Making Sense of Audio Vibration for Liquid Height Estimation in Robotic Pouring

Hongzhuo Liang, Shuang Li, Xiaojian Ma, Norman Hendrich, Timo Gerkmann, Fuchun Sun, Jianwei Zhang

Nov. 06, 2019 at Macau, China

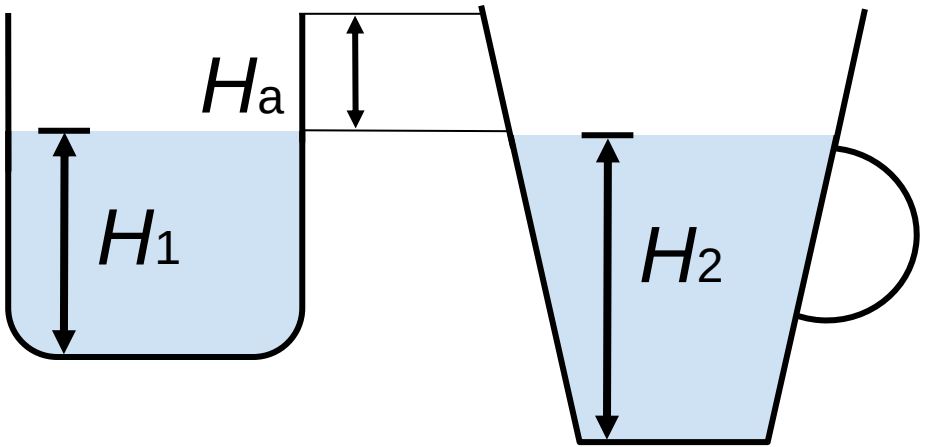
Why Audio?

- Visual sensing methods cannot generalize to occluded situations.
- Force / torque sensors are expensive and cannot measure the liquid height in an unseen target container.

