

Article

Face Memorization Using AIM Model for Mobile Robot and Its Application to Name Calling Function

Masahiko Mikawa * , Haolin Chen and Makoto Fujisawa

Faculty of Library, Information and Media Science, University of Tsukuba, 1-2 Kasuga, Tsukuba, Ibaraki 305-8550, Japan; sikon766@gmail.com (H.C.); fujis@slis.tsukuba.ac.jp (M.F.)

* Correspondence: mikawa@slis.tsukuba.ac.jp

Received: 7 October 2020; Accepted: 13 November 2020; Published: 19 November 2020



Abstract: We are developing a social mobile robot that has a name calling function using a face memorization system. It is said that it is an important function for a social robot to call to a person by her/his name, and the name calling can make a friendly impression of the robot on her/him. Our face memorization system has the following features: (1) When the robot detects a stranger, it stores her/his face images and name after getting her/his permission. (2) The robot can call to a person whose face it has memorized by her/his name. (3) The robot system has a sleep–wake function, and a face classifier is re-trained in a REM sleep state, or execution frequencies of information processes are reduced when it has nothing to do, for example, when there is no person around the robot. In this paper, we confirmed the performance of these functions and conducted an experiment to evaluate the impression of the name calling function with research participants. The experimental results revealed the validity and effectiveness of the proposed face memorization system.

Keywords: social human–robot interaction; face memorization; face classification; user acceptance; AIM model; sleep function

1. Introduction

In recent years, social robots used in public spaces have been studied and developed actively. We are studying an information service robot with a face in a snowy cold region, as shown in Figure 1. When the robot wandered among pedestrians on a sidewalk, many children were glad to meet it. However, although the robot has a face with eyes and an exterior body made with soft material like a snowman, some adults obviously avoided the robot, and it seemed that the robot made them feel uncomfortable.

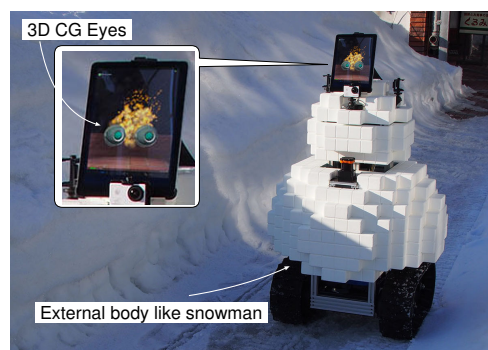


Figure 1. Information service robot on a sidewalk.

Thus, there have been several studies to investigate people’s impressions of a robot. Bartneck et al. investigated how robot design influences people [1]. Several emotion models for social robots and

robot eye-gazing methods for human–robot interactions have been reviewed in [2,3] respectively. Influences of robot head movements were investigated in [4–6]. Huang et al. proposed the friendly behavior design method for interactive robots based on a combination of response time, approach speed, distance and attentiveness [7]. Shulte et al. developed a mobile robot with the mechanical face that could express emotion using face expressions and voice [8]. It was shown that speaking was effective at attracting people through experiments in [9].

Kanda et al. or Ishiguro et al. conducted human–robot interaction experiments in several environments: an elementary school, an office and a shopping mall [10–13]. The mobile robots used in these experiments could call a person’s name by using information embedded in the RFID tag, and it was described in [11] that, “Many children were attracted by the robot’s name-calling behavior.” Chen et al. [14] conducted experiments in which a robot called out to a person by name, and investigated the reaction time, and showed that there was a significant difference between calling out a person by name and not calling out a person by name. Like these studies, name calling is one of the important functions for social robots. However, the studies are still not enough, and there is no experiment that quantitatively evaluates the impression the robot gives to a person by name calling.

We studied a name calling function using a face memorization system. In order to build the face memorization system, face detection, face classification and classifier (re-)training functions, and speech recognizer and text-to-speech functions are required. Various methods for the face detection and the face classification, which are important functions in this study, have been proposed. Broadly classified, there are 2D image-based methods that use only features extracted from an image, and 3D based methods that use depth information obtained from a RGB-D camera or LiDAR in addition to image features. The 3D based methods such as [15–17] are more robust than the 2D image-based methods because they use depth information too. However, there are two problems: one is the cost of the device for measuring the depth information, and the other is the weakness of the device to sunlight and strong light because it uses infrared or laser light to measure the depth information. On the other hand, many 2D image-based methods have been proposed [18,19], and ordinary cameras can be used. Since it is difficult to say which is better at the present time, in this study, we adopted the 2D image-based method, which allows using an inexpensive, ordinary camera, for the robotic system shown in Figure 1.

In addition to the face detection and classification, it is necessary for the social robot to memorize the face and name of a stranger that the robot meets for the first time. However, it takes a long time to (re-)train the face classifier; moreover, the number of calculations is large. On the other hand, the interaction between a person and the robot is not always busy. For example, when there are no people around the robot, the robot has nothing to do.

Our robot has a sleep–wake function based on our proposed mathematical activation-input-modulation (AIM) model. Multiple processes for controlling the robot system run in parallel, and the AIM model transitions between sleep and wake states depending on the stimuli detected by an external sensor, and controls the frequency of execution of these processes. In the wake state, the face detection and classification functions work hard during detecting persons’ faces. When the robot detects a stranger, it stores her/his face images with her/his name after getting her/his permission. When there is no person around the robot, it begins sleeping through the operation of the AIM model. There are two sleep states: rapid eye movement (REM) sleep and non-REM sleep. The face classifier is re-trained by adding newly collected face images in the REM sleep state just like a person dreams, and the execution frequencies of almost all of information processes are reduced in the non-REM sleep state. This means that the operation of the AIM model can decrease battery draining, and as the result, the operating time of the mobile robot can be increased.

In summary, the features of our social robot system are that it has the function to call a person by her/his name and the function to (re-)train a stranger’s face and name during the sleep state controlled by the mathematical AIM model. The face memorization system controlled by the sleep function is shown in Section 2. The experimental system with the face memorization function and

the effects of the function for the mobile robot system are shown in Section 3. Finally, in Section 4, we describe the results of an experiment to evaluate the impression of the name calling function on research participants, and discuss them.

2. Face Memorization System for a Mobile Robot

The basic idea of the face memorization system has been proposed in our previous work [20]. Although there are some overlaps, since there are some modifications and additions, the detailed system is described in this section.

