Real-time Perception of Non-verbal Human Feedback in a Gaming Scenario

Salah Al-Darraji  
Robotics Research Lab.  
Department of Computer Science  
University of Kaiserslautern  
Gottlieb-Daimler-Str. 48  
67663 Kaiserslautern, Germany  
http://rrlab.cs.uni-kl.de  
saleh@cs.uni-kl.de

Zuhair Zafar  
Robotics Research Lab.  
Department of Computer Science  
University of Kaiserslautern  
Gottlieb-Daimler-Str. 48  
67663 Kaiserslautern, Germany  
http://rrlab.cs.uni-kl.de  
zafar@cs.uni-kl.de

Karsten Berns  
Robotics Research Lab.  
Department of Computer Science  
University of Kaiserslautern  
Gottlieb-Daimler-Str. 48  
67663 Kaiserslautern, Germany  
http://rrlab.cs.uni-kl.de  
berns@cs.uni-kl.de

Social interactive robots require sophisticated perception and cognition abilities to behave and interact in a natural human-like way. The precise perception of behaviour of an interaction partner plays a crucial role in social robotics. Human feedback is one of the most important behavioural cues of human during conversation. Feedback is a signal that human sends to the interlocutor, consciously or unconsciously, that can convey four communicative functions. These functions: continuation, perception, understanding, and acceptance of the message, activate or inhibits different behaviours of human. Past systems focused mainly on the scripted scenarios irrespective of interaction partners feedback behaviour. This paper proposes a system for perception and interpretation of the interaction partner feedback from non-verbal cues. The proposed work has been tested in a 20 question gaming scenario in real time and it has shown promising results.

1. INTRODUCTION

Since the last decade, the concept of robots being part of human life is not limited to fictional film stories any more. The success of industrial robots paved way for starting a new era of social robotics. With the advent of advanced technology, a lot of task specific prototypes have already been manufactured. If these robots have to be part of our lives and help us like any other human, then they must have the same functionality as them. They should have not only the basic functionality but also more complex cognitive characteristics. The objective of social robots is to interact with humans in a human-like way and understand their needs, their emotions, their limitations and their surroundings.

The most crucial aspect in human-robot interaction is to recognize non-verbal cues. Psychological studies shown that non-verbal content is more than 60 per cent of our face-to-face communication scenarios (Noller 2006). These non-verbal cues are often uncontrolled and spontaneous; they explain the verbal content and describe the emotional message behind it. One of the most important type is facial expressions which shows the emotional state of human. Other include head gestures, hand gestures, body gestures, pitch and tone of voice, gaze and even the appearance of the person (DeLamater et al. 2014).

To date, numerous systems have been reported so far that use non-verbal content of human for human-robot interaction (HRI). The study in (Akalin et al. 2014) implements a robot sign language tutor system for children with communication impairments. The robot recognizes different flash cards and then performs the same gesture in a teaching phase. In examination phase, robot performs a gesture and the child has to show relevant flash card and if it is correct, robot smiles. Yoon and Chi (2006) implement a rock, paper, scissor game to play with the robot. In (Chao et al. 2011), they implement “Simon says”, a turn-taking imitation game, that uses non-verbal cues like gaze and motion and also verbal content. They collect experimental data and use minimum necessary information to characterize a predictive human response delay. They find out that human response time is varied with respect to the communication channel. It is observed that...
generally, humans don’t speak till the robot finishes speaking but they respond earlier when they use non-verbal content. In (Han et al. 2012), non-verbal cues are used for human-robot interaction. Head gestures, object recognition, human tracking and hand gestures are used for interaction with the robot. Robot uses WikiTalk system to interact with the interlocutor and share the information. Interlocutor is also capable to interrupt the robot as a feedback using robot’s tactile sensor when he/she is not interested to hear the long monologue. In (Jevtic et al. 2015), three different types of interaction modalities have been presented for human-robot interaction. Direct physical interaction: where robot is displaced from its position. Person-following robot interaction: where robot follows the human keeping a safe distance. Lastly, pointing interaction: which is used to instruct the robot through pointing gesture to a target position. Raising left hand gesture is used as a feedback to stop the robot at any given time.

