2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 1-5, 2023. Detroit, USA

# Measuring People's Boredom and Indifference to the Robot's Explanation in a Museum Scenario

Rei Nagaya Stela H. Seo Takayuki Kanda

*Abstract*—**To personalize the robot guide experience, the robot needs to detect a person's indifference and adjust its explanation toward the person's interest in topics. However, detecting the person's indifference is challenging in a museum, as we cannot use a bulky wearable or facial expression recognition due to unexpected light condition or standing position. We propose to observe people's behaviors and movements on detecting people's indifference. To prove its feasibility, we invited 11 participants to our in-lab museum-like environment. Our robot explains exhibits while videorecording the interaction. Then, we asked participants to watch the recordings and report when they felt bored or indifferent to the explanation. We labelled their movement and matched them to their report so that we know which behaviors and movements hint the person's indifference. We used the decision tree and random forest methods to understand the common pattern when people are indifferent during the explanation in a museum scenario. From our observation experiment, we found that if the listener nods their heads many times or looks at the exhibit for a long time, they are likely interested in the topic, fewer overall movements or looking elsewhere hint that the listener may be indifferent, and if the explanation goes longer than three minutes, the listener would be likely bored.**

## I. Introduction

Various kinds of robots have emerged in our daily life. One is a guiding robot to provide the information while moving with a user [1] or positioning as a kiosk-type [2]. Guide robots can provide three major benefits: filling the reducing number of human employees, providing flexible time schedule, and personalizing their guide for each visitor (Figure 1).

For personalizing a visitor's experience, one challenge for a robot is to understand the visitor's attention level. For example, expert museum guides catch subtle changes of listeners' reactions, understand listeners' interest and indifferent in various topics, and adjust their storytelling and explanation to provide much satisfying tour experience. However, while a robot can support human guides and provide novel experience to visitors, they have a difficulty of understanding people's indifferent level. Understanding people's indifference to a robot's explanation is a key factor for a robot to provide personalized, interesting, and satisfying experience to visitors in a museum.

There are various techniques to understand people's emotional states including interest and indifferent levels (e.g., machine-learning based facial expression recognition [3] or engagement detection using electroencephalography [4]).

R. Nagaya is with Social Informatics, Kyoto University, JAPAN. (e-mail: nagaya.rei.62r@st.kyoto-u.ac.jp).

S. H. Seo is with Social Informatics, Kyoto University, JAPAN. (corresponding author, e-mail: stela.seo@i.kyoto-u.ac.jp).

T. Kanda is with Social Informatics, Kyoto University, JAPAN. (e-mail: kanda@i.kyoto-u.ac.jp).

Figure 1. A guide robot in a museum provides personalized explanations to a visitor. This can be achieved by understanding the visitor's interest or indifference in real-time.

However, they are not always accessible or feasible, especially in our goal scenario – museum guide – as the visitor would move around and the light condition is not guaranteed (i.e., depending on the exhibition, the light condition varies). As such, understanding people's interest and indifferent level becomes challenging in a museum scenario. To overcome this difficulty and the limitation in a museum setting, we propose to observe and utilize people's behavior and movement data.

We conducted a data collection experiment in an in-lab museum environment and analyzed the collected data to understand which behavior or movement is closely linked to people's indifference to the robot's explanation. In summary, if they are nodding many times or looking at the exhibit for a long time, then they are likely interested in the explanation. If a visitor makes fewer overall physical movements or looks at somewhere else other than the robot or the exhibit, then they likely feel bored or indifferent. Unless the explanation is adapted (i.e., personalized), the listener will likely feel bored if an explanation goes longer than three minutes.

Our findings indicate the possibilities of detecting people's indifference to robot explanations based on their behaviors and movements in a museum. This can help the development of future robot guides adjusting their explanations to each visitor by detecting the visitor's interest and indifferent level to robot explanations in real-time. As such, museum visitors would enjoy personalized and satisfying experience with guide robots.

