Designing Enjoyable Motion-Based Play Interactions with a Small Humanoid Robot

Martin Cooney • Takayuki Kanda • Aris Alissandrakis • Hiroshi Ishiguro

M.Cooney (✉) • T. Kanda • H. Ishiguro

Advanced Telecommunications Research Institute International IRC/HIL

2-2-2 Hikaridai, Keihanna Science City, Kyoto, Japan

e-mail: mcooney@atr.jp

T. Kanda

e-mail: kanda@atr.jp

H. Ishiguro

e-mail: ishiguro@atr.jp

A. Alissandrakis

University of Hertfordshire, Hatfield, AL10 9AB, U.K.

e-mail: a.alissandrakis@herts.ac.uk

M. Cooney • H. Ishiguro

Department of Systems Innovation, Osaka University, 1-3 Machikaneyama Toyonaka Osaka 560-8531 Japan

Abstract Robots designed to co-exist with humans in domestic and public environments should be capable of interacting with people in an enjoyable fashion in order to be socially accepted. In this research, we seek to set up a small humanoid robot with the capability to provide enjoyment to people who pick up the robot and play with it by hugging, shaking and moving the robot in various ways. Inertial sensors inside a robot can capture how the robot's body is moved when people perform such "full-body gestures". Unclear is how a robot can recognize what people do during play, and how such knowledge can be used to provide enjoyment. People's behavior is complex, and naïve designs for a robot's behavior based only on intuitive knowledge from previous designs may lead to failed interactions. To solve these problems, we model people's behavior using typical full-body gestures observed in free interaction trials, and devise an interaction design based on avoiding typical failures observed in play sessions with a naïve version of our robot. The interaction design is completed by investigating how a robot can provide "reward" and itself suggest ways to play during an interaction. We then verify experimentally that our design can be used to provide enjoyment during a playful interaction. By describing the process of how a small humanoid robot can be designed to provide enjoyment, we seek to move one step closer to realizing companion robots which can be successfully integrated into human society.

Keywords *Interaction design for enjoyment • Playful human-robot interaction • Full-body gesture recognition • Inertial sensing • Small humanoid robot*

1 Introduction

We are on the brink of an age in which we will be able to see robots co-existing and interacting with people in our homes, our workplaces, and the streets and stores of our cities. Robots will perform "troublesome" tasks [19], as well as service-related tasks requiring memorization or fine perceptual capabilities [40]. However, a robot will be expected to do more than performing tasks: it will also be required to interact with people in an enjoyable way. Enjoyment has been shown to be a key constituent of functionality and work [8, 38], and to serve an important role in technological artifacts gaining social acceptance [45, 51], a finding which has been replicated in the field of Human-Robot Interaction (HRI) [20]. It would be useful to know how enjoyment can be provided. Toward this, we propose a way in which a companion robot can provide enjoyment through the act of play.

A "companion robot" need not be only the object of play but can act proactively, as a play partner, based on its own intentions; entertainment need not be its primary function (we do not assume people will limit play behavior to such robots); and in some circumstances a robot should not play (e.g., when its task is serious or urgent for a user [17]). "Enjoyment", the goal for this study, is a "pleasant experiential state . . . at the heart of entertainment" [47] and an important optimization criterion [31] related to "satisfaction" and usability (ISO 9241-11); it is often accompanied by a feeling of reward, clear goals, the matching of task challenge with a user's skill level, and perception of control [9, 10]. Lastly, "play" is a typical behavior which is "free" and absorbing [22], and which people of all ages engage in and enjoy for its own sake [12].

Motion plays an intrinsic role in play with artifacts: people move toys such as dolls, teddy bears, puzzle pieces, and construction set blocks; other toys such as balls, swings, teeter-totters, hula hoops and jump ropes furthermore provide immediate and fine feedback through their own motion when moved. In HRI as well, many robots also interact non-verbally via motion, e.g. [25, 29, 35]. To investigate this important scenario, a platform was set up to engage in such non-verbal, motion-based play, Sponge Robot [6].

A humanoid form was selected to provide a rich, natural interface for interaction and elicit ingrained behaviors through the metaphor of the robot as being like a human [3]. During one-on-one play with a small humanoid robot, we expect people to pick up the robot and hug it, shake it, and move its full body around in various ways (e.g., as in Fig. 1). An inertial sensor placed inside the robot's body can be used to capture such *full-body gestures* performed by people toward the robot. Thus, the goal of this paper is to show how a small humanoid robot can provide enjoyment in such a motion-based play scenario.

The challenge lies in modeling people's complex perceptions (i.e., finding how people play with a small humanoid robot and what strategies for associating play affordances and a robot's behaviors will be perceived by people as enjoyable). This is a difficult problem; in the words of Dautenhahn and her colleagues, human social behavior is "very complex and subtle", and designing a robot also suitable in terms of its perception by people who are interacting with it is "one of the most challenging open issues" [11].

To solve this problem, we observe and analyze interactions. *Typical gestures* which we identify are recognized online using Support Vector Machines (SVMs) and heuristic thresholds. We adapt intuitive knowledge from previous robot designs by analyzing typical failing points in interactions; this leads us to test candidate strategies for a robot's responses and suggestions, and propose guidelines.[1] By overcoming this challenge, we can realize our goal for this study, providing enjoyment through play, and move one step closer toward smooth integration of robots into society.

1 Parts of this work have been previously presented [6, 7]

Fig. 1. Enjoyable play interaction with a small robot

The rest of this article is organized as follows. Section 2 describes some related work, and Section 3 introduces our platform for playful interaction, Sponge Robot. Section 4 describes our approach to modeling and recognizing people's behavior. We propose interaction design guidelines in Section 5 which are fleshed out in Section 6, and verified in Section 7. Section 8 provides discussion, and Section 9 summarizes contributions.

2 Related Work

Designing a robot capable of providing enjoyment in playful interactions involves making decisions with regard to embodiment, sensing capability, behaviors, and "interaction strategy" (what a robot should do and when).

Many previous achievements have addressed the first three: embodiment, sensing capability, and behaviors. Robots have been modeled on humans (e.g., E.M.A., QRIO [41, 42], Jingle Bell Rock Santa, My Real Baby, Nao, and Robosapien), pets such as dogs (AIBO) [14], cats (Dream Cat), extinct or rare creatures such as dinosaurs (Pleo) or seals (Paro) [48], and toys such as teddy bears (Huggable) [39] or balls (Roball) [35, 36]. One very successful design, AIBO, can see, hear, and feel touch, and features over a thousand behaviors. What has not been described is how such sensed input and behaviors can be associated to provide enjoyment.

Some general principles have also been suggested which could be important in designing a playful interaction with a robot: lifelikeness (Fujita, for AIBO, [14]), robustness (Salter et al. for Roball [35, 36]), rhythm (Kozima et al., for Keepon [25]), and timing and delays (Robins and his colleagues, for Kaspar [34]). These general principles however do not indicate requirements for a robot's behavior during an interaction which, if unsatisfied, could result in failure to provide enjoyment: i.e., an interaction strategy for enjoyment.

Such a strategy may be specified in terms of a set of *guidelines* for how to structure a robot's behavior. Motivation for this formulation is based on the fact that guidelines are useful for complex interactions in which not all possible user actions can be foreseen [28]. As well, a balance can be struck between being too general (missing problems) or too specific (difficult to apply) by providing both general principles and specific implementation descriptions as in [32].

As a first step to form such an interaction strategy, previous designs were examined for similarities which could reveal intuitive knowledge. A robot's appearance and behavior were found to be related; also, the "style" of interaction is turn-based and not too complex. For example, Dream Cat meows and purrs, small humanoids Robosapien and E.M.A. walk and dance, and the ball-shaped Roball spins and rolls. Users playing with such robots can press buttons or perform other simple actions to trigger motions (e.g., E.M.A., Robosapien, or Tickle Me Elmo); robots can also idle to appear lifelike (e.g., Paro, Dream Cat, or Pleo). These two rules-of-thumb, although not sufficient for providing enjoyment, were used to find an interaction strategy based on user feedback and analysis of failed interactions.

In summary, what has been missing was a clear reference specifically targeting enjoyment which tells a designer what to do and what not to do when developing a companion robot for play. The new contribution of this work is that it reports general guidelines for how to *structure an interaction* for enjoyment, which are tested via experiment[1]; with this knowledge, it becomes possible to know how to create robots capable of engaging in enjoyable motion-based play interactions.

1 Two earlier conference papers first described our approach for recognizing gestures [6] and how this information can be applied [7]. The current article chronicles the whole design process from start to finish in one place, providing additional details, and offers new results needed to support our finding that the proposed solution is effective toward providing enjoyment.

3 Sponge Robot

The platform developed for providing enjoyment through play, Sponge Robot (shown in Fig. 1 and 2 and described in detail in our previous paper [6]), is a typical small humanoid robot with limited sensing and motion capabilities, based on a commercial kit (Robovie-X from VStone Co., Ltd.). Its system architecture comprises several basic components shown in Fig. 3; *recognition and design* (depicted on the top right-hand side of the figure) represents the vital knowledge we seek to acquire from this study.

During play, a user interacts with the robot. Data from the robot's inertial sensor are sent wirelessly to an external laptop and processed to recognize the user's gestures. This information is used to decide the robot's behavior: e.g., "responding" to a human's action, "suggesting" a way to play through proactive motion, or changing the robot's internal state (its intention or desire to play in a certain way). Finally, wireless commands to perform motions or sounds are relayed back to Sponge Robot, which seeks to provide enjoyment by moving using a typical number of degrees of freedom for a small humanoid robot (13, with two in each arm, four in each leg, and one in its head) and playing sounds using a speaker located in its abdomen.

Overall, the platform was made to be soft and light to facilitate motion-based interactions, and its humanoid form was intended to elicit rich communicative behavior [3], although exactly how people seek to play with a small humanoid robot was unknown and had to be investigated.

4 Recognizing People's Play Behavior

Data collection was performed in two phases: first, to reveal the types of play gestures a person performs, and second, to collect inertial sensor data for these gestures.

4.1 Modeling using Typical Full-body Gestures

To clarify how people seek to play with a small humanoid robot (i.e., what a robot should recognize), users were asked to freely interact with Sponge Robot and the gestures they performed were recorded. In total, 17 young adult Japanese users played with the robot for five minute sessions each.

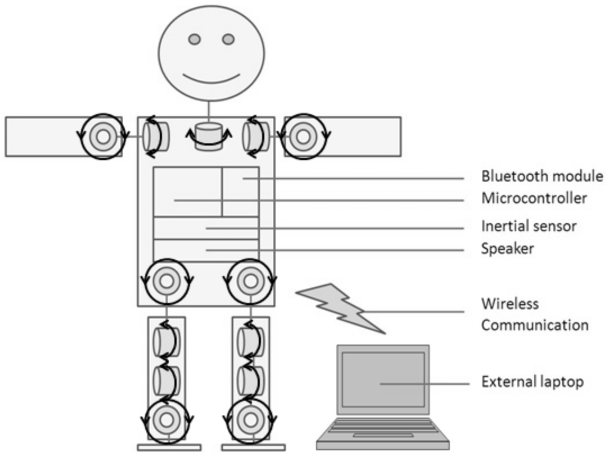

Fig. 2. Sponge Robot's hardware: degrees of freedom (DOFs) are represented as cylinders with curved arrows indicating the directions of rotation; and the location of components such as the inertial sensor and speaker are indicated.

