

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Learning to Understand Expressions of Approval and Disapproval through Game-Based Training Tasks

Anja Austermann and Seiji Yamada

The Graduate University for Advanced Studies (SOKENDAI)

National Institute of Informatics

Japan

1. Introduction

One of the most important factors for infants' learning is positive and negative feedback from their caregiver. In a similar way, learning robots can also use the feedback from their user as a basis for learning and adapting to the user's preferences. A popular example of learning through user feedback is reinforcement learning with a human teacher (Thomaz & Brezeal, 2006), but many applications of learning by positive and negative examples with assistance from a user or behavior adaptation and refinement (Kim & Scassellati, 2007) require an understanding of the user's expression of approval and disapproval. This paper focuses on enabling a robot to learn to understand natural, multimodal approving or disapproving feedback given in response to the robot's moves.

Humans express approval and disapproval toward a robot through different channels, such as words, prosody, gestures, facial expressions and touch. Most work on understanding approval and disapproval has been done with single-modal approaches based on prosodic information from speech signals such as intonation, pitch, tempo, loudness and rhythm (Breazeal, 2002) (Kim & Scassellati, 2007). However, we assume that integrating multiple modalities improves the reliability of the recognition and allows the system to adapt to the individual preferences of the user. We determined the modalities to implement in our system through a user-study. It is described in detail in (Austermann & Yamada, 2008). We found, that speech was by far the most frequently used modality, when giving feedback to an AIBO robot. 78.37% of all feedback was given by speech. It was followed by touch, which was used for 20.92% of the feedbacks. Gesture was applied for giving instructions, but did not play a significant role for giving feedback and was only used in 0.71% of the cases. Therefore, in addition to prosody, we focus on the contents of the speech utterances as well as on interaction through the touch sensors of the robot. We did not integrate the recognition of facial expressions, because we wanted the users to move around freely and interact naturally. A facial expression recognizer would have restricted the users' movements by requiring them to look straight into a camera.

In order to learn to interpret user feedback, our system utilizes a biologically-inspired two-staged learning method which is modeled after basic learning processes in humans and

Source: Advances in Human-Robot Interaction, Book edited by: Vladimir A. Kulyukin,
ISBN 978-953-307-020-9, pp. 342, December 2009, INTECH, Croatia, downloaded from SCIYO.COM

animals. It combines unsupervised training of Hidden Markov Models (HMMs), which models the stimulus encoding occurring in natural learning and clusters similar observed user feedbacks, with an implementation of classical conditioning that associates the trained HMMs with either approval or disapproval. The combination of supervised and unsupervised learning as well as specifically designed training tasks allow our system to learn interaction without requiring any transcriptions of training utterances and without any prior knowledge on the words, language or grammar to be used. As a model of the top-down processes, which occur in human learning, we use the associations learned in the conditioning stage to integrate context information when selecting the best HMM for retraining. This is done by adding a bias on models, that are already associated with approval or disapproval depending on what feedback is expected based on the state of the training task.

Adaptation of a robot to a user is done in a training phase before actually using the robot. The training tasks are designed to allow the robot to anticipate and explore the user's feedback. During the training phase, the robot solves special training tasks in cooperation with the user. The tasks are modeled to resemble simple games. The training phase is inspired by the Wizard-of-Oz principle, aiming at giving the user the feeling that the robot adequately reacts to his or her commands in a stage, where the robot actually does not understand the user. However, the training can be performed without remote controlling the robot because remote controlling would be infeasible for actually training a newly bought service robot. Instead, the tasks are designed to ensure that the robot and the user share the same understanding of whether a move is good or bad. This way, the robot is able to anticipate the user's feedback and instructions and can explore its user's expressions of approval and disapproval by deliberately executing good or bad moves. As a result, natural, situated feedback can be observed and learned.

In the experiment, we use "virtual" tasks. The robot plays on a computer-generated game board which is projected from the back to a white screen. This way, we do not need to rely on the potentially erroneous processing of sensor data for determining the state of the task. Further explanations on the training tasks are given in section 3.

2. Related work

There has been a great deal of work on adapting robots to their users, understanding human expression of affect and emotion (Breazeal, 2002) (Thomaz & Breazeal, 2006) processing natural language (Iwahashi, 2004) (Kayikci et al., 2007) and learning through human feedback in recent years. One example that is particularly related to our work is presented in (Kim & Scassellati, 2007). Kim and Scassellati described an approach to recognize approval and disapproval in a Human-Robot teaching scenario and used it to refine the robot's waving movement by Q-Learning. They employed a single-modal approach to discriminate between approval and disapproval based on prosody.

Learning the connections between words and their meanings through natural interaction with a user has been researched upon in the field of language acquisition

Iwahashi described an approach (Iwahashi, 2004) to the active and unsupervised acquisition of new words for the multimodal interface of a robot. He applied Hidden Markov Models to learn verbal representations of objects and motions, perceived by a stereo camera. The learning component used pre-trained HMMs as a basis for learning and the robot interacted with its user in order to avoid and resolve misunderstandings.

Kayikci et al. (Kayikci et al., 2007) utilized Hidden Markov Models and a neural associative memory for learning to understand short speech commands in a three-staged recognition procedure. First, the system recognized a speech signal as a sequence of diphones or triphones. In the next step, the sequences were translated into words using a neural associative memory. The last step employed a neural associative memory to finally obtain a semantic representation of the utterance.

In the same way as the approaches, outlined above, our learning algorithm attempts at assigning a meaning to an observed auditory or visual pattern using HMMs as a basis. However, our system is not trying to learn the meaning of individual words or symbols, but focuses on learning patterns expressing a feedback as a whole. Moreover, our proposed approach is not limited to a single modality but tries to integrate observations from different modalities.

For learning associations between approval or disapproval and the HMM representations of the observed user behavior, classical conditioning is used in our system. Mathematical theories of classical conditioning were extensively researched upon in the field of cognitive psychology. An overview can be found in (Balkenius & Moren, 1998). The relation of classical conditioning to the phase of learning word meanings in human speech acquisition has been postulated in the book "Verbal Behavior" by B. F. Skinner (Skinner, 1957) and has been adopted and modified by researchers in the field of behavior analysis. An explanation of the processes involved in learning word meanings by conditioning is described by B. Lowenkron in (Lowenkron, 2000).

There have been different approaches to use classical conditioning for teaching a robot, such as in (Balkenius, 1999). However, to our knowledge our proposed approach is the first one to apply classical conditioning to acquire an understanding of speech utterances and integrating multimodal information about user behavior in Human-Robot-Interaction.

3. Training tasks

We propose a training method that allows the robot to explore and provoke approving and disapproving feedback from its user. Our learning algorithm does not depend on the way, training data is recorded. However, we found in an exploratory study (Austermann & Yamada, 2007) that natural feedback, given during actual interaction with a robot in a similar task differs from feedback that a user would record in advance. Therefore, we implemented a training method that uses "virtual" games and allows the robot to explore its user's way of giving feedback and learn actual, situated feedback during realistic interaction.

The robot is supposed to learn to understand the user's feedback in a training phase. This implies that by the time of the training it cannot actually understand its user. However, in order to ensure natural interaction, it needs to give the user the impression that it understands him or her by reacting appropriately. This is done by designing the training task in a way, that the robot can anticipate the user's feedback by knowing which moves are good or bad. If the task ensures, that the user can easily judge whether the robot performed a good or a bad move, the robot can expect approving feedback for good moves and disapproving feedback for bad moves. This way the robot can deal with instruction from the user without actually understanding his or her utterances and can freely explore and provoke its user's approving and disapproving feedback. Our training phase consists of training tasks which were designed based on this principle. The tasks are based on easy

games suitable for young children. In the experiments, the participants were asked to teach the robot, how to correctly play these games using natural feedback.