However, in all of the above presented state-of-the-art HRI systems, one major shortcoming is human feedback. In human-human interaction, we use non-verbal feedback channels to know the internal state of the interlocutor and then accordingly, if needed, modify our action/reaction. Similarly, in case of robots, there is a need to have such feedback system, which provides information of internal state of human to the robot. In this domain of study, we develop a behaviour based feedback system that enables robot to recognize non-verbal cues to interact with interlocutor and play a 20 questions game, see Fig. 1. Hand and head gestures are used to interact with robot while facial expressions and head movements are used as feedback channel. Robot analyses the emotional state of the human over the course of the game as a feedback and reacts accordingly. This type of interaction is natural and spontaneous as compared to the scripted scenarios. In the following sections, we describe the game and our feedback system in detail.

The rest of the paper is organized as follows: psychological insight, perception and interpretation of human feedback are presented in sections 2, 3, and 4, respectively. Section 5 discusses the conducted experiments. Conclusion and future work is discussed in section 6.

2. FEEDBACK CUES

Humans use feedback in their conversation to exchange information about four communicative functions. These functions are continuation, perception, understanding and attitudinal reactions. Any expression, verbal or non-verbal, can be regarded as a feedback if it serves one or more of these functions.

In order to communicate successfully, according to Allwood (2001), two participants should establish a contact with each other. It is necessary that participants show their ability and willingness to continue in the interaction. Continuation signal tells the interlocutor about the desire of continuing in the interaction regardless of the contents of the message. Two types of continuation signal can be distinguished: You go on and I go on (Cerrato 2007).

Once the contact is established, one participant possibly produces a message. The receiver should be able and willing to perceive the message and he may or may not understand the message. Understanding the perceived message means that the contents of the message is interpreted in the same way that the speaker intends to. Once the receiver understands the message, he/she may give attitudinal and behavioural reaction according to the acceptance of the message. Attitudinal reaction can be positive, negative or neutral.

Interactive robots that deals socially with humans should be able to interpret the feedback during conversation. The interpretation of feedback cues is a real challenge for robot and sometimes even for human. The current work uses the observations of Allwood (2001) and his interpretation of feedback cues during cooperation scenarios among humans. He used a classification scheme by tracking some of the related gestural articulators and body movements of the participants. The tracked gestural articulators are head, face, eyes, body postures, and arms and hands. He introduces feedback functions for these gestural articulators. Based on his interpretations and findings, a feedback interpretation tree with his feedback functions can be built. This interpretation tree classifies the feedback into four types as shown in Fig. 2. The lower level of the tree is the observations of various body movements that are related to each feedback function.
3. PERCEPTION OF HUMAN FEEDBACK CUES

Perception of human feedback via non-verbal cues is a necessary skill that interactive robot should have. It enables the robot to select and change the conversation scenario according to the interaction partner’s behaviour, mood and personality. The current work perceives non-verbal behaviour of human via vision channel. Hand gestures, head gestures and facial expressions are used to perceive feedback of human. The perceived cues are fused together in order to generate a single feedback using behaviour-based network.

3.1. Hand Gesture Recognition

Hand gestures, like other non-verbal cues, convey great deal of information. They emphasize verbal content and also give hints about internal state of human. They enable humans to express their feelings or sometimes also are used to express basic information like numbers one, two and so on. Hand gestures can be speech-related or speech-independent. For example, parenting a child requires speech related hand gestures while managing traffic by a traffic policeman requires speech independent hand gestures.

Most of these approaches use skin detection to localize hand.