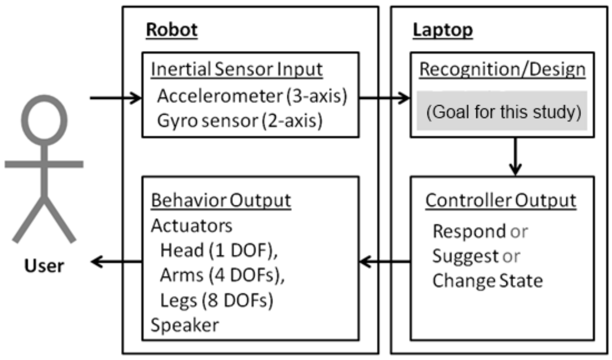

Fig. 3. System architecture: showing the goal of the current study in relation to process flow during a playful interaction.

4.1.1 Labeling

During sessions, the experimenter noted in simple language what users did. Afterwards, recorded video footage was checked to ensure the experimenter's written notes were complete. The notes were then codified using short labels (e.g., "Walk" to refer to a gesture in which a user shuffled the robot back and forth as if it were walking). In this process, notes and labels were attributed by the experimenter based on the observed meaning of gestures, and not on actual physical movements. Thus, moving the robot like a child might play with a toy airplane was labeled "Airplane" whether the robot was swung through the air from left to right, or vice versa, or diagonally, or even up or down. Inertial data were also recorded.

4.1.2 Robot Set up

During data acquisition, the robot did not move, in line with other studies identifying how people touch a teddy bear robot [24] and a cat-like robot [52]. Reasons for using a motionless robot include the difficulty in showing all possible motions a robot could perform and lack of a clear indication for how a subset should be chosen, the results of some pre-trials we conducted in which we found that a robot's movements could stop participants from performing gestures, and our expectation that even robots capable of motion will sometimes be motionless.

The robot was also switched on and its limbs were locked (joints could not be bent). This was because we wished to investigate "full-body" affordances yielded by a robot's small humanoid form—how people will move the full body of a robot—without making assumptions about the robot's degrees of freedom which could make the results less applicable to other cases. Likewise, no props (tools or objects) were made available to participants. This was because there is no way to offer every kind of prop that could be used in playing with a robot, and to avoid a possible source of confusion which could bias results.

4.1.3 Full-body gestures

As a result of observation, we found that people's behavior was complex. For example, in one brief 15 second period, a user turned the robot first one way then the other, hugged the robot, transported it to his knee, lifted and dropped it, and laid the robot face down. However, some typical patterns were observed: 13 gestures were performed by multiple users. We refer to these gestures as "*typical gestures*" (defining "typical" as meaning "performed by more than one participant"). Their frequency of occurrence suggests that others seeking to play with a small humanoid robot will be likely to perform them, and that these gestures should be recognized.

Some gestures such as Inspect, Up Down, Lay Down, and Stand were frequently observed; in contrast, some gestures were performed only by a single user, such as Ball Games or Rub Head with Robot. We decided to select as our classification target all typical gestures (those performed by at least two participants). These gestures are described in Table 1.

The typical gestures observed provide insight into people's perception of the robot. Some gestures such as Rock Baby and Hug seemed like those which might be targeted toward a small human (an infant), whereas others such as Fight or Upside-down seemed targeted more toward a robotic toy. Interviews confirmed that some participants felt that Sponge Robot resembled a human baby, whereas others referred to it as just a robot, toy or machine. This coincides with observations in previous work that some people are more inclined to perceive human-like traits in robots than others [43]. Additionally, the observation of various playful gestures suggested the validity of our premise that Sponge Robot could be used as a platform for interactive play.

4.2 Inertial Data Acquisition

Initially obtained inertial data varied greatly in number of samples per class, which could adversely affect gesture recognition accuracy, and were not in a convenient form for training a recognition system. Therefore, new data were obtained from an additional 21 participants (young adults in their 20s) as follows.

During data acquisition, participants sat on "tatami" floor mats in front of the robot but were allowed to stand and act freely (see Fig. 4), and were instructed to perform the 13 target gestures. A separate monitor to one side ran a simple clock program to allow identification of when gestures started and ended. Sessions lasted approximately 15 minutes. Before starting, participants were handed a sheet with a list of typical gestures and descriptions (as in Section 4.1) and given simple instructions. However, users were not told how they should perform gestures.

We were also interested in how the robot's movements would affect recognition accuracy; therefore, participants were asked to repeat the target gestures for four different robot motion conditions, which are shown in Fig. 5 and described below:

a) No motion—the robot's joints were stiff and the robot was in a neutral pose, with arms outstretched and legs together.

b) Idling—slight, but continuous motion; Gaussian noise was applied to the robot's servo positions.

Table 1 Typical Full-body Gestures

	Gesture label	Description	Percentage of participants who performed this gesture
1	Inspect	look at parts of the robot from various angles	100%
2	Up Down	move the robot up and down	94%
3	Lay Down	lower the robot to a horizontal orientation	94%
4	Stand	raise the robot to a vertical orientation	88%
5	Balance	balance the robot (e.g. on one hand)	64%
6	Walk	make the robot look like it is walking	59%
7	Airplane Game	make the robot look like it is flying	47%
8	Dance	make the robot do a little dance	47%
9	Upside-down	turn the robot upside-down	47%
10	Rock Baby	cradle and rock the robot like a baby	41%
11	Back and forth	shake the robot back and forth	35%
12	Fight	make the robot fight	35%
13	Hug	hug the robot	29%

Gestures performed by only one participant were: Hop, Climb wall, Rub head with robot, Ball games, Massage robot, Give robot first-aid, Press ear to robot, Rotate on head, Show robot to others, Hide-and-go-seek, and Travel up person.

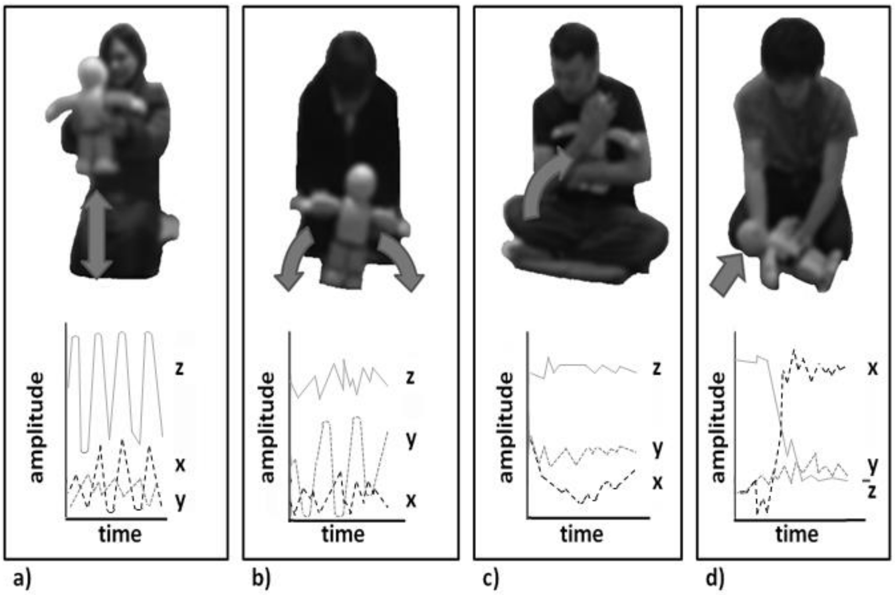

Fig. 4. Gestures being performed (top), with inertial data (below): a) Up Down, b) Dance, c) Hug, d) Lay Down; accelerometer x axis is dark and dashed, y axis is medium shaded and dotted, z axis is light and solid lined

c) Try to turn—a sudden motion; the robot quickly tucked in one arm and raised one leg to create an unbalanced state.

d) Flap arms and legs—a large motion; the robot made a motion which could interfere with a participant’s ability to grasp the robot.

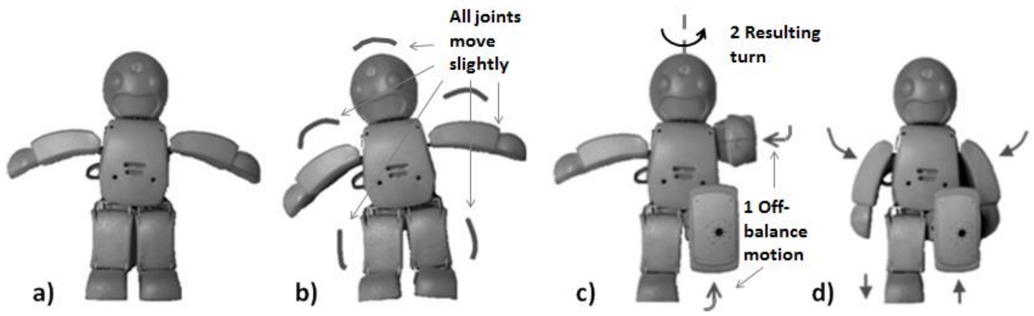

Fig. 5. The four motion conditions investigated: a) No motion, b) Idling (all joints are perturbed by Gaussian noise), c) Try to turn (a sudden motion of the robot's left arm and leg cause the robot to lose balance and fall to its left), d) Flap arms and legs

As in the initial data acquisition, the robot was powered up in all conditions, and its joints could not be moved by participants. Both continuous and non-continuous motions were tested because we wanted to see the effect of different kinds of motions. The idling motion (b) was continuous, performed by the robot for the entire duration of the condition; the other two motions (c) and (d) were triggered to occur once during each gesture. We predicted that the robot's movements would affect gestures, reducing recognition accuracy.

After obtaining the data, 1328 gesture instances were manually labeled using video recordings of the sessions.

4.3 Recognition System

To recognize full-body gestures, data must be parsed for discrete gesture events and classified using a recognition technique.

For the former problem, a fixed-size window of three seconds was selected, which seemed sufficient to capture characteristic information for the target gestures. This means gestures are expected to last a few seconds, but not that the system must wait the full time. Gesture recognition can take place each time a new data point has been added to the window (with a delay of around 80ms). Thus, for short gestures, the probability output for that gesture is likely to go high before the full three seconds has passed, and the system does not need to wait the entire time. This timing depends on the training samples and how the gestures are temporally defined; e.g. when does "Hug" start? Does the gesture start when the robot is picked up? When the robot is raised and (usually) tilted slightly backward? When the robot is tilted forward (and first comes into contact with the person's chest)? Or just before the robot is tilted backward (and released from physical contact)? These decisions affect when probability output goes high, and when the system can recognize a gesture.

