How can robots better serve food? (Group 7)

Chang, Feng Yi
fc4wa@virginia.edu

Ma, Jiachen
jm7gz@virginia.edu

Wang, Alan
ahw9f@virginia.edu

1 Introduction

In recent years, there is a new trend in applying robotics, artificial intelligence, and information technology to food serving area for improvements of food nutrition selection, customer service, and cost efficiency [1]. Food serving robots can have many distinct advantages over human serving robots, from availability during nonstandard hours of operations, to issues of health and hygiene. There have been many innovative researches of applying the intelligence robots in food serving area [2, 3]. One could argue that the assessment of vending machine [2] is the predecessor to a modern robot server. Vending machines, in many cases, are often the only possible choice to provide unhealthy dietary food and beverage for workers and visitors [4, 5]. [3] discussed an idea of E-Restaurant that consists of robotic waiters to serve customers. The robotic waiters are able to deliver food to each table after customer ordering. Therefore, food serving robots that are able to provide feedback of food selection to consumers according to the calculated food calories are much needed.

This paper is devoted to address the aforementioned challenges, the contributions of this paper are as follows:

- Establishing a general food serving method for the Nao robot to give food suggestions based on calorie count with the combination of facial features, such as age, gender and emotion state, to make a recommendation to users.
- Obtaining users' interaction data between every different events over time to track its changes over the iterations.
- Demonstrating the accuracy, comfort, and perceptions of the food serving interaction through surveying the participant.

The results indicate that the food serving method is applicable since the testers are able to interact with the robot without designers' guidance. The calorie estimate is perceived as more accurate than the food ingredients. Additionally, the tester's facial classification was biased towards middle aged males. The rest of the paper is organized as follows: Section 3 reviews the previous studies. Section 4 provides methods we developed in our food serving robot, as well as the robot algorithm and experimental design applied in this paper. The experiments for the effectiveness and improvement of the proposed methods are demonstrated in Section 5. Section 6 discusses the quantitative and qualitative results we analyzed through the captured data. Section 7 interprets the experiment results and possible future works are summarized in Section 8.

2 Related works

In this section, we highlight the advancement of food serving machines from vending machine to robot servers and advocate for emerging areas for Human-Robot Interaction in the food industry.

Studies of food serving via vending machines have been shown to have distinct impact over the purchasing and consumption behavior of certain populations. Specifically, [6] have shown that over 40% of children and adolescents consumed at least one competitive food daily while at school. This demonstrates that the introduction of vending machines have had a large and salable impact on the daily food consumption for a variety of age groups. However, traditional vending machines generally cannot travel to the client's destination, nor can it apply additional features such as give suggestions for food or use contextual information to carry out a simple conversation as a waiter might do.

The modern robot servers boasts numerous additional functionality compared to traditional vending machines. Nowadays, the robots could detect surroundings in multiple ways. The modern robots are able to count calories of foods [7, 8] and conduct facial prediction from humans [9] with the help of computer vision application. Hua et. al analyzed calorie labels on menus to inform consumers and encourage healthy choices [10]. The findings suggest that fact-based messages may be particularly helpful for consumers. An intelligent food choice method is proposed in [1], which the food serving robots can help consumers select foods by predicting customers' dietary preference, such as ingredients, price, types of spices, etc. By training robots to transform a series of data such as calories of foods, age of a person, gender of a person, and emotion of a person to corresponding response, it is beneficial to make robots better interact with humans. [11] proposed a knowledge-based hybrid decision model for nutrition management. For preference prediction, the proposed model data consists of age, sex, body mass index, region, chronic disease, and food preferences. The model selects a suitable candidate food according to the health condition of the user and provides a recommendation for N foods using the Top-N of the user's food preferences. Probably robots would suggest the rough food calorie when people are ordering meals. In addition, robots could tell a joke to customers when they look depressed. Hence, it is desired to have a systematic method for food serving robots to serve consumers, calculate food calories and provide feedback to customers.