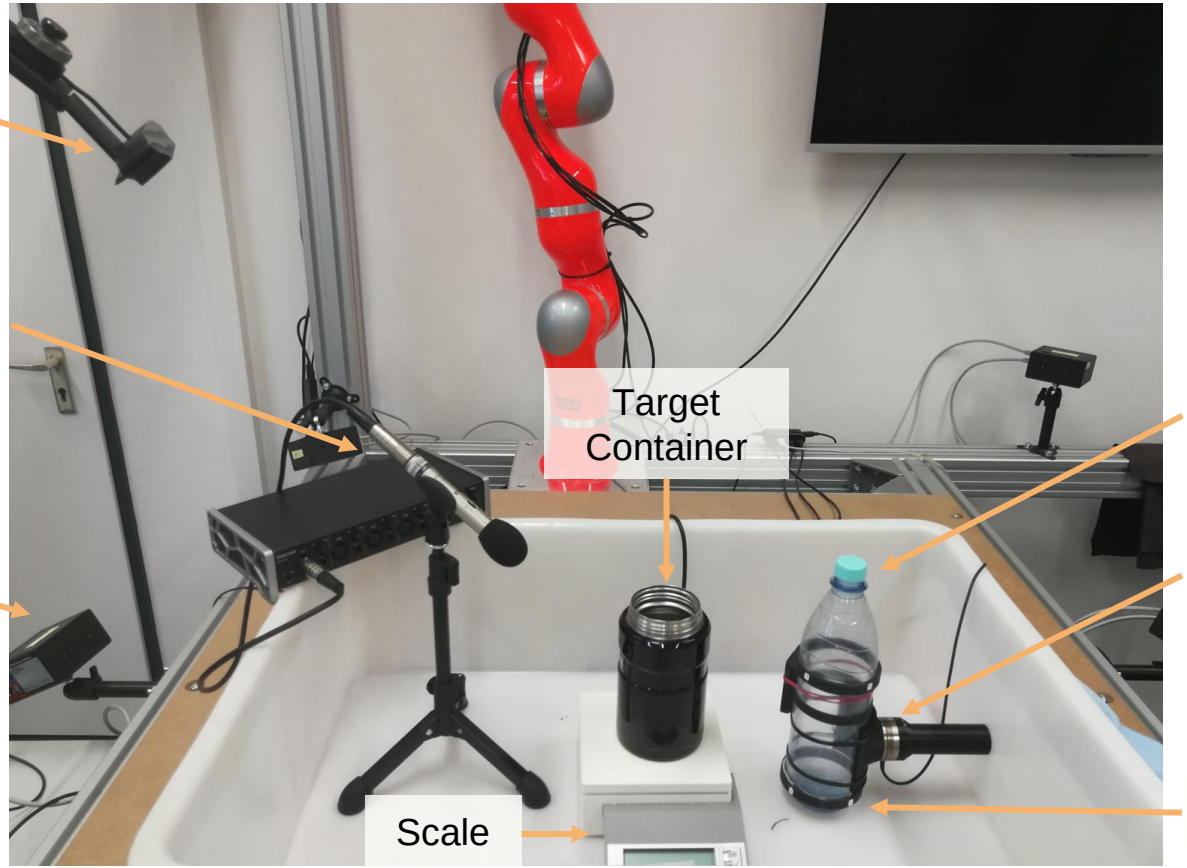
Task definition

- Given a desired **length of air column H_a**, the robot estimates the current length of air column h_a from the audio input and stop pouring when $h_a \leq H_a$.

- Length of air column H_a has a direct relationship with **resonance frequency** of the air.



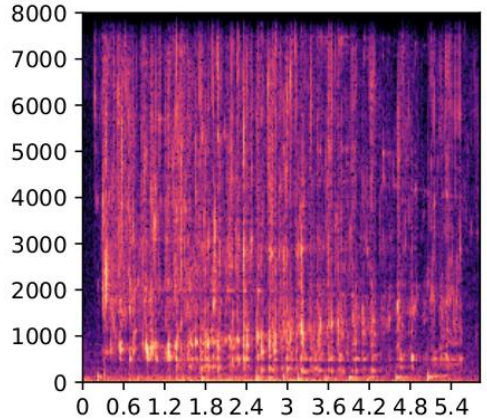
Multimodal pouring dataset



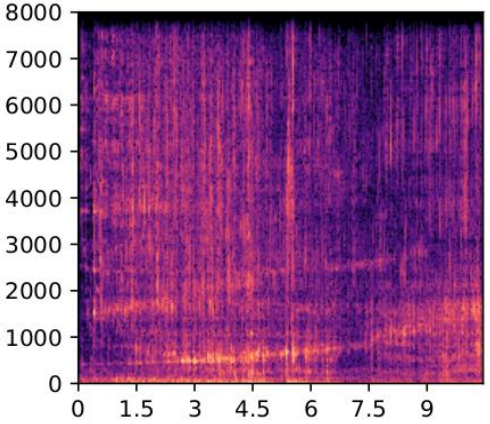
An example of dataset collection

- The dataset was collected by humans in a quiet environment.
- Pouring during the training only involved water.

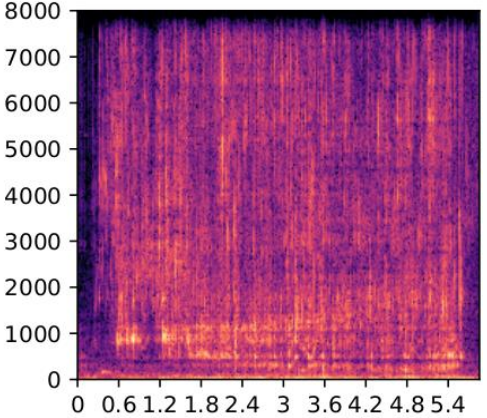
Audio data process



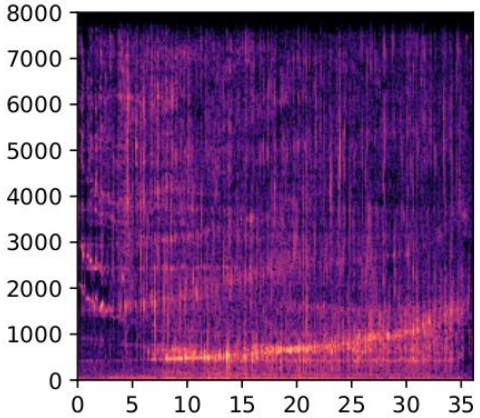
Pouring into glass



Pouring into thermos (without spout)