In order to classify windows, Support Vector Machines (SVMs) were used. SVMs are a useful supervised machine learning method [53]. This approach involves calculating a boundary defined by points known as "support vectors" which maximizes inter-class separation while allowing for some errors; this boundary is then used to label new data as belonging to either of two classes. Some decisions must be made when using this approach. For problems involving more than two classes, multiple binary SVMs may be combined as "one-vs.-all", which involves using a single SVM to discriminate each one class from all others, and "one-vs.-one", which uses a SVM for each pair of classes. A suitable "kernel" function can be chosen to map data which may not be linearly separable into higher dimensions in the hope that the projected data will become more easily separable. In addition to parameters learned by the algorithm, the designer must also choose "hyper-parameters" which define characteristics of the boundary.

For our case, SVMs were implemented using a freely available library, LIBSVM ("A Library for Support Vector Machines") [4, 5]. A one-vs.-one system was chosen for accuracy at the cost of using more binary classifiers than a one-vs.-all system. The Radial Basis Function (RBF) kernel was chosen due to its applicability to nonlinear problems, avoidance of numerical problems by constraining kernel coefficients to be between zero and one, and the small number of hyper-parameters which must be found (only C and γ, where C describes the degree to which the boundary can accept misclassified samples, and γ describes the influence of support vectors over the data space) [4]. The tool "grid.py" provided with LIBSVM was called automatically from our code to perform a grid search to find values for C and γ for each fold during cross-validation. For the entire dataset, values of $C = 8$ and $\gamma = 0.5$ were selected.

Additional processing is required to use the system for online detection of a person's behavior in interactions: classification should occur when a person actively plays with the robot, and the system should be able to avoid misclassifications due to brief occasional noise in the constant stream of SVM output (e.g., occurring between gestures). Class labels are used as a basic filter to recognize when a user is actively playing or not interacting much (e.g., Hug and Dance, versus Balance and Inspect). An additional check, using codebook vectors for typical orientations such as lying or standing and thresholds calculated by recording inertial data when the robot moved, allows the system to recognize at any time step if a person is moving the robot or the robot is moving by itself.

Also, some additional thresholds specify the number of times a class label should be recognized to be confident that a gesture has occurred, thus allowing for stable performance. Other more complex approaches exist for detecting when gestures occur in time stream data, but the solutions developed seemed sufficient for our problem, allowing the system to determine when a gesture is performed.

4.4 Features

In addition to a recognition technique, important information for recognizing gestures had to be calculated from the raw inertial data, but it was not evident from the literature what features would be good; the approach followed was to compare different types of features to find a good feature type, and then select best features from this group.

For the former problem, three typical candidate feature types were investigated (shown in Table 2), each from a different domain: time, frequency, and time/frequency. Time-domain statistics features included mean axis values (also used for Roball [35, 36]), standard deviation, and the overall changes between first input value and last input value for each axis. Cyclic wave-patterns in the inertial data for Walk, Dance, and Up-Down, etc., suggested the usefulness of frequency-based features such as Discrete Fourier Transform (DFT) coefficients, which have been used for Huggable in [39]. As well, speed of calculation, simplicity, and ability of wavelet analysis to capture information from both frequency and time domains were reasons to consider the use of Haar Transform coefficients (e.g., used for pattern finding in [46]).

These three candidate feature types were tested using a wrapper-based feature selection algorithm. This yielded a cross-validation accuracy score for each full group of features, which was used to rank feature groups. As a result, the "statistics" group was found to perform the best for our task (shown in Table 3).

Table 2 Feature types

	Domain	Type	Previous work using such features
1	Time	Statistics (e.g., mean, standard deviation, etc.)	Roball [35, 36]
2	Frequency	Discrete Fourier Transform (DFT)	Huggable [39]
3	Time/Frequency	Haar Wavelet Transform	Object detection [46]

Types of features from three different domains were evaluated as a first step to find good features

Table 3 Comparing feature types

Feature Type	Cross-validation accuracy (%)
Statistics	74.3
DFT coefficients	62.1
Haar coefficients	51.4

Individual features were then selected to increase accuracy, prevent over-fitting, and better understand which properties of the data capture the differences between gestures. For every statistics feature type such as "mean", our set initially held five features: three for each accelerometer axis and two for each axis of the gyro sensor. Eliminating related features (three accelerometer or two gyro features) from this initially full set gave a slight improvement in cross-validation accuracy. This resulted in the 19 features shown in Table 4.

4.5 System Performance

To verify that the system would be capable of supporting the intended type of playful interaction, leave-one-out cross-validation was performed using the data in the case for which the robot was not moving. The system was found to be capable of recognizing the set of 13 typical gestures with an accuracy of 77%. Accuracy was calculated by dividing the total number of true positives by the total number of gesture samples.

Detailed recognition results are shown in the confusion matrix in Fig. 6. The system performed well in recognizing some gestures such as Stand, Upside-down, and Back and Forth (indicated in the outlined cells in the figure), but did not do so well for other gestures such as Walk, Inspect, and Fight (highlighted darkly); also, Walk and Inspect were often confused for Balance (highlighted lightly).

For these cases, imperfect recognition was due to the complexity of the task: participants were not told or shown how to perform gestures and as such came up with many interpretations. Fig. 7 shows examples of such confusion. In Fig. 7a, a user simply pushes the robot for Walk (thus appearing similar to Balance to the inertial sensor). As shown in Fig. 7b, users turned the robot left, right, and down,

Table 4 Feature types

	Statistics Feature	Number of Features*
1	Mean	3 (Accelerometer)
2	Standard Deviation	5 (3 Accelerometer, 2 Gyro)
3	Change from start to finish	3 (Accelerometer)
4	Median	3 (Accelerometer)
5	Minimum	3 (Accelerometer)
6	Maximum	2 (Gyro)

*The three features used for accelerometer correspond to the x, y, and z axes; the two features for gyro correspond to the x and y axes

Ground truth / Classifier Output

Ground truth \ Classifier Output	Inspect	Up Down	Lay Down	Stand	Balance	Walk	Airplane	Dance	Upside-down	Rock Baby	Back and Forth	Fight	Hug
Inspect	49	0	3	3	18	9	6	0	0	9	0	3	0
Up Down	0	89	2	0	0	0	3	0	0	3	0	0	3
Lay Down	3	0	86	0	0	0	3	0	0	5	0	3	0
Stand	0	0	0	98	0	0	2	0	0	0	0	0	0
Balance	0	0	2	0	81	10	0	0	0	0	0	5	2
Walk	0	0	0	0	35	41	0	13	0	0	0	11	0
Airplane	8	0	3	3	0	0	83	0	0	3	0	0	0
Dance	3	0	0	0	0	3	0	80	0	0	0	9	5
Upside-down	2	0	0	0	0	0	0	0	98	0	0	0	0
Rock Baby	12	3	0	0	0	3	0	3	0	64	3	3	9
Back and Forth	0	0	2	0	0	0	0	0	0	0	98	0	0
Fight	3	3	0	0	8	3	0	16	0	3	3	58	3
Hug	0	0	0	0	9	0	0	0	0	15	0	12	64

Fig. 6. Confusion matrix for recognizing the 13 target gestures: the three highest true positive scores are shown in outlined cells, the three lowest accuracy scores are shown darkly highlighted, and the greatest sources of confusion are shown lightly highlighted

or even set it down to look at it for Inspect (which would also seem like Balance). Fig. 7c-e and f-h show users performing very different motions for Fight and Hug.

We were also interested in how a robot's own motions might affect recognition accuracy because a robot is expected to itself move during playful motion-based interaction. Leave-one-out cross-validation could not be directly performed for comparison because the motion dataset was three times larger than the non-motion dataset, and the number of samples available for training could affect recognition accuracy. To compare the two cases, the motion set was made to be the same size as the non-motion set by random sampling without replacement, and the process was repeated ten times with resulting accuracies averaged in order to avoid lucky or unlucky draws. This yielded a 21% drop in accuracy: 77% to 56%.

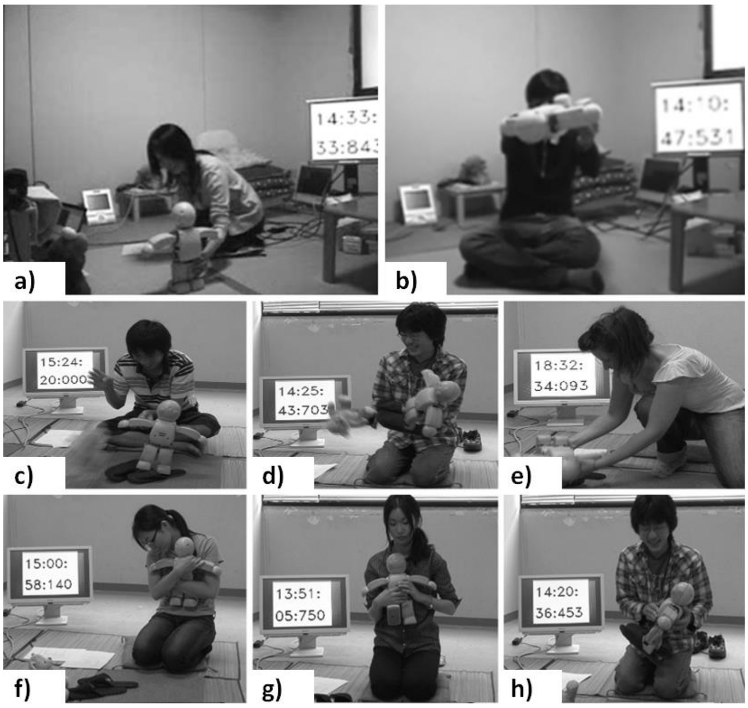

Fig. 7. Confusion in recognition results stemmed from the many ways people chose to interpret and perform gestures: e.g., a) a participant pushes the robot instead of moving its legs for Walk, b) a participant turns the robot in all directions for Inspect, c-e) participants make the robot punch, kick, and body-slam for Fight, f-h) participants hug the robot's front, back, or side incompletely

Most strongly affected were Balance (-74%), Hug (-40%), and Walk (-33%), which were often confused with "noisier" gestures such as Fight.

The reason for this disparity is that the robot's movements in some cases interfered with users' gestures and introduced noise into the inertial data detected by the sensor. An example of this may be seen in Fig. 8. At the top of the figure, a participant performs the gesture Stand which is interrupted by the robot's motion halfway through: (a) the participant starts to raise the robot, (b) a large motion performed by the robot interferes with the gesture (the robot almost slips out of participant's grasp), and (c) the participant recovers, completing the gesture. The bottom of Fig. 8 shows that data from the inertial sensor are significantly affected as a result of such motion; (d) and (e) show data for the non-motion and motion cases respectively.