3 Methodology/Experimental Design

For our experiment, we wanted to ways that a robot might be helpful to a user who wants to be more nutrition conscious. To test this, we tracked the timing between each interaction scenario, and or how a user's interaction with the robot for those scenarios over time would increase or decrease over time based on their learned experience interacting with a robot. Additionally, we wanted to capture the human component of their interaction, so we deployed a survey after 3 of their interactions to see learn about the results. Figure 1 below demonstrates an overall system diagram of our workflow.

A couple of key things to note about the settings in our experiment: 1) We turned off autonomous life mode when the sound in the lab kept drawing the nao robot to look elsewhere, making it hard to focus the camera 2) For the main question and answer from the nao, we use the speech recognition box with the default sensitivity settings as our primary communication method with the user 3) Users see the processes of the computer as the computer analyzes the information. This serves as a secondary indicator that informs the user about his or her food alignment, face alignment, and communication timing 4) Note the event logger. It will be further explained in the data collection section (section 4) 5) Needing to keep the robot on the ground for safety reasons also made the angle the camera somewhat awkward for the participants, and 6) we compensated this by showing the user the direct camera feed to the computer, and additionally made the *Nao* robot stand up as much as we can.

The experiment is conducted as following: the user is not given any directions as to how to

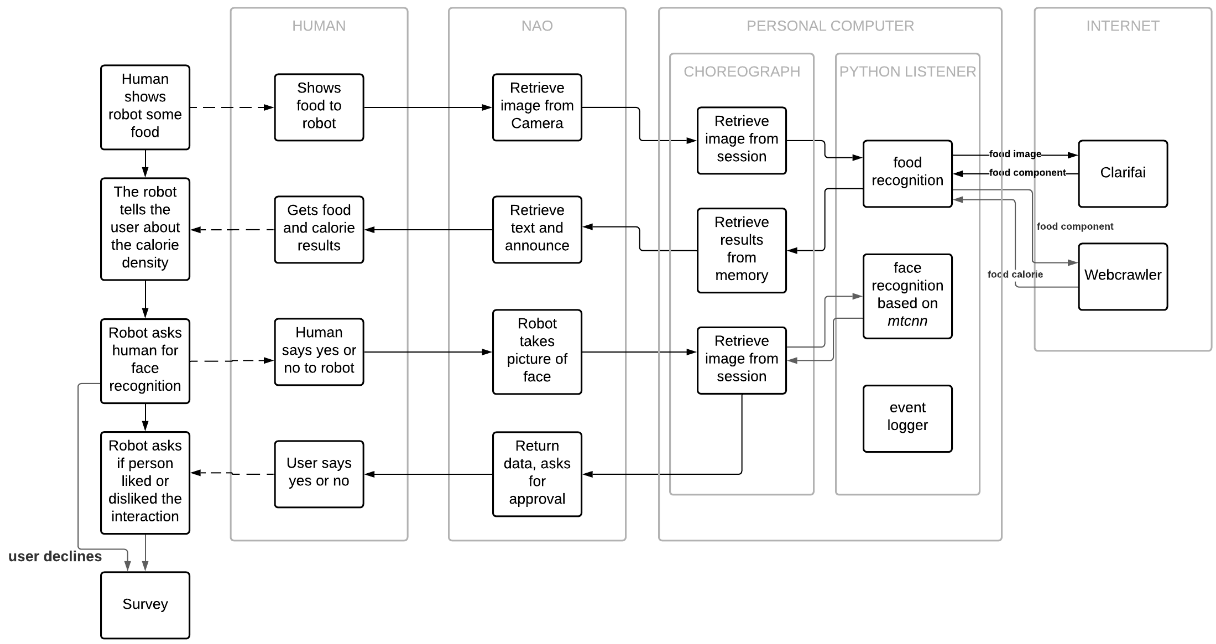

Figure 1: Overview of Nao Workflow

interact with the robot, *unless they indicate they cannot move forward.* Each user given a random assortment food item to test, and are asked to run three experiments one after the other with different fruit. We plan to see if there are any learning effects, which would theoretically decrease the amount of time it takes to interact with the robot. Additionally, we wanted to test a further feature that takes the learned calorie count and combines it with facial features that detect age, gender, and emotion state to make a recommendation. Our goal is to uncover whether or not there might be any positive or negative interactions that can point to the feasibility of having an autonomous nutritionist. Figure 2 below demonstrates our experimental set up.