2.1. System Outline

An outline of our proposed face memorization system is shown in Figure 2. First, when a mobile robot detects a person, it identifies whether it is a stranger or not based on a face classifier. When the robot classifies her/him as a stranger, it asks permission for collecting her/his facial images and name. Second, it is necessary to re-train the face classifier by adding newly collected face images and names of strangers, but the computational cost for re-training is high and it takes much time to re-train it. Thus, our proposed system executes the re-training process when there is no person around the robot. Multiple processes are executed in parallel in the system, and their behaviors are controlled by the activation-input-modulation (AIM) model described in the following subsection. Since it is unnecessary to process external sensor data in real-time when there is no person around the robot, the robot sleeps depending on a state of the AIM model. Specifically, the execution frequencies of processes for external sensors decrease, and the re-training process begins to be executed just as though the robot dreams. Last, when the robot detects a person again, it wakes up by the operation of the AIM model. In other words, the face classification process begins to be executed again. When the robot detects a known person, it can call her/him by her/his name.

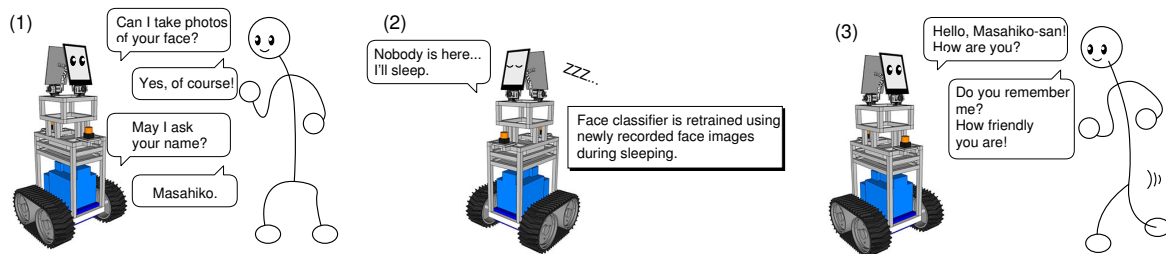


Figure 2. Calling name function based on face memorization.

This system described above is controlled by a finite state machine shown in Figure 3. The system has four states: awoken, relaxed, non-REM sleep and REM sleep. The state transitions are described in detail in the next sub-section. While the robot detects a person, the system is in the waking state. The robot calls the name of her/him it remembers, or collects her/his facial images when she/he is a stranger. After the robot detects no person, the system shifts to the sleep state through the relaxed state. In the REM sleep state, the face classifier is (re-)trained.

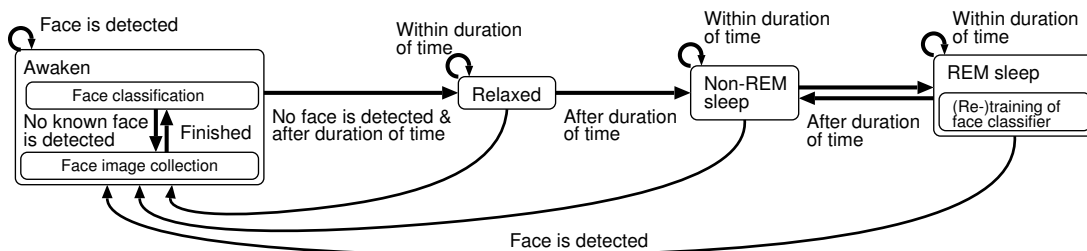


Figure 3. Finite state machine of system.

2.2. Mathematical AIM Model

The AIM state-space model proposed by Hobson [21] is shown in Figure 4. Humans’ consciousness states are expressed depending on levels of the following three elements. Activation controls processing power. Input switches information sources. Modulation switches external and internal information processing modes. In an waking state, external information obtained by external sensory organs is processed actively. A human dreams in rapid eye movement (REM) sleep, and it is said that stored internal information is processed actively. In non-REM sleep, input and output gates are closed and the processing power declines totally.

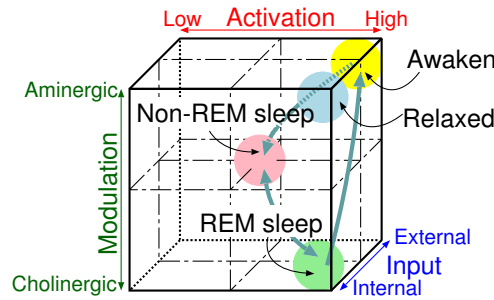


Figure 4. AIM model.

We have designed a mathematical AIM model [22,23] for controlling sensory information processing systems of an intelligent robot based on the AIM state space model. Since the details of the mathematical AIM model have been described in these previous works, only the summary is explained in this paper.

The block diagram of the mathematical AIM model is shown in Figure 5. The mathematical AIM model controls ratios among the external and internal information processors. The element *S* calculates stimuli, such as changes of perceptual information, extracted from sampled data. The element *A* decides an execution frequency of each information processor. The element *I* decides parameters used when stimuli are calculated in *S*. The element *M* decides an execution frequency for each sensory data capture. Each element consists of two sub-elements. The subscripts *ex* and *in* mean that the sub-elements relate to external and internal respectively.

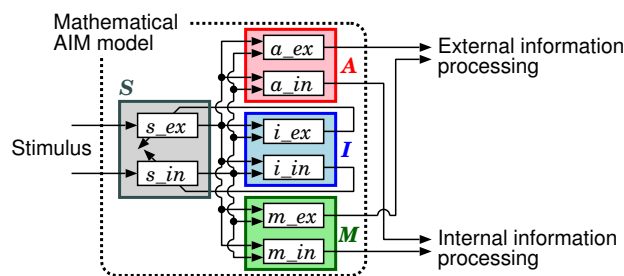


Figure 5. Mathematical AIM model.

An example of the variations of the sub-elements with time is shown in the upper figure of Figure 6. While an external stimulus is detected, the levels of the elements related to the external are higher than those related to the internal. After the external stimulus becomes lower than a threshold, the state is shifted to relaxing. After a certain time, first, the state is shifted to the non-REM sleep. Then, each level increases and decreases periodically. When again a stimulus is detected, the state shifts to the waking state. The lower figure of Figure 6 shows the variations of *A*, *I* and *M* calculated from their sub-elements with time.