In summary, evaluation showed that the developed gesture recognition system was capable of recognizing 13 typical full-body gestures with an average accuracy of 77%. The robot's movements did affect recognition accuracy, but the effect was not prohibitive; the system is not required to be perfectly accurate to engage

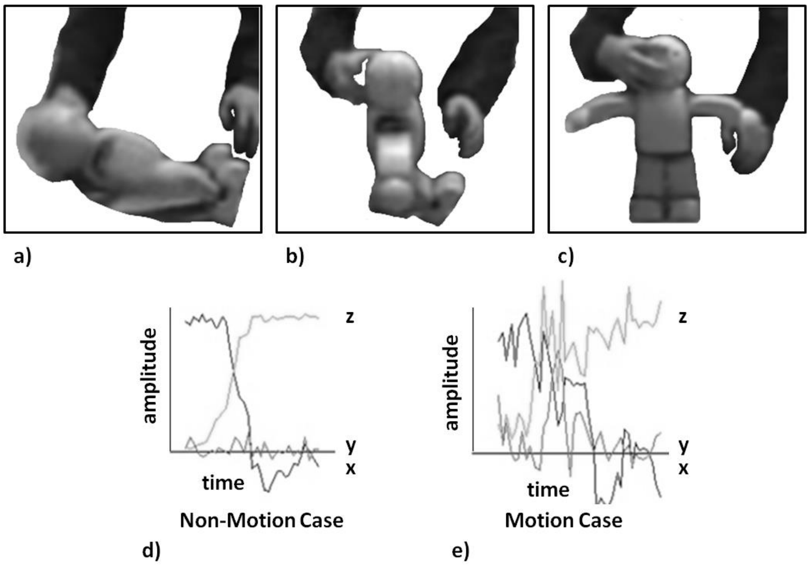

Fig. 8. A robot is expected to move in a playful motion-based scenario, but a robot's own

in the type of playful interaction which is our goal for this work, as long as contingency and consistency are perceived.

5 Designing an Enjoyable Interaction

To interact, a robot requires not only recognition capability, but also behaviors (e.g., motions or sounds) and an interaction strategy to associate recognition input with behavior output: design guidelines for structuring an interaction to provide enjoyment were developed by observing interactions with users.

5.1 Initial Design

It was not initially clear if it would be difficult to provide enjoyment, and, if so, what approach could be followed to yield an appropriate design. To answer these questions, a prototype version of our robot was prepared to hand to users.

This initial design was based on intuitive knowledge drawn from previously designed robots, as discussed in Section 2: the robot's behavior was made to be plausible given its appearance, and the interaction style was kept simple (turn-based with idling). For plausible content, behaviors used by other small humanoid robots were implemented, such as walking and dancing motions (present in E.M.A., Jingle Bell Rock Santa, and Robosapien), as well as laughter and crying sounds, which were used for enjoyable effect in QRIO, Tickle Me Elmo and Pleo. Motions were designed to be smooth, not overly large, and otherwise infant-like.

To implement a simple turn-based interaction style, the robot was set to suggest and respond through its motions when a user performs a gesture, laughing if its suggestions are followed, and crying otherwise. Laughter and crying sounds were limited to one sound each. The robot also was designed to simply suggest the most recent gesture the user performed, which could positively reinforce the user's behavior and create an enjoyable mood. As well, the robot was set to idle if a user did not interact (to appear as if it had a life of its own). Instructions before playing with the robot were kept simple to avoid boring participants or telling them how to play: e.g., "play freely with the robot", or "the robot can perceive when its body is moved".

5.2 First Interactions with Users

When the robot was given to participants to play with one-on-one, we were surprised to observe that interactions were volatile and unstable, and often failed. Fig. 9 shows several specific examples of failures in initial interactions. In (a), Sponge Robot's suggestion motions and sounds are not noticeable to a bored user, who slowly turns the flailing robot to observe its backside. Sponge Robot strikes a participant on the wrist in (b) as part of a motion response to being rotated; this participant started the interaction with a bright smile but no longer seemed enthusiastic at the end. (c) shows a participant tapping Sponge Robot, trying to stop it from idling and get some reaction—any reaction—from the robot. In (d), a user lunges to save the robot from suddenly falling.

The failure of the initial design to provide enjoyment to solitary users suggested that a more appropriate design was required. However, users' feedback on how to improve the system did not indicate how to achieve such a design: comments often contradicted (e.g., the robot should be more docile or wilder) or involved extra modalities such as vision or speech. Thus, participants' suggestions did not elucidate how to improve our design.

On the other hand, it was clear when interactions failed, and feedback could be obtained regarding causes. Therefore, typical failure "patterns" were identified from observation, corresponding causes were found by interviewing users, and a model was proposed to avoid failures. The usefulness of a similar pattern-finding methodology for HRI has been expounded in [23], and applied in [37].

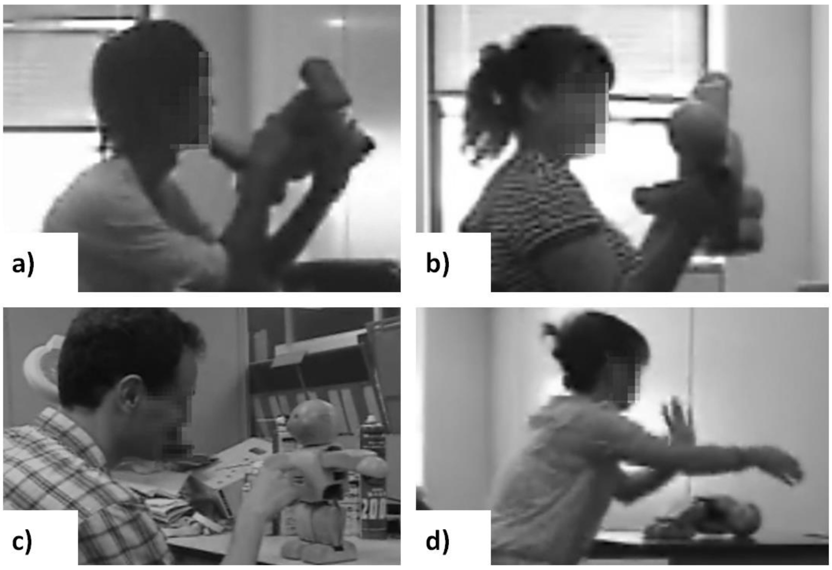

Fig. 9. Some scenes showing failures in initial interactions

5.3 Identifying Typical Failure Patterns

Pattern-finding was conducted by adopting a bottom-up approach. First, failing points in initial interactions were identified by the experimenter; failures which occurred in a similar form for more than one user were then grouped, yielding seven preliminary typical patterns [7]. Some similarities existed within these preliminary patterns; recoding resulted in three typical failure patterns: “Meaningless motions”, “Robot ignores me”, and “Just moving the robot”. In “Meaningless motions”, the robot’s output is problematic, whereas in “Robot ignores me”, the robot fails to recognize the user’s input; “Just moving the robot” may result from input or output problems. This similarity arises as a result of the complexity of people’s behavior and, as will be seen below, the existence of a set of common reasons underlying failures.

Knowledge of such causes for failures is required to avoid them; therefore, users’ comments were analyzed. Typical failure patterns, and reasons derived from users’ feedback, are described next.

5.3.1 Meaningless (“bug”) motions

Users often did not respond as intended to the robot’s motions, claiming the robot seemed inconsistent or like a “bug” in its own world, ignoring them and just doing its own motions; sounds were one-pattern, faded into the background, or seemed “cat-like” (although they had been obtained from videos of real babies on

YouTube). Other users saw meaning in the robot's motions, but interpreted them in a negative way: e.g., "When I tried to hug the robot, it pushed me away".

The cause was that users did not understand the intended meaning of the robot's motions. They could not tell if the robot was supposed to be reacting or suggesting, laughing or crying (especially when only a single sound was emitted, without any accompanying motion). The extent of this problem surprised us: one user asked to watch and describe the robot's motions was only correct 25% of the time. For example, a dancing motion was interpreted as emphatic speaking using beat gestures, a walking motion was seen as a request for assistance, and hugging motions were seen as the robot wanting to shake hands or lie down (because the robot raised its arms and leaned forward). For cases in which users interpreted the robot's motions negatively, unclear response motions which could be interpreted badly (e.g., a sudden kicking motion which punished rather than rewarding the user), and suggestions which did not convey a desire to play together from the robot were the cause.

5.3.2 Robot ignores me

We were surprised to see some users hardly interacted at all: e.g. only picking up the robot and putting it back down on the desk. Other users did try to interact and seemed to have an idea of what they would like to do, but moved the robot too slightly or slowly to trigger a reaction, and avoided large gestures such as laying the robot down on the table.

The cause in the former case was complete failure of the robot's responses, suggestions, and the instructions. Users did not derive more reward from observing the robot's responses than its idling motions (e.g., flexing its arm or shifting its weight), suggestions did not entice them to move the robot, and explicit explanation was required for what the robot *cannot* sense, as some participants eschewed the prescribed modality in favor of seeking to communicate via touch, sound, or vision. In the latter case, apprehension that the robot could be damaged, or somehow hurt users with a sudden motion, contributed to this failure; such a feeling is not unreasonable considering that young adult Japanese are not typically used to interacting with robots, and that some people may feel anxious toward robots [33] or worried about holding moving creatures such as cats or dogs. The robot's suggestions did not convey that users could move the robot freely, and lack of responses (feedback from the robot) perpetuated such behavior. Instructions also did not convey that the robot was robust.

5.3.3 Just moving the robot

Users who did interact often moved the robot without a clear purpose and in an incomplete way: e.g., pausing with the robot partially turned. Other users moved the robot continuously and mostly ignored the robot's suggestion and response motions.

Lack of meaningful responses/feedback perpetuated users' behavior, suggestions were not understood, and instructions did not help. For the latter case in particular, the turn-based style for the responses and suggestions failed and users did not understand the pattern of playing "with" the robot.

5.3.4 Comparison with other studies

The failure patterns identified in the current work were compared with four other patterns found in another HRI study, which involved a large wheeled robot approaching a human [37]. Close analogies were found for three of these patterns: "Unreachable" (no interaction occurs) and "Unsure" (a person tries to interact but receives no reaction) correspond to *Robot ignores me*, and "Unaware" (a person does not observe the robot) resembles *Just moving the robot*. However, we did not find a match for "Rejected" (a person does not want to interact) because the users in our study were hired to play with the robot, whereas the other study was based on a field experiment in which participants were free to avoid interacting. This close correlation indicates that the identified failing points may be applicable to other contexts and other robots (e.g., those mentioned in Section 2). Therefore, an approach for avoiding typical failures was desired.