4 Experiments, Data Capture, and Analysis

For our data capture, we wrote a time stamping code (the event logger) that directly communicates with the Nao memory in order to track the time taken for every stage. Our custom box implemented in Nao essentially alters a known variable name, which our python listener picks up and runs its operation, before switching back to a ready mode. We plan to identify which stages of the interaction takes the most amount of time, and also which one do people learn to make the fastest improvements across the three steps. Additionally, we conducted a survey to collect on a likert scale different dimensions of the interaction. The survey intends to capture aspects of the interaction that the user has found troublesome or annoying.

4.1 Population

For our recruitment process, we tried to recruit first from our classmates and then from passerby's of UV's Engineer Way. Due to corrupted data or incomplete interaction scenarios, we had to discard one female graduate and student and one male graduate students' timing data. Table 1 summarizes the final population of our study:

Figure 2: Images of our experimental Setup. On the left, the physical layout of the experiment. On the right, an example image of the nao's view of food. (Note that we switched the orientation of the robot in the right image to reduce glare.)

Table 1: Participant Population

Participant ID	Gender	Distinction	Class	Timing Data	Survey Data
1	Male	In HRI Class	Graduate Student	Yes	Yes
2	Male	In HRI Class	Graduate Student	Yes	Yes
3	Male	In HRI Class	Graduate Student	Yes	Yes
4	Male	Stranger	Undergraduate Student	Yes	Yes
5	Male	Stranger	Undergraduate Student	Yes	Yes
6	Female	Stranger	Undergraduate Student	Yes	Yes
7	Female	In HRI Class	Graduate Student	No	Yes
8	Male	In HRI Class	Graduate Student	No	Yes

We could not remove participant 7 and 8's data from the survey data, because they were completed anonymously. As such, we will report the aggregate but then reflect on the user's individual comments in the discussion sections.

5 Results

The results section is divided into the qualitative section and the quantitative section. Our group prioritized for the acquisition of non-class-related users over the demonstration of statistical competence, since our sample size is small and potentially anecdotal evidence might bring more insight for the reader from this interaction.

5.1 Quantitative

Based on our event logger code, we we were able to generate a timeline of the user interaction between every different event of the interaction. Figure 3, Participant 4 was removed from this graph as an outlier, but similar to participant 3 spent the longest amount of time during second interaction. Generally, the amount of time between trial 1 and 2 is reduced significantly, while the amount of time changes between 2 and 3 are about less drastic.

Figure 4 below shows the timeline for participant 6. The timeline indicates the existence of the

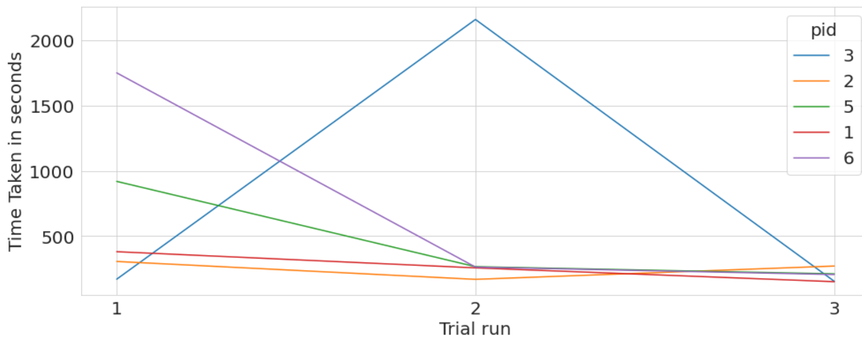

Figure 3: Learning Effect Graph

learning effect in our testing, as the time to interact with the Nao specifically for the food showing section has been reduced over the runs.