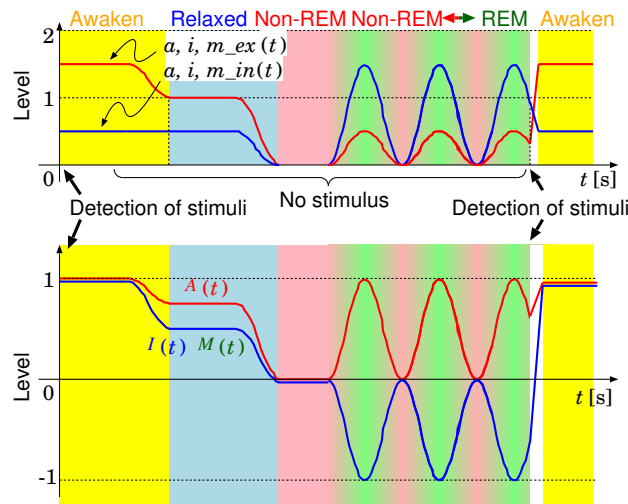


Figure 6. Variations of A , I , M and their sub-elements with time.

2.3. Processes for Face Memorization Function Controlled by the AIM Model

Figure 7 shows relations between the mathematical AIM model and processes for the face memorization function. A microphone is used for collecting persons’ names by voice recognition. A camera is used to detect people’s faces, and collect images of strangers’ faces. When a stranger’s face is detected, the robot asks for her/his name using text-to-speech function and collects her/his face images and name with her/his permission.

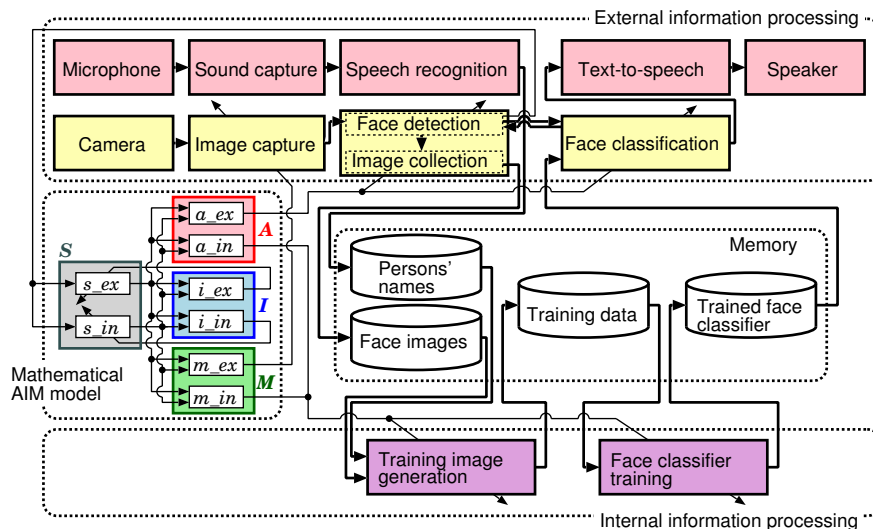


Figure 7. Block diagram of the AIM model and processes for face memorization function.

The face memorization function consists of external information processing and internal information processing. In the external information processing system, image data that require real-time processing are handled. In the internal information processing system, the following kinds of data are treated: When the data are difficult to process in real-time, (taking too much time). When it is not necessary to process in real-time.

The external information processing system consists of the image capture, the face detection, the face image collection and the face classification functions. When this system detects a person’s face, it classifies whether she/he is a stranger or not. When she/he is a stranger, the robot asks for her/his name, and at the same time collects her/his face images. When she/he is an acquaintance, since the robot knows her/his name based on the classifier, the robot can call her/him by name. Since it is necessary to execute these processes in real-time, they are executed in the waking state.

The internal information processing system has functions related to memories, and consists of a function for generating training images from both the face images and their names newly stored in the waking state and a function for (re-)training the face classifier using the generated training images of faces. It is difficult to execute both the training face image generation and face classifier (re-)training in real-time.

An execution frequency of each process is controlled depending on a value of the sub-element of the AIM model shown in Figure 6. The execution frequencies of the face detection and classification are decided based on the value of a_{ex} . For example, when a_{ex} becomes higher, the external information processing frequency increases. The frequency of the external sensory data capture is decided based on m_{ex} . The training image generation and the face classifier training are executed depending on a_{in} .

2.4. Face Memorization

The face memorization system mainly consists of the following five functions. (1) A face detection function is used to detect people's faces and crop and collect face images. (2) A face classification function can estimate class probabilities based on a trained model. (3) A face classifier training function is used to re-train the face classifier with previously the collected face images, newly added face images and the people's names. (4) A speech recognition function is used for simple conversation and to collect people's names. (5) A text-to-speech function is used to have a conversation and call a person by her/his name.

In order to crop a face area from a captured image, the multi-task cascaded convolutional neural network (multi-task CNN) proposed by Zhang et al. [24] is used in this paper. Although we tried the Haar Cascades classifier in OpenCV and the Face Detector in Dlib, the multi-task CNN is better than these two detectors regarding computational time, detection rate for an image with a complex background and robustness for changes of face direction or face brightness.

The FaceNet proposed by Schroff et al. [18] and the FaceNet model pre-trained by Sandberg [25] are used to extract a 512-dimensional feature vector for a face image. The support vector machine algorithm in scikit-learn [26] is used to implement a face classifier.

Julius [27] and Open JTalk [28] are used for Japanese speech recognition and text-to-speech respectively.

2.5. Consciousness State Transition

The current state of the robot system is determined depending on the combination of the values of A , I and M as shown in Table 1. Moreover, there are two kinds of modes in the waking state. One is a face image collecting mode, and the other is a face classification mode. The system is normally in the face classification mode. When the face classifier finds a stranger's face, the mode translates to the face image collecting mode. Figure 8 shows both a state transition diagram and values of execution frequencies related to the image processing. The execution frequency of the face detection does not become less than 1 (fps) even in the REM or non-REM sleep; as a result, the system can wake up from sleep when a human face is detected in a captured image.

Table 1. Relations among AIM values and consciousness states.