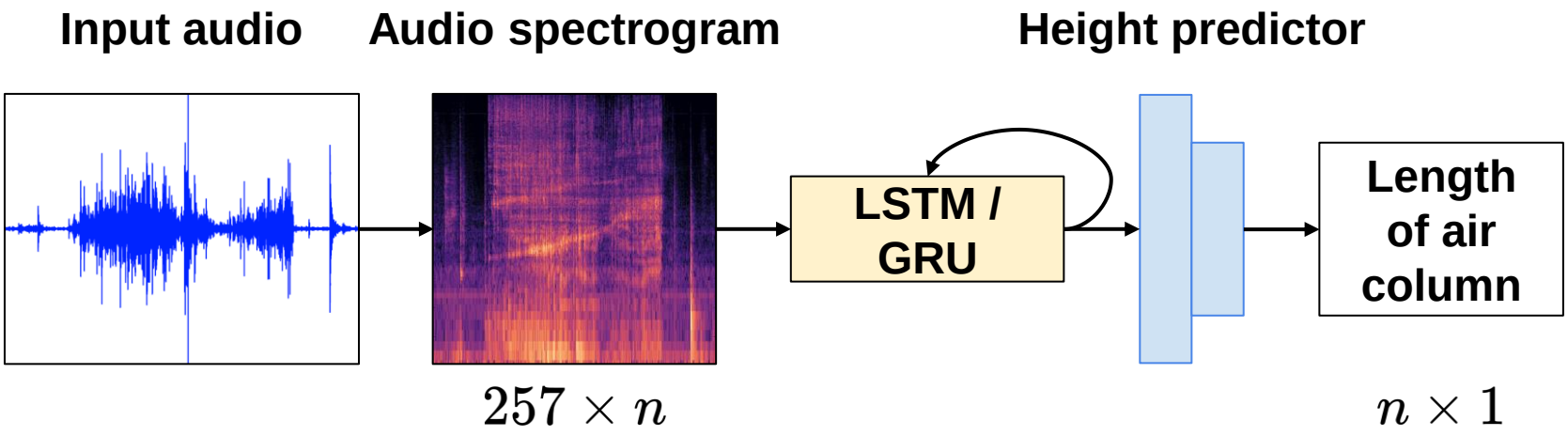


Pouring into mug



Pouring into thermos (with spout)

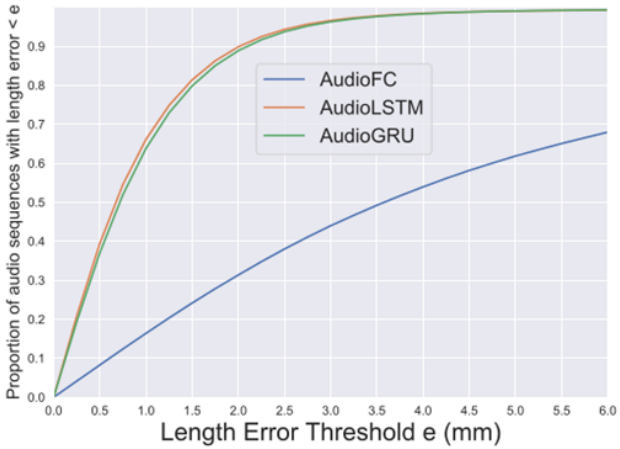
PouringNet



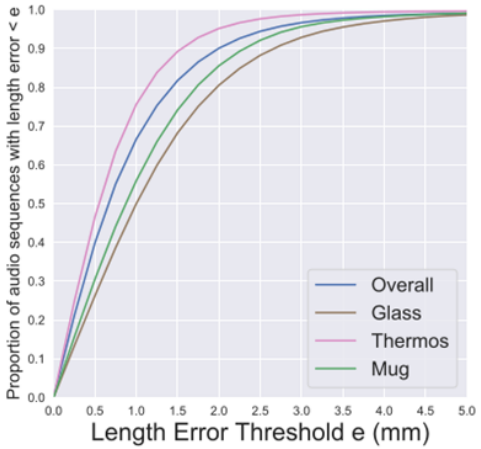
- L_{height} : mean squared error loss
- L_{mono} : keep the output monotonously decreasing along the time
- Overall loss:

$$L_{audio}(\theta) = L_{height} + \alpha \cdot L_{mono}$$

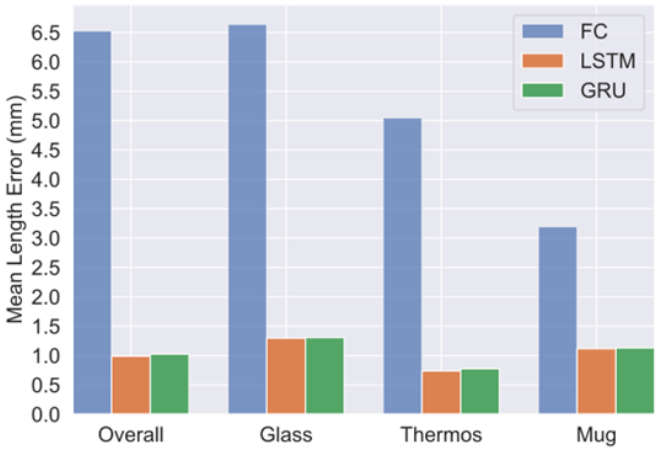
PouringNet evaluation



(a)



(b)



(c)

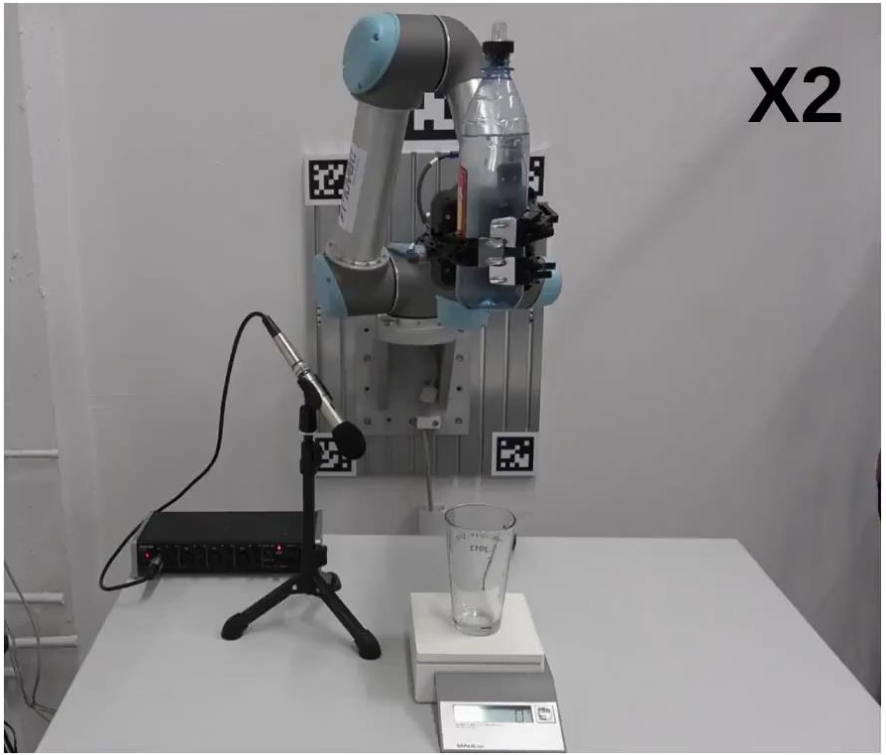
Two evaluation metrics:

- the fraction of audio sequences whose length predictions error are below a threshold;
- the average length prediction error of each target container.