However, with the introduction of RGB-D (colour and depth data) devices, localization of hand has become much easier. NITE middleware library provides 3D position of hands in real time using depth data. Our hand gesture recognition module uses this information, segments hand window from the colour image and discards those regions whose depth is greater than hand depth. Scale Invariant Feature Transform (SIFT) features are extracted on the segmented hand images. Bag-of-Features (BoF) approach is used alongside K-means clustering algorithm to form bag-of-words (BoW) vector for each image where Support Vector Machine (SVM) is used for classification of hand gesture (Zafar and Berns 2016). Fig. 3 shows the working of hand gesture recognition module.

3.2. Head Gesture Recognition

Another non-verbal aspect, which this paper considers, is human head movement or head gestures. Humans have the ability of interpreting these movements quickly and effortlessly, however,
it is regarded as a difficult challenge in computer systems and robotics. Detecting human head movement requires estimating head pose over the time. For example, head nodding is the deviation of the pitch angle of the head, whereas head shaking is the deviation of the yaw angle.

In order to build a reliable human-robot interaction system, a robust head pose estimation algorithm is needed. The current work utilizes head pose estimation used in (Saleh et al. 2013; Saleh and Berns 2015). Eight head gestures are recognized in this work. These gestures are nodding, shaking, tilting, looking ahead, looking left, looking up and looking down. Head nodding and shaking are dynamic gestures, in which the head pose is changing over time, whereas the others are static gestures in which the head pose is nearly not changed.

Head nodding gesture can be detected, by the robot, as a sequence of poses where the pitch angle of the head exceeds the threshold \( \theta_p \) in both directions. The speed of the nods depends on the number of nods and the time in which the nods occurred.

\[
   r_n = \frac{N}{t}
\]  

Where

- \( N \) is the number of nods;
- \( t \) is the time in which the nods are occurred;
- \( r_n \) is the nodding rate.

\[
   \text{Nodding Speed} = \begin{cases} 
   \text{slow} & \text{if } r_n < \rho_n \\
   \text{fast} & \text{if } r_n \geq \rho_n 
   \end{cases}
\]  

Where \( \rho_n \) is the nodding speed threshold. Head shaking is detected in the same way but on the yaw angle instead of pitch.

Static gestures can be detected by calculating the duration of head poses in different directions. If the duration of a specific direction exceeds 80 per cent of a specific period of time, then the corresponding gesture is regarded as active. Otherwise, the gesture is regarded as inactive.

### 3.3. Facial Expression Recognition

The facial expressions of humans play an important role in inter-human interaction. They are the primary source of expressing emotions and feelings. Humans express more feelings through facial gestures than any other body movements. The emotional state of human can also be reflected on the face through facial expressions. Psychologists addressed diverse communicative functions for facial expressions such as giving feedback, opening and closing communication channels and complementing verbal cues.

The proposed method uses deep neural network, specifically convolutional neural network (CNN), to extract and classify features from the input data automatically. Deep learning has been widely used recently and has shown promising results in various areas. Fig. 4 depicts the architecture of the convolutional neural network that is used to recognize the six basic facial expressions as in (Al-Darraji et al. 2016).

The CNN receives a human face as a 32 × 32 gray image and outputs the confidences of seven facial expressions (including neutral). It uses two convolutional layers each with its own sub-sampling layer (max-pooling). The convolutional layer applies a set of learnable filters on the input image. Using more than one convolutional layer enables to extract features on different levels. Each convolutional layer extract higher level features than the previous layers. For \( m \times m \) filter \( w \), the output unit \( x_{ij}^{l} \) of the convolutional layer \( l \) is calculated using equation 3.

\[
   x_{ij}^{l} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} w_{ab} f_{(i+a)(j+b)}^{l-1}
\]

Figure 4: Architecture of deep neural network for facial expression recognition. It contains two convolutional layers each with its maximum sampling layer and one fully connected layers with 80 neurons that is connected to the output layer of 7 neurons.
turn is connected to seven output neurons, which represents six emotions in addition to neutral. The activation function that is used in all layers is \( \tanh \) function. Mean squared error (MSE) is used as a loss function and stochastic gradient descent function for the optimization.