State	Value of A	Value of I	Value of M
Awaken	$0.75 < A \leq 1.0$	$0.5 < I \leq 1.0$	$0.5 < M \leq 1.0$
Relaxed	$0.5 < A \leq 0.75$	$0.0 < I \leq 0.5$	$0.0 < M \leq 0.5$
Non-REM	$0.0 < A \leq 0.5$	$-0.5 < I \leq 0.0$	$-0.5 < M \leq 0.0$
REM	$0.5 \leq A \leq 1.0$	$-1.0 \leq I \leq -0.5$	$-1.0 \leq M \leq -0.5$

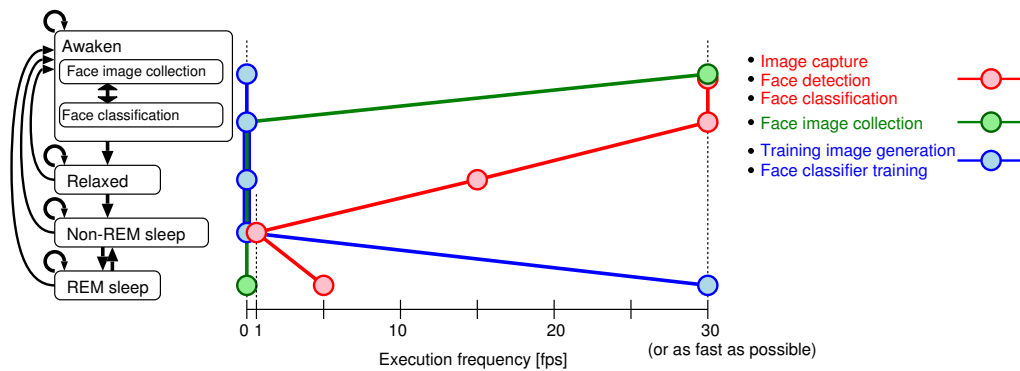


Figure 8. Execution frequency of each process.

3. Experimental System

3.1. System Configuration

Figure 9 shows the mobile robot system. The wheeled mobile robot, Pioneer 3-DX (OMRON Adept Technologies, LLC., Pleasanton, CA, USA), was equipped with a USB camera and a tablet PC. A note PC (Ubuntu 16.04, Intel Core i7-2630QM, Memory 8 GB) for controlling the system was put on the mobile robot. The PC was supplied electric power from internal batteries of the mobile robot through a DC/AC converter.

A robotic face with two eyes was displayed on the tablet PC; the direction of eyes often changed randomly in a horizontal direction, and the eyes sometimes blinked.



Figure 9. Mobile robot.

Software libraries are shown in Table 2. Robot Operating System (ROS) [29] is used for communications among processes, TensorFlow [30] is used for detecting faces and models of face recognition, scikit-learn [26] is used for training the face classifier, OpenCV [31] is used for several image processing tasks, Julius [32] is used for speech recognition and Open JTalk [28] is a Japanese text-to-speech system.

Table 2. Versions of software libraries.

Library	ROS	TensorFlow	Scikit-Learn	OpenCV	Julius	Open JTalk
Version	Kinetic Kame	1.12.0	0.20.1	3.3.1	4.3.1	1.07

3.2. Implementation of the Face Classifier

Figure 10a–c shows some face classification results. As shown in Figure 10a, the face classifier works well when a face is rotated obliquely. As shown in Figure 10b,c, there was no influence of the presence or absence of a pair of glasses, and a person who was not included in the trained classifier could be classified as a stranger. Table 3 shows the computational times of the face detection and the classification.

As described in the Section 2.4, the face image (re-)training function consists of the preprocessing where face images are cropped and stored by the multi-task CNN, the feature vector transformation by the FaceNet and the (re-)training by the SVM. Table 4 shows execution speeds of the preprocessing, the feature vector transformation and the re-training with or without the mathematical AIM. In total, 7000 face images of 35 persons were collected for these calculations. Of those, 3500 face images of 35 persons were used for re-training the face classifier, and the rest were used for evaluating the re-trained face classifier.

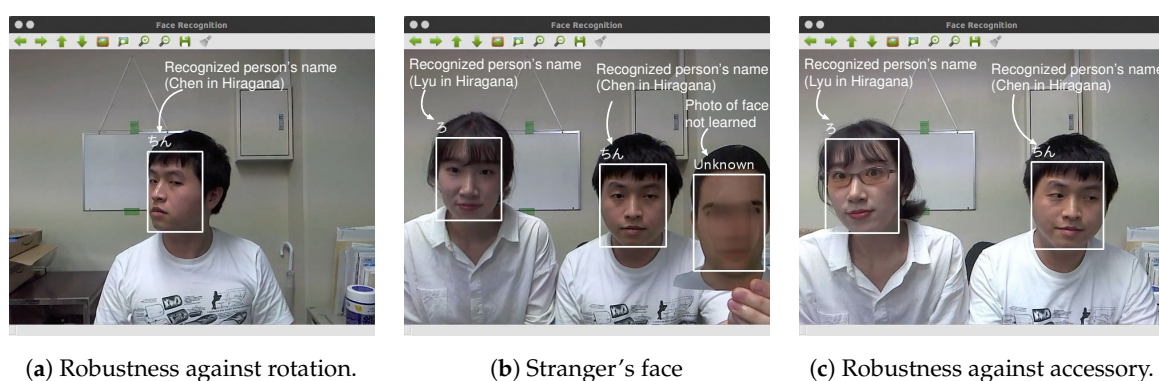


Figure 10. Face classification examples.

By using the AIM model, since the system changes its states between the REM sleep where these processes related to the re-training are executed and the non-REM sleep where most processes do not work periodically, it takes longer to finish re-training the classifier than the calculation without the AIM model. However, we believe that it is not problem, because the re-training is executed while the system has nothing to do and sleeps. The number of persons for the classification evaluation was 35 in this paper. Although 35 was not so many, the classification results could achieve 100%.

Table 3. Computational times of face detection and face classification.

	Computational Time
Face detection	0.14 [s]
Face classification	0.07 [s]
Face detection & classification	0.21 [s] (=4.77 [fps])

Table 4. Computational time for re-training face images.

	With AIM	Without AIM
Preprocessing	1589.4 [s]	1040.1 [s]
FaceNet	791.7 [s]	356.1 [s]
SVM	7.88 [s]	7.23 [s]
Precision	100.0 [%]	100.0 [%]

3.3. Effectiveness of AIM Model

It is important for a mobile robot to save its battery. In order to show another effectiveness of the AIM model, we conducted a simple experiment wherein the only executed processes were the face detection and the face classification. However, the computational loads of these processes are high,

and the execution frequency was about 4.8 (fps) in the experimental system, as described in Section 3.1. The variations of battery voltage with time were measured under two conditions: with the AIM model and without the AIM model.

Figure 11 shows the battery voltage variations with time with/without AIM. In the case of without the AIM model, since the face detection and classification processes were executed as fast as possible, the voltage decreased quickly. In the case of with the AIM model, although we woke the system by showing our faces in front of the camera about every five minutes, since the execution frequency of the image processing became lower through the operation of the AIM model in the sleeping mode, the electricity consumption could be cut.