## II. Related Work

Personalization in the museum is suggested and applied by previous pioneers and field experts especially with human museum guides. The modern technology allows advanced personalization and recommendation (e.g., by storing people's profiles) to provide engaging and satisfying experience to users in

2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) | 978-1-6654-9190-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/IROS55552.2023.10341996

various systems [5]–[7]. In museum scenarios, many work have been done for virtual tour experience instead of in-person experience. We believe that similar personalization can be done by utilizing a robot's capabilities. In addition, the use of robots is promising as robots can provide better experience when we compare it to the traditional audio guide [8], [9]. Thus, we propose to use the robot to provide personalized and satisfying experience to museum visitors by bringing the personalization techniques suggested for virtual museum tours.

Main characteristics of the museum guide robots are identifying and greeting a visitor, presenting exhibits, and expressing farewell [10]. Many researchers pointed various techniques for successful social human-robot interactions [9]–[11]. In addition, researchers explored people's perception on the guide robots with personalized explanations [5] and how to achieve social navigations for guiding in a museum [1]. While the techniques and research lead us to believe the deployment of advanced museum guide robots even in tomorrow, there is one important but missing factor for the guide robot: detecting a person's interest or indifference. For a museum guide robot to provide personalized explanations to a visitor, it needs to understand the person's interest or indifference to the exhibit topics or explanations.

Understanding a person is an important topic in various field; one we are specifically interested in is a person's engagement level. As it is closely related to human's brain, electroencephalography (EEG) could be a good option to collect and sense the relevant data [4], [12]. However, the downside of this technique is the device. That is, the device is often bulky and hard to move around while wearing it. Another option is to use cameras to detect the person's facial expressions, as they would show the person's current feeling [3]. However, as the light conditions may change for different exhibits, in a museum, RGB-camera is not reliable. Depth cameras would not be impacted much; however, depending on their installation locations and a visitor's standing pose or orientation, it may not capture the visitor's face well enough. For example, if the camera is mounted on a robot, people behind of others would not be visible to the robot. If it is mounted up on ceiling, it would not be able to capture people's face clearly. Thus, instead of focusing on people's facial expressions, we looked at the visitor's body movement which can be reliably detected by ceiling mounted depth cameras.

There is a body of work on detecting people and their kinematics using multiple depth cameras [13], [14]. Further researchers explored possibilities of detecting people's emotional states from their pose data [15]. However, major challenge for the museum scenario is that the person may not make a large movement (other than moving away from the exhibits). Therefore, collectively, in this work, we explore the possibilities of detecting people's indifferent to museum guide robot's explanations using people's behaviors and movements which can be detected by ceiling-mounted depth cameras.

## III. Data Collection

To understand how a person's movement is correlated with their indifference to robot's explanations, we collected people's behavior and movement data in a robot-guided museum scenario. We prepared three different exhibits in a room and spaced them apart so that a person has enough room to move away when they feel bored (Figure 2 and Figure 3). Generally speaking, many museums have tangible items, drawings, and pictures. As such, we prepared two Buddhist statues and one picture of Buddhist statue to increase the feeling of being in a museum. In addition, we expect many people would lose their interest quickly enough while listening to Buddhist history.