### 4. HUMAN BEHAVIOUR INTERPRETATION

The Perception of human behaviour requires human tracking, monitoring different parts and interpreting these findings over a period of time. In this work, interpretation of human behaviour is achieved by using behaviour-based network. This section will explain the behaviour-based network, which is normally used in robots control.

#### 4.1. Integrated behaviour-Based Control (iB2C)

Integrated behaviour-Based Control (iB2C)\(^1\) is a framework developed at the Robotics Research Lab of the University of Kaiserslautern for the realization of behaviour-based control systems (Proetzsch et al. 2010a)(Proetzsch et al. 2010b). Using behaviour-based approach enables building of complex control networks by combining simple behaviour modules. In this work, iB2C is used for perception of human behaviours and feedback. The basic component of iB2C is the behaviour module, Fig. 5. A behaviour module can be characterized by the triples:

\[
B = (f_a, f_r, F)
\]

where \( f_a \) represents the activity function, \( f_r \) means the target rating function and \( F \) is the transfer function of the behaviour. These three functions calculate the output signals, activity \( \vec{a} \), target rating \( r \) and the output vector \( \vec{u} \) depending on the input signals of a behaviour: the stimulation \( s \), the inhibition \( \vec{i} \) and the input vector \( \vec{e} \). These functions, their input and outputs are described briefly in the following. More details can be found in (Proetzsch et al. 2010a).

4.1.1. **Stimulation** \( s \)

It can be used to adjust the influence of the behaviour or to allow some other behaviours to control this behaviour. \( s = 1 \) means full stimulation and \( s = 0 \) means no stimulation, values in between refer to a partially stimulated behaviour.

4.1.2. **Inhibition** \( \vec{i} \)

A behaviour can be inhibited by several other behaviours. The inhibition \( i \in [0, 1] \) of a behaviour is defined as: \( i = \max_{j=0,...,n-1} i_{kj} \), where \( n \) denotes the number of inhibiting behaviours. The inhibition has the inverse effect of stimulation, \( i = 0 \) refers to no inhibition and \( i = 1 \) refers to full inhibition.

4.1.3. **Activation** \( a \)

The activation of a behaviour represents the relevance of a behaviour in the behaviour network. It is calculated depending on the stimulation \( s \) and the inhibition \( \vec{i} \), with \( a = s \cdot (1 - \vec{i}) \). The value of activation is between 0, no activation and 1, full activation.

4.1.4. **Activity** \( a \)

The activity \( a \in [0, 1] \) of a behaviour represents its influence on the current system state. An activity of 1 represents a fully active behaviour whereas an activity of 0 represents a completely inactive behaviour.

4.1.5. **Target Rating** \( r \)

The target rating \( r \in [0, 1] \) of a behaviour indicates the behaviour’s satisfaction. A target rating of 0 refers to a satisfied behaviour, a target rating of 1 to a completely unsatisfied one. In the current work, the target rating signal is not used.

4.1.6. **Transfer Function** \( F \)

The output vector \( \vec{u} \) of a behaviour can be calculated by the transfer function \( F \). The transfer function \( F(\vec{e}, \vec{i}) \) is defined as follows:

\[
F : \mathbb{R}^m \times [0, 1] \rightarrow \mathbb{R}^n, F(\vec{e}, \vec{i}) = \vec{u}
\]

Another basic module is the **Fusion behaviour**. It receives the output vectors from different behaviour modules as well as other behaviour signals and outputs a data vector depending on the fusion function. There are three implemented functions: maximum fusion, weighted fusion and weighted sum fusion. A set of behaviours can be grouped into a behaviour group that can be regarded as one module.

