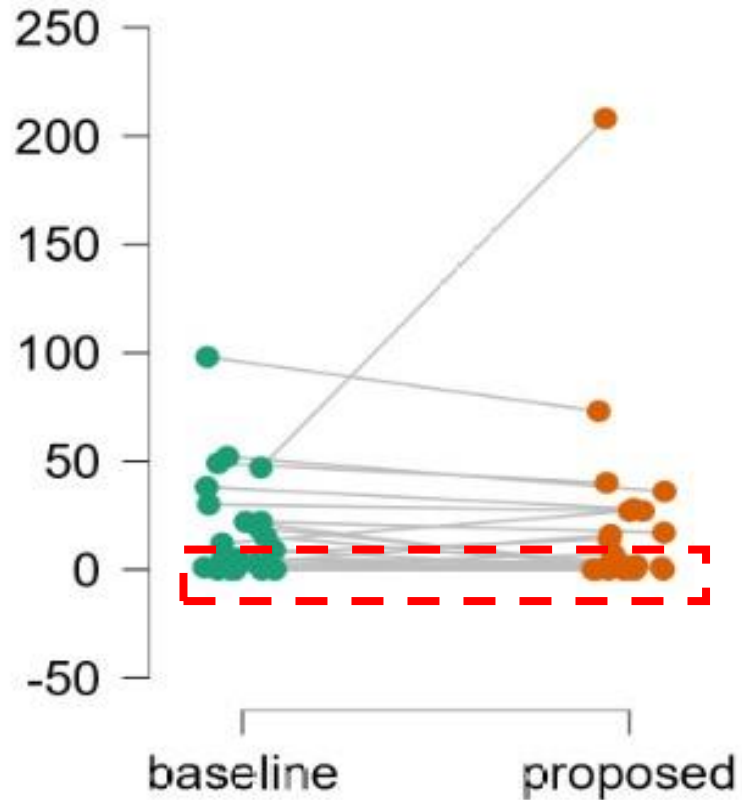
- Done in Japan
- How people think about required politeness?
- How is intent recognition for other languages?

Ethical consideration

- Workers cannot choose their words
- Workers cannot engage in other tasks
- Who decide what is appropriate?

（ Response to aggression
Own voice vs robot's voice ）

No driving!



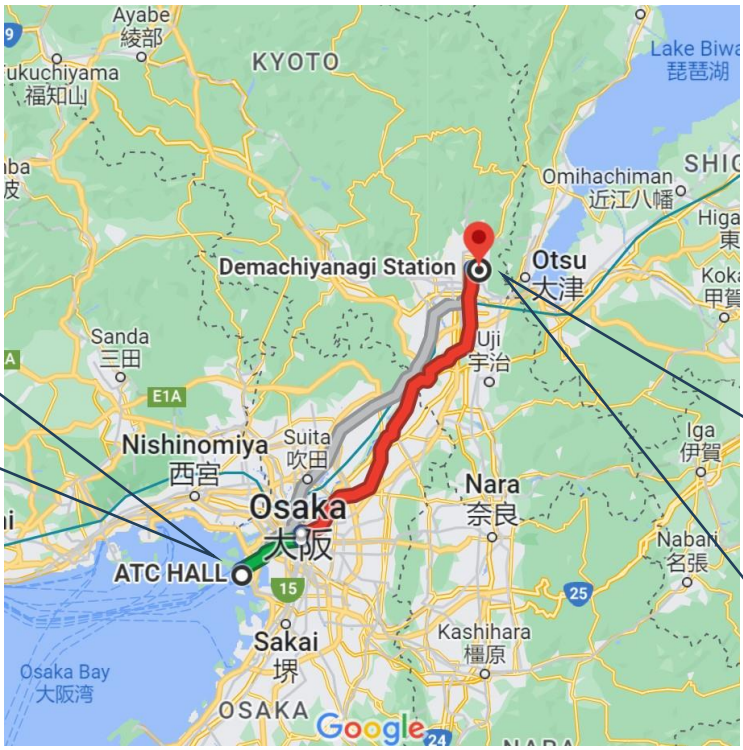
(No significant difference)

Many participants did not approach customers!!

- Small corridor
- No background noise
- Only one customer at a time

Field experiment

ATC shopping mall, Osaka



approx. 70km

HRI Lab, Kyoto



Learn from good operators

- Larger area => operators must drive to reach customers
- Are some operators better than others?
- Can we learn from good operators to support bad operators?

After first round of in house tests (with experienced operators)

- No operator could do it smoothly
- Requests for better interface
- Requests for automation

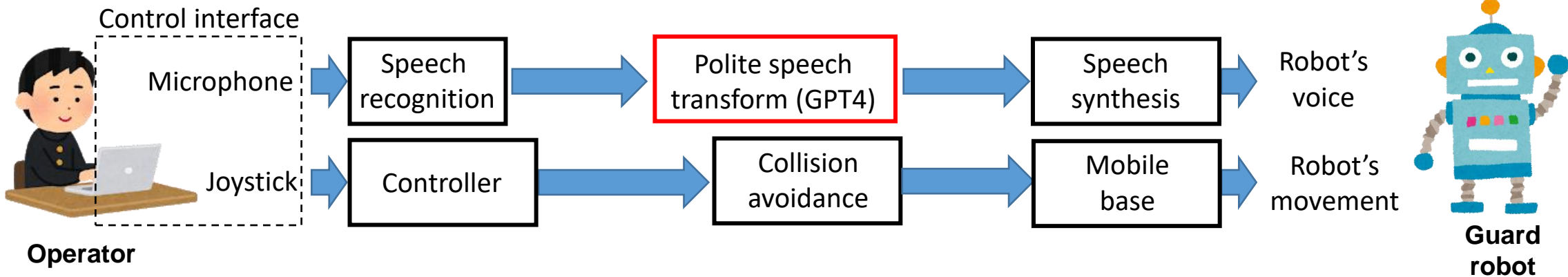
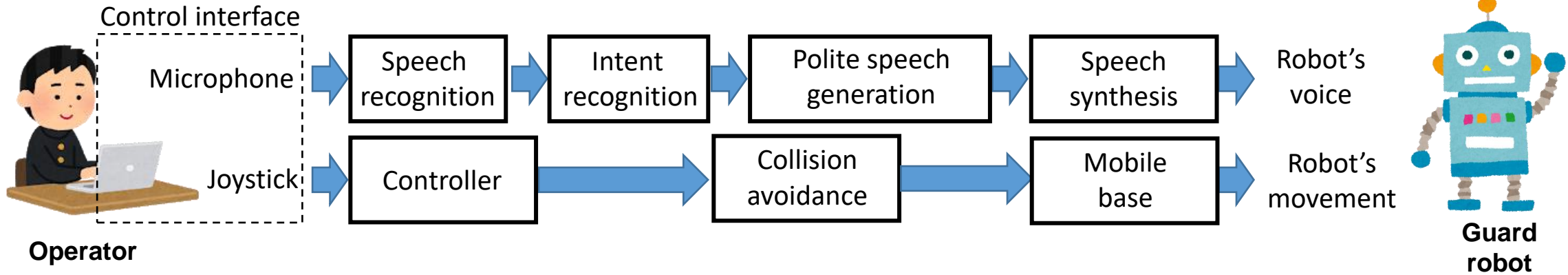


Automate the approach using the intent

Conclusion

- The proposed support system decreases the workload for the conversation.
- Users are still unable to appropriately approach the customers
- We have to support the navigation!

Bonus: Generative AI



Works relatively well but it is slow

(Use it for out of intent set utterances)

Thank you.
(Q&A later)