Learning Reward Modalities for Human-Robot-Interaction in a Cooperative Training Task

Anja Austermann[1], Seiji Yamada[1, 2]

[1] The Graduate University for Advanced Studies (SOKENDAI), 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430 Japan, e-mail: anja@nii.ac.jp
[2] National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430 Japan, e-mail: seiji@nii.ac.jp

Abstract—**This paper proposes a novel method of learning a users preferred reward modalities for human-robot interaction through solving a cooperative training task. A learning algorithm based on a combination of adaptable pre-trained Hidden Markov Models and a computational model of classical conditioning is outlined. In a training task, where the desired outcome is known by an AIBO pet robot as well as its human instructor, the robot can freely explore human reward behavior. By this method, the robot is able to learn situated, user-specific reward behavior in the different modalities such as gestures, speech and interaction using the robot's built-in sensors. After the training phase, the learned reward behavior can be used as a basis for reinforcement learning of more complex tasks. A preliminary experimental study is presented, which investigates on the effects of restricting possible reward modalities, when teaching a pet robot. The results of the experiments suggest that being able to provide reward freely makes users give more reward compared to a scenario, where reward modalities are restricted. Moreover, the experiments showed that even if a restriction in possible reward modalities is introduced, users tend to give reward that does not conform to the restriction.**

I. INTRODUCTION,

Most applications in the area of service-robots and entertainment-robots require the robot to learn new behaviors or adapt known behaviors to a specific situation, with the help of the user. Different approaches, such as learning by demonstration [1], as well as learning by tutelage [2] can be found in literature. One issue that almost automatically comes up, when a human has to teach a robot is the need to give feedback to the robot telling it, whether it is performing an action correctly or incorrectly.

While feedback, either from a human teacher or from the environment is the key part of any reinforcement learning, it can also be used in various other approaches to supervised learning, providing a binary classification of actions or objects into groups of positive and negative examples.

Although efforts have been made, to characterize typical human reward contingency, such as in [3] and to construct robots or virtual characters that can be taught by a human using reinforcement-learning, the type of reward used throughout the studies is usually a *fixed* stimulus, such as a clicker used for training dogs [4], input over the robot's touch sensors or by clicking buttons to give reward to a virtual character [3].

In other studies, such as [2] a natural language interface processes different kinds of predefined utterances, including positive and negative feedback. In order for reward to be understood by the robot, it must be uttered in a predefined way. Up to now, little attention has been paid to the way people naturally provide verbal, prosody-based and gesture-based reward to support robot learning. Our research objective is, to develop a system that learns to understand natural human reward behavior and use it as a basis for learning.

This paper presents some preliminary experimental results, suggesting that the reward modalities which can be used to instruct a robot have a strong influence on the human's reward behavior and that restrictions in possible reward behavior, such as the restriction to certain sentences to be used for positive/negative reward, or the restriction to using the robot's touch sensors only, for reinforcing the robot, significantly hamper the interaction and make the user give less feedback to the robot. Moreover, the experiments show that it is hard for a user to stick to a designated reward behavior only, even when he is told to do so.

We describe an ongoing study that aims at enabling a robot to acquire an understanding of its user's preferred ways of giving positive and negative reward. This is done by accomplishing a cooperative training task in which the robot can intentionally provoke the user to express different reward behaviors. Details of the training task are given in section III of this paper. We are utilizing a pet robot AIBO. AIBO has the advantage that its dog-like appearance makes humans intuitively interact with it in a similar way like they would interact with a real dog [5]. However, the implementation does not rely on this kind of dog-like interaction but adapts itself to the actually observed user behavior and can therefore be used with other kinds of robots as well.

The robot learns the user's preferred reward modalities instead of having the user learn how to give reward by looking it up in a manual. In such a system, the user can freely and naturally teach a robot in his/her preferable reward modality. This makes human-robot training task more efficient and reduces the user's cognitive load.

The user is allowed to reward or punish the robot freely by verbal utterances, gestures and by touching the robots sensors. The speech utterances are analyzed on the phonetic level as well as on the prosodic level.