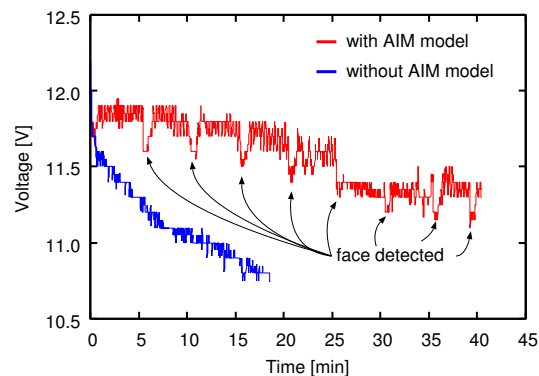


Figure 11. Variations of battery voltage with time.

4. Experimental Results

Previous studies have shown that a robot's behaviors can affect the impression that a person gets from the robot. Traeger et al. reported that behaviors of a social robot influenced not only the communication between the robot and humans, but also the communication between humans [33]. When a robot and a person pass each other in a corridor, the experiments in [34] have confirmed that the impression that the robot's actions give to the person is different depending on whether the direction of the robot's face and gaze is controlled or not.

Therefore, various factors may affect the experiments under realistic experimental conditions. Thus, in order to properly investigate the effectiveness of the name calling function using the face memorization system, we chose a simple experimental environment to evaluate research participants' impressions.

4.1. Outline of Experiments

In order that the mobile robot have the ability to call a person by her/his name, we developed the face memorization system. We conducted some experiments under the following two conditions:

- (1) The robot calls a research participant using her/his name during a conversation.
- (2) The robot calls the participant using the pronoun "you" during a conversation.

That is to say, the purpose of the experiments was to investigate what kind of impression the robot would give a person by using her/his name.

In this experiment, we chose the simplified Old Maid as the card game to play during a conversation. The situation and the rules are simple: the research participant has two trump cards, among which the joker is included, and the robot estimates which card the joker is. However, if an experimenter explains to the participants the purpose of these evaluation experiments in advance, the participants pay attention the robot's way of addressing them too much, and as a result, they may not be able to give fair evaluations. Thus, in this study, before the experiments, the experimenter gave the participant a false explanation as follows:

- (a) The robot has a face memorization function.
- (b) The robot has face expression detection and analysis functions, and has an ability to estimate the position of the joker based on a human's facial expressions.
- (c) The purpose of this experiment is not to win Old Maid, but to let the robot guess which of two cards is the joker correctly.

After all the experiments, the experimenter told the participants the true experimental purpose.

The research participant had a meeting with the robot three times in a set of evaluation experiments. On the first meeting, the participant's name and face images were collected by the robot. Upon the second and third meetings, she/he played the card game with the robot and evaluated her/his impression of interaction with the robot after each meeting. The robot used the participant's name during one meeting, and it called them by the pronoun "you" during the other meeting. The experiments were conducted with a counterbalanced measures design.

4.2. Procedure of Experiments

The detailed procedures of the experiments are described below.

- (1) An experimenter explains the fake experiment's purpose to a research participant in a waiting space, and takes her/him to an experimental space shown in Figure 12 described in the following subsection.
- (2) The mobile robot that stands by in the experimental space says hello to the participant.
- (3) The robot asks permission to collect her/his name and face images for experimental records.
- (4) After the robot finishes collecting the face images, the participant returns to the waiting space once.
- (5) The experimenter explains to the participant that she/he has the second and third meetings with the robot and plays the card game six times in one meeting after this explanation. Then the participant enters the experimental space.
- (6) After sitting in a chair, she/he shuffles two cards and sets them in a card holder on a table. The robot answers which of two cards is the joker. This transaction is repeated six times in one meeting.
- (7) After the second meeting, the participant returns the waiting space again, and answers a questionnaire given by the experimenter.
- (8) The procedure (6) is repeated once more as the third meeting.
- (9) After the third meeting, the participant returns the waiting space again, and answers another questionnaire given by the experimenter.
- (10) The experimenter tells the subject the true purpose of the experiments.

The detailed dialogue between a participant and the robot is shown in Appendices A.1 and A.2.

Here, in order to evaluate participants' impressions properly according to the purpose of this study, these experiments were conducted under the following conditions. Basically, we wanted to avoid bad influences because of failures of the image or voice recognition.

- (i) All the experiments were conducted based on the Wizard of Oz (WOZ) method [35]; the mobile robot was teleoperated by a human operator.
- (ii) The robot could get the correct position of the joker five out of six times but not the fourth time. This was realized by using a hidden camera.

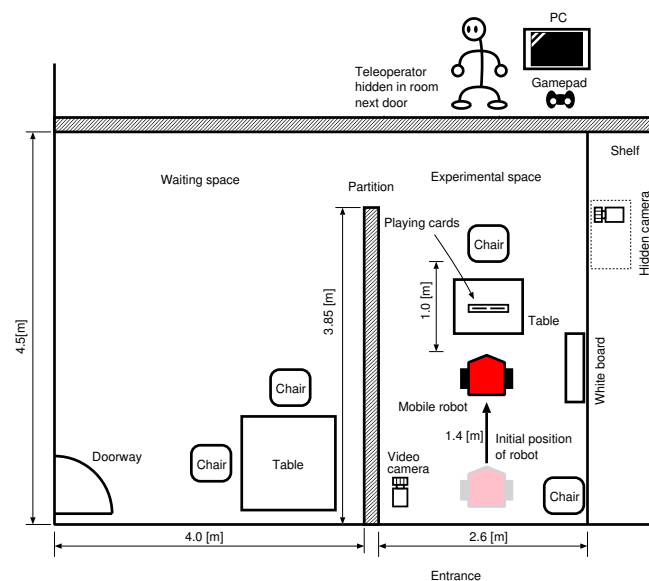
Each participant evaluated her/his impression using the Godspeed Questionnaire [36]. It is said that, "The Godspeed Questionnaire Series (GQS) is one of the most frequently used questionnaires in the field of human-robot interaction" [37]. Table 5 shows the application of the five Godspeed questionnaires with a 5-point Likert-scale.

Table 5. The Godspeed Questionnaire Series.