5.4 Design Guidelines

To design resolution strategies, reasons for failures were grouped into categories. Four categories were found to be related to each failure: motions, responses, suggestions, and instructions. Then, guidelines to avoid failures were assembled for each category, based on participants' feedback, previous work, and our own ideas. It is known that guidelines may suffer from being too general (missing many problems) or too specific (extraneous details may hinder application) [28]. A balance was struck, as in [32], by providing a list of general guidelines and specifying how to apply them; these guidelines are described below and summarized in Table 5. Guidelines comprise both components-development and systems-related items, and relate to system architecture as follows: Guidelines 1, 2, and 3 relate to motion creation and planning (control logic). Guideline 4 is implemented alongside motion creation, and comes into play before interactions.

Regarding novelty, some guidelines are largely obvious but important (e.g., timely motions). Others make sense, but may easily be overlooked (e.g., providing opportunities for an unskilled user to move all of a robot's degrees of freedom). Some are not at all obvious (strategies for offering reward or goals for the interaction), and are therefore tested and clarified in Section 6.

5.4.1 Meaningful Motions

Enjoyable motions should first be meaningful. The temptation to use any motions which yield enjoyment should be avoided, as this may result in detrimental misunderstandings over the course of an interaction, such as were noted in Section 5.3.1. A simple mapping does not exist to distinguish meaningful from meaningless motions; therefore we propose a heuristic-based approach as follows: iterative design of motions related to typical user behaviors with testing by users to ensure that intended meanings are conveyed, accompaniment of appropriate sound streams, set-wise assessment by the designer (of uniqueness in different motions, homogeneity in similar motions, and comprehensiveness), timeliness, and non-repetition. First, naïve users who have never seen the robot should be directly shown implemented motions and asked to describe what the robot is doing, and designs refined until a sufficient level of consistency is achieved. The temptation, for simplicity, to avoid using sounds or to use sounds without motions should be avoided; sound streams should accompany motions typically accompanied by sound in the real world (such as yawning, snoring, laughing, or crying) or they will not be understood. Motions should be evaluated by the designer not only in isolation, but as *motion sets*. Similar motions with different meanings (e.g., walking and dancing) should be clearly differentiated; motions which can be triggered in close succession (e.g., such as the constituents of a progressive response) should be viewed together by the designer to check for

Table 5 Interaction Design

	General Guidelines	Concrete Items
1	**Meaningful Motions**	Semantics testing, sound stream pairing, set-level assessment, timeliness, and non-repetition
2	**Rewarding Responses**	*Maximum or Progressive**
3	**Inspiring Suggestions**	*Shifting or Persisting**
4	**Fulfilling Instructions**	Disclosing non-evident capabilities (sensory-behavioral and robustness)

* item which requires additional clarification

unintended slight differences which could convey cues. Furthermore, users will also not understand if they cannot easily cause the robot to move its head, arms, or legs during an interaction; the set of motions designed should incorporate noticeable movements of all of the robot's major appendages. On the systems-level, users will miss or not understand motions which do not start quickly. Also, motions should be "broken up" to avoid repeating a pattern too many times; motion primitives can be used as a basis for encoding and varying motions [27].

For the current work, motions were implemented for seven full-body gestures which seemed to be clearly understood (naïve users were able to state what the robot was doing with an accuracy of 90%). Video clips were used to check how the robot's motions would look like together. Also implemented were energetic, "happy-seeming" motions and wild flailing motions to accompany laughter and crying; volumes were adjusted based on feedback that loud sounds were "scary" and quieter sounds were difficult to hear, before confirming that naïve users found the robot's laughter to be enjoyable.

5.4.2 Rewarding Responses

A robot should "reward" the user for interacting. Large, fast responses to a user's actions can be used to this end; these convey to the user, "What you did really affected me!" As well, users who experienced positive moments in the interaction (e.g. a suggestion to hug from the robot, then a positive reaction after doing so) expressed high opinions of the robot, indicating that such motions can also provide reward.

What is not clear is if responses should always be large and positive; previous work indicated two possible strategies for structuring responses. Robotic toys like Tickle Me Elmo (which was aimed toward children) tend to always show large, loud, "maximum" responses which are exciting and which can provide much enjoyment for short-term interactions. On the other hand, robots intended for longer-term, caring interactions like Paro (which was intended for use with the elderly) typically have some capability for “progressive” responses which are adapted based on the user’s behavior.

Therefore, two designs were implemented. In the first, *maximum reward*, the robot gives its largest and most positive reactions when the user does what the robot wants: in our case, moving its arms, legs, and head while using the loudest portion of its laughter stream. In the second design, *progressive reward*, the robot provides progressively larger and more positive reactions when the user does what the robot wants: first moving only its arms or legs, then its arms and legs, and finally its arms, legs and head, while laughing quietly at first, then with greater volume. The last motion is the maximum reward response. Fig. 10a shows such a reaction, in which the robot flaps its arms, legs, and head, while emitting laughter.

5.4.3 Inspiring Suggestions

Proactive motions performed by a robot can “inspire” users during an interaction by conveying ways to play (“Let’s do this!”) and the robot’s capabilities (“I can play like this”). An example of such a suggestion is shown in Fig. 10b, in which a robot shows it can dance and invites the user to join it in playing in this way. Suggestions should offer less reward than responses (by being smaller, incomplete, or otherwise discordant) to encourage the user to move the robot instead of merely watching. On the systems-level, suggestions can be performed when a user is not interacting; these should be quickly cancelled when a user interacts (it should not be assumed that users will give the robot a “turn”). (For the current study, codebook vectors and activity thresholds were used to determine when the robot was standing and not being moved.) Also, suggestions should not be performed only once or they will not be noticed; and, the robot’s first suggestions should be easy to understand and respond to (corresponding to gestures which can be recognized with few false positives or negatives).

Unclear is how a robot can communicate its desires and capabilities throughout an interaction: a robot could seek to quickly present various options and capabilities, or repeat a particular option several times for clarity. The literature did not provide any indication which option might be better. This resulted in two designs: *shifting suggestions* in which a robot seeks to express various ways to play (selected randomly) and *persisting suggestions*, in which the robot repeats a suggestion more than once. In the latter case, suggestions are shown in a

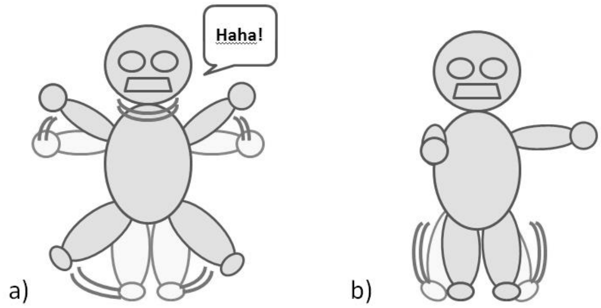

Fig. 10. Motion examples: (a) a large, positive response, (b) a suggestion ("Shall We Dance?")

predefined "easiest-first" order (Up-Down, Lay-Down, Dance, Walk, and Hug); the robot requests a handshake to indicate when its intention changes; and intentions change if a certain maximum time has gone by (two minutes), if the user has responded to the robot's suggestions and completely explored its responses, or if the user turns the robot upside-down for a significantly long time. It should be noted that a bow could have been used in place of a handshake to enhance usability and enjoyment in consideration of our target demographic group's culture; a handshake was chosen for other reasons (we wished to encourage active touching, also ensuring a user understands the robot's intention has changed, and this motion could be detected with the robot's inertial sensor).

5.4.4 Fulfilling Instructions

Future robots will no doubt have greater recognition and expressive capabilities, and people will also have more experience interacting with them; in the meantime, particularly for novel ways of interacting, it is vital that some information is conveyed through instructions to ensure users have all the information they need. This missing information which is not conveyed in motions (responses or suggestions) can be found by interviewing users: e.g., asking what they would tell the next user. In our case, we found we should tell users which modalities a robot can sense, what can be done to the robot without breaking it, and that the robot can respond and suggest.

6 Interaction Structure for Reward and Suggestions

The previous section left two key questions unanswered regarding strategies for providing reward and suggesting ways to play during the interaction. An experiment was conducted to answer these questions and complete the proposed interaction guidelines; additionally, users' comments were analyzed to gain insight into why users perceived enjoyment.

6.1 Participants

20 paid Japanese participants who had never played with our robot before participated in the experiment (9 females and 11 males; average age 20.3 years, SD=2.1 years). These were not the same participants whose feedback was used in developing the design in Section 5.

6.2 Conditions

A 2x2 within-participants factorial design was used, with two factors: the robot's reward and suggestions.

6.2.1 Reward factor

a) Maximum reward condition: the robot moves its arms, legs, and head while using the loudest portion of its laughter stream when a user does what the robot wants

b) Progressive reward condition: the robot moves its arms or legs, then arms and legs, then arms, legs, and head, laughing with increasing volume when a user does what the robot wants.

6.2.2 Suggestions factor

a) Shifting suggestions condition: the robot "scrolls" through various suggestions, when standing on the desk.

b) Persisting suggestions condition: the robot does not show all of its suggestions immediately, but repeats a single suggestion for a while, when standing on the desk.

These four conditions (described in more detail in Section 5) were combined to construct four different robot designs for participants to play with. In all cases, the gestures the robot could recognize were the same. Each participant experienced all four design conditions. The order of the conditions was counterbalanced.

6.3 Procedure

Participants sat at a desk in a partitioned-off space and played four times, each time with a different version of the robot. During the experiment, participants could hold the robot, put it down on the desk, place it on their laps, play with it on the floor, or hold it while standing or moving around the room. The robot performed suggestions when placed upright on the desk or responded if the participants did something to the robot. Participants informed the experimenter when they would like to stop playing.

Before the first session, participants were given instructions, including a handout with contents similar to what might be found on a toy package. For example, "The robot recognizes when its body is moved!", "If you put the robot on the table, the robot will suggest something", and "Please don't throw the robot or set fire to it". Participants were also shown a short video clip of the robot's legs shaking and told this was a hardware bug (servo jittering sometimes occurred due to the large torque affecting the robot's hip motors). To mitigate novelty and order effects, participants were given the chance to hold the robot and trigger a motion by turning the robot. However, we did not indicate the gestures the robot could recognize or how participants should play. After each session, participants filled out a questionnaire. After the fourth session, a short interview was conducted.

6.4 Measures

A questionnaire was used to obtain subjective measurements for perceived enjoyment, as well as factors we thought might contribute to enjoyment. Although perceived enjoyment had been assessed in a previous robotics study [20], the questionnaire used was targeted toward the context of a verbal conversation involving gaze behavior and could not be applied.

Participants answered the following questions using a scale from one to seven:

· *How to Play* – Did you understand/not understand how to play with the robot?

· *Perceived Variety* – How rich/not rich in variety were the robot's responses to your actions?

· *Control* – Did you feel/not feel a sense of control (like you were controlling the flow and contents of the play the way you wanted to)?

· *Intentions* – Did you understand/not understand what the robot was trying to do?

· *Enjoyment* – Was playing with the robot enjoyable/not enjoyable?