An issue that we became aware of during preliminary experiments is the very limited ability of the AIBO robot to physically manipulate its environment and to move precisely. The possibility of not detecting errors, such as failing to pick up or move an object, poses a risk for misinterpreting the current status of the task and learning incorrect associations. So we decided to implement the training task in a way that the robot can complete it without having to directly manipulate its environment. We use a “virtual playfield” which is computer-generated and projected from the back to a white screen. The robot shows its moves by motion and sounds. It retrieves information directly from the game server using the AIBO Remote Framework. This way we can ensure that the robot is able to assess its current situation instantly, anticipate the user's next feedback or instruction correctly and associate the observed behavior correctly with approval or disapproval.

The following tasks were selected to be used in our experiments, because they are easy to understand and allow a user to evaluate every move instantly. We selected four different tasks in order to see whether different properties of the task, such as the possibility to provide not only feedback but also instruction, the presence of an opponent or the game-based nature of the tasks influence the user's behavior. We implemented them in a way that they require little time-consuming walking movement from the robot.

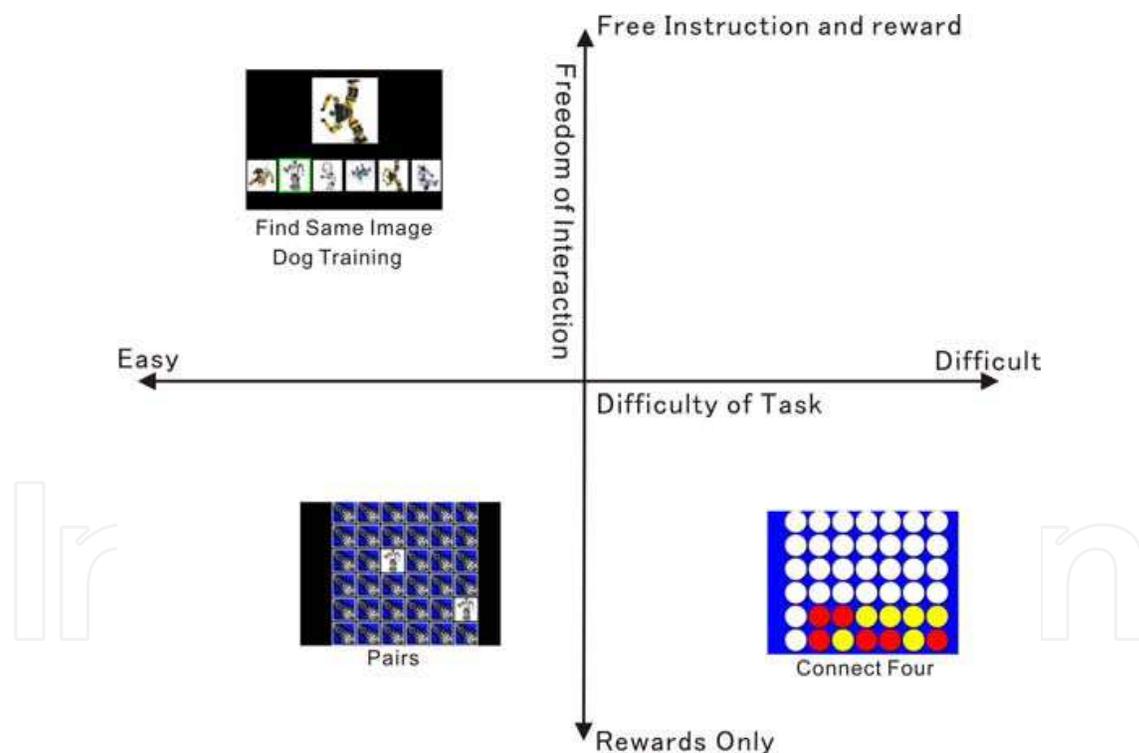


Fig. 1. Properties of the different training tasks

We selected and implemented the different training tasks in a way, that they cover two dimensions which we assume to have an impact on the interaction between the user and the robot.:

- Easy - Difficult: Training tasks can range from ones, that are very easy to understand and evaluate for the user, to tasks where the user has to think carefully to be able to evaluate the moves of the robot correctly.

- **Constrained - Unconstrained:** In the most constrained form of interaction in our training tasks, the user is told to only give positive or negative feedback to the robot but not to give any instructions. In an unconstrained training task, the user is only informed about the goal of the task and asked to give instructions and reward to the robot freely. The positions of the different tasks in the two dimensions can be seen in Figure 1. There is one task for each of the combinations “easy/constrained”, “easy/unconstrained” and “difficult/constrained”. The reason, why there is no task for the combination “difficult/unconstrained” is that that in such a situation, the user behavior becomes too hard to predict, so that the robot cannot reliably anticipate positive or negative reward. Screenshots of the playfields can be seen in Figure 2

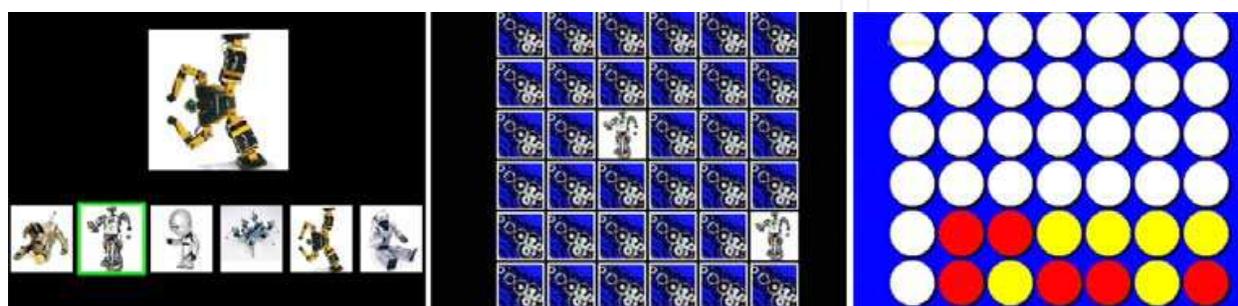


Fig. 2. Game screens of the “Virtual” Training Tasks. (left: Picture Matching, right: Pairs)

3.1.1 Picture matching

On the easy/unconstrained end of the scale, there is the “Find Same Images” task. In this task, the robot has to be taught to choose the image that corresponds to the one, shown in the center of the screen, from a row of six images. While playing, the image that the robot is currently looking or pointing at is marked with a green or red frame to make it easier for the user to understand the robot's viewing or pointing direction. By waving its tail and moving its head the robot indicates that it is waiting for feedback from its user. In this task the user can evaluate the move of the robot very easily by just looking at the sample image and the currently selected image. The participants were asked to provide instruction as well as reward to the robot freely without any constraints to make it learn to perform the task correctly. The system was implemented in a way that the rate of correct choices and the speed of finding the correct image increased over time.

3.1.2 Pairs

As an easy/constrained task, we chose the “Pairs” game. In this task, the robot plays the classic children's game “Pairs”: At the beginning of the game, all cards are displayed upside down on the playfield. The robot chooses two cards to turn around by looking and pointing at them. In case, they show the same image, the cards remain open on the playfield. Otherwise, they are turned upside down again. The goal of the game is to find all pairs of cards with same images in as little draws as possible. In this task the user can evaluate easily whether a move of the robot was good or bad by comparing the two selected images.

The participants were asked not to give instruction to the robot, which card to chose but to assist the robot in learning to play the game by giving positive and negative feedback only.