Evaluation of different target containers

1) Known cups

Desired : 40.0000 mm
Real : 39.3915 mm
Network: 38.9372 mm

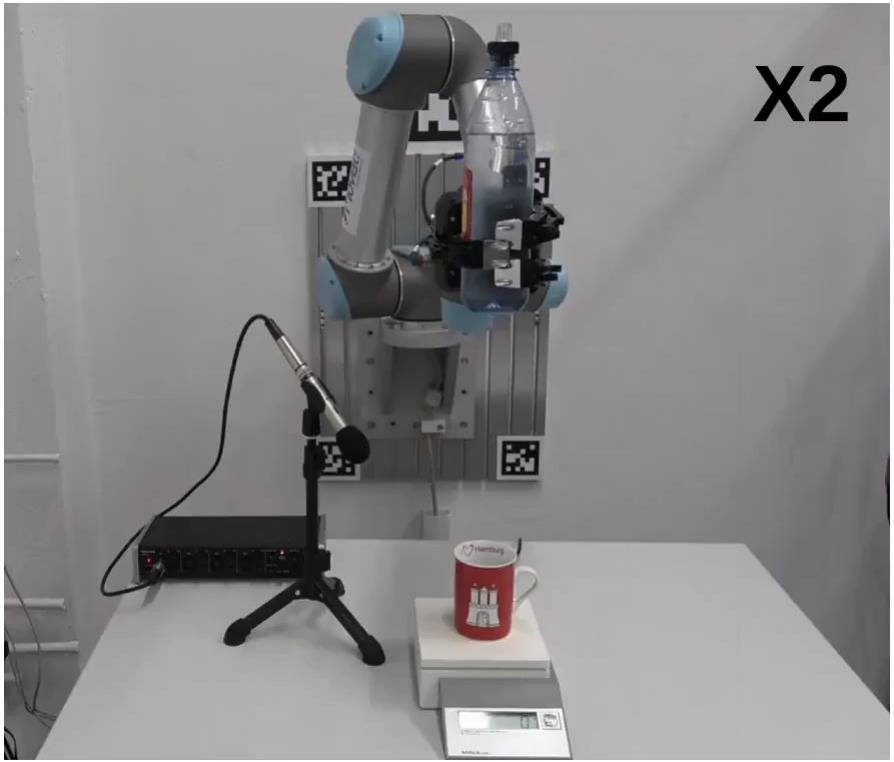


Evaluation of different target containers

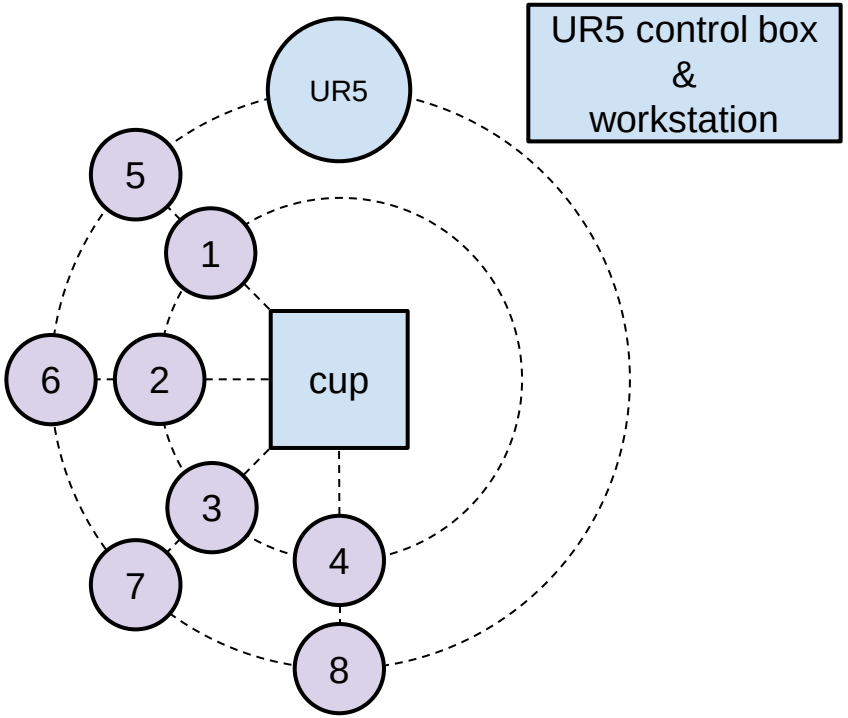
2) Unknown cups

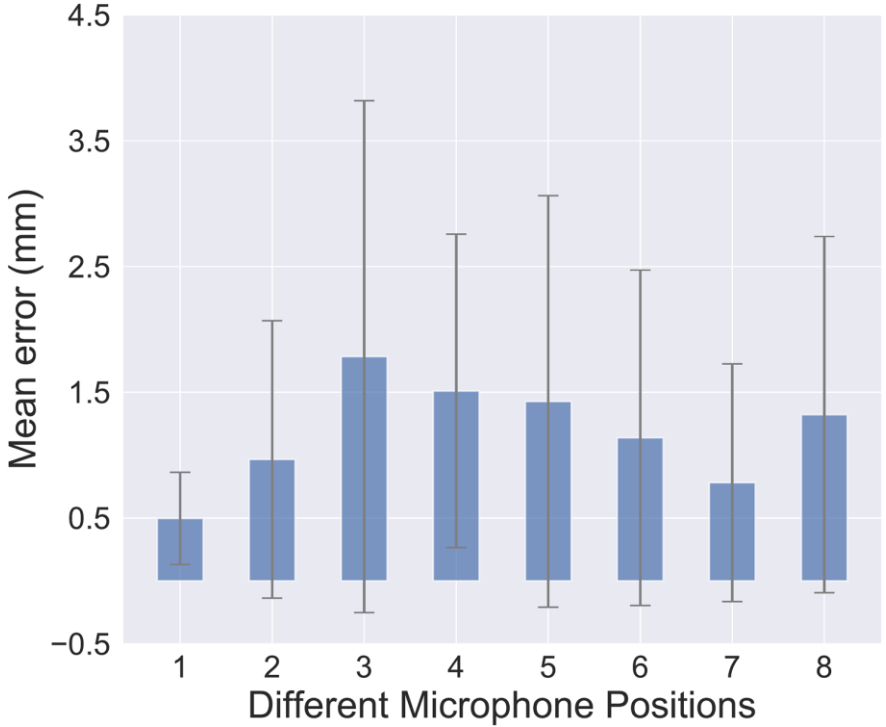
Desired : 40.0000 mm
Real : 39.8560 mm
Network: 39.2886 mm

```
Current length of the air column is 43.184803009, go on pouring
Current length of the air column is 42.1836242676, go on pouring
Current length of the air column is 42.1836242676, go on pouring
Current length of the air column is 42.1836242676, go on pouring
Current length of the air column is 42.1836242676, go on pouring
Current length of the air column is 38.9372825623, go on pouring
[INFO] [1551723117.436503]: Enjoy your drink!
[INFO] [1551723121.437580]: Network 38.9372825623mm, Real 39.391525759mm, De
sire 40.0mm, Scale 0.288g
^C(pouring) [lhz@tams223, load: 0.25] (Mo Mär 04 - 19:12:05)
~/code/audio_pouring $ python demo.py --cuda --bottle=6 --cavity-height=40
```