6.5 Hypothesis and Predictions

Our hypothesis was that more expressive feedback (varying responses) would be desirable, as robots' motions tend to be limited and always seeing the same response could be boring; and that naïve users would appreciate clear suggestions from the robot, given the novelty of the interaction. Specific predictions were as follows:

Prediction 1: Users would perceive most richness of variety and most enjoyment using the progressive reward condition.

Prediction 2: Participants would best understand how to play, and perceive the most enjoyment with the persisting suggestions condition.

6.6 Results

To analyze the data, a two-way repeated measures analysis of variance (ANOVA) was conducted with two within-subject factors: the robot's reward and suggestions. Questionnaire and ANOVA results are shown in Fig. 11.

We found that progressive reward significantly increased perceived variety ($F(1, 19)=6.0$, $p=.024$, $\eta^2=.240$), and its effect was nearly significant for understanding how to play ($F(1, 19)=4.0$, $p=.059$, $\eta^2=.175$). However, it did not significantly contribute to perceived control, understanding of the robot's intentions and enjoyment (Control: $F(1, 19)=1.1$, $p=.30$, $\eta^2=.055$; Intentions: $F(1, 19)=1.9$, $p=.18$, $\eta^2=.092$; Enjoyment: $F(1, 19)=1.7$, $p=.21$, $\eta^2=.082$). Our first prediction that users would perceive most richness of variety and most enjoyment using the progressive reward condition was only partially supported. For the robot's suggestions, participants rated the persisting suggestions condition significantly higher for all measured items than the shifting suggestions condition (How to Play: $F(1, 19)=37.3$, $p<.001$, $\eta^2=.663$; Perceived Variety: $F(1, 19)=45.2$, $p<.001$, $\eta^2=.704$; Control: $F(1, 19)=10.2$, $p=.005$, $\eta^2=.35$; Intentions: $F(1, 19)=19.0$, $p<.001$, $\eta^2=.50$; Enjoyment: $F(1, 19)=71.1$, $p<.001$, $\eta^2=.789$). Our second prediction, that participants would best understand how to play, and perceive the most enjoyment with the persisting suggestions condition, was

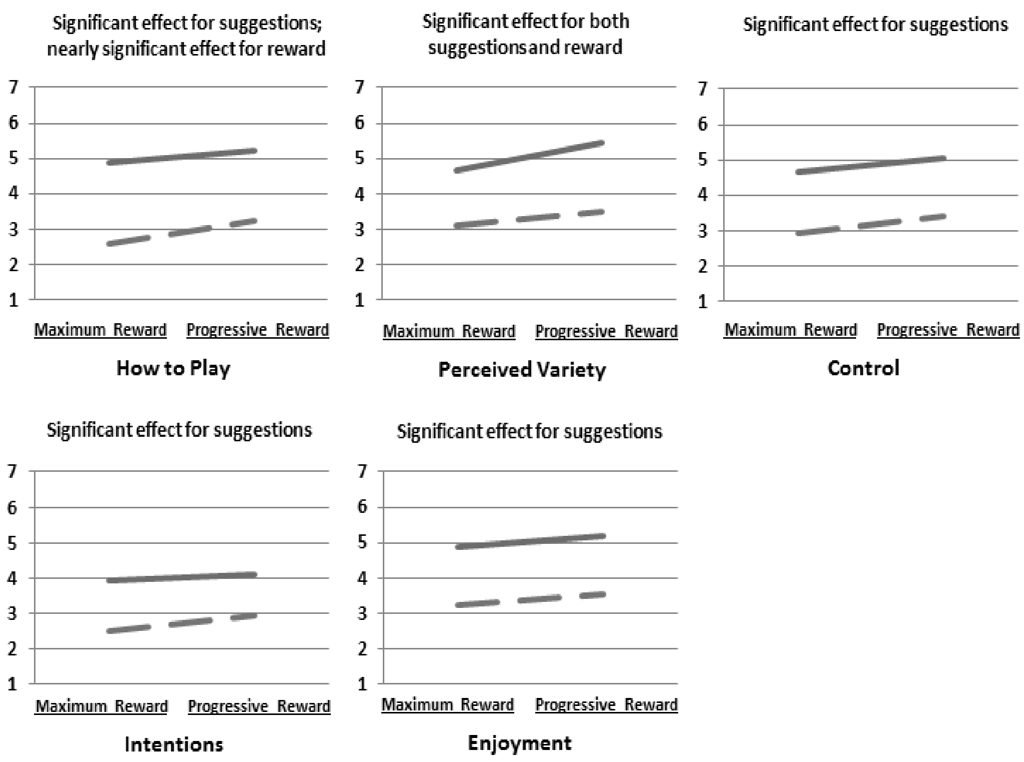

Fig. 11. Average questionnaire scores and ANOVA results for reward and suggestions (solid line: persisting suggestions, dashed line: shifting suggestions)

supported. None of the interaction effects were significant: How to Play: $F(1, 19)=.609$, $p=0.445$, $\eta^2=.031$; Perceived Variety: $F(1, 19)=.717$, $p=.408$, $\eta^2=.036$; Control: $F(1, 19)=.39$, $p=.54$, $\eta^2=.020$; Intentions: $F(1, 19)=.012$, $p=.91$, $\eta^2=.001$; Enjoyment: $F(1, 19)=.000$, $p=1.000$, $\eta^2=.000$).

To gain insight into these results, users' comments during the interviews were analyzed. With regard to reward strategies, one user mentioned that the robot's progressive motions and the changing volume of its voice helped to know how to play, and another user said she lost her confidence in how to play when the robot always gave the same responses; this may explain the nearly significant effect of the progressive condition on understanding how to play. For variety, three users who often triggered the robot's responses mentioned observing variation in the progressive case, two noting that the largeness of response depended on how many times a gesture had been done, and one correctly indicating the implemented three-stage motion progression (arms, to arms and legs, to arms legs and head); another three users mentioned that in the maximum case the robot did the same motions. However, for enjoyment, two users thought the robot only moved its arms in the progressive reward condition, claiming that the robot's responses in the maximum reward condition were richer because they involved the robot’s whole body. We think if users played more than once with the robots or for a longer time that they would trigger the robot’s desired responses more often and experience less enjoyment from always seeing the same response in the maximum reward condition.

Regarding the unexpectedly large difference in users’ evaluations of the two suggestion conditions, users indicated that the shifting suggestions condition was difficult to understand and that they had trouble triggering responses. Eight users mentioned they had difficulty understanding the robot and what it wanted. As well, five users noted the robot sometimes didn't respond to their actions. Three users mentioned the robot's responses were similar each time, two users claimed they did not hear the robot's voice as much, one mentioned not knowing if the robot was happy, and one said the robot's responses were not so large, all indicating that these users did not often observe the robot's large positive responses. But why might this have been the case? Users’ comments indicate that the pace of the interaction was too fast. Three users mentioned that the robot was moving and suggesting much, one said there was more variation, and one mentioned the robot's motions changed too fast; five users also mentioned that they did not at first understand what to do and required time. In the persisting suggestions condition, by contrast, participants were able to try the same gesture several times and confirm that their actions did elicit a contingent response from the robot. One user mentioned that it took two to three tries for him to confirm a connection between his gestures and the robot’s responses. This better understanding of the pattern of interaction helped participants know how to play, allowed them to see more of the robot’s repertoire of motions, and by doing so, provided more enjoyment.

6.7 Sources of Enjoyment

In order to gain insight into the principles associated with users’ perceptions of enjoyment, further analysis of users’ comments during the interviews was conducted. As a result, two sources of enjoyment were found: *playing with the robot in various ways*, and *making the robot “happy”*.

6.7.1 Enjoyment from playing in various ways

One example of a user’s play experience is shown in Fig. 12 and described in Table 6. “Bob” (not his real name) does not immediately understand the robot’s suggestion or how he can play. Due to the robot’s persisting suggestions, however, he has a chance to see the suggestion several times and understand it, and is consequently able to succeed at triggering the robot’s responses to being raised high. But how is this related to his perception of enjoyment?

Comments made by users during their interviews provide some insight. Two users reported that they liked understanding how the robot wanted to play (the robot’s suggestions), and six said they were happy when the robot reacted to something they did (the robot’s responses). A large number of users (14) furthermore recounted specific moments during the interaction which they had enjoyed most: four users like “Bob” enjoyed raising the robot high, four liked the robot’s Stand suggestions (the robot doing push-ups, sit-ups or trying to get up on its own), four mentioned the robot’s dancing, one indicated the robot’s unhappy struggling for Upside-down, and one liked laying the robot down. Thus, it seems like different users enjoyed different ways of playing, and that the design provided enjoyment by offering opportunities to successfully play with the robot in various ways.

6.7.2 Enjoyment from making the robot happy

The robot’s large positive responses to users’ actions also made many users feel as if they had made the robot “happy”. 14 users, without being asked by the experimenter, said the robot had seemed happy, and another four mentioned that the robot had laughed when they played with it. Four users explicitly mentioned that the robot’s happiness had been a result of their actions. Those users who described the robot as happy were asked by the experimenter what had made them feel this way: 13 replied voice or laughter, and seven recounted the robot's body movements (e.g., the robot's hands “flapping”). Finally, the users were asked how the robot’s happiness had affected them. Although two users reported simply observing the robot’s state (“Oh, it's happy”), nine users claimed to have felt happy, good, or to be having fun, and one user felt a sense of achievement. Therefore, the design also appears to have provided some enjoyment in an altruistic sense.

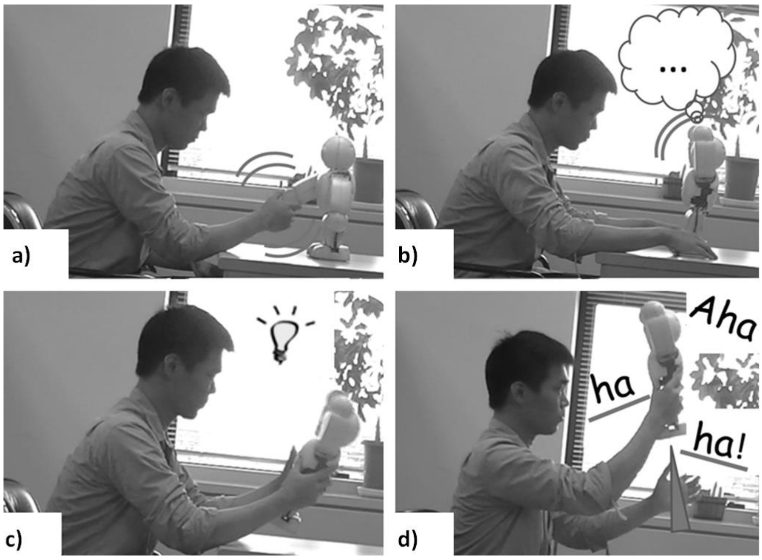

Fig. 12. Playing with the completed system: a) shaking hands at the start b) watching a suggestion from the robot c) interpreting the meaning of the motion d) performing the robot's desired gesture

Table 6 Playing with Sponge Robot

A participant, whom we will call Bob, cautiously watches the little yellow robot on the desk in front of him. All of a sudden, Sponge Robot thrusts its hand forward toward Bob. (Fig. 12a) He reaches out, touching the robot's soft hand gingerly—and the robot shakes his hand up and down, with a small laugh. (Fig. 12b) Sponge Robot points both of its arms toward the ceiling. Bob picks up the robot, turning it in his hands a few times. (Fig. 12c) He appears unsure what the robot's motion means or what it might want him to do, but he decides to try something. (Fig. 12d) Bob hesitatingly raises the robot… Sponge Robot guffaws loudly! Bob has found what the robot wanted him do! He lifts the robot again, this time with more confidence, and then once more, and each time the robot seems even happier, waving its arms and legs and head and howling with laughter.

7 Evaluation

The experiment in the previous section resulted in a completed design, but it remained to be shown that the proposed design really was an improvement over the initial design attempted. A final experiment was conducted to acquire results to support our findings that our system with the proposed design is perceived as more enjoyable than one with a design based only on intuitive knowledge from previous designs.

7.1 Participants

21 paid Japanese users (8 females and 13 males; average age 21.8 years, SD=3.0 years) participated in the experiment.

7.2 Conditions

Participants each experienced the two conditions below, in a counterbalanced order.

a) Naïve design condition: the simple initial design for the robot described in Section 5.1.

b) Proposed design condition: the proposed design with progressive reward and persisting suggestions described in Section 6.2.

To refresh the reader's memory, important concepts from these previous sections have been summarized in Table 7.

7.3 Procedure

Participants played with the robot as described in Section 6.3: they were free to sit at a desk or stand while playing, filled out a questionnaire after each session, and were interviewed at the end. The only difference from before was that the experimenter prompted participants to finish a session if ten minutes passed.

7.4 Measures

The same measures described in Section 6.4 were used ("How to Play", "Perceived Variety", and "Enjoyment").

7.5 Hypothesis and Predictions

We predicted that our proposed interaction design would be evaluated higher than the naïve design for each measured item (for the reasons given in Section 5).

7.6 Results

Questionnaire results for the second experiment are shown in Fig. 13. A one-way repeated measures analysis of variance (ANOVA) was conducted with one factor: robot design (naïve or proposed). Participants rated the proposed interaction design condition significantly higher than the original naïve condition for all three measured items (How to Play: $F(1,20)=26.5$, $p<.001$, $\eta^2=.570$; Perceived Variety: $F(1,20)=23.3$, $p<.001$, $\eta^2=.538$; Enjoyment: $F(1,20)=18.0$, $p<.001$, $\eta^2=.473$).

Table 7 Important Concepts from Previous Sections

Concept (Section)	Summary
Naïve design (5.1)	Based only on intuitive knowledge from previous designs: plausible behavior and simple turn-based interaction flow with idling The robot performs infant-like motions, suggesting and responding after a user performs a gesture; it laughs if its suggestion is met and cries otherwise. The robot performs an idling motion when not interacting and always suggests the last gesture performed by the user. Instructions were kept simple.
Proposed design (5.4)	*Meaningful motions*, *rewarding responses*, *inspiring suggestions*, and *fulfilling instructions*.
Progressive reward (6.2)	The robot moves its arms or legs, then arms and legs, then arms, legs, and head, laughing with increasing volume when a user does what the robot wants.
Persisting suggestions (6.2)	The robot repeats a single suggestion for a while when standing and not being moved by the user.

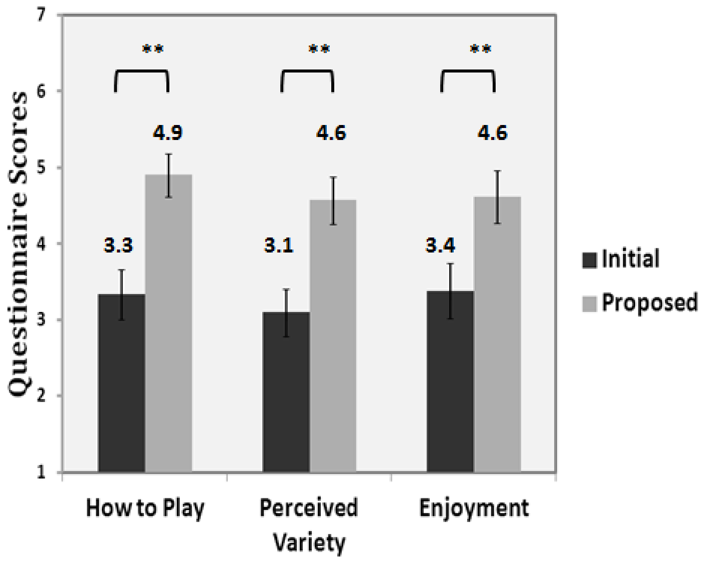

Fig. 13. Averaged scores for each questionnaire item and design condition

In summary, our prediction was supported, showing that the proposed interaction design did provide a more enjoyable interactive play experience than a naïve design.

8 Discussion

8.1 Summary

Important points noted so far regarding providing enjoyment through play with a robot are summarized below:

· Robots should be capable of recognizing people's play behavior; such behavior can be modeled by identifying *typical gestures*. In this research, 13 typical full-body gestures are recognized with an accuracy of 77% using a fixed-sized data window, Support Vector Machines (SVMs), and statistical features.

· A design based only on intuitive knowledge from previous robots (comprising appearance-related behavior and simple turn-based interaction flow with idling) leads to frequent failures in interactions (the robot's motions do not achieve their purpose, and the user interacts insufficiently or without specific aim). To avoid such typical failures, motions should be meaningful, responses rewarding, suggestions inspiring and instructions fulfilling.

· A *progressive reward* strategy for responses increases perceived variety; *persisting suggestions* increase understanding, perceived variety and enjoyment; and users find enjoyment in playing with a robot in various ways and making it happy.

· Using the proposed design described above (instead of a naïve design based only on intuitive knowledge from previous designs) results in increased understanding, perceived variety, and enjoyment.

8.2 The Bigger Picture

We return now for a brief moment to reexamine the existing literature in HRI from the perspective afforded by our study. It's known that people attribute meaning to a robot's behavior. For example, Kozima et al. stated that children interpret "Keepon's responses, although merely simple gestures and sounds, as having communicative meanings" [25]; Michalowski et al. found users ascribe meaning and intentions to their puppet's behavior [29]. It has also been reported that people judge a robot's capabilities not only once but a number of times over the course of an interaction [30]; this suggests intuitively that a robot should be made as understandable as possible in order to avoid clashes of expectations and reality. Nonetheless, both positive and negative examples of such clashes are described. Salter et al. noted one amusing incident in which a robot already on the ground confused a child by asking to be put down; this led to a higher degree of engagement and interaction [36]. Turkle et al. observed that when children expected but did not receive an answer from a robot, some of the children acted gently, whereas others acted in an angry fashion [44].

For motion-based play interactions with a small humanoid robot, our study confirmed the intuition that clarity and understanding are fundamental, important, and difficult-to-attain when seeking to providing enjoyment, and added to this basic idea. User feedback in Section 5 suggested that clarity and understanding are not limited to any one facet of interaction design, but apply to designing motions tested for meaning, responses which show the robot is happy to interact, suggestions which inform and facilitate, and instructions. Results from Section 6 indicated that clear suggestions promoted understanding and significantly increased perceived enjoyment (more than the factor of variety, which we also thought would be important). Experimental results from Section 7 showed that users will not necessarily understand an interaction with a naïve design and highlighted the importance of understanding toward providing enjoyment in the designated context of this study.

These findings could apply to some extent in other contexts as well. It is known in general outside of robotics that people may find patterns where there are none [15, 21]. And, the effects of such clashes can be generally difficult to predict; for example, participants paid less to perform a boring task may describe the task more favorably due to perceived *cognitive dissonance* [13]. Yet, the importance of understanding ("clear goals") is described in Csíkszentmihályi's pioneering work investigating enjoyment in human science [9, 10], although he does not emphasize this above other principles. We suggest that in HRI, the limited communicative capabilities robots possess make this principle even more important than in human-human interactions.

8.2 Limitations and Future Work

Limitations should be noted with respect to demographic group and the form of play investigated. All studies were performed with young adult Japanese who moved the body of a small humanoid robot; other forms of play, such as playing with a ball, were not studied. Nonetheless, we believe the current study offers some insight into how robots in general can provide enjoyment through playful interactions, for the following reasons. The proposed design is scalable and may accommodate more gestures, motions and sounds. Second, the design is not dependent on hardware and can be implemented on robots with different degrees of freedom and sensorimotor capabilities. Furthermore, people will touch complex robots (e.g., capable of speech or vision), for the same reasons that they touch other people: touch is a fundamental modality [1, 16] with powerful and interesting interpersonal effects [50], which is vital for trust, cooperation, and group functioning [26]. More complex robots should not lack the fundamental capability to interact within this simpler context. Thus, we believe the presented work provides a basis for generalization to other interaction scenarios with more complex robot embodiments.

How specifically could the proposed design be used to increase enjoyment from existing products (e.g., one of the robots listed in Section 2)? As described earlier in Section 2 and Section 5.1, existing products, like our initial design, cannot react to many of people’s communicative behaviors and feature simple interaction styles; by recognizing people’s behaviors and structuring an interaction, enjoyment can be provided, as was seen in Section 7. As one trivial but specific example, a robot could laugh when a child picks it up (without the child having to press a button), or clearly communicate a desire to be lifted up (rather than only making a sound or moving and relying on the user to think of a way to play with the robot). Such “social intelligence” has also been shown [20] to yield increased enjoyment, and allows a transition from passive play in which the human must take the initiative (“working to play”) to more active play in which the robot may also assume a proactive role.

Future work will extend the proposed design by investigating how it fares with complex robots capable of vision or speech, larger numbers of participants from a wider range of ages and cultures, and longer-term interactions (when initial novelty effects have faded). Standardized questionnaires for assessing system usability (SUS questionnaire) [2], attractiveness (AttrakDiff) [18], and people's affective appraisals of interactions (PANAS) [49] will provide further insight from various perspectives and facilitate between-study comparison. Recognizing a person’s level of enjoyment in motion-based data and other channels (e.g., smiling or laughing) in conjunction with a suitable cycling process strategy will allow a robot to adapt its behavior to deliver truly enjoyable interactions.

9 Conclusion

In summary, this article reported on how a robot can provide enjoyment through play. The contribution which enables this is a clear set of guidelines for how to structure a robot's behavior during an interaction (see Section 8.1). This knowledge can be used by designers to create robots capable of engaging in enjoyable play interactions, possibly toward integrating robots into domestic and public environments.

Specifically, this article first described how people behave toward a small humanoid robot when seeking to play with it, and introduced a method for recognizing *typical behavior* with 77% accuracy. Next, the article found that a naïve interaction design based only on intuitive knowledge from previous designs is not sufficient for providing enjoyment. Three typical failure patterns and their causes were identified, and four guidelines for enjoyment were proposed: *meaningful motions*, *rewarding responses*, *inspiring suggestions*, and *fulfilling instructions*. An experiment was conducted to complete the design and answer questions for how a robot can provide reward and suggest ways to play. As a result, we found that clear *persisting* suggestions from the robot have a strong positive impact on a playful interaction; that *progressive* reward can increase perceived variety; and that people playing with our robot found enjoyment in *playing in various ways* and *making the robot "happy"*. A second experiment validated the proposed design, showing that the guidelines can be used to provide enjoyment.

Acknowledgements We'd like to thank everyone who helped with this project. This research was supported by the Ministry of Internal Affairs and Communications of Japan and JST, CREST.

References

1. Barnett K (1972) A theoretical construct of the concepts of touch as they relate to nursing. In: Nursing Research, 21, pp 102-110
2. Brooke J (1996) SUS: a quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland AL, Usability Evaluation in Industry. London: Taylor and Francis
3. Brooks AG, Arkin RC (2007) Behavioral overlays for non-verbal communication expression on a humanoid robot. In: Auton. Robots 22(1), pp 55-74
4. Chang C, Hsu C, Lin C (2003) A Practical Guide to Support Vector Classification. In: Taipei 106, Taiwan
5. Chang C, Lin C (2001) LIBSVM: a library for support vector machines. Internet: http://www.csie.ntu.edu.tw/cjlin/libsvm
6. Cooney MD, Becker-Asano C, Kanda T, Alissandrakis A, Ishiguro H (2010) Full-body Gesture Recognition Using Inertial Sensors for Playful Interaction with Small Humanoid Robot. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2010, pp 2276-2282
7. Cooney MD, Kanda T, Alissandrakis A, Ishiguro H (2011) Interaction Design for an Enjoyable Play Interaction with a Small Humanoid Robot. In: 2011 IEEE-RAS International Conference on Humanoid Robots, Humanoids 2011, pp 112-119
8. Cramer H, Mentis H, Fernaeus Y (2010) Serious work on playful experiences: a preliminary set of challenges. In: 'Fun, seriously?' Workshop at CSCW 2010
9. Csíkszentmihályi M, Rathunde K (1993) The measurement of flow in everyday life: Towards a theory of emergent motivation. In: Jacobs JE, Developmental perspectives on motivation.

Nebraska symposium on motivation, Lincoln: University of Nebraska Press
10. Csíkszentmihályi M (1975) Beyond Boredom and Anxiety: experiencing flow in work and play. San Francisco, CA: Jossey-Bass
11. Dautenhahn K, Nehaniv CL, Walters ML, Robins B, Kose-Bagci H, Mirza NA, Blow M (2009) KASPAR - A Minimally Expressive Humanoid Robot for Human-Robot Interaction Research. In: Special Issue on "Humanoid Robots", Applied Bionics and Biomechanics 6(3), pp 369-397
12. Ellis MJ (1973) Why people play. Prentice-Hall
13. Festinger L (1957) A theory of cognitive dissonance. Stanford University Press
14. Fujita M (2004) On Activating Human Communications with Pet-type Robot AIBO. In: Proc. of the IEEE, 92(11), pp 1804-1813
15. Garfinkel H (1967) Studies in Ethnomethodology. Prentice-Hall
16. Geldard FA (1960) Some neglected possibilities of communication. In: Science, 131, pp 1583-1588
17. Goetz J, Kiesler S, Powers A (2003) Matching robot appearance and behavior to tasks to improve human-robot cooperation. In: 2003 IEEE International Workshop on Robots and Human Interactive Communications, RO-MAN 2003, pp 55-60
18. Hassenzahl M, Burmester M, Koller F (2003) AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: Mensch & Computer, Interaktion in Bewegung, pp 187-196 (in German).
19. Hayashi K, Shiomi M, Kanda T, Hagita N (2010) Who is Appropriate? A Robot, Human and Mascot Perform Three Troublesome Tasks. In: 2010 IEEE International Workshop on Robots and Human Interactive Communications, RO-MAN 2010, pp 348 - 354
20. Heerink M, Kröse BJA, Wielinga BJ, Evers V (2008) Enjoyment, Intention to Use And Actual Use of a Conversational Robot by Elderly People. In: Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction, HRI 2008, pp 113-119
21. Heider F, Simmel M (1944) An experimental study of apparent behavior. In: American Journal of Psychology, 57, pp 243–259
22. Huizinga J (1950) Homo Ludens: a study of the play element in culture. Roy Publishers, the Beacon Press, Boston
23. Kahn PH. Freier NG, Kanda T, Ishiguro H, Ruckert JH, Severson RL, Kane SK (2008) Design Patterns for Sociality in Human-Robot Interaction. In: Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction, HRI 2008, pp 97-104
24. Knight H, Chang A, Breazeal C et al. (2009) Robot Design Rubrics for Social Gesture Categorization and User Studies with Children. In: HRI Workshop on Acceptance of Robots in Society
25. Kozima H, Michalowski MP, Nakagawa C (2009) Keepon: A Playful Robot for Research, Therapy, and Entertainment. In: International Journal of Social Robotics 1(1), pp 3-18
26. Kraus MW, Huang C, Keltner D (2010) Tactile communication, cooperation, and performance: An ethological study of the NBA. In: Emotion, 10, pp 745-749
27. Kulic D, Ott C, Lee D, Ishikawa J, Nakamura Y (2012) Incremental Learning of Full Body Motion Primitives and their Sequencing through Human Motion Observation. In: The International Journal of Robotics Research, 31(3), pp 330-345
28. Lewis C, Rieman J (1993) Task-Centered User Interface Design: A Practical Introduction. Internet: HCI Bibliography: Human-Computer Interaction Resources http://www.hcibib.org/tcuid/index.html
29. Michalowski MP, Sabanovic S, Michel P (2006) Roillo: Creating a social robot for playrooms. In: 2006 IEEE International Workshop on Robots and Human Interactive Communications, RO-MAN 2006, pp 587 - 592
30. Movellan JR (2005) An Infomax Controller for Real Time Detection of Social Contingency. In: Proceedings of the 4th International Conference on Development and Learning (ICDL05)
31. Nakatsu R, Rauterberg M, Vorderer P (2005) A New Framework for Entertainment Computing: From Passive to Active Experience. In: Kishino F, Kitamura Y, Kato H and Nagata N (Eds.), Entertainment Computing-ICEC, Lecture Notes in Computer Science (LNCS), 3711, Berlin: Springer, pp 1-12
32. Nielsen J, Molich R (1990) Heuristic evaluation of user interfaces. In: Proc. ACM CHI'90 Conf, pp 249-256
33. Nomura T, Suzuki T, Kanda T, and Kato K (2006). Measurement of Negative Attitudes toward Robots, Interaction Studies, 7(3), pp 437-454
34. Robins B, Dautenhahn K, Boekhorst RT, Nehaniv CL (2008) Behaviour Delay and Robot Expressiveness in Child-Robot Interactions: A User Study on Interaction Kinesics. In: Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction, HRI 2008, pp 17-24

35. Salter T, Michaud F, Dautenhahn K, Letourneau D, Caron S (2005) Recognizing interaction from a robot‘s perspective. In: 2005 IEEE International Workshop on Robots and Human Interactive Communications, RO-MAN 2005, pp 178-183
36. Salter T, Michaud F, Letourneau D, Lee DC, Werry IP (2007) Using Proprioceptive Sensors for Categorizing Human-Robot Interactions. In: Proceedings of the 2nd ACM/IEEE International Conference on Human Robot Interaction, HRI 2007, pp 105-112
37. Satake S, Kanda T, Glas DF, Imai M, Ishiguro H, Hagita N (2009) How to approach humans?: strategies for social robots to initiate interaction. In: Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction, HRI 2009, pp 109-116
38. Shneiderman B, Plaisant C (2004) Designing the User Interface: Strategies for Effective Human-Computer Interaction, (4th Edition). Pearson Addison Wesley
39. Stiehl W, Breazeal C (2005) Design of a Therapeutic Robotic Companion for Relational, Affective Touch. In: 2005 IEEE International Workshop on Robots and Human Interactive Communications, RO-MAN 2005, pp 408-415, Best Paper Award
40. Takayama L, Ju W, Nass C (2008) Beyond Dirty, Dangerous and Dull: What Everyday People Think Robots Should Do. In: Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction, HRI 2008, pp 25-32
41. Tanaka F, Cicourel A, Movellan JR (2007) Socialization between Toddlers and Robots at an Early Childhood Education Center. In: Proceedings of the National Academy of Sciences of the USA (PNAS), pp 17954-17958
42. Tanaka F, Movellan JR, Fortenberry B, Aisaka K (2006) Daily Hri Evaluation at a Classroom Environment: Reports from Dance Interaction Experiments. In: Proceedings of the 1st ACM/IEEE International Conference on Human Robot Interaction, HRI 2006, pp 3-9
43. Turkle S (2006) A Nascent Robotics Culture: New Complicities for Companionship. AAAI Technical Report Series
44. Turkle S, Breazeal C, Scasselati B, Dasté O (2004) Encounters with Kismet and Cog: Children‘s Relationships with Humanoid Robots. In: 2004 IEEE-RAS International Conference on Humanoid Robots, Humanoids 2004
45. Venkatesh V (2000) Determinants of perceived ease of use: integrating control, intrinsic motivation, and emotion into the technology acceptance model. In: Information Systems Research, 11, pp 342–365
46. Viola P, Jones M (2001) Robust real time object detection. In: IEEE ICCV Workshop Statistical and Computational Theories of Vision, Vancouver, Canada
47. Vorderer P, Klimmt C, Ritterfeld U (2004) Enjoyment: At the heart of media entertainment. In: Communication Theory, 14, pp 388-408
48. Wada K, Shibata T (2007) Living With Seal Robots-Its Sociopsychological and Physiological Influences on the Elderly at a Care House. In: IEEE Transactions on Robotics, 23(5), pp 972-980
49. Watson D, Clark LA, Tellegan A (1988) Development and validation of brief measures of positive and negative affect: The PANAS scales. In: Journal of Personality and Social Psychology 54(6), pp 1063–1070
50. Williams LE, Bargh JA (2008) Experiencing physical warmth promotes interpersonal warmth. In: Science, 322, pp 606-607
51. Yi MY, Hwang Y (2003) Predicting the use of web-based information systems: self-efficacy, enjoyment, learning goal orientation, and the technology acceptance model. In: Int J Hum-Comput Stud 59(4), pp 431-449
52. Yohanan S, MacLean KE (2012) The Role of Affective Touch in Human-Robot Interaction: Human Intent and Expectations in Touching the Haptic Creature. In: I. J. Social Robotics 4(2), pp 163-180
53. Vapnik V, Golowich S, and Smola A (1996) Support vector method for function approximation, regression estimation, and signal processing. In: Advances in Neural Information Processing Systems, 9, pp 281–287