3.1.3 Connect four

As a difficult/constrained task, we selected the “Connect Four” task. In the “Connect Four” game, the robot plays the game “Connect Four” against a computer player. Both players take turns to insert one stone into one of the rows in the playfield, which then drops to the lowest free space in that row. The goal of the game is, to align four stones of one's own color either vertically, horizontally or diagonally.

The participants were asked to not to give instructions to the robot but provide feedback for good and bad draws in order to make the robot learn how to win against the computer player. Judging whether a move is good or bad is considerably more difficult in the “Connect Four” task than in the three other tasks as it requires understanding the strategy of the robot and the computer player.

3.1.4 Dog training

We have implemented the “Dog Training” task as a control task in order to detect possible differences in user behavior between the virtual tasks and “normal” Human-Robot-Interaction. Like the “Find Same Images” task covers the dimensions easy/unconstrained. The user can easily evaluate the robot's behavior and use his/her way of giving instruction and reward freely without restrictions. In the “Dog Training” task, the participants were asked to teach the speech commands “forward”, “back”, “left”, “right”, “sit down” and “stand up” to the robot. The “Dog Training” task is the only task that is not game-like and does not use the “virtual playfield”. Only in this task the robot was remote-controlled to ensure correct performance.

4. Learning method

We use a biologically inspired approach for learning to classify approval and disapproval using speech, prosody and touch. Our learning method consists of two stages, modeling the stimulus encoding and the association processes, which are assumed to occur in human learning (Burns et al., 2003) (Lowenkron, 2000) (Werker et al., 2005) of associations and word meanings. Details about the biological background of this work are given in section 4.1.

The first learning stage, the feedback recognition learning, is based on Hidden Markov Models. It corresponds to the stimulus encoding phase in human associative learning. Separate sets of HMMs are trained for speech and prosody. The models are trained in an unsupervised way and cluster similar perceptions, e.g. utterances that are likely to contain the same sequence of words or similar prosody. Touch is handled in a different way, because the data returned by the AIBO remote framework does not suffice for HMM based modeling.

The second stage is based on an implementation of classical conditioning. It associates the HMMs which were trained in the first stage with either approval or disapproval, integrating the data from different modalities. As users have different preferences for using speech, prosody and touch when communicating with a robot, the system has to weight the information, coming in through these different channels depending on the user's preferences. Classical conditioning can deal with this problem by emphasizing cues that frequently occur in connection with approving or disapproving feedback for a certain user. It allows the system to weight and combine user inputs in different modalities according to the strength of their association toward approving or disapproving feedback. The data structure, resulting from the learning process, is shown in Figure 3.

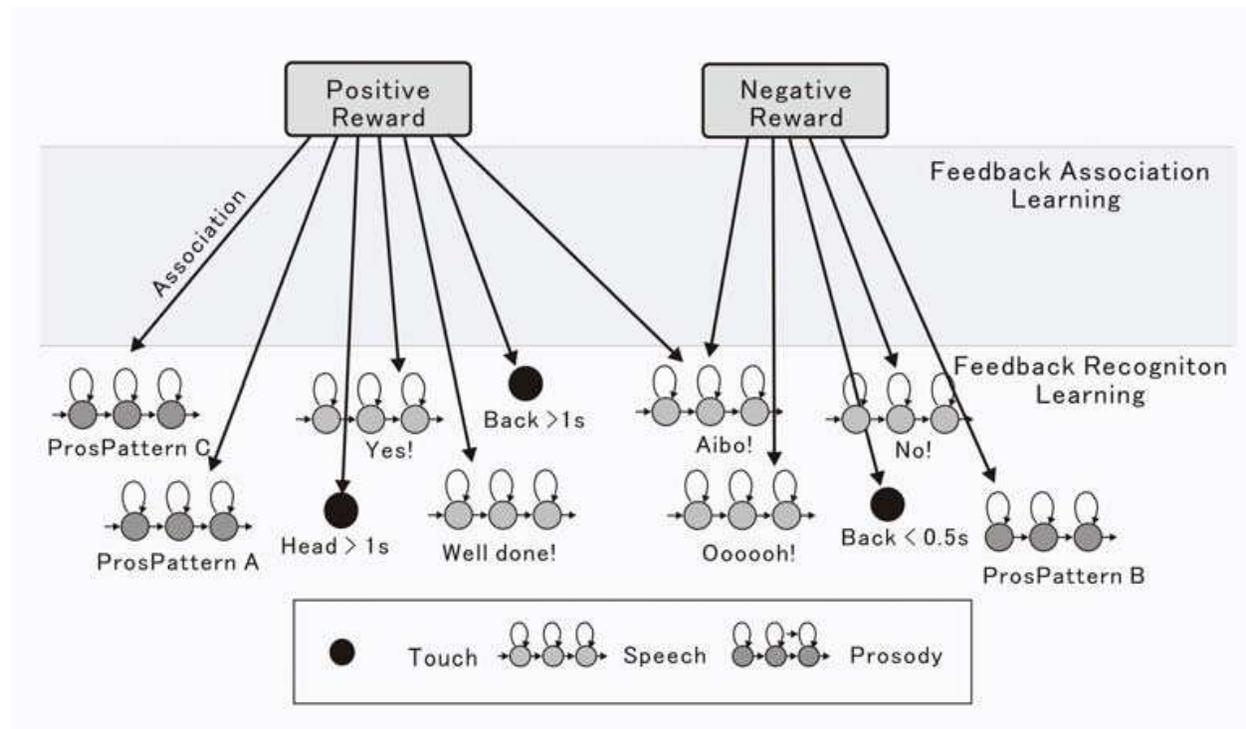


Fig. 3. Data structure, that is learned in the training phase.

4.1 Biological background

Our approach towards understanding feedback from a human is inspired by the biological and psychological processes which are found in human associative learning, speech perception and speech acquisition. However, we do not claim to implement an accurate model of all processes which occur in natural associative learning and understanding of elementary utterances. Instead, we focused on the concepts which appeared most relevant to our research objective of learning to understand human feedback for a robot.

4.1.1 Stimulus encoding for associative learning

Before a human or animal can establish an association between a stimulus and its meaning, the physical stimulus needs to be converted into a representation that the brain can deal with. This process is called *stimulus encoding* (Eysenck & Keane, 2005). Stimulus encoding also enables the brain to abstract from the concrete individual stimuli - which always differ to some extent - to attain a common representation. Evidence of these two stages has been found in experiments on classical conditioning as well as infant word learning (Eysenck & Keane, 2005) (Werker et al., 2005).

For speech, the process of phonological encoding develops and refines in the first months of an infant. Experiments found, that infants' speech acquisition starts from acquiring a proper way of encoding speech-based stimuli (Werker et al., 2005) several months before they are actually able to learn the meaning of words by associative learning.

We adopt this separation between the stimulus encoding and the learning of associations between stimuli and their meanings for our learning algorithm. We combine a stimulus encoding phase based on unsupervised clustering of similar perceptions and an associative learning phase using classical conditioning as a supervised learning method. This allows our system to learn the meaning of feedback from the user during natural interaction

because the learning algorithm does not require any explicit information, such as transcriptions of the user's utterances or gestures for stimulus encoding. It only needs the information of whether an utterance means approval or disapproval to associate the HMMs with their correct meanings. This information is given through the training task.

4.1.2 Classical conditioning

The theory of classical conditioning, which was first described by I. Pavlov (Pavlov, 1927) and originates from behavioral research in animals. It models the learning of associations in animals as well as in humans. 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 (Balkenius & Moren, 1998). In our system, the concepts of approving or disapproving feedback are modeled as *US*. They can, for instance, be interpreted as a positive or negative signal from a reward function used in reinforcement learning. The models of the user's utterances, prosody patterns and touches are *CS* which are associated with approval or disapproval during the feedback association learning phase.