Section	Items		
Anthropomorphism	Fake	-	Natural
	Machinelike	-	Humanlike
	Unconscious	-	Conscious
	Artificial	-	Lifelike
	Moving rigidly	-	Moving elegant
Animacy	Dead	-	Alive
	Stagnant	-	Lively
	Mechanical	-	Organic
	Artificial	-	Lifelike
	Inert	-	Interactive
Likeability	Apathetic	-	Responsive
	Dislike	-	Like
	Unfriendly	-	Friendly
	Unkind	-	Kind
	Unpleasant	-	Pleasant
Perceived Intelligence	Awful	-	Nice
	Incompetent	-	Competent
	Ignorant	-	Knowledgeable
	Irresponsible	-	Responsible
	Unintelligent	-	Intelligent
Perceived Safety	Foolish	-	Sensible
	Anxious	-	Relaxed
	Agitated	-	Calm
	Quiescent	-	Surprised

4.3. Experimental Environment

Figures 12 and 13a,b show an experimental environment. The waiting space and the experimental space were divided by a partition. Only one research participant entered this place at once. The experimenter always sat on a chair in the waiting space. An operator for the WOZ was in a room next door, and teleoperated the mobile robot in the experimental space with a gamepad and a monitor. The operator could watch the entire experimental room and the position of the joker set on the card stand through a hidden camera shown in Figure 13a. Figure 13b shows one scene during an experiment.

**Figure 12.** Layout of experimental environment.

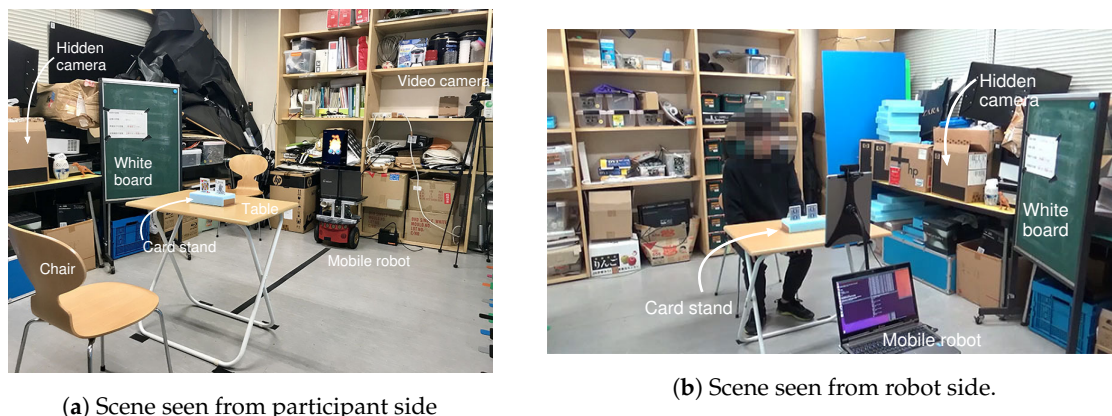


Figure 13. Scene example of evaluation experiment.

4.4. Evaluation Results

A total of 22 participants in their twenties were recruited in this evaluation experiment. The number of females was 10, the number of males was 12 and the average age was 22.1 years old. They were all undergraduate or graduate students in the University of Tsukuba, Japan, and some were international students.

The standard deviations of the results of the impression evaluation using the Godspeed Questionnaire are shown in Figure 14, and the results of t-tests are shown in Table 6. Although all the averages under conditions where the robot called participants by their names were higher than those under conditions wherein it called them by the pronoun “you”, there was a significant difference in only the likeability. It can be said that the robot could give a better impression to a person by calling her/him by her/his name during conversation.

Table 6. Evaluation results.

Scale		Call by name	Call by pronoun
Anthropomorphism	Mean	3.155	3.009
	Standard deviation	0.617	0.683
	Significance probability		0.195
Animacy	Mean	3.568	3.409
	Standard deviation	0.521	0.621
	Significance probability		0.226
Likeability	Mean	4.273	4.009
	Standard deviation	0.544	0.606
	Significance probability		0.039 (*)
Perceived Intelligence	Mean	4.036	3.927
	Standard deviation	0.368	0.511
	Significance probability		0.117
Perceived Safety	Mean	3.727	3.636
	Standard deviation	0.935	0.959
	Significance probability		0.701

* $p < 0.05$.

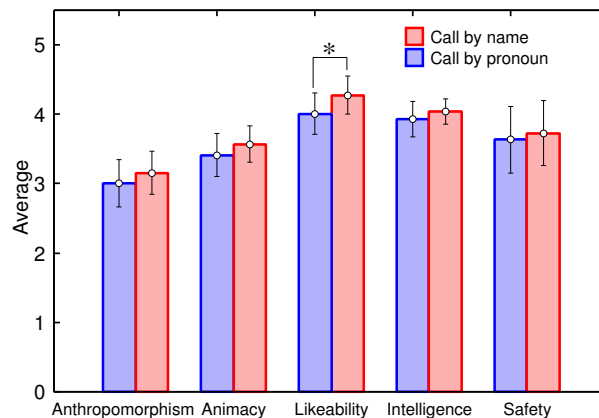


Figure 14. Evaluation results.

4.5. Discussion

We could confirm the basic validity and effectiveness of the name calling function using our proposed face memorization system through the experiments with research participants.

At the same time, the following facts were revealed from the interviews with the research participants or free descriptions in the questionnaire after the third meeting in the experiment.

- (1) Some participants did not notice the difference between the second and third meeting with the robot.
- (2) Some participants believed that the robot really had the ability to estimate the position of the joker based on human facial expressions. Since the robot could estimate correctly five out of six times, they had the impression that this robot's ability was awesome.
- (3) Since the way of the robot's speaking using the text-to-speech software was polite and humble, some participants had good impressions of the robot.
- (4) The two CG eyes of the robot shown in Figure 9 sometimes blinked and the line of sight was changed. However, since the movements were not synchronized with the conversation, there were some research participants who answered the questionnaire by saying that "it was impersonal".

Including these points, the weaknesses of this paper are shown below.

- (a) Despite the simple experimental environment, the contents of the experiment and the behaviors and appearance of the robot might have some influence on the impression evaluation.
- (b) The experiments were conducted based on the WOZ in order to avoid malfunctions in image processing and voice recognition.
- (c) The number of persons whose faces were memorized by robots was still small.

Thus, in the future work, we believe that it is necessary to conduct more experiments under the following two types of environments. One is an experimental environment that is simple but less affected by the appearance of the robot and the tasks, and the other is a more realistic experimental environment for evaluating the usefulness of the entire system. In addition, since the progress of face detection and face classification is moving rapidly, it is necessary to always adopt the latest algorithm in order to improve the performance of the system.