The realization of the algorithm for learning the association between certain user behaviors and positive/negative rewards is described in detail at the beginning of *section III* of this paper. The algorithm consists of a low level learning part, based on Hidden Markov Models (HMM) for speech, prosody and gesture recognition, and a high level part based on a mathematical model of classical conditioning, to learn the associations between a certain reward, such as "punish-

978-1-4244-1635-6/07/$25.00

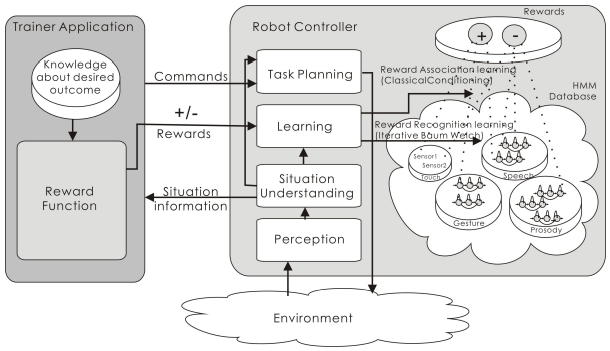

Fig. 1: Overview of the system

ment/negative" or "praise/positive" and the user's verbal and gesticulatory behavior when giving that specific kind of reward, which is represented by the corresponding HMMs.

II. Related Work

Recently, there have been several studies on constructing robots that can be taught by humans in a natural way and investigating the way that humans like to teach robots or other artificial creatures, such as virtual characters.

Lockerd et al. described an experimental setting for assessing human reward behavior and its contingency [3][6]. The participants of the study could give positive as well as negative reward to teach the virtual character Sophie to bake a cake in the "Sophie's World" scenario. Reward could be given by an interactive reward interface that allowed the user to assign any reward on a scale from -1 to +1 either to a certain object or to the world state. The character learned from a human teacher by this kind of reinforcement. In their experiments they found a strong bias towards positive reward and discovered a phenomenon that they described as anticipatory rewards, positive rewards that were assigned to an object that the character has to use in a later step. This kind of reward can be interpreted as guidance for the character.

Other studies focused on teaching a dog-like creature, either an AIBO robot or a virtual character by clicker training, a method, that originates from training real animals, where the sound of a so-called "clicker" is used to give positive reward to the animal.

Kaplan et al. presented a method to train complex tasks to an AIBO robot by performing clicker training [4]. Blumberg et al. showed how to teach a synthetic dog-like character by clicker training based on reinforcement learning [7]. However, his study mainly focused on adapting reinforcement learning to work with a human teacher and on an adequate way to model the state- and action-space for reinforcement learning.

Yamada et al. described an approach to mutual adaptation between a human and an AIBO type robot based on classical conditioning using the Klopf neuron model [5]. While the robot learned to interpret the human's commands, the human found out in the course of the experiments, which stimuli the robot understood in what way.

III. Basics and Outline of the System

The robot learns the association between the concepts "positive reward" and "negative reward" and the corresponding user behaviors. We utilize HMMs to recognize the rewards and apply a mathematical model of classical conditioning to calculate the association.

The theory of classical conditioning was first described by I. Pavlov and originates from behavioral research in animals. In classical conditioning, an association between a new, motivationally neutral stimulus, the so-called conditioned stimulus (*CS*), and a motivationally meaningful stimulus, the so-called unconditioned stimulus (*US*), is learned. The unconditioned stimulus produces an unconditioned reaction (*UR*) as a natural behavior. After completing training, which is done by repeatedly presenting the conditioned stimulus just before the occurrence of the unconditioned stimulus, the conditioned stimulus is able to evoke the same reaction, when it is presented alone. This reaction is called the conditioned reaction. Pavlov found this relationship while he was doing experiments investigating the gastric function of dogs and measuring the amount of their salivation in response to food. At first the dog did not show any reaction to the tone of a bell (*CS*) but when the dog was given food (*US*), it salivated (*UR*). After repeatedly ringing the bell just before feeding the dog, the tone of the bell alone was able to make the dog salivate.

This example serves as a model for our implementation of classical conditioning for learning human reward behavior. An overview of the system is given in Fig. 1. In our implementation, the software is separated into the robot control software itself and a trainer program, which is tailored towards a specific training task, such as a game, or the task described in the section "Experiments". It possesses all information, needed to solve the task and evaluates the current situation that the robot is in. Based on its knowledge about the training task and its desired outcome, it sends commands and provides reward signals to the robot. These reward signals serve as an unconditioned stimulus for the conditioning based learning component.

In analogy to the dog model, the reward from the trainer application can be interpreted as some immediately painful or pleasant signal (*US*). The robot software is able to learn the association between the user's behavior (*CS*) and the reward from the trainer program (*US*). In Fig. 1, positive/negative rewards, which can be given by the trainer application, are denoted by +/-. The associations between the rewards and the HMMs for reward recognition, which are learned by conditioning, are shown as broken lines connecting rewards and HMMs in the right part of the image. After the training phase, the conditioned stimuli that correspond to the user's behavior when punishing or praising the robot, can be used instead of the explicit positive/negative reward given by the trainer application to control the behavior of the robot.

One advantage of using conditioning as an algorithm for learning the associations between positive/negative reward and the user's corresponding behaviors is its rather quick convergence, depending on the learning rate. The learning is inspired by the training methods used for real dogs, where it is widely assumed, that dogs do not understand natural speech based on vocabulary and grammar, as humans do, but are able to associate, for instance, the sound of a command with some action that yields reward when being performed and "understand" a limited set of commands or word-to-object associations in that way.

A: Reward recognition learning

A set of pre-trained HMMs for each of the three modalities speech, speech prosody and gesture is created from pre-recorded audio and video data in order to minimize the need for training samples from each individual user.

The HMM-set for speech recognition contains biphone models as well as pre-trained sequences for the most commonly used words and utterances. The models are based on MFCC-feature-vectors extracted from the recorded speech data. The HMM-set for prosody recognition is based on features like the pitch and energy contour of the speech signal [8][9] that are typically used for recognizing emotion or affective intent in speech. First trials are done with standard left-right HMMs but there is some evidence in literature [10] that deals with emotion recognition from speech, suggesting that ergodic HMMs may be better suited for recognizing prosody. The HMM-set for gesture recognition bases on features describing the relative position of the hands and the face of a person. A similar approach is described in [11].

For each of the three above described modalities, Hidden Markov Models are pre-trained, using an implementation of the Baum-Welch-algorithm. The pre-trained HMMs are stored separately for each modality in a HMM database.

During the training phase with human instruction, observed reward behavior in each of the three modalities, is first processed by the Viterbi algorithm, in order to match it against the pre-trained HMM-Models in the HMM-database. In this stage, the matching is done on isolated "word" level: The full utterance is assumed to correspond to one HMM in the database and matched against every single HMM in the HMM database. The output of the Viterbi algorithm is the best-fitting HMM along with a confidence value.

If the confidence value is above a threshold, then the HMM is trained with the observed utterance/prosody/gesture and the application proceeds to the reward association learning phase. If the confidence value output by the Viterbi algorithm is below the threshold, the Viterbi Recognition is executed again. This time, it is used as a continuous recognition based on an EBNF-like grammar describing the possible HMM-sequences for recognition. The sequence of HMMs that results from this execution of the Viterbi recognition is merged into a new HMM. The new HMM is trained with the utterance/prosody/gesture and inserted into the HMM database for reuse.

The HMM that results from this low-level classification and learning stage, that is, the HMM providing the most accurate available model of the observed utterance/prosody/gesture serves as an input for the high level learning.

B: Reward association learning

The reward association learning phase is based on the theory of classical conditioning, described at the beginning of this section. There are several mathematical theories, trying to model classical conditioning as well as the various effects that can be observed when training real animals using the conditioning principle [12]. The models describe how the association between an unconditioned stimulus and a conditioned stimulus is affected by the occurrence and co-occurrence of the stimuli.

In our study, the Rescorla-Wagner model, which was developed in 1972 and has served as a foundation for most of the more sophisticated newer theories, will be used for the first trials, but more complex models may be employed later on. In the Rescorla-Wagner model, the change of associative strength of the conditioned stimulus A to the unconditioned stimulus $US(n)$ in trial n, $\Delta V_{A(n)}$, is calculated as in (1).

$$\Delta V_{A(n)} = \alpha_A \beta_{US(n)} (\lambda_{US(n)} - V_{all(n)}) \quad (1)$$

α_A and $\beta_{US(n)}$ are the learning rates dependent on the conditioned stimulus A and the unconditioned stimulus $US(n)$ respectively, $\lambda_{US(n)}$ is the maximum possible associative strength of the currently processed CS to the $US(n)$. It is a positive value if the CS is present when the US occurs, so that the association between US and CS can be learned. It is zero if the US occurs without the CS. In that case, $\Delta V_{A(n)}$ becomes negative. Thus, the associative strength between the US and the CS decreases. $V_{all(n)}$ is the combined associative strength of all conditioned stimuli towards the currently processed unconditioned stimulus. The equation is updated on each occurrence of the unconditioned stimulus for all conditioned stimuli that are associated with it.

In our study, the learning rates for conditioned and unconditioned stimuli are fixed values for each modality but can be optimized freely. They determine how quickly the algorithm converges and how quickly the robot adapts to a change in reward behaviour. The maximum associative strength is set to one, in case the corresponding CS is present, when the US occurs, zero otherwise. The combined associative strength of all conditioned stimuli towards the unconditioned stimulus can be calculated easily by summarizing the pre-calculated association values of all the CS towards the US.

The reward association learning is practically realized by maintaining a table, containing the associative strength between each HMM (CS) and the representations of "positive reward" and "negative reward" (US). In order to avoid the HMM database for low level learning to grow too large, pruning is employed to remove complex HMMs that are not reused. When the associative strength of one of the HMMs to its corresponding reward falls below a threshold, the HMM is removed from the database.

C: The training task

The robot learns its user's preferred reward modalities in a cooperative training task. Both, the robot and the user know the desired outcome of the task. For example, the task could be outlined as a first step to adapt the robot to its user in the user-manual of a newly bought pet-robot. The user's job is, to instruct the robot in a way, that it is able to solve the training task. He also needs to give helpful feedback to the robot. The user, who is not aware of the robot knowing the desired outcome of the task, can instruct the robot freely by using speech, gesture and touch.

In the training task, the possible order of steps leading to the goal must be fixed, so that the robot knows what to do and whether it can expect positive or negative rewards, at any given time without understanding its user's instructions. Moreover, a suitable task must contain goals and sub-goals and allow the robot to perform different kinds of desired actions and mistakes in order to make the user show different

Fig. 2: Experimental setting

kinds of reward behavior. A very basic task, that fulfils these requirements is picking up a certain object and delivering it to a designated place.

As an additional requirement to the execution of the task, the percentage of correct behavior needs to increase in the course of the training, to give the user the impression that the robot is actually learning, in order to keep up his motivation to carry on with the training.

For the time being, the system interprets all utterances and gestures from its user that occur in a situation, where reward is expected, as rewards. Although the contents of the utterance may be the repetition of an instruction, its prosody may contain information on whether the user approves of AIBO's performance, or not, i.e. positive/negative reward.

IV. Implementation

As previously mentioned, the system, presented in this paper is currently under development. The robot, used in this study, AIBO, is a four-legged dog-shaped pet-robot that has roughly the size of a cat. The control of the robot is based on the AIBO remote framework [13]. For the HMM based low level learning stage, the Hidden Markov Model Toolkit HTK [14] is used. As the built-in camera of the AIBO is not able to observe the scene and the user at the same time and poses additional difficulties to the processing by constantly changing its position and viewing direction, three additional cameras are used: one camera for recording the scene and a pair of cameras for stereo-based gesture recognition. In order to avoid noise for the speech and prosody recognition, a close-talking microphone is used for speech recording.

V. Experiments

In a preliminary experimental study, we investigated on the effects of a restriction in reward modalities, like it is implemented in most studies, focusing on reinforcement learning with a human teacher. Based on this information, we attempt to draw conclusions concerning the feasibility and usefulness of learning reward modalities in advance by the means of solving a cooperative training task, prior to the actual implementation.

In the experiments, a total of 109 minutes of video data were gathered from four participants, interacting with an AIBO pet robot. Altogether 141 rewards were given by the participants, 64 of which were positive, 77 negative. All four participants were male graduate students aged 25-35 and experienced computer users but had no previous experience in interacting with entertainment robots or service robots. Throughout this preliminary study the robot was fully remote-controlled, and a Wizard-of-Oz-scenario was applied.

A: Experimental Setting and Instruction

The experimental setting can be seen in Fig. 2. The AIBO was equipped with a shovel attached to the front of its body, to enable it to move objects easily without having to pick them up with its mouth.

The participants were told to instruct the robot to approach one of four differently colored bricks and move it to its appropriate place. The places, that the robot had to deliver the objects to, were marked by four A3 sized sheets of paper with differently colored frames. Each participant received a deck of cards showing which object to move and where it should be placed. One such card, representing the instruction to put the yellow object onto the red sheet of paper, is shown in Fig. 3. Every participant was given a short introduction on the concept of reinforcement and was told to give positive and negative reward to the robot, i.e. to praise and punish the robot, when necessary.

Three different reward principles were introduced to each participant throughout the course of the experiment. The first reward principle was introduced before the first trial, the following two reward principles were introduced after three successful trials and after six successful trials. The order of reinforcement principles was changed for each user to avoid sequence effects. The three different reinforcement principles will be referred to as “touch”, “recorded” and “free” reward throughout the remainder of this paper.

The following instructions were given to the participants describing the different reinforcement principles:

Touch: If you want to give *positive reward* to the robot, touch its *head sensor*.

If you want to give *negative reward* to the robot, touch its *back sensor*.

Recorded: Please decide in which way you want to give *positive/negative reward*. You can choose any combination of spoken words, gestures as well as touching the head and back sensors of the robot.

Before starting to instruct the robot, please have your choice of behavior for positive/negative reward recorded. Please stick to the chosen behavior whenever you want to give positive/negative reward to the robot. Please do not change your reward behavior after the beginning of this experiment

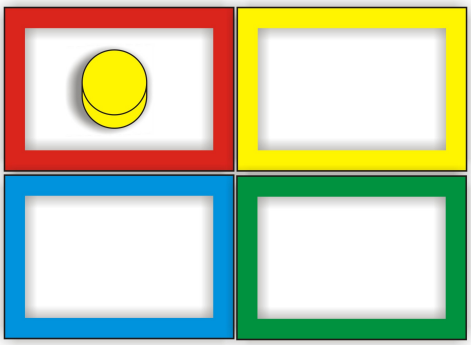

Fig. 3: Example instruction card

Free: The choice of positive/ negative reward is up to you. Please give the kind of feedback you like, depending on the situation freely using voice, gestures or by touching the robot's head and back sensors.

B: Robot control in the experiment

During the experiments we remote-controlled the AIBO to perform the task of moving an object to its designated place, but make some mistakes in the fulfillment of the task in order to receive positive as well as negative reward from the user. The following actions were considered positive because they led to the correct fulfillment of the task. Therefore, the actions are likely to result in a positive reward from the user:

- **picking up** the correct object
- **delivering** the object to the correct target
- **recovering** from a mistake

The robot made three different types of mistakes. Those actions were considered negative, as they hamper or slow down the fulfillment of the task. Therefore, negative reward is expected from the user:

- **"clumsy" mistake:** The robot tries to pick up an object with its shovel but misses it. The robot tries to deliver an object to a target but misses it.
- **"misunderstanding" mistake:** The robot walks into the wrong direction towards an object or target.
- **"lazy mistake":** The robot sits or lies down or walks very slowly.

As the different kinds of mistakes were expected to have an effect on the reward behavior of the participants, the three types of mistakes committed by the robot were balanced throughout the trials.

The reward from the user in the "touch" and "recorded" setting could be classified into one of four categories:

- **correct reward:** the designated reward was given. Repetition of correct reward, such as saying "good dog ... good dog" was still considered correct reward.
- **plus-reward:** the designated reward and some additional reward were given to the robot. (e.g. touching the robots' head sensor and saying "good robot", in the "touch" reward principle)
- **incorrect reward:** the designated reward was not given but replaced by some different reward
- **no reward:** Although one of the above described positive/negative actions was performed by the robot, no reward was given at all.

Only utterances/actions containing an implicit or explicit evaluation like "okay", "good", "bad", "no", etc. were counted as positive/negative rewards. Repetition and rephrasing of commands were analyzed separately.

C: Questionnaire

After the experiment, the participants were asked to fill a short questionnaire on their experience and on their subjective rating of the different reinforcement principles.

Marks from 1 to 5 could be given to rate the statements in the questionnaire, 1 meaning "absolutely agree", 5 meaning "absolutely disagree. The results can be found in Table 1. The values in each column are mean and standard deviation

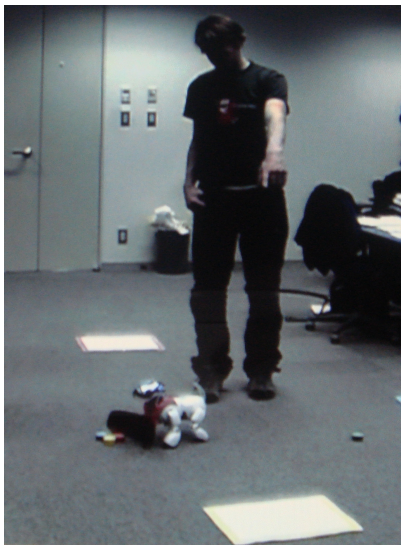

Fig. 4: Image captured from one of the videos of the experiment

of the given responses. In the free response part, the participants could utter their comments concerning the experiments freely. One interesting remark, which was put forward by two of the four participants, was that, in real world, they would not want to give reward to a robot for just doing its job, especially, if it is just fulfilling a rather simple service task. This may be a cause for the bias towards negative rewards, found in the experiments.

D: Results

The users' interaction with the robot was recorded on video and analyzed after the experiments. As the number of participants in this preliminary study is rather small a thorough statistical analysis is not possible. Thus, more systematic experiments with a larger number of participants will be necessary. However, the results can be understood as a rough direction of what effects we have to expect when developing more or less restricted reward modalities for a robot.

The most surprising result turned out to be that even though the participants were explicitly told to stick to the designated reward behavior and to give no different kind of reward in the "touch" and "recorded" reinforcement principle, none of the participants actually provided reward in the designated way only. Out of 44 rewards given by the participants with the "touch" reinforcement principle, there were 19 correct rewards, 17 plus-rewards, most of which adding a speech utterance such as "very good!" to touching the sensors and 8 incorrect rewards, mostly providing a speech utterance

Table 1: Results from the questionnaire

touch	recorded	free
I was able to instruct the robot in a natural way		
4.35 (0.6)	3.25 (0.3)	1.25 (0.3)
I would like to give feedback to a real world service robot in the same way		
4.5 (0.6)	2.0 (0)	1.5 (0.3)
Throughout the experiment, I was always sure about my next step to instruct the robot		
2.5 (1.2)	1.75 (0.6)	1.5 (0.4)

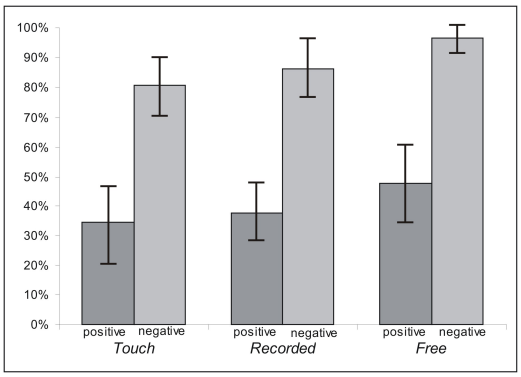

Fig. 5: Percentage of positive and negative behaviors of the robot that result in positive/negative reward

only instead of touching the sensors.

Out of the 48 rewards given by participants with the "recorded" reinforcement principle, there were 23 correct rewards, 18 plus-rewards and 7 incorrect rewards. The recorded preferred rewards differed between all four users, but were all including verbal utterances. One of the users additionally decided to clap his hands for positive reward, like a clicker, and one participant decided to underline verbal positive reward by touching AIBO's head sensor.

The effect of the different reward principles on the frequency of giving rewards can be seen in Fig. 5. The diagram compares the percentage of positive and negative behaviors of the robot, like picking up an object or "misunderstanding" an instruction, in response to which a reward was given by the user, between the three reinforcement principles. The percentage of positive reward is lower because the users typically did not give positive reward to all of the positive actions performed by the AIBO. The only positive action that was always rewarded, was delivering the correct object to the correct place but the amount of positive reward given for approaching the correct object and recovering from errors differed largely between the different participants. This is also the reason for the larger standard deviation in case of positive rewards. However, even the total numbers of 64 positive and 77 negative rewards given throughout all the experiments show a bias towards negative rewards.

In case of "touch" reward, there was reward given for 34.7 percent of the positive and 80.5 percent of the negative actions performed by the robot. In case of "recorded" reward, both reward percentages are slightly higher with 37.6 percent positive and 86.3 percent negative rewards. When it comes to "free" reward, the percentages both increase by roughly 10 percent compared to "recorded" reward. When reward could be given freely, 47.9 percent of the robot's positive and 96.4 percent of the robot's negative actions were rewarded.

An additional finding from the experiments was that apart from giving explicit reward, the users tended to repeat their previous instructions in situations when either positive/negative reward was expected, either instead of giving a reward or in addition to a reward. This behavior was not noticeably affected by the choice of the reward principle.

The kind of mistake, committed by the robot, also affected the choice of reward behavior given by the participants. This was not only visible in the "free" reward principle but also became obvious in the "touch" and "recorded" reward principles, where especially "lazy" mistakes, which were probably rather unexpected to the participants, tended to cause punishment behavior different from the designated one.

However, in order to make the training task, applied for this experiment, feasible for an autonomous training of the robot, some restrictions, especially in posture and position of the human instructor need to be considered. Apart from that, the amount of interaction-free task-execution time is rather high due to the slow walking speed of AIBO. Optimizations to the task that allow more frequent interaction are needed to gather the necessary amount of training data in a short time.

VI. Conclusion

This paper proposed an approach to deal with variable human reward behavior by cooperatively solving a training task and an approach to learn to understand human reward expressed by speech, prosody and gesture. A two-staged learning algorithm, which can be applied to the observations of the robot during the training task, is outlined.

The findings from the experiments indicate that users react quite sensitive to restrictions in acceptable reward behavior, what becomes visible in a decrease of rewards given as well as the utilization of incorrect rewards. These findings suggest that there is a need for techniques that allow a robot to process reward given freely by the user.

VII. References

[1] B. Wrede, M. Kleinhagenbrock, J. Fritsch, "Towards an Integrated Robotic System for Interactive Learning in a Social Context", Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, Oct, 2006

[2] C. Breazeal, A. Brooks, J. Gray, G. Hoffman, J. Lieberman, H. Lee, A. Lockerd, and D. Mulanda. "Tutelage and collaboration for humanoid robots." International Journal of Humanoid Robotics, 1(2), 2004

[3] A. L. Thomaz, G. Hoffman, and C. Breazeal. "Reinforcement Learning with Human Teachers: Understanding how people want to teach robots." In Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 2006.

[4] F. Kaplan, P.-Y.-Oudeyer, E. Kubinyi, A. Miklosi "Robotic clicker training." Robotics and Autonomous Systems 38(3-4), 2002, 197-206

[5] S. Yamada, T. Yamaguchi, " Training AIBO like a dog - preliminary results", Proceedings of the IEEE Conference Robot and Human Interactive Communication, 2004. ROMAN 2004, 431- 436

[6] A. L. Thomaz, and C. Breazeal. "Reinforcement Learning with Human Teachers: Evidence of feedback and guidance with implications for learning performance." In Proceedings of the 21st National Conference on Artificial Intelligence (AAAI), 2006.

[7] Bruce Blumberg, Marc Downie, Yuri Ivanov, Matt Berlin, Michael Patrick Johnson, Bill Tomlinson, "Integrated Learning for Interactive Synthetic Characters" Proceedings of the SIGGRAPH 2002

[8] C. Breazeal, Recognition of Affective Communicative Intent in Robot-Directed Speech Autonomous Robots Volume 12 , Issue 1, January 2002, 83 - 104

[9] Albino Nogueiras, Asuncion Moreno, Antonio Bonafonte, José B. Marino, "Speech Emotion Recognition Using Hidden Markov Models" Proceedings of Eurospeech 2001

[10] T. L. Nwe, S. Foo, S. Wei; L. De Silva, "Speech emotion recognition. using hidden Markov models", Speech communication 41,4, 2003

[11] Jie Yang, Yangsheng Xu, "Hidden Markov Model for Gesture Recognition", The Robotics Institute, Carnegie Mellon University, CMU-RI-TR-94-10, 1994

[12] C. Balkenius and J. Morn. "Computational models of classical conditioning: a comparative study." Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior, 1998

[13] AIBO Remote Framework http://openr.aibo.com

[14] S. Young et al., "The HTK Book" HTK Version 3, 2006 http://htk.eng.cam.ac.uk/