For our task of learning multimodal feedback patterns, the most relevant properties of classical conditioning are blocking, extinction and second-order-conditioning as well as sensory preconditioning:

Blocking

Blocking occurs, when a *CS1* is paired with a *US*, and then conditioning is performed for the *CS1* and a new *CS2* to the same *US* (Balkenius & Moren, 1998). In this case, the existing association between the *CS1* and the *US* blocks the learning of the association between the *CS2* and the *US* as the *CS2* does not provide additional information to predict the occurrence of the *US*. The strength of the blocking is proportional to the strength of the existing association between the *CS1* and the *US*. For the learning of multimodal interaction patterns, blocking is helpful, as it allows the system to emphasize the stimuli that are most relevant. For instance, if a certain user always touches the head of the robot for showing approval, and sometimes provides different speech utterances together with touching the robot, then blocking slows down the learning of the association between approval and these speech utterances if there is already a strong association between touching the head sensor and approval. This way, the more reliable cues are emphasized.

Extinction

Extinction refers to the situation, where a *CS* that has been associated with a *US*, is presented without the *US*. In that case, the association between the *CS* and the *US* is weakened. (Balkenius & Moren, 1998) This capability is necessary to deal with changes in user behavior and with mistakes, made during the training phase, such as a misunderstanding of the situation by the human and a resulting incorrect feedback.

Sensory preconditioning and second-order conditioning

Sensory preconditioning and second-order conditioning describe the learning of an association between a *CS1* and a *CS2*, so that if the *CS1* occurs together with the *US*, the association of the *CS2* towards the *US* is strengthened, too. (Balkenius & Moren, 1998) In sensory preconditioning, the association between *CS1* and *CS2* is established before learning the association towards the *US*, in second-order conditioning, the association between the *US* and *CS1* is learned beforehand, and the association between *CS1* and *CS2* is learned

later. Secondary preconditioning and second-order conditioning are important for our learning method, as they enable our system to learn connections between stimuli in different modalities. They also allow the system to continue learning associations between stimuli given through different modalities even when it could not determine whether the robot's move was good or bad, as long as new stimuli, such as new or commands are presented together with stimuli that are already known and associated to a feedback. E.g. a new positive speech feedback is uttered with a typical, known positive/negative prosody pattern.

4.1.3 Top-down and bottom-up-processes in speech understanding

Human perception is not an unidirectional process but involves bottom-up and top-down processes. (Eysenck & Keane, 2005). The *bottom-up processes* are triggered by the physical stimuli, such as audio signals received by the inner ear or light hitting the retina. The *top-down processes*, on the other hand, are based on the context in which a specific stimulus occurs. The context is used to generate expectations about which perceptions are likely to occur. Both, bottom-up and top-down processes, work together in human perception of audio-visual signals to determine the best explanation of the available data.

The interplay of bottom-up processes and top-down processes in speech perception has been investigated in detail by psychologists (Eysenck & Keane, 2005). W. F. Ganong found, that if a person heard an ambiguous phoneme, such as a mixture between "d" and "t", and one of the possible phonemes made a correct word, while the other one didn't, such as "drash"/"trash", the participants were more likely to identify the ambiguous phoneme as the one, that belonged to a correct word. C.M. Connine found that the meaning of the sentence, that an ambiguous phoneme is presented in, has an influence on its identification. These findings suggested that perception is not only driven by the physical stimulus but also depends on expectations generated from the context. Figure 4 shows an overview of bottom-up and top-down processes in human speech perception.

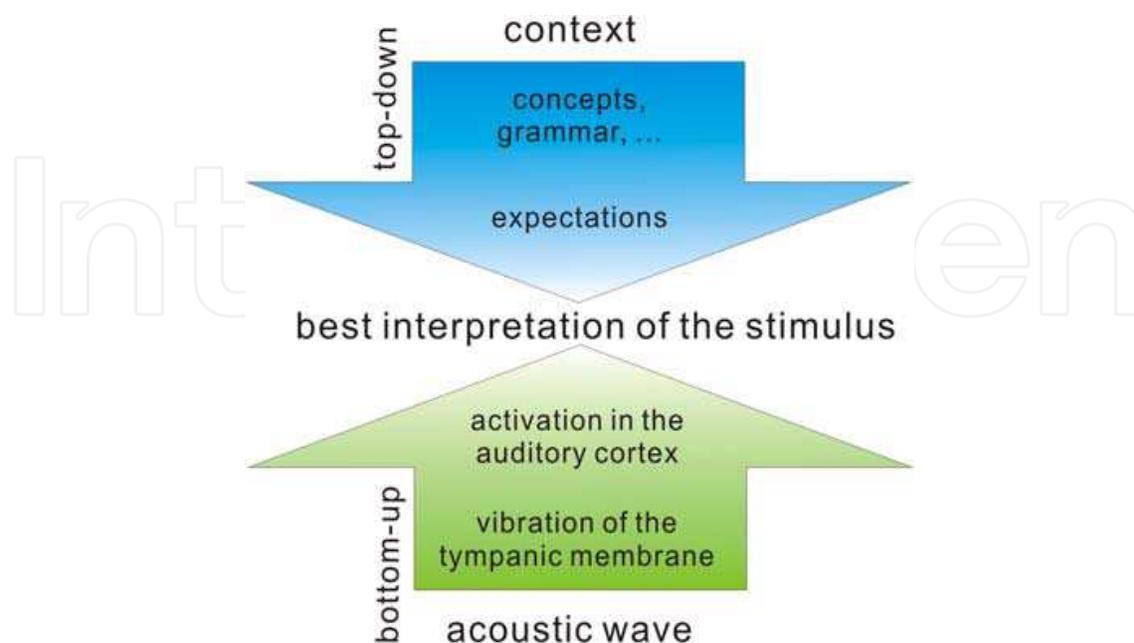


Fig. 4. Bottom-Up and Top-Down Processes in Speech Perception.

In our system, top-down processes are used to improve the selection accuracy when choosing an HMM for retraining. They generate an expectation on which utterances or prosodic patterns are likely to occur, using context information. The context information is calculated from the state of the training task, which suggests whether positive or negative reward is expected, and the learned associations between HMMs and positive or negative feedback. This way HMMs, that have previously been associated with either positive or negative reward, become more likely to be recognized, when another positive or negative reward is expected.

4.2 Feedback recognition learning

The Feedback Recognition learning stage of our learning algorithm clusters and learns the robot's perceptions of the user's feedback. It is based on Hidden Markov Models for speech as well as for prosody and a simple duration-based model for touch.

For each feedback, given by the user, the best matching speech, prosody and touch models are determined according to the methods, described in 4.2.1 to 4.2.3. Then, the most closely matching models are retrained with the data corresponding to the observed feedback. When retraining has finished, the models are passed on to the feedback association learning stage where they are associated with either approval or disapproval based on the situation, that the robot was in, when perceiving the feedback.

In our work, HMMs are employed for the low-level modeling of perceptions. As a standard approach for the classification of time series data, HMMs are widely used in literature. The use of Mel-Frequency-Cepstrum-Coefficients (MFCC) for HMM-based speech recognition is described in (Young et al., 2006). Appropriate feature-sets for emotion and prosody recognition are outlined in (Breazeal, 2002) and (Kim & Scassellati, 2007). We use these tried and tested feature-sets as an input for the HMM-based low-level learning phase.