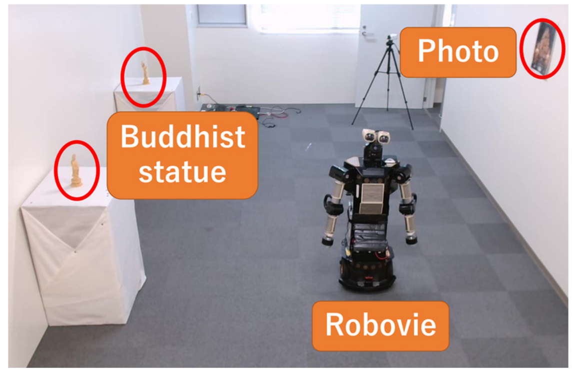


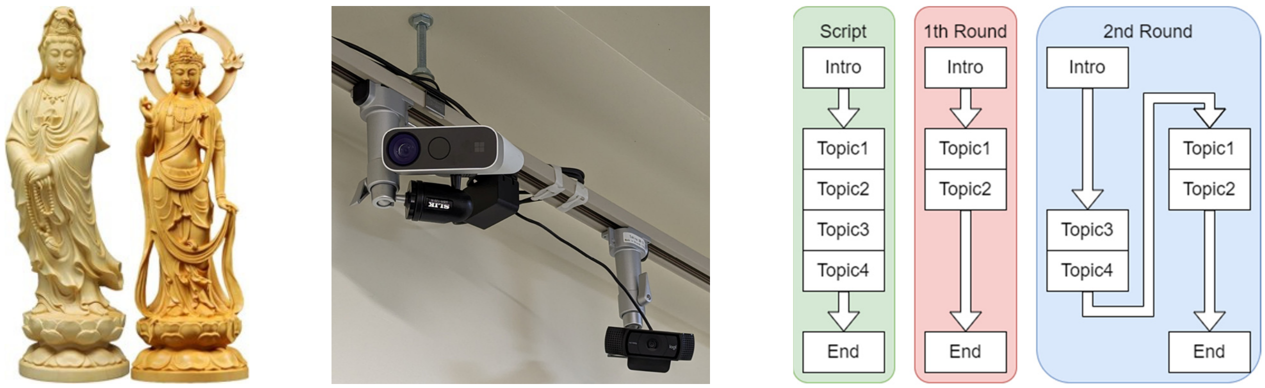

Figure 2. The data collection experiment setup. There are three exhibits, two Buddhist statues and one picture of a Buddhist statue. The robot moves to an exhibit, looks at the visitor, and explains the exhibit.

To increase the intractability with a visitor and simulate actual museum guide robot, we used a teen-sized humanoid robot, Robovie. The robot can navigate the room while avoiding obstacles (including the visitor), make gestures, and look at the visitor while synthesizing speeches with a kid-like voice. To track people and their movement, we installed multiple ceiling-mounted depth cameras and grid them in a network (a.k.a., sensor network). The sensor network can provide people's location, orientation, and pose data in real-time.

We let a participant enter the room and follow the robot's instruction which starts with a greeting. Then, the robot explains an exhibit (an explanation about an exhibit consists of multiple topics, and it is about 10-minute long as we expect most people lose their interest in Buddhist history in less than 10 minutes). We instruct the participant to leave or move away anytime when they feel bored or indifferent. Then, the robot starts its explanation of next exhibit. This continues until the robot finishes its explanations of all three exhibits (a session). After a short break in between sessions, we repeat the session in total of three times. Hence, for each participant, we have three sessions with three exhibits (as long as the time allows).

We carefully crafted our scripts for a robot to provide its explanation on each exhibit without any command phrases such as "look here" as this would impact people's behaviors. In addition, to prevent the participant from listening to the exact same explanations, a robot continues its explanation where it left off for the next session (Figure 3). This would decrease the impact of getting indifferent by listening to the same explanation in next sessions.

While a participant is in the room, we captured their movement and videorecorded the interactions. Soon after all sessions, we watched the scene together with the participant and asked them to tell us at which point they felt bored or indifferent. Since the indifference is a person's internal state, it is difficult to measure accurately other than asking the person in-situ. That is, only the person would know the ground truth. Thus, we decide to ask participants to express their feelings while watching the video together.




A. Two Buddha statues

B. A depth camera from the collection of sensors

C. The robot's explanation logic. In this example, the robot continues from topic 3 in the second session

Figure 3. For data collection, we prepared three Buddha related items (two physical ones and one picture), ten ceiling-mounted depth cameras, and explanation scripts with in-house software to smoothly continue the robot's explanation from where it left off.

We recruited 11 participants from the general public (20 ~ 50 years old, 2 males and 9 females). We gave honorariums of 3000 JPY to them for their 2-hour participation. This study is approved by the institution's ethics board. From the data collection sessions, we collected 95 discrete datasets. Since there were 11 people, 3 sessions, and 3 exhibits, the total datasets should have been counted 99; however, there were three datasets (one session) missing. For one participant, we could only get one partial and two complete datasets due to a technical issue, and we did not have enough time to redo a session.