### 4.2. Interpretation of Human Feedback

After detecting human visual cues such as head pose, hand static gesture and facial expression, these cues should be tracked over time to interpret

---

\(^1\)http://agrosy.cs.uni-kl.de/en/research/ib2c/
human behaviour. Behaviour modules and groups are used to interpret human behaviours. The main behaviour modules and groups are movement detector, hand gesture, head gesture, facial expression and feedback.

The first behaviour module in the network is the movement detector which detects all the movement (displacement) in human face, hands or body in any of 3D space. It inhibits the next behaviour modules and groups to prevent any misinterpretation of the corresponding articulators during movements. For example, hand gesture detection is not robust during the hand movement from the initial to the final position and should be avoided until it reaches the final position. Fig. 6 depicts the relation between movement detector and other behaviours.

Tracking head poses of a human over a period of time enables detection of head gestures such as nodding, shaking, tilting and looking (gazing). When the head pose is not changing for a specific slice of time, it means that human is looking at something (static gesture). Any change in the pose breaks the static gestures and may activate dynamic gestures. Implementing gestures as behaviours helps to select one gesture at a time and enables some gestures to inhibit other gestures which have different properties. For example, changing pitch angle of the head activates the nodding behaviour which in turn inhibit static gesture behaviours because changing head angles breaks the static gestures. Fig. 7 shows the head gesture behaviour group.

Our perception of human non-verbal feedback is based on the analysis of Allwood (2001) as described in Fig. 2. For each feedback type, a behaviour module has been implemented. These behaviour modules calculate their activities according to the input human information. The activity of a module increases as the time of the corresponding cues increase. The output of each behaviour is the feedback code which is then fused using “Winner Takes All” fusion behaviour to give four signals that represent four types of communicative feedback.

Afterwards, these four feedbacks are fused to generate the final feedback. It is important to notice that some feedbacks can serve more than one type. For example, positive acceptance feedback necessarily means understanding, perception and continuation feedback. But continuation feedback doesn’t necessarily mean understanding or acceptance of what has been said. Fig. 8 depicts the feedback behaviour types and their fusion using behaviour-based network.

5. EXPERIMENTATION AND EVALUATION

In order to validate our findings, a gaming scenario is implemented in which human plays a game with a humanoid robot. In this section, we describe the game, our humanoid robot and in the end evaluate the performance of the system.

5.1. 20 Questions Game-play

In order to validate our behaviour based feedback system, we have designed a popular interactive game of 20 questions\(^2\). In this game, user has to think of something and then robot asks around 20 questions to guess that thing. Our version of game enables user to play it only through his/her visual cues. Hand gestures are used for answering multiple choice questions. For Yes and No answers, head nodding and shaking can also be used. For every question, there are 5 possible answers; yes, no, unknown, irrelevant and sometimes.

The game process includes four important steps; (a) importing a question from website\(^2\), (b) getting the answer from the user, (c) sending back the answer to the website, and (d) again fetching a new question.

\(^2\)http://www.20q.net
Real-time Perception of Non-verbal Human Feedback in a Gaming Scenario

Al-Darraji • Zafar • Berns

In order to implement this, a Finite State Machine (FSM) has been designed. In the first state, the robot looks for human face and skeleton. As soon as it finds an interaction partner, it explains the game and asks the user to think of something. A busy state is added in FSM, which prevents robot to perceive while it is speaking. Through empirical studies, a wait state with time 2 seconds is used after every question. If the robot doesn’t perceive in this time frame, then it repeats the question and waits for the answer. If the user is found looking other than the robot during a perceiving state for a specific amount of time, then the robot would recognize this ignoring behaviour and would ask the user whether he wants to continue playing. Similarly, if a person is not understanding a question for a specific amount of time, then the robot would repeat its question. Furthermore, it also recognizes the acceptance behaviour of human. For example, when robot guesses what the user has thought and says “I am guessing that it is a piano”, user would either show negative acceptance by shaking his/her head or wrinkling of eyebrows or show positive acceptance by nodding his/her head.