4.2.1 Speech utterances

To model speech utterances our system trains a user-dependent set of whole-utterance HMMs based on the observed feedback utterances. As a basis for creating utterance models it uses an existing set of monophone HMMs. As the robot learns automatically through interaction, no transcription of the utterances is available. Therefore, an unsupervised clustering of perceived feedbacks that are likely to correspond to the same utterance is necessary. This is done by using two recognizers in parallel. One recognizer tries to model the observed utterance as an arbitrary sequence of phonemes. The other recognizer uses the already trained utterance models to calculate the best-matching known utterance. Every time a feedback from the user is observed, first the system tries to recognize the utterance with both recognizers. Matching is done by HVite, an implementation of the Viterbi Algorithm included in the Hidden Markov Model Toolkit (HTK) (Young et al., 2006). The recognizers return the best-matching phoneme sequence and the best matching utterance out of the utterance models that have been generated up to that point. In addition to that, a confidence level is output by the system for both recognition results.

The confidence levels, which are calculated by HVite as the log likelihood per frame of both results, are compared to determine whether to generate a new model or retrain an existing one. Typically, for an unknown utterance, the phoneme-sequence based recognizer returns a result with a noticeably higher confidence, than the one of the best matching utterance model. For a known utterance, the confidence corresponding to the best-matching utterance

model is either higher or similar to the best-matching phoneme-sequence. Therefore, if the confidence level of the best-fitting phoneme sequence is worse than the confidence level of the best-fitting utterance model or less than 10^{-5} better, then the best-fitting utterance model is retrained with the new utterance.

If the confidence level of the best-matching phoneme sequence is more than 10^{-5} better than the one of the best-fitting whole-utterance model, then a new utterance model is initialized for the utterance. The new model is created by concatenating the HMMs of the recognized most likely phoneme sequence. The new model is retrained with the just observed utterance and added to the HMM-set of the whole-utterance recognizer. So it can be reused when a similar utterance is observed. An overview of the training for speech is shown in Figure 5.

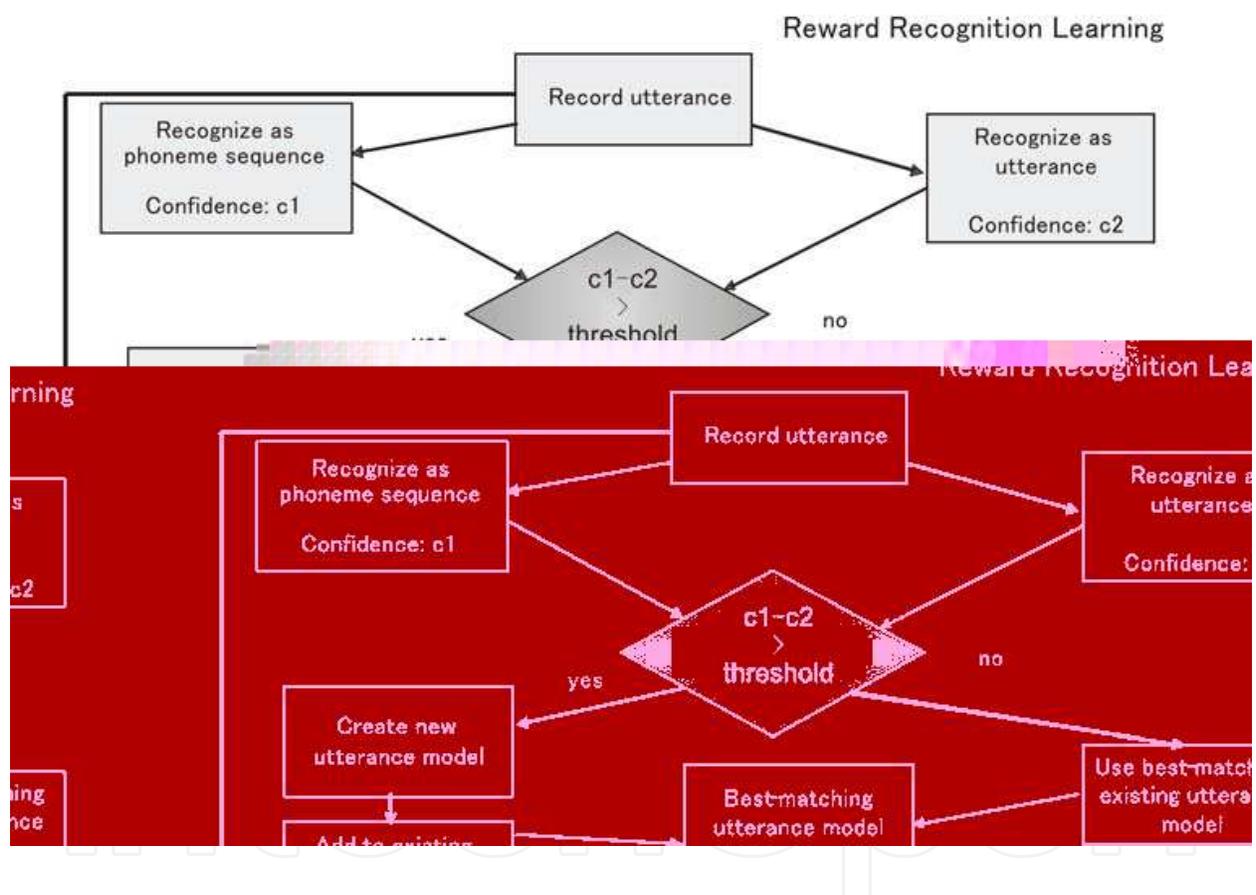


Fig. 5. Algorithm for Recognizing Speech.

The HMM-set for the phoneme-sequence recognizer contains all Japanese monophones and is taken from the Julius Speech Recognition project. We use a simple grammar for the phoneme recognizer that permits an arbitrary sequence of phonemes, not restricted by a language dependent dictionary. A sequence of phonemes may have an optional beginning and ending silence and contain short pauses. The grammar of our utterance model allows exactly one utterance with an optional beginning or ending silence.

During the training phase, utterances from the user are detected by a voice activity detection based on energy and periodicity of the perceived audio signal.

4.2.2 Prosody

We also employ HMMs for recognizing the prosody of speech utterances. The HMMs for interpreting prosody are based on features extracted from the speech signal. First, the signal is divided into frames of 32 ms length with 16 ms overlap. For every frame, the system calculates the pitch, using the YIN Algorithm (Cheveigne & Kawahara, 2002), the overall log energy as well as the frequency spectrum.

Based on this data, a feature vector is calculated consisting of the pitch, the pitch difference to the previous frame, the energy, the energy difference to the previous frame and the energy in frequency bands 1..n. The sequence of feature vectors is written to a file in HTK format to be used for training the HMMs.