## IV. Extracting Attributes

While boredom or indifference are rather continuous state (e.g., a bit boring, really boring, etc.), as an initial step of our exploration, we asked participant to report their indifference in a binary state. That is, the participant is going to tell us when they felt indifferent (i.e., indifference to which topic), and we assume that they did not feel indifferent for the other topics.

For the participants' behaviors and movements, one researcher labelled all of them (e.g., gestures, nodding, moving away) by hand while watching all videorecords. We matched these labels to participants' reports on their indifference to topics (i.e., topic classification). Participants did not speak during the robot's explanations. Note that we labelled behaviors and movements, as our scenario assumes that participants' facial expression is not recognizable due to various light settings and their masks. Additionally, while we believe that these can be detected automatically, we labelled them manually as this work is not about automating human behavior detections.

Out of all 546 topics, we found 188 topics marked as negative (the non-marked are positive). From the researcher's labelling, we got 6170 behavior and movement labels in total. We explain movement types below.

### A. Head Orientation

We think that people's head orientation has rich information especially in relation to their interest or indifference.

| | |
|---|---|
| turn to the robot | the listener turns their head to the robot from somewhere else |
| turn to the exhibit | the listener turns their head to the exhibit from somewhere else |
| look up | the listener looks up including slightly oblique orientation |
| look away | the listener turns their head to somewhere else other than the robot and the exhibit while not looking up |

### B. Head Movement

In addition to the head orientation, we think people's head movement provides some hints on their interest level.

| | |
|---|---|
| nod deeply | the listener nods their head deeply (make a large movement) |
| small nod | the listener makes a small nod |
| nod once | the listener nods their head once |
| nod many times | the listener nods their head many times |
| tilt their head sideways | the listener tilts their head while facing forward |

### C. Peek at the Exhibit

We observed some peeking actions. We separated the labels based on how big the movement is and peeking direction.

| | |
|---|---|
| peek lightly | the listener gets their head closer to the exhibit |
| take a closer look | the listener gets their head and torso closer to the exhibit |
| tilt their body sideways | the listener tilts their body sideways and peek at the exhibit's side |

### D. Arms or Hands

We observed people's hand- or arm- motions while listening to the robot's explanation.

| | |
|---|---|
| touch their cloth | the listener touches or fixes their cloth |
| touch their head | the listener touches their face or hair |
| play with hands | the listener touches their fingers or hands or plays with their hands |
| cross arms | the listener crosses their arms in front of them |

| | |
|---|---|
| place both hands behind | the listener places their hands behind |

### E. Body

We observed people's body movements without moving from one place to another. As we think they are an important hint for the person's indifference, we labelled them as well.

| | |
|---|---|
| move their legs | the listener moves their legs or crosses them without moving from the position |
| swing their body | the listener lightly swings their torso side-ways |
| twist their body | the listener lightly twists their torso |

### F. Locomotion

We labelled both moving during the explanation and leaving after the explanation. The leaving denotes the end of a round for the exhibit, we did not use this label for identifying indifference, but used as the end mark.

| | |
|---|---|
| move closer | the listener moves closer toward the exhibit |
| change position | the listener changes their standing position without changing their distance to the exhibit (e.g., small sidewalks) |
| move away | the listener moves away from the exhibit while facing it |
| leave | the listener faces elsewhere and moves away from the robot and the explaining exhibit |

### G. Time-based Rate

We normalized the topic length so that the timestamp. Then, we prepared five more attributes based on the time.

| | |
|---|---|
| elapsed time | the elapsed time since the robot starts explaining the exhibit |
| rate for looking at the robot | how long the listener looked at the robot within the topic |
| rate for looking at the exhibit | how long the listener looked at the exhibit within the topic |
| rate for looking up | how long the listener looked up within the topic |
| rate for looking elsewhere | how long the listener looked elsewhere within the topic |

The sum of the rates of looking somewhere is 100 percent. For example, if the listener looked at the robot for 75 percent of time and at the exhibit for 25%, then the rate for looking up and elsewhere would be zero.