5.2. Robotic Platform

Robothespian, a humanoid robot has been used in our study for evaluation of our behaviour based feedback system. It is equipped with a backlit projected face, arms, hands and torso. It can speaks via its built-in speech synthesis module in English and German language. The face makes use of projective technology to express almost any facial expression using action units. The head is able to move sideways for about ±45°. ASUS Xtion Pro is installed on the chest of robot. Additionally, a high definition camera is also installed on the head. The whole arm has 14 degrees of freedom, where hands are able to perform nearly all gestures. Robothespian has its own processor that can handle all the movements of joints. For perception, a standalone system, Intel Core i7 running at 3.40GHz, has been used to process the RGB-D data and run the finite state machine. Figure 1 shows a subject interacting with the robot, Robothespian.

5.3. Performance Evaluation

5.3.1. Perception Modules Evaluation

In order to evaluate the whole feedback system, a reliable recognition of low level perception modules is critical. As already mentioned in previous sections, our feedback perception system uses low level percepts to recognize subject’s behaviour in a game scenario. This section provides experimental results for two main perception modules namely, facial expressions and hand gestures. Table 1 shows the confusion matrix of six basic facial expressions and a neutral expression. The Angry and Fear expressions have lower recognition rate due to the

Table 1: Confusion matrix of facial expression recognition.

<table>
<thead>
<tr>
<th>Detected as</th>
<th>Ne</th>
<th>An</th>
<th>Di</th>
<th>Fe</th>
<th>Ha</th>
<th>Sa</th>
<th>Su</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>95.2</td>
<td>0.0</td>
<td>0.0</td>
<td>4.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Angry</td>
<td>0.0</td>
<td>81.0</td>
<td>9.5</td>
<td>4.8</td>
<td>0.0</td>
<td>4.8</td>
<td>0.0</td>
</tr>
<tr>
<td>Disgust</td>
<td>0.0</td>
<td>0.0</td>
<td>95.2</td>
<td>4.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Fear</td>
<td>0.0</td>
<td>9.5</td>
<td>0.0</td>
<td>81.0</td>
<td>0.0</td>
<td>0.0</td>
<td>9.5</td>
</tr>
<tr>
<td>Happy</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>94.1</td>
<td>0.0</td>
<td>5.9</td>
</tr>
<tr>
<td>Sad</td>
<td>4.8</td>
<td>0.0</td>
<td>4.8</td>
<td>0.0</td>
<td>0.0</td>
<td>90.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Surprise</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>4.8</td>
<td>0.0</td>
<td>0.0</td>
<td>95.2</td>
</tr>
</tbody>
</table>

Figure 8: Human feedback interpretation using behaviour-based network.
fact that both of these expressions are quite similar to each other with minor dissimilarities, which makes the recognition task difficult for certain subjects. Table 2 shows the recognition rates of five different hand gestures when tested in real time. Gesture three and gesture four have less than 90 per cent accuracy because of the fact that both of these gestures are sometimes confused with each other as both have similar type of shape. Our hand gesture module reports an average accuracy of around 95 per cent (Zafar and Berns 2016) and facial expression module reports an average accuracy of 91 per cent (Al-Darraji et al. 2016).

<table>
<thead>
<tr>
<th>Gestures</th>
<th>Images</th>
<th>Correctly Recognized</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>150</td>
<td>149</td>
<td>99.3%</td>
</tr>
<tr>
<td>Two</td>
<td>150</td>
<td>145</td>
<td>96.7%</td>
</tr>
<tr>
<td>Three</td>
<td>150</td>
<td>133</td>
<td>88.6%</td>
</tr>
<tr>
<td>Four</td>
<td>150</td>
<td>134</td>
<td>89.3%</td>
</tr>
<tr>
<td>Five</td>
<td>150</td>
<td>150</td>
<td>100%</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td>94.78%</td>
</tr>
</tbody>
</table>