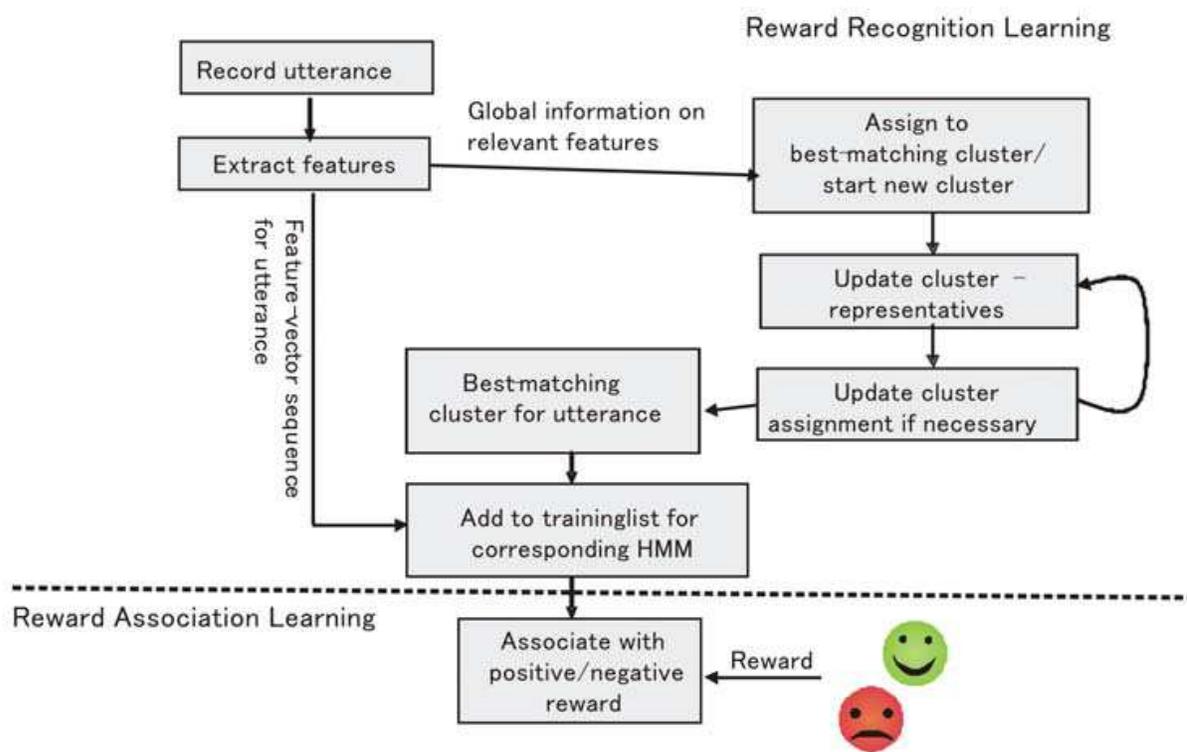


Fig. 6. Algorithm for Learning Prosody.

Additionally, the algorithm calculates some global information based on all frames belonging to one utterance. These are the average, minimum and maximum pitch and energy, the range and standard deviation of pitch and energy as well as the average difference between two adjacent frames of pitch as well as energy. For determining, which HMM is trained with which utterances, the system relies on these global features which have proven to be effective for speech emotion and affect recognition (Breazeal, 2002) (Kim & Scassellati, 2007). A variation of the k-means algorithm which optimizes the number of clusters k between two and ten is used for clustering utterances with similar global features. One HMM is trained for each cluster.

To associate the HMMs with approval or disapproval, every utterance is recognized using the trained HMMs to get the best matching model. This model is then passed to the feedback association learning stage. Figure 6 shows an overview of our prosody recognition.

4.2.3 Touch

We decided not to use HMMs to model touch but a simple duration based model because the output of the touch sensors of the AIBO robot does not suffice for HMM-based modeling. It is binary and does not contain any information on the force applied when touching the sensors. Moreover, the refresh rate when using the AIBO remote framework is quite low.

Therefore, we classified touches of the head sensor and of the back sensor depending on their duration:

- short: less than 0.5 seconds
- medium: between 0.5 seconds and 1 second
- long: one second or longer

Typically, short touches were observed when the user was hitting the robot, while medium and long touches corresponded to caressing or stroking the robot. However, many participants in our user study employed touch only for expressing approval.

4.3 Feedback association learning

In the feedback association learning phase, an association between the HMM or touch pattern model obtained from the feedback recognition learning and either approval or disapproval is created or reinforced. The information of whether the model should be associated with approval or with disapproval is obtained from the current state of the task. If the last move of the robot was a good one, the model, which represents the perceived user feedback, is associated with approval. If the last move was a bad one, it is associated with disapproval.

4.3.1 The Rescorla-Wagner-Model

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. The models describe how associations between unconditioned stimuli and conditioned stimuli are learned. In this study, the Rescorla-Wagner model (Rescorla & Wagner, 1972) is used. It was developed in 1972 and most of the more sophisticated newer theories are based on it. In the Rescorla-Wagner model, the change of associative strength of the conditioned stimulus A to the unconditioned stimulus $US(n)$ present 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 this 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 feedback behavior. 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 association values of all the CS towards the US, that have been calculated in the previous runs of the feedback recognition learning.

The major drawback of the Rescorla-Wagner-Model is that it is not able to model the effects of second-order-conditioning and sensory preconditioning directly. We dealt with this issue by running a second pass of the Rescorla-Wagner-algorithm to learn associations between simultaneously occurring CS. In this second pass, the CS1 serves as the US for the conditioning of CS2. In a third pass of the algorithm, we update the relation between the US and all CS2, that have an association to the actually occurred CS1, using a new learning rate $\alpha_{A\text{ second}}$, which is calculated as the product of the original learning rate α_A and the associative strength between the CS1 and the corresponding CS2

4.4 Integration of top-down-processes

Without top-down processes, all HMMs are equally likely to be selected for retraining in the feedback recognition learning phase. The selection of the best-matching model depends only on the perceived signal while the context is not taken into account.

In order to improve the selection of the best-matching speech and prosody models for retraining, we integrated an implementation of top-down processes, which are also present in human audio-visual perception. (Eysenck & Keane, 2005) It uses the associations, learned in the feedback association learning phase to generate expectations about which stimuli, modeled by HMMs, are most likely to occur in a given context.

Knowing through the state of the training task, whether a positive or negative feedback is expected from the user in a given situation, the system uses the learned association matrix to assign a positive or negative bias to each of the existing HMMs. We calculate the bias B_A for an HMM A from the difference of the associative strength V_A of the HMM A towards the expected feedback and the associative strength of it towards the opposite feedback. In case of positive feedback, the factor would be calculated as in (2):

$$B_A = a V_{A,\text{positive}} - b V_{A,\text{negative}} \quad (2)$$

The constants a and b , which can have values between 0 and 1, determine the impact of the excitatory and inhibitory influences on the calculated bias. A high value a makes the system reuse known HMMs, which are already associated to the present stimulus. A high value b makes the system avoid HMMs, which are already associated to a different stimulus. We found that moderate values for a and high values for b produce best results. In our experiment, we used the values $a=0.2$ and $b=0.8$. The bias B_A is used, if the feedback recognition learning determines that there is more than one HMM that models the stimulus well enough to be a candidate for retraining. In this case, the biases modify the confidence factors returned by the Viterbi algorithm. The biases B_A and the normalized confidence factors C_A are weighted as shown in (3) to select the best HMM for retraining.

$$D_A = c B_A - (1-c) C_A \quad (3)$$

Using this method, HMMs, which are already associated with either positive or negative feedback, become more likely to be selected when a similar feedback is expected. Depending

on the constant c associations of one HMM with positive and negative reward at the same time, are more or less likely. A value of $c=0.8$ has turned out to increase the quality of the HMM selection, while still allowing HMMs for ambiguous utterances to be associated with both, positive and negative reward.

5. Experimental evaluation

We experimentally evaluated the training method and training tasks as well as the learning algorithm. Ten persons participated in the study. All of them were Japanese graduate students or employees at the National Institute of Informatics in Tokyo. Five of them were females, five males. The age of the participants ranged from 23 to 47. All participants have experience in using computers. Two of them have previous experience in interacting with entertainment robots. Interaction with the robot was done in Japanese. During the experiments, we recorded roughly 5.5 hours of audio and video data.