## V. Analysis of the Interaction Data

Since there were more numbers of positive labels than negative labels, we adjusted the class weight based on the difference. We calculated the weight by

$$W = \frac{NTotalLabels}{2 * NClassLabels}$$

where *NTotalLabels* is the total number of labels and *NClassLabels* is either the number of positive or negative labels. That is the weight of the negative class is higher than the weight of the positive class.

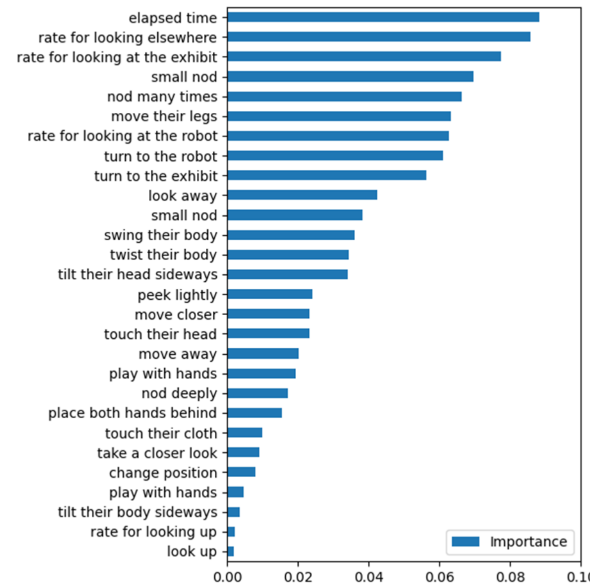



Figure 4. The feature importance from the random forest.

### A. Cross Validation

We set each participant data to be the test set and evaluate our model by cross validation (i.e., leave-one-out cross-validation). As a result, the accuracy is 68.2%, the recall is 0.413, the specificity is 0.817, the F-measure is 0.463, respectively. We use model based only on distribution of positive or negative label as baseline model because we want to get knowledge about action each feature on identification. Accuracy by the baseline model is 54%, so our model is 14.2% higher.

### B. Evaluation of the Feature Importance

To understand the importance of each feature (i.e., each label), we feed our data to the random forest method with Gini importance (Figure 4). We can visually understand how much the feature contributes to splitting the samples and is calculated by amount of lowered Gini impurity by the feature. From the results, we can see that the elapsed time is the most important attribute followed by how long the listener looks at the exhibit or elsewhere. Figure 5 shows partial dependence plot (PDP) of *elapsed time*. This express marginal effect of *elapsed time* to output of our model. In this case, the higher partial dependence gets, the higher the probability of negative label is. From this, we can see that our participants' indifference changes over time. Especially, we can see that their feeling of indifference increases until about 3-minute mark, but decreases afterward (i.e., the degree of interesting is proportional to *elapsed time*). We believe that this is because participants were able to get away from the explanation in this experiment. From our records, we noticed that they stayed and listened to the guide for at least 3 minutes even if they feel indifferent.

Among the listener's movement labels, their head movements (i.e., *small nod*, *turn to the robot*, *turn to the exhibit*) are considered as important movements than others. Nodding could be considered as positive feedback in human-human interaction. Additionally, since both of turning face and the time rate of looking somewhere are important, we should focus on

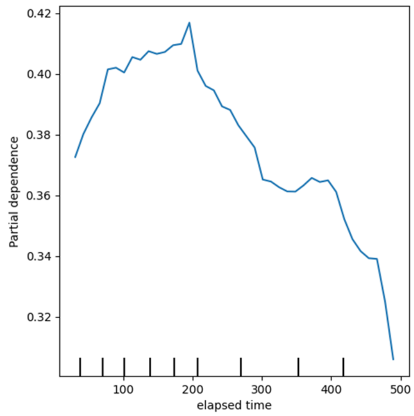



Figure 5. Partial dependence plot of *elapsed time*. Partial dependence increases up to 200 seconds and decrease since then.

head reaction much more. On the other hand, body and hand movements do not have much influence in the model except moving their legs. It shows the general knowledge which listeners stir when they are not interested can only be noticeable with their legs.