5. Conclusions

We proposed a face memorization system with a sleep-wake function for a social mobile robot, and applied it to a name calling function. When the robot meets a stranger, although it stores her/his face images and name in storage, since it takes long time to re-train a face classifier, the re-training of the classifier is not executed immediately, but is executed in a sleep state where there is no person around the robot and there is no sensory information to process. We conducted two kinds of experiments

in this paper. One was that battery voltage variations with time were measured with/without the sleep–wake function by the AIM model. Although the experimental conditions were simple, the battery running time using with the AIM model was a few times longer than that without the AIM model. The other was that we conducted an impression evaluation of the name calling function with research participants. Research participants evaluated their impressions of the robot that called them by their names through conversation during a simple card game. The Godspeed Questionnaire was used for the impression evaluation; it was confirmed that there was a significant difference in only the likeability. It can be said that the robot could give a better impression to a person by calling him/her by her/his name during conversation. The experimental results revealed the validity and effectiveness of the proposed face memorization system.

Author Contributions: Conceptualization, H.C., M.M. and M.F.; methodology, H.C., M.M. and M.F.; software, H.C.; validation, H.C., M.M. and M.F.; formal analysis, H.C., M.M. and M.F.; investigation, H.C., and M.M.; resources, M.M.; data curation, M.M.; writing—original draft preparation, H.C., and M.M.; writing—review and editing, M.M.; visualization, H.C., and M.M.; supervision, M.M.; project administration, M.M.; funding acquisition, M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by JSPS KAKENHI grant number JP18K04041.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Appendix A. Dialogue between a Research Participant and the Mobile Robot

Appendix A.1. First Meeting: The Robot Collects the Participant’s Face Images and Name

P and R mean a participant and the robot respectively.

- (1) R: Hello. A person has appeared at last. Since there has been nobody until now, I had been bored. I am glad to meet you. Then, let me introduce myself. I am the mobile robot, PIO. I appreciate that you joined this experiment today. I believe that the experimenter has already given you the instructions of this experiment, but for confirmation, I will explain again. I would like you to have three meetings with me. In this first meeting, let me collect your name and face images, as it is necessary for card games in the next meetings. Let me begin by collecting your name and face images. Are you ready?
- (2) P: Sure.
- (3) R: First of all, I’ll ask your name. Would you tell me your name?
- (4) P: XX. (The participant answers her/his name.)
- (5) R: XX-san (The Japanese word, *san*, means Mr. or Ms. in English). OK, I understood. Next, let me collect your face images. Please keep moving your head from side to side and up and down until I give you a sign. It is going to take about ten seconds. Now, are you ready?
- (6) P: Yes.
- (7) R: Then I’ll begin. Please start to move your head slowly. (The robot starts to collect the face images).
- (8) R: Collecting your face images has finished. Thank you for your cooperation. Would you please return the waiting space until the next meeting?

Appendix A.2. Second or Third Meeting: The Robot Calls a Participant by Her/His Name or Using the Pronoun “You”

- (1) R: Hello. Today, I am going to play the card game Old Maid with [XX-san/you]. I will explain to you how to play it. First, please pick both of these two cards on this table, and shuffle them under the table. Secondly, please put the cards into the card stand in light blue. Lastly, when you are ready, please tell me “I am ready.”
Then, XX-san, would you shuffle these two cards?

- (2) P: I'm ready.
- (3) R: Please stare at my eyes. I'll undertake the analysis.
(The robot holds for five seconds to pretend to consider).
- (4) R: The analysis has been completed. The joker is on the right (left) side. Is it correct?
- (5) P: (Except the 4th turn) Correct.
(In the 4th turn) Incorrect.
- (6) R: (Except the 4th turn) Well, I am glad the answer is correct.
XX-san, thank you for your cooperation/Thank you for your cooperation.
(In the 4th turn) Too bad.
In spite of XX-san's/your great effort, my analysis didn't work well. (Proceed to the next)
- (7) R: (In the 1st, 2nd, 3rd or 5th turn) Let's do our best next time. Let's move on to the next turn.
Then, pick the cards up and shuffle them please. (Return to (2))
(In the 4th turn) I'll do my best next time. Then, pick the cards up and shuffle them please.
(Return to (2))
(In the 6th turn, proceed to the next (8))
- (8) R: The game is over. I'll show you the results. I could get five correct answers in six turns.
Thanks to XX-san/you, the results were great. I really appreciate it.
Let's wrap this up. I was very glad to play the card game with XX-san/you. I really appreciate it. Would you please return to the rest area?

References

1. Bartneck, C.; Kanda, T.; Mubin, O.; Mahmud, A. Does the Design of a Robot Influence Its Animacy and Perceived Intelligence? *Int. J. Soc. Robot.* **2009**, *1*, 195–204. [[CrossRef](#)]
2. Paiva, A.; Leite, I.; Ribeiro, T. Emotion Modelling for Social Robots. In *The Oxford Handbook of Affective Computing*; Oxford University Press: Oxford, UK, 2014; pp. 296–419.
3. Admoni, H.; Scassellati, B. Social Eye Gaze in Human-robot Interaction: A Review. *J. Hum.-Robot Interact.* **2017**, *6*, 25–63. [[CrossRef](#)]
4. Wang, E.; Lignos, C.; Vatsal, A.; Scassellati, B. Effects of Head Movement on Perceptions of Humanoid Robot Behavior. In Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction (HRI2006), Salt Lake City, UT, USA, 2–3 March 2006; pp. 180–185.
5. Mikawa, M.; Yoshikawa, Y.; Fujisawa, M. Expression of intention by rotational head movements for teleoperated mobile robot. In Proceedings of the 2018 IEEE 15th International Workshop on Advanced Motion Control (AMC2018), Tokyo, Japan, 9–11 March 2018; pp. 249–254.
6. Lyu, J.; Mikawa, M.; Fujisawa, M.; Hiiragi, W. Mobile Robot with Previous Announcement of Upcoming Operation Using Face Interface. In Proceedings of the 2019 IEEE/SICE International Symposium on System Integration (SII2019), Paris, France, 14–16 January 2019; pp. 782–787.
7. Huang, Ch.; Iio, T.; Satake, S.; Kanda, T. Modeling and Controlling Friendliness for An Interactive Museum Robot. In Proceedings of the Robotics: Science and Systems (RSS2014), Berkeley, CA, USA, 12–16 July 2014.
8. Schulte, J.; Rosenberg, C.; Thrun, S. Spontaneous, short-term interaction with mobile robots. In Proceedings of the 1999 IEEE International Conference on Robotics and Automation, Detroit, MI, USA, 10–15 May 1999; pp. 658–663.
9. Torta, E.; van Heumen, J.; Cuijpers, R.H.; Juola, J.F. How Can a Robot Attract the Attention of Its Human Partner? A Comparative Study over Different Modalities for Attracting Attention. In Proceedings of the 4th International Conference on Social Robotics, Chengdu, China, 29–31 October 2012; pp. 288–297.
10. Kanda, T.; Hirano, T.; Eaton, D.; Ishiguro, H. A Practical Experiment with Interactive Humanoid Robots in a Human Society. In Proceedings of the Third IEEE International Conference on Humanoid Robots, Karlsruhe-Munich, Germany, 30 September–3 October 2003.
11. Kanda, T.; Sato, R.; Saiwaki, N.; Ishiguro, H. A Two-Month Field Trial in an Elementary School for Long-Term Human-Robot Interaction. *IEEE Trans. Robot.* **2007**, *23*, 962–971. [[CrossRef](#)]