5.1 Instruction and experimental setting

The participants were instructed to teach the robot in the different training tasks described in section 3. They received explanations of the rules of the game-tasks including whether or not they were expected to give instruction or only feedback. We asked them to use speech, gesture and touch freely in their preferred way and showed them the location of the touch sensors of the AIBO robot, as well as the stereo cameras and the microphone. The experimental setting is shown in Figure 7 and screenshots of the video taken during the experiments is shown in Figure 8.

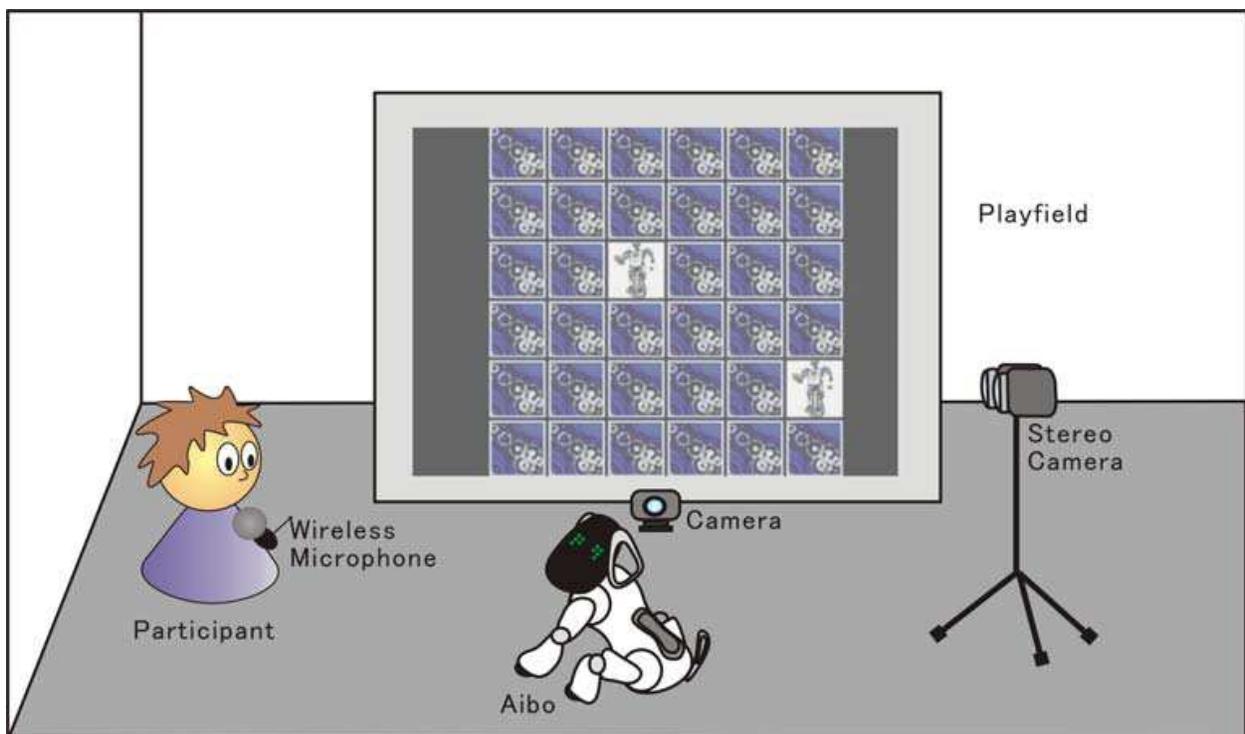


Fig. 7. Overview of the Experimental Setting.

5.2 Results

We evaluated the performance of the learning algorithm offline with the data recorded within the above described experimental setting. The system was trained and evaluated in a user-dependent way using 10-fold cross evaluation. The average accuracy of our system for classifying between approving and disapproving feedback given by one user based on speech, prosody and touch was 95.97%. The standard deviation between users was 3.30%. As the feedbacks given by the participants showed a slight bias toward approval, the confusion matrix, shown in Table 1 gives a more detailed overview over the performance of our recognizer.

	Positive(actual)	Negative (actual)
Positive (recognized)	52.50%	1.76%
Negative (recognized)	2.27%	43.47%

Table 1. Confusion Matrix

Using speech only we reached a recognition rate of 83.53% with a standard deviation of 8.30%. Using prosody only, the recognition rate was 84.27% with a standard deviation of 8.57%. For touch the recognition rate was 88.17% with a rather high standard deviation of 11.77% as the usage and frequency of touch varied strongly between users. All single-modality recognition rates are considerably lower than the recognition rate for combined feedbacks shown above. This result underlines that combining stimuli given through different modalities is crucial for a reliable recognition.

Without the integration of top-down processes for speech, the recognition rate for speech as a single modality drops to 77.42% with a standard deviation of 13.12%. This degrades the overall recognition rate to 93.95% with a standard deviation of 5.33% if top-down processes are not used.

5.3 Questionnaire based evaluation

In a questionnaire, we asked the participants to evaluate their experience throughout the four different training tasks. The participants could rate their agreement with the statements, shown in table 1, on a scale from 1 to 5, where 1 was the best and 5 the lowest rating.

In the Dog Training task, the robot was remote controlled to react to the user's commands and feedback in a typical Wizard of OZ-Scenario. However, in the Same Image task, the user's instructions and feedback were not actually understood by the robot but anticipated from the state of the training task. This did not have a negative impact on the participants' impression that the robot understood their feedback, learned through it and adapted to their way of teaching, compared to the Wizard of Oz scenario.

The lowest ratings were given for the "Connect Four" task. As the robot's moves could not be evaluated as easily, as in the other tasks, the participants were unsure, which rewards to give and therefore did not experience an effective teaching situation.

5.4 Feedback given by the participants

We analyzed the feedback, given by the participants to find typical similarities and differences in the interaction with AIBO between different users. As for the modalities used for giving reward, we found a strong preference for speech-based reward. Among 2409

	Picture Matching	Pairs	Connect Four	Dog Training
Teaching the robot through the given task was enjoyable	1.81 1.04	1.90 0.83	1.81 0.89	1.63 0.81
The robot understood my feedback	1.27 0.4	1.81 0.74	2.90 0.85	1.81 0.30
The robot learned through my feedback	1.36 0.59	2.81 0.93	3.45 0.95	1.54 0.69
The robot adapted to my way of teaching	1.45 0.66	2.63 1.05	3.45 1.04	1.64 0.58
I was able to teach the robot in a natural way	2.18 0.96	2.09 0.86	2.54 1.12	1.64 0.69
I always knew, which instruction or reward to give to the robot	2.00 0.72	2.09 0.86	2.90 1.02	1.91 0.83

Table 1. (First value: average, second value: standard deviation)

stimuli used for giving reward, 1888 (78.37%) were given by speech, 504 (20.92%) were given by touching the robot and 17 (0.71%) were given by gestures. For the different users, the percentage of speech-based rewards ranged from 52.25% to 97.75%. Gestures were frequently employed by the participants for giving instructions, but we almost did not observe gestures being used for giving positive or negative reward.



Fig. 8. Experiment scenes: 1: Picture Matching, 2: Pairs, 3: Connect Four, 4: Dog Training

Typically, multiple rewards were given for a single positive or negative behavior of the robot. Counting only the rewards given while the robot signaled that it was waiting for feedback after an action, 3.43 rewards were given for one action on average, usually including one touch reward and one to four utterances. One utterance was counted as one reward. Repetitions of an utterance were counted as multiple rewards. In case of touch, one or multiple contacts with the robot's touch sensors were counted as one reward, as long as the participant kept his/her hand close to the sensor.