### C. *Decision Tree*

After confirming the baseline accuracy from the random forest with our data, we created a decision tree with all our data. In this process, to prevent the tree from being meaninglessly repetitive, we set the maximum depth to be 10 and used the cost-complexity pruning method. This method uses a parameter α to how determine the size of tree. The alpha decides the penalty of having a big tree (i.e., the bigger alpha means the higher penalty, thus the method prunes more branches).

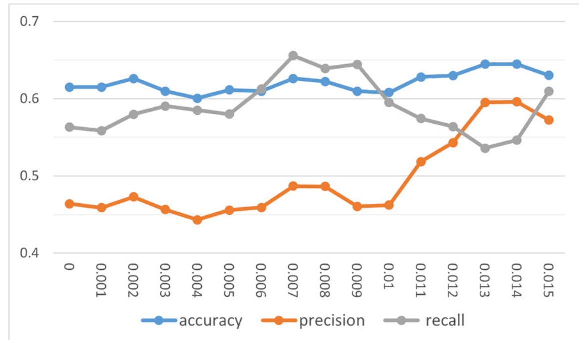



Figure 6. The result decision tree's accuracy, precision, and recall for each α from 0 to 0.015. X-axis is α value while Y-axis is the accuracy in the scale of 0 to 1 inclusive.

We explored the tree's accuracy, precision, and recall rate for each α value to determine appropriate α (Figure 6). The precision increases when α is higher than 0.011. However, this is because we prune too many branches and finally tree has only one node. When α is 0.007, it shows the highest recall rate. Thus, we think α = 0.007 is the appropriate parameter. The final decision tree can be seen in Figure 7.

From the decision tree, we can say that features of head (*rate for looking elsewhere*, *small nod*, *look away*) are mainly focused on identification as before. There is *peek lightly* in second node, but its importance is not high as it only splits 9 samples from 87 samples. *Elapsed time* is not used in the high-level branch even if their importance is high in the random forest analysis (Figure 4). This means *elapsed time* acts as an auxiliary in our decision tree.