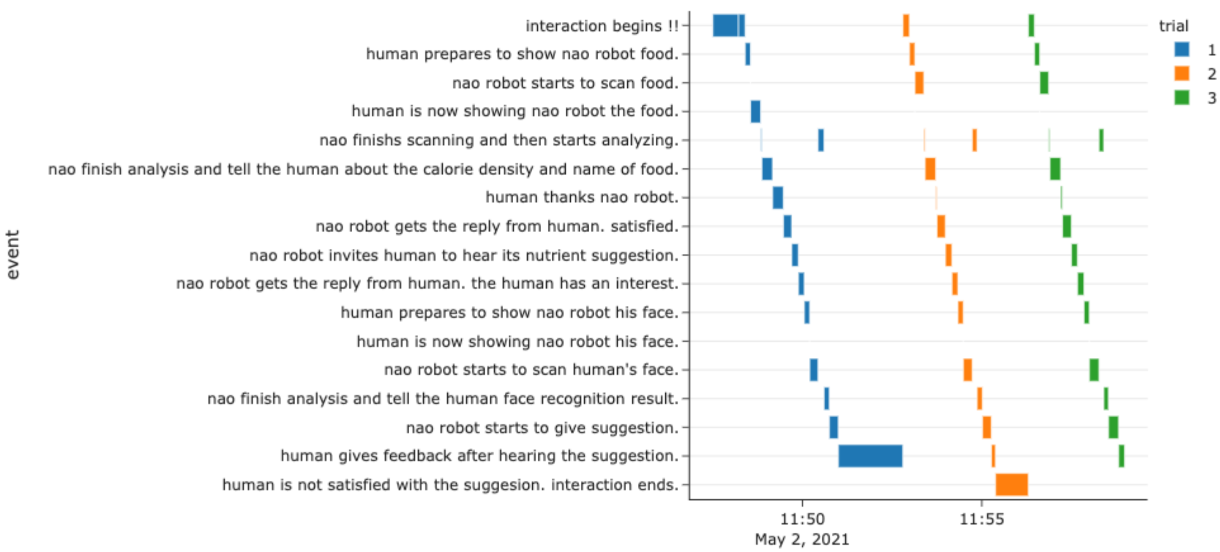

Figure 4: Participant 6 Timeline

Interestingly, there is an additionally learning effect that is not that of the participant getting faster at accomplishing the task, but rather the tester getting better at resetting the experiment. This is shown in the reduced delay between the end of one experiment to the start of the other. One can observe this general trend in the Appendix, where the timelines for the rest of the participants are recorded.

5.2 Qualitative

Figure 5 below shows the aggregated response of all 8 of our participants. The aggregate demonstrates that no one is really happy with the age estimates of our classifier (every one was rated at least an average of about 8 years older), and surprisingly users find the calorie estimate more

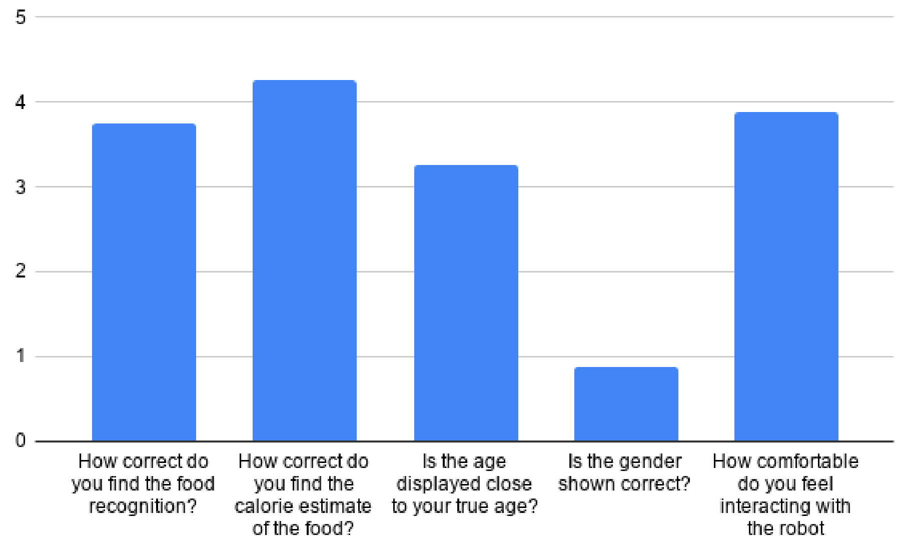

Figure 5: Survey Results Aggregate (note that the gender question is a binary one between yes (1), and no (0)

accurate than the food estimate, since the two are supposed to be related.

6 Discussion

Interesting to note, while participants are very quick to understand that an *apple* is not a *ball*, they are less good at judging what a calorie count actually means. While taking the survey, most of the participants noted that they didn't really know what the calories mean. Perhaps unsurprisingly, mislabelling a participants gender has a tremendous negative effect on the user's overall mood, even if they might have found it interesting at the time of the facial recognition. The same issues comes from misrepresenting a user's age. Our study has shown us that having the robot's facial recognition systematically under detect the user's age might lead to a more amicable interaction. Additionally, not including gender recognition as part of the interaction is also advised, as there is a significant backlash to the attitude of the interaction if the gender is misattributed. Our study has also shown that it is not the cloud computing that takes the most amount of computing time, but rather the voice interactions mental model misalignments that take the most amount of time. Allowing the voice listener to interrupt the current routine and go to the next step if the user has recognized the patter and no longer wants to hear an explanation can reduce the annoyance a user might feel over the interaction, in addition to giving feedback about not understanding the words.

7 Future work

Many of the participant points out the clunky nature of the communication, while commenting on how the video interactions were relatively smooth. This fact is probably due to their being an interface that gives feedback to the user about the internal model of the robot. This suggests that if the robot's interpretation of the user's voice is let know to the participant, that might improve the speed of the interaction.

References

[1] Xinke Li et al. "Application of Intelligent Recommendation Techniques for Consumers' Food Choices in Restaurants". In: *Frontiers in Psychiatry* 9 (Sept. 2018), p. 415. DOI: 10.3389/fpsyt.2018.00415.

[2] Melissa A Matthews and Tanya M Horacek. "Vending machine assessment methodology. A systematic review". In: *Appetite* 90 (2015), pp. 176–186.

[3] Karan Kaushal et al. "The E-Restaurant". In: *2016 Ninth International Conference on Contemporary Computing (IC3)*. 2016, pp. 1–5. DOI: 10.1109/IC3.2016.7880205.

[4] Urška Rozman et al. "Sweet, Fat and Salty: Snacks in Vending Machines in Health and Social Care Institutions in Slovenia". In: *International Journal of Environmental Research and Public Health* 17.19 (2020), p. 7059.

[5] Colin E Champ, Nick A Iarrobino, and Christopher P Haskins. "Hospitals lead by poor example: An assessment of snacks, soda, and junk food availability in veterans affairs hospitals". In: *Nutrition* 60 (2019), pp. 70–73.

[6] Mary Kay Fox et al. "Availability and consumption of competitive foods in US public schools". In: *Journal of the American Dietetic Association* 109.2 (2009), S57–S66.

[7] Yanchao Liang and Jianhua Li. "Computer vision-based food calorie estimation: dataset, method, and experiment". In: *Arxiv* (2017), pp. 1–7.

[8] Quin Thames et al. "Nutrition5k: Towards Automatic Nutritional Understanding of Generic FoodQuin". In: *Arxiv* (2021), pp. 1–12.

[9] Sun Weiran. "Real-time Age, Gender and Emotion Prediction from Webcam with Keras and OpenCV". In: 2020.