12. Mitsunaga, N.; Miyashita, T.; Ishiguro, H.; Kogure, K.; Hagita, N. Robovie-IV: A Communication Robot Interacting with People Daily in an Office. In Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2006), Beijing, China, 9–15 October 2006; pp. 5066–5072.
13. Kanda, T.; Shiomi, M.; Miyashita, Z.; Ishiguro, H.; Hagita, N. An affective guide robot in a shopping mall. In Proceedings of the 2009 4th ACM/IEEE International Conference on Human-Robot Interaction, La Jolla, CA, USA, 11–13 March 2009; pp. 173–180.
14. Chen, X.; Williams, A. Improving Engagement by Letting Social Robots Learn and Call Your Name. In Proceedings of the Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, Cambridge, UK, 23–26 March 2020; pp. 160–162.
15. Vezzetti, E.; Marcolin, F.; Tornincasa, S.; Ulrich, L.; Dagnes, N. 3D geometry-based automatic landmark localization in presence of facial occlusions. *Multimed. Tools Appl.* **2018**, *77*, 14177–14205. [[CrossRef](#)]
16. Lin, T.; Chiu, C.; Tang, C. Rgb-D Based Multi-Modal Deep Learning for Face Identification. In Proceedings of the 2020 IEEE International Conference on Acoustics, Speech and Signal, Barcelona, Spain, 4–8 May 2020; pp. 1668–1672.
17. George, A.; Mostaani, Z.; Geissenbuhler, D.; Nikisins, O.; Anjos, A.; Marcel, S. Biometric Face Presentation Attack Detection With Multi-Channel Convolutional Neural Network. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 42–55. [[CrossRef](#)]
18. Schroff, F.; Kalenichenko, D.; Philbin, J. FaceNet: A Unified Embedding for Face Recognition and Clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 815–823.
19. Deng, J.; Guo, J.; Xue, N.; Zafeiriou, S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4685–4694.
20. Chen, H.; Mikawa, M.; Fujisawa, M.; Hiiragi, W. Face Memorization System Using the Mathematical AIM Model for Mobile Robot. In Proceedings of the 2019 IEEE/SICE International Symposium on System Integration (SII2019), Paris, France, 14–16 January 2019; pp. 651–656.
21. Allan Hobson, J. *The Dream Drugstore*; The MIT Press: Cambridge, MA, USA, 2001.
22. Mikawa, M.; Yoshikawa, M.; Tsujimura, T.; Tanaka, K. Intelligent Perceptual Information Parallel Processing System Controlled by Mathematical AIM Model. In Proceedings of the 2007 7th IEEE-RAS International Conference on Humanoid Robots, Pittsburgh, PA, USA, 29 November–1 December 2007; pp. 389–403.
23. Mikawa, M.; Yoshikawa, M.; Tsujimura, T.; Tanaka, K. Librarian Robot Controlled by Mathematical AIM Model. In Proceedings of the ICROS-SICE International Joint Conference, Fukuoka, Japan, 18–21 August 2009; pp. 1200–1205.
24. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Process. Lett.* **2016**, *23*, 1499–1503. [[CrossRef](#)]
25. Sandberg, D. Face Recognition using Tensorflow. Available online: <https://github.com/davidsandberg/facenet> (accessed on 2 October 2019).
26. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
27. Lee, A.; Kawahara, T.; Shikano, K. Julius—An Open Source Real-Time Large Vocabulary Recognition Engine. In Proceedings of the 7th European Conference on Speech Communication and Technology (EUROSPEECH 2001), Aalborg, Denmark, 3–7 September 2001; pp. 1691–1694.
28. Open JTalk. Available online: <http://open-jtalk.sourceforge.net> (accessed on 7 October 2020).
29. ROS (Robot Operating System). Available online: <http://wiki.ros.org> (accessed on 7 October 2020).
30. TensorFlow. Available online: <https://www.tensorflow.org> (accessed on 7 October 2020).
31. Bradski, G. The OpenCV Library. *Dr. Dobb's J. Softw. Tools* **2000**, *25*, 120, 122–125.
32. Lee, A.; Kawahara, T. Julius-Speech/Julius: Release 4.5. 2019. Available online: <https://doi.org/10.5281/zenodo.2530396> (accessed on 7 October 2020).
33. Traeger, M.L.; Strohkorb, S.S.; Jung, M.; Scassellati, B.; Christakis, N.A. Vulnerable robots positively shape human conversational dynamics in a human–robot team. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 6370–6375. [[CrossRef](#)] [[PubMed](#)]
34. Mikawa, M.; Lyu, J.; Fujisawa, M.; Hiiragi, W.; Ishibashi, T. Previous Announcement Method Using 3D CG Face Interface for Mobile Robot. *J. Robot. Mechatron.* **2020**, *32*, 97–112. [[CrossRef](#)]

35. Kelley, J.F. An Empirical Methodology for Writing User-friendly Natural Language Computer Applications. In Proceedings of the the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, USA, 12–15 December 1983; pp. 193–196.
36. Bartneck, C.; Croft, E.; Kulic, D. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int. J. Soc. Robot.* **2009**, *1*, 71–81. [[CrossRef](#)]
37. Weiss, A.; Bartneck, C. Meta analysis of the usage of the Godspeed Questionnaire Series. In Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2015), Kobe, Japan, 31 August–4 September 2015; pp. 381–388.

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).