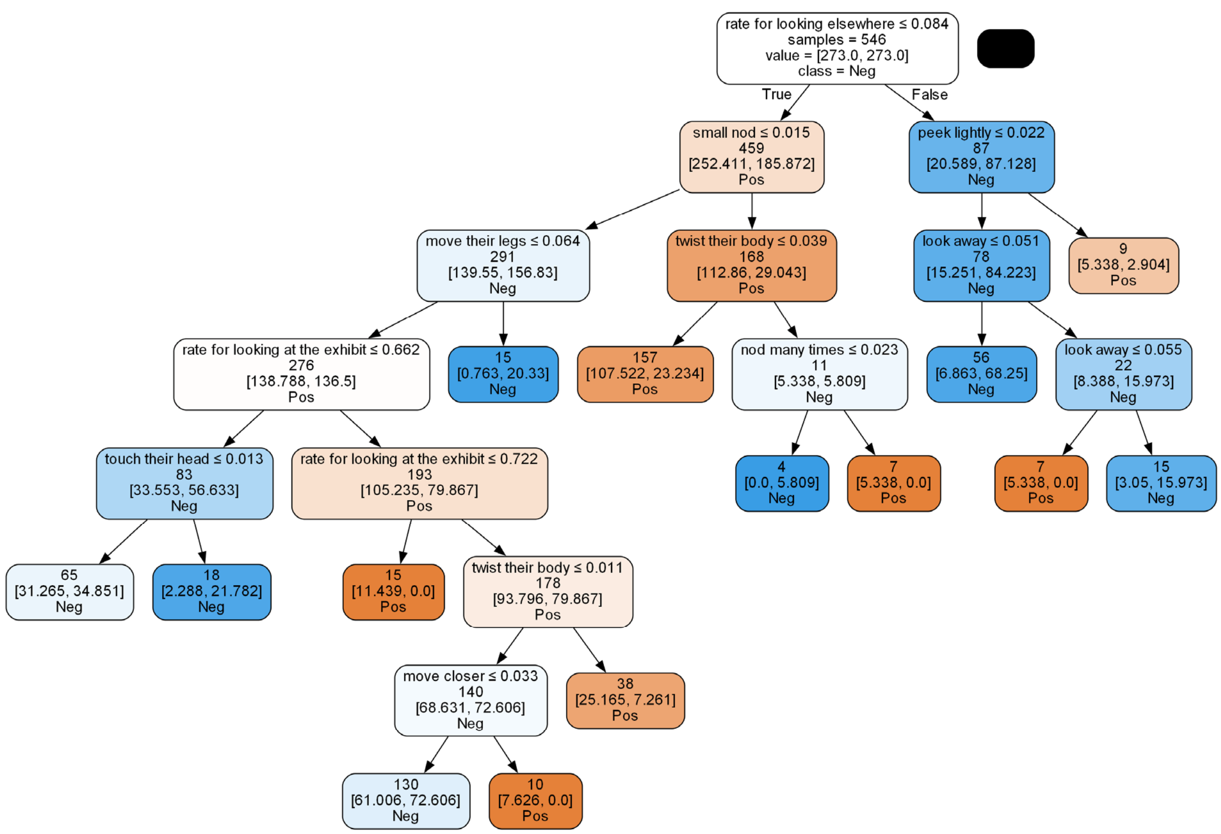



Figure 7. The final decision tree with the cost-complexity pruning parameter α=0.007.

Looking at the nodes after second level, we can see the most of them lead to negative when there is a fewer number of features. We can also see that a listener makes fewer overall physical movements when they are bored from the features except *elapsed time* and *rate for X* which is a frequency. Further, by observing each node, we can assume that a person likely felt indifferent when they look elsewhere for 10% or longer time for a topic. It is the same for few *small nods*. *Move their legs* splits just a few samples but strong tendency to negative appears when that frequency is higher than 0.064.

## VI. Discussion and Limitations

One benefit of using the decision tree and random forest methods is to see the feature importance and understand the logic by following the nodes. From there, while we assumed that the participant's head orientation may be one key aspect of detecting their indifference, we found that the rate (how long they spent time doing it) is important in the random forest. This hints that the simplest model could be detecting the person's look direction. That is, by knowing the person's look direction, we may be able to achieve some level of indifference detections. However, we are not sure how this would work compared to our decision tree. We need to confirm our decision tree's performance and explore other detection methods.

Another key point is the person's head movement. In general, more than their whole-body movements or hand movements, head movements are noticeably important. For example, if the person nods many times, it can be a hint that the person is interested (rather, the person does not feel indifferent). However, there is one caveat before making a conclusion from our findings – culture.

Our experiment was conducted in Japan. There is a stereotype that many Japanese nods their head to signify the speaker that they are listening to the speaker. In addition, we found that they make a fewer overall physical movements as they get bored. We are not sure if this also has a cultural influence. We need further investigation in other cultures.

In this work, we observed one participant's interaction with one robot. As a group, people's social dynamics changes, and this may influence their behaviors and movements around the robot explaining the exhibit. As such, this work's findings are limited, and we note that further investigation is required.

## VII. Conclusion

In this paper, we argued that museum robots should be able to make a personalized guide by detecting visitors' interest and indifference. To achieve this goal, they must be able to detect in which topic the visitor feels indifferent and adjust their ongoing explanation accordingly. We described in-lab experiment to collect people's behavior and movement data in a museum-like environment. We analyzed the data, presented the results, and discussed our thoughts and limitations.

In summary, we found several behaviors and movements are likely linked to a person's interest and indifference to the robot's explanation.

1. The more time a listener spends looking at the exhibit or the robot, the more likely they are interested in the explanations (i.e., a listener gazes away from the exhibit and the robot when they feel indifferent).
2. A listener makes fewer overall physical movements when they are bored; if they are interested in, they may make many small nods.
3. A listener stays for about three minutes before leaving even if the explanation is not interesting.

This shows that we can detect a person's indifferent level from simple human behaviors and movements. In near future, we would like to develop a personalization module by detecting a person's indifference and tuning ongoing explanations in real-time in museum scenarios to increase the visitors' engagement and satisfaction.

## Acknowledgment

This work was supported by JST Moonshot R and D under Grant Number JPMJMS2011, Japan.

## References

[1] K. Ichihara, T. Hasegawa, S. Yuta, H. Ichikawa, and Y. Naruse, "Waypoint-Based Human-Tracking Navigation for Museum Guide Robot," *J of Robotics and Mechatronics*, Oct. 2022.
[2] X. Z. Tan, E. J. Carter, S. Reig, and A. Steinfeld, "Go That Way: Exploring Supplementary Physical Movements by a Stationary Robot When Providing Navigation Instructions," in *ACM SIGACCESS Conf on Computers and Accessibility*, Oct. 2019.
[3] Samadiani *et al.*, "A Review on Automatic Facial Expression Recognition Systems Assisted by Multimodal Sensor Data," *Sensors*. 2019.
[4] M. Hassib, S. Schneegass, P. Eiglsperger, N. Henze, A. Schmidt, and F. Alt, "EngageMeter: A system for implicit audience engagement sensing using electroencephalography," in *Conf on Human Factors in Computing Systems*, May 2017.
[5] L. Aroyo, R. Brussee, L. Rutledge, P. Gorgels, N. Stash, and Y. Wang, "Personalized Museum Experience: The Rijksmuseum Use Case," in *Museums and the Web 2007*, 2007. Accessed: Feb. 26, 2023. [Online]. Available: https://ir.cwi.nl/pub/12427
[6] D. Kosmopoulos and G. Styliaras, "A survey on developing personalized content services in museums," *Pervasive Mob Comput*, Jul. 2018.
[7] M. A. A. Khan, V. Dziubak, and A. Bunt, "Exploring Personalized Command Recommendations based on Information Found in Web Documentation," in *Int Conf on Intelligent User Interfaces*. 2015.
[8] A.-M. Velentza, D. Heinke, and J. Wyatt, "Museum robot guides or conventional audio guides? An experimental study," *Advanced Robotics*. Dec. 2020.
[9] K. Yamazaki *et al.*, "Revealing Gauguin: engaging visitors in robot guide's explanation in an art museum," in *SIGCHI Conf on Human Factors in Computing Systems*, Apr. 2009.
[10] N. Gasteiger, M. Hellou, and H. S. Ahn, "Deploying social robots in museum settings: A quasi-systematic review exploring purpose and acceptability," *Int J Adv Robot Syst*, Nov. 2021.
[11] M. Hellou, J. Lim, N. Gasteiger, M. Jang, and H. S. Ahn, "Technical Methods for Social Robots in Museum Settings: An Overview of the Literature," *Int J Soc Robot*, Oct. 2022.
[12] Y. Abdelrahman, M. Hassib, M. G. Marquez, M. Funk, and A. Schmidt, "Implicit Engagement Detection for Interactive Museums Using Brain-Computer Interfaces," in *Int Conf on Human-Computer Interaction with Mobile Devices and Services Adjunct*, Aug. 2015.
[13] D. Brscic, T. Kanda, T. Ikeda, and T. Miyashita, "Person Tracking in Large Public Spaces Using 3-D Range Sensors," *IEEE Trans Hum Mach Syst*, Nov. 2013.
[14] D. Bršćić, R. W. Evans, M. Rehm, and T. Kanda, "Using a Rotating 3D LiDAR on a Mobile Robot for Estimation of Person's Body Angle and Gender," *Sensors*, Jul. 2020.
[15] F. Ahmed, A. S. M. H. Bari, and M. L. Gavrilova, "Emotion Recognition From Body Movement," *IEEE Access*, 2020.