[10] Sophia V. Hua et al. "A Qualitative Study of Parents With Children 6 to 12 Years Old: Use of Restaurant Calorie Labels to Inform the Development of a Messaging Campaign". In: *Journal of the Academy of Nutrition and Dietetics* 120.11 (2020), 1884–1892.e4. ISSN: 2212-2672. DOI: https://doi.org/10.1016/j.jand.2020.05.018. URL: https://www.sciencedirect.com/science/article/pii/S2212267220305438.

[11] Joo-Chang Kim and Kyungyong Chung. "Knowledge-based hybrid decision model using neural network for nutrition management". In: *Information Technology and Management* (2019), pp. 1–11.

A Appendix

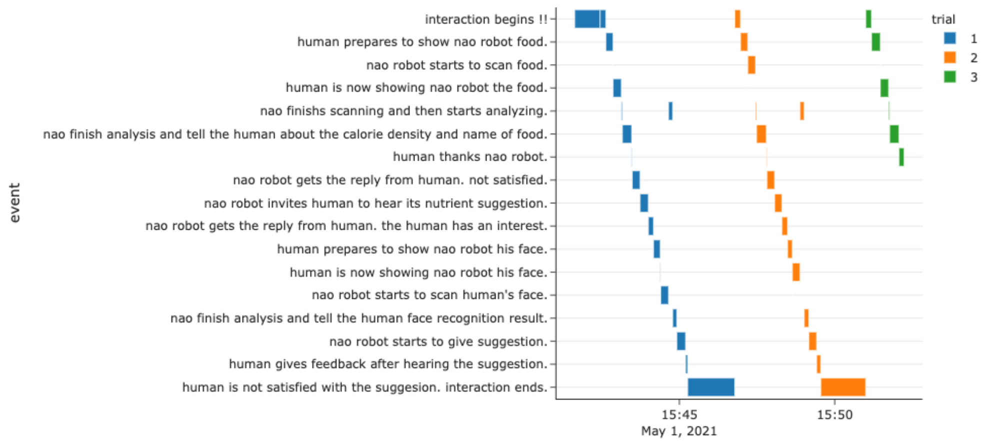

Figure 6: Participant 1 Timeline

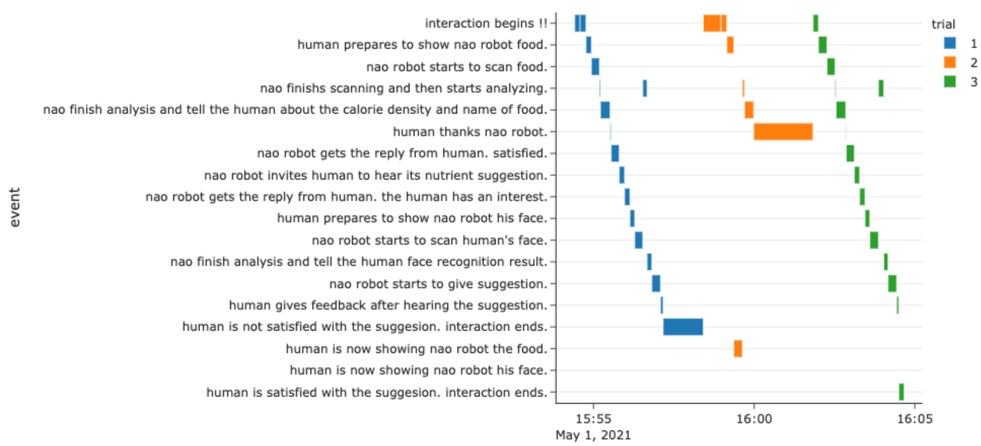

Figure 7: Participant 2 Timeline

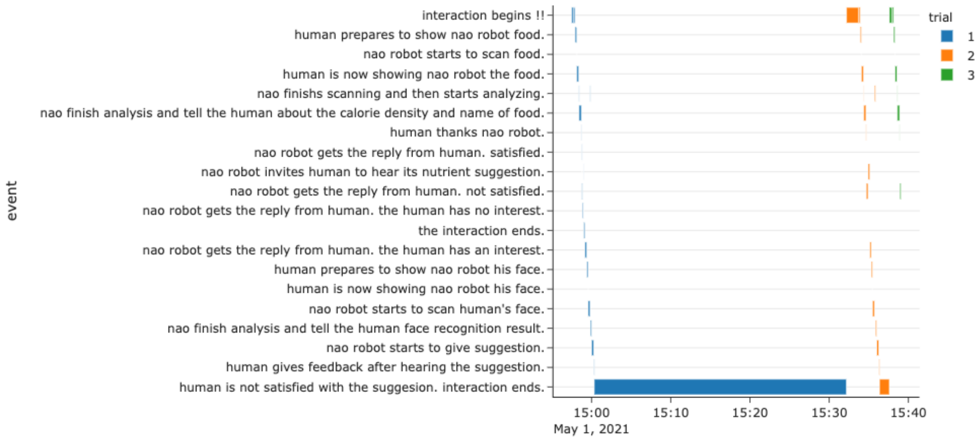

Figure 8: Participant 3 Timeline

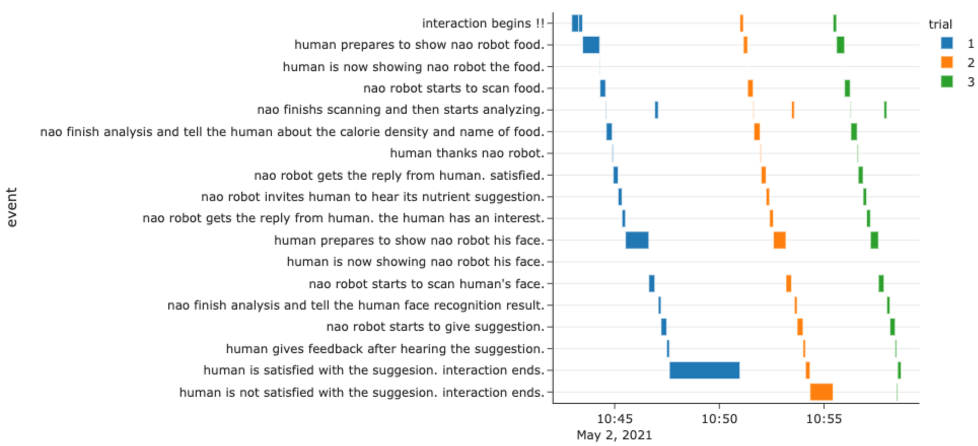

Figure 9: Participant 4 Timeline

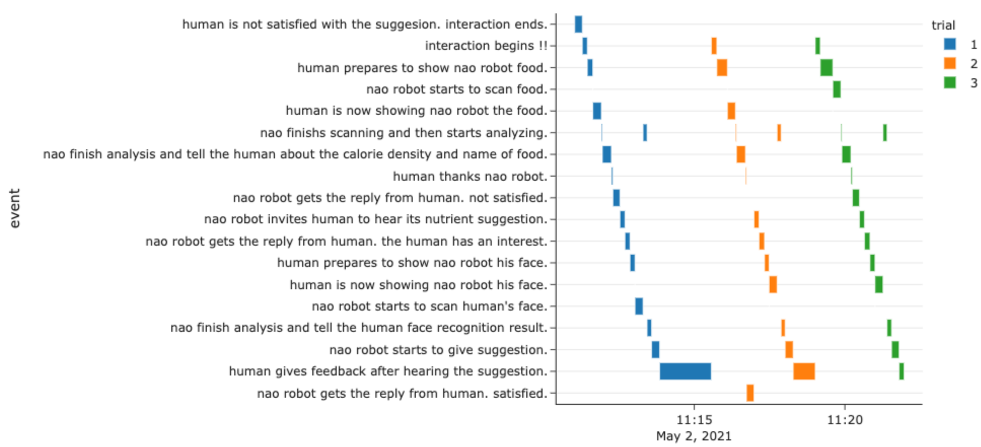

Figure 10: Participant 5 Timeline