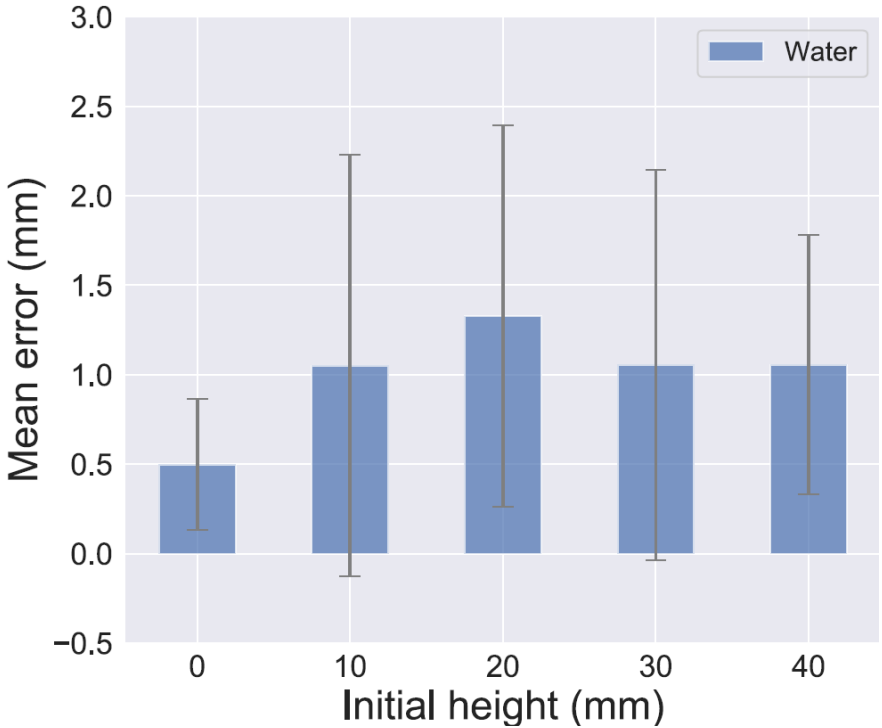


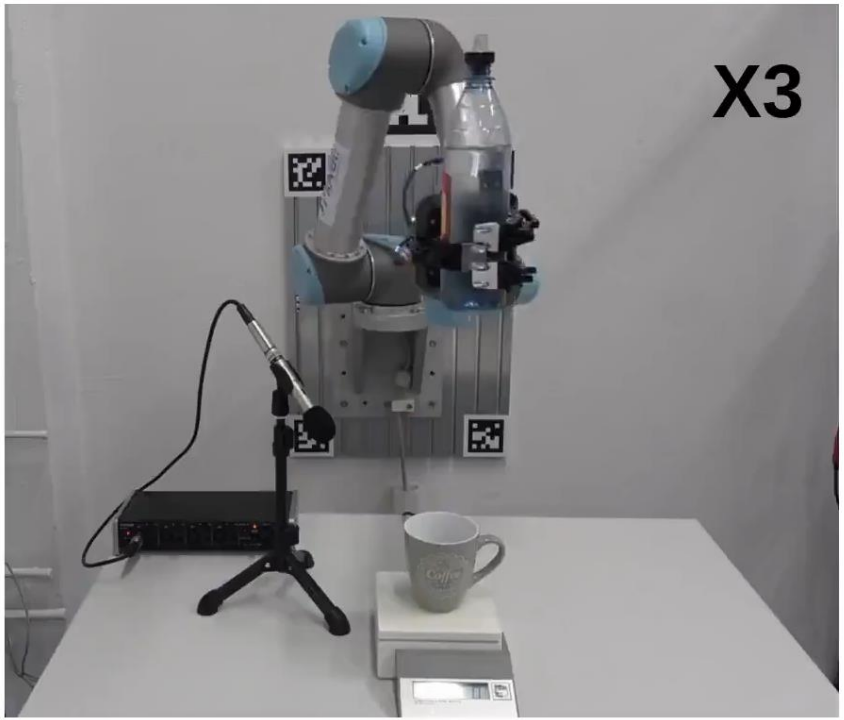
Evaluation of varying microphone positions



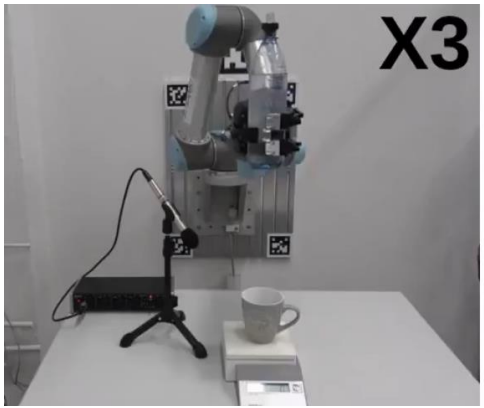


Evaluation of varying initial liquid height



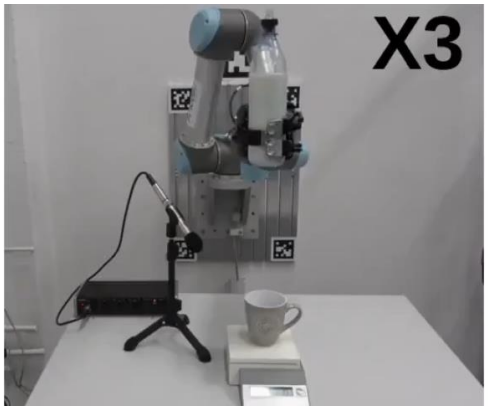


Evaluation of different types of liquid



Carbonated water

Mean error: 1.96 mm



Milk

18.54 mm

Orange juice

4.29 mm

- Our model can generalize to common household liquids, *e.g.,* carbonated water and orange juice.
- It has failed to work well on liquids with high viscosity, *e.g.,* milk, oil.

Conclusion and future work

- This paper presents a real-time system used for estimating the liquid height by **audio vibration**.
- Code and dataset are available at **https://lianghongzhuo.github.io/AudioPouring**.
- We plan to extend our approach to more noisy environments, *e.g.*, human voice and ego-noise of the robot.
- Making use of the force, motion trajectories, and visual data from our multimodal dataset would be an interesting research direction.

Thanks for your attention

ACKNOWLEDGMENT

This research was funded by the German Research Foundation (DFG) and the National Science Foundation of China (NSFC) in project Crossmodal Learning, DFG TRR-169/NSFC 61621136008. And it was partially supported by project STEP2DYNA (691154).