5.3.2. Game Scenario Evaluation

For experimentation, 5 different subjects play 20 questions game with the robot. As few subjects don’t know how to play the game, robot explains the game and game-play at the beginning. To answer multiple choice question, basic hand gestures are used. An information card is placed beside the robot for the user to consult which hand gesture represents which answer option. For evaluation of the experiments, subject plays the game with the robot. System writes the activation of all the behaviours in a separate file along with head, hand, face gestures and the question for every frame. A Human expert analyses the subject behaviours in the recorded videos. He observes the feedback behaviours whenever a question has been asked by the robot and fills in the questionnaire. He also takes notice of subtle changes in the behaviour, for example lack of interest, understanding or not understanding behaviours etc. At the end of the experiments, the recorded file by the system is compared with expert questionnaire. Fig. 9 shows plots of different behaviours recognized by the system (blue line) along with behaviours detected by human expert (red line).

Generally, the system recognized key moments (activations) in all the behaviours, which can be seen from the plots. In Fig. 9b, which shows whether a subject is not interested in the game, system recognized that this behaviour is active between 80ms and 100ms. Exactly in the same time frame, human expert also detected the inactivity of this behaviour. Similarly, in Fig. 9a, our system follows the human results. Here, one important statistic can be seen, that even a small value of activeness of this behaviour, in this case 0.2, suggests that this behaviour is active. Similarly, in perceiving and not perceiving behaviour, as soon as the value activity threshold increases above 0.2, these behaviours are active as shown in Fig. 9c and Fig. 9d. However, in Fig. 9e, which depicts understanding behaviour, the activity threshold is above 0.6 to be recognized as understanding behaviour. While system missed don’t understanding behaviour couple of times and activate this behaviour at wrong time frames. This incorrect recognition of this behaviour depends mainly on shaking and wrinkling of eyebrows. After analysing the facial expression and action units, we find out that action unit 4 (AU4), which is wrinkling of eyebrows, is active when eyebrows are clearly and visibly wrinkled. Since humans can detect small subtle changes with ease, hence faint changes in the appearance, in this case eyebrows, are detected in contrast to our facial expression module. For acceptance and denial don’t accept behaviour as shown in Fig. 9g, system recognizes these behaviours precisely in accordance with human expert results.

After recognizing feedback behaviour, the most important step is the fusion of all these feedback behaviours. It has to be noted that at any time instant, the system can recognize multiple feedback behaviours in the same way as humans detect these cues. However, human takes the most active behaviour from all of them. Similarly, a fusion system has been designed that takes all the behaviours as input and outputs the feedback behaviour with maximum activity as shown in Fig. 8. The robot, using this feedback behaviour, adapts itself and reacts differently based on human feedback. Fig. 1a shows experimentation setup, where a subject plays a game and Fig. 1b shows that uninterested behaviour of human.

6. CONCLUSION AND FUTURE WORK

In literature, several studies have been done on human robot interaction through non-verbal communication so far. However, automatic recognition of human feedback during human robot interaction is totally ignored. In this domain of study, we present a real time automated system, which is able to recognize human feedback behaviours. Using psychological studies, a human feedback behavioural model has been generated, which is based on four major communicative feedbacks, i.e., continuation, perception, understand, and acceptance. For recognition of feedback behaviours, facial expressions and head gestures are used. In order to validate our claims, a 20 question game has been
Real-time Perception of Non-verbal Human Feedback in a Gaming Scenario
Al-Darraji • Zafar • Berns

Figure 9: Feedback types. A Comparison between system activity and human observations.
implemented. Using hand gestures, a person can answer the questions while the robot analyses the behavioural cues to recognize feedback behaviours and accordingly adapts itself by visualizing human behaviour. The system shows promising results for initial experiments. In our future work, we would like to revise our fusion module and instead of using a maximum activation function, we propose to use a learning algorithm such as neural network that makes the fusion depending on the previous history and activation of other feedback behaviours.

REFERENCES