The favorite verbal feedback differed between the users especially in case of positive reward. None of the utterances, used for positive feedback, appeared within the first six most frequently used utterances for all ten participants. On average, each person shared his/her overall most frequently used positive feedback with one other person. In case of negative reward, the feedback, given by the participants was more homogenous. The most frequently used feedback - "wrong" (chigau) - was shared by eight out of ten persons. For the two remaining persons, it was the second and third most frequently used feedback utterance.

As for the variability of the feedback, given to the robot by an individual user: On average, participants used 12.3 different verbal expressions to convey positive feedback and 13.4 different expressions to express negative feedback. However, this number varies strongly between individuals: One person always used the same utterance for giving positive feedback and a second utterance for giving negative feedback while the person with the most variable feedback used 30 different expressions for giving positive and 28 different expressions for giving negative feedback. 55.61% of all verbal feedback was given by the participants using their preferred feedback utterance. 88.73% of a user's verbal feedback was given using one of his/her six most frequently used positive/negative utterances, so understanding a relatively small number of different utterances suffices to cover most of a participant's verbal feedback.

For positive feedback, four out of ten participants had one preferred utterance which did not vary between the four training tasks. In case of negative reward, this was true for five people. For eight out of ten participants in case of positive reward and six participants in case of negative reward, their overall most frequently used feedback utterance was among the top three feedback utterances in each individual task. In the cases, where the preferred feedback was not the same in all tasks, it typically differed for the "Connect Four" task, while in the three other tasks, including the "Dog Training" control task similar feedback was used as described above. As in the "Connect Four" task it was difficult for the users to judge, whether a move was good or bad in order to provide immediate reward, feedback tended to be very sparse and tentative like "not really good" (amari yokunai), "Is this good?" (ii kana?) or "good, isn't it" (ii deshou).

6. Discussion and outlook

In this paper, we described and evaluated a method for learning a user's feedback for human-robot-interaction. The performance based on interpreting speech, prosody and touch feedbacks from a human can be considered sufficiently reliable for being used to teach a robot, for example, by reinforcement learning.

One potential drawback of our approach is that the robot has to complete a training phase with every user who wants to interact with it in order to adapt to the user's way of giving rewards. However, a typical pet robot or entertainment robot only interacts with a very

limited number of persons in a household and usually interacts with the same users frequently and for a long time. Therefore, we assume that user-specific adaptation is desirable even though it needs some initial training effort.

Currently the learning algorithm works offline using the data gathered in the training tasks to generate HMM sets and associations. Main issues that need to be targeted for implementing an online version of the algorithm are the clustering of the training samples for prosody as well as the incremental re-training of the HMMs for speech and prosody.

Our method has only been evaluated for learning feedback, so far. However, it can be used without changes for learning object names as well as simple, non-parameterized commands. However, extensions to the current algorithm are necessary if the robot needs to learn commands with parameters, such as "Can you put {the red ball} {in the box}". While gesture was not necessary for recognizing approval or disapproval, gesture recognition will be helpful to understand commands, so integrating gesture as an additional modality is the current priority of our ongoing research.

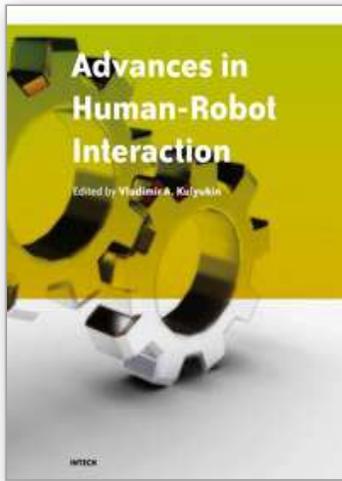
One important question that remains open after the study is the similarity of user behavior between virtual tasks and real world tasks. Although we did not observe differences in user feedback between the virtual game tasks and the dog training in our experiment this does not necessarily mean that it is generally possible to train a robot for a real world task using a virtual task. This question will be targeted in a follow-up study.

7. References

- A. Austermann, S. Yamada (2008). "'Good Robot, Bad Robot' - Analyzing Users' Feedback in a Human-Robot Teaching Task", *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 08)*, pp. 41-46
- C. Balkenius and J. Moren (1998). "Computational models of classical conditioning: a comparative study." *Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior*, pp. 348 - 35
- C. Balkenius and J. Moren (1999). "Dynamics of a Classical Conditioning Model", *Autonomous Robots* 7, 41-56
- B. Burns, C. Sutton, C. Morrison, P. Cohen (2003), "Information Theory and Representation in Associative Word Learning", *Epigenic Robotics (EpiRob 2003)*
- C. Breazeal (2002). "Recognition of Affective Communicative Intent in Robot-Directed Speech" *Autonomous Robots* Volume 12, Issue 1, 83 - 104
- A. de Cheveigne and H. Kawahara (2002): "YIN, a fundamental frequency estimator for speech and music", *The Journal of the Acoustical Society of America*, Vol. 4, pp. 1917 - 1930
- M. W. Eysenck, M. T. Keane (2005). "Cognitive Psychology - A Student's Handbook", Psychology Press
- N. Iwahashi (2004). "Active and Unsupervised Learning for Spoken Word Acquisition Through a Multimodal Interface", *RO-MAN 2004, 13th IEEE international workshop on robot and human interactive communication*, pp. 437 - 442
- E. S. Kim, B. Scassellati (2007). "Learning to Refine Behavior Using Prosodic Feedback", *Proceedings of the 6th IEEE International Conference on Development and Learning (ICDL 2007)*, pp. 205-210

- Z. K. Kayikci, H. Markert, G. Palm (2007): "Neural Associative Memories and Hidden Markov Models for Speech Recognition", *IJCNN 2007 Conference Proceedings*, pp. 1572 - 1577,
- B. Lowenkron (2000). "Word meaning: A verbal behavior account", Presented at the annual convention of the Association for Behavior Analysis, Washington DC, May
- I. P. Pavlov (1927). "Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex" (translated by G. V. Anrep), Oxford University Press
- R. Rescorla, A. Wagner (1972). "A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement.", *Classical Conditioning II: Current Research and Theory* (Eds Black AH, Prokasy WF) New York: Appleton Century Crofts, pp. 64-99
- B. F. Skinner (1957). "Verbal Behavior". Copley Publishing Group, Acton
- A. L. Thomaz, and C. Breazeal (2006): "Reinforcement Learning with Human Teachers: Evidence of feedback and guidance with implications for learning performance" *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI 2006)*
- J. F. Werker, H. Henny, Yeung (2005). "Infant Speech Perception Bootstraps Word Learning", *Trends in Cognitive Sciences*, Vol. 9, No. 11, pp. 519-527
- S. Young et al., (2006) "The HTK Book", HTK Version 3, <http://htk.eng.cam.ac.uk/>

IntechOpen



Advances in Human-Robot Interaction

Edited by Vladimir A. Kulyukin

ISBN 978-953-307-020-9

Hard cover, 342 pages

Publisher InTech

Published online 01, December, 2009

Published in print edition December, 2009

Rapid advances in the field of robotics have made it possible to use robots not just in industrial automation but also in entertainment, rehabilitation, and home service. Since robots will likely affect many aspects of human existence, fundamental questions of human-robot interaction must be formulated and, if at all possible, resolved. Some of these questions are addressed in this collection of papers by leading HRI researchers.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Anja Austermann and Seiji Yamada (2009). Learning to Understand Expressions of Approval and Disapproval through Game-Based Training Tasks, *Advances in Human-Robot Interaction*, Vladimir A. Kulyukin (Ed.), ISBN: 978-953-307-020-9, InTech, Available from: <http://www.intechopen.com/books/advances-in-human-robot-interaction/learning-to-understand-expressions-of-approval-and-disapproval-through-game-based-training-tasks>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2009 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen