Affect Sensing in an Affective Interactive e-Theatre for Autistic Children

Li ZHANG

School of Computing

Teesside University, UK

l.zhang@tees.ac.uk

Abstract:

In this paper we describe an interactive emotional social virtual framework for autistic young people age 11 – 14 to learn verbal and non-verbal emotional expression in role-play situations. In order to provide an interactive learning environment with some degree of automatic metaphorical understanding from open-ended text input, new developments on affect detection for the processing of several metaphorical languages have been presented. The emotional gesture and facial animation have been created for users' avatars and activated by affective states detected from users' text input. Emotion appraisal and affect detection functions have been embodied in an intelligent conversational agent, who interacts with human users for drama improvisation. We have also conducted user testing with 24 autistic children to evaluate the overall framework. The work has the potential to improve autistic young people's ability in language learning and their social connection with other people. It contributes to the conference themes on natural language processing, interactive affective social interfaces and user studies of intelligent interfaces.

Keywords:

Affect detection; emotional expressive animation; metaphor

1. Introduction

Autism generally have difficulties with social communication and interaction. Especially, they find it difficult to understand facial expression, metaphor, etc. They appear to be insensitive to emotion and suffer from various learning disabilities. Although autism is not curable, research showed the conditions could be improved dramatically if treated when the diagnosed individuals are very young. However, current manual special care for autistic children is very costly. Moreover, due to the fact that the characteristics of autism vary from one to another, few existing autism software could be capable enough to fulfill different users' requirements.

While there have been numerous computerised training systems for helping autistic children none have combined rich social behaviour with interactivity. Many systems have involved training children to recognise emotion. These systems (for example the Mind Reading and Transporters systems developed by the University of Cambridge [3]) have tended to use static images and videos to display the emotions and do not simulate any sort of interaction. This means that, while players are taught to recognise emotions, they do not learn appropriate responses and interactions. There have also been interactive virtual worlds systems such as the AS Interactive system by the University of Nottingham [7], but this does not provide rich verbal and non-verbal behaviour that can help learn subtle social cues. KASPAR [14], a child-sized humanoid robot, with minimum facial expressions and a few body movements, has been developed for autistic children to encourage imitation and turn-taking during interaction. However, it is restricted to limited social communication skills with a desire of a good coverage of social affective communication skills on verbal and non-verbal aspects in future developments.

Our project provides an interactive affective social environment with real-time subtle affective cues in verbal and non-verbal communication for autistic children age 11 to 14. The system has the potential to help autistic children to learn emotional expression in body language and verbal communication (such as using metaphor) and improve their ability in language learning and their social connection with other people. We have created an interactive improvisational chat-room like environment with user avatars. In order to enhance emotional communication for autistic young people, affect detection including metaphor interpretation, emotion appraisal and simple emotional facial and gesture animation have been employed for the development of an intelligent conversational agent who interacts with human users in role-play situations. The intelligent agent detects 25 affective states in users' text input and informs users' avatars to perform simple emotional facial and gesture animation for some particular detected affective states. Users' avatars are cartoon-like 2D characters and are capable of expressing 8 particular emotional states (happy, anger, fear, sad, surprise, caring, threatening, and neutral) and 1 non-emotional state (greeting). The overall system adopts client/server architecture. The intelligent agent plays as a special AI client, while other human users

perform as normal clients. The communication between the server and all the clients uses XML stream. Human users may log into the system remotely and they may be assigned to different chat rooms which relate to different scenarios. There are two scenarios used in our current application: School Bullying and Skin Cancer, which are both highly emotionally charged topics. Users may need to play characters in each scenario after they log into the system and only their characters' names are shown on the screen. Thus users are anonymous. We borrow basic affect detection component developed in our previous work [24] for this application. New developments on interpretation of food, affect, size and animal metaphor have been integrated into the component for the purpose of conducting metaphorical communication effectively with autistic young people, who normally find it very hard and challenging to use metaphorical language.

The paper has been organized in the following way. We report technically relevant work on affect sensing and interactive storytelling systems in section 2, new developments on metaphor interpretation in section 3, system architecture, emotion appraisal of the AI agent and emotional expressive animation in section 4. Finally we report user studies with autistic young people and draw conclusions in section 5.

2.   Relevant work

Textual affect sensing is a rising research branch for natural language processing. ConceptNet [16] is a toolkit to provide practical textual reasoning for affect sensing for six basic emotions, text summarization and topic extraction. Shaikh et al. [20] provided sentence-level textual affect sensing to recognize evaluations (positive and negative). They adopted a rule-based domain-independent approach, but they haven't made attempts to recognize different affective states from open-ended text input.

Although Façade [17] included shallow natural language processing for characters' open-ended utterances, the detection of major emotions, rudeness and value judgements is not mentioned. Boucouvalas [4] demonstrated an emotion extraction module embedded in an Internet chatting environment. It used a part-of-speech tagger and a syntactic chunker to detect the emotional words and to analyze emotion intensity for the first person (e.g. 'I' or 'we'). Unfortunately the emotion detection focused only on emotional adjectives, and did not address deep issues such as figurative expression of emotion (discussed below). Also, the concentration purely on first-person emotions is narrow. There has been relevant work on general linguistic clues that could be used in practice for affect detection (e.g. Craggs and Wood [8]).

There is also well-known research work on the development of emotional conversational agents. Egges

et al. [9] have provided virtual characters with conversational emotional responsiveness. Elliott et al. [10] demonstrated tutoring systems that reason about users' emotions. They believe that motivation and emotion play very important roles in learning. Virtual tutors have been created in a way that not only having their own emotion appraisal and responsiveness, but also understanding users' emotional states according to their learning progress. Aylett et al. [2] also focused on the development of affective behaviour planning for the synthetic characters. Cavazza et al. [6] reported a conversational agent embodied in a wireless robot to provide suggestions for users on a healthy living life-style. Hierarchical Task Networks (HTN) planner and semantic interpretation have been used in this work. The cognitive planner plays an important role in assisting with dialogue management, e.g. giving suggestions to the dialogue manager on what relevant questions should be raised to the user according to the healthy living plan currently generated. The user's response has also been adopted by the cognitive planner to influence the change of the current plan. The limitation of such planning systems is that they normally work reasonably well within the pre-defined domain knowledge, but they will strike when open-ended user input going beyond the planner's knowledge has been used intensively during interaction. The system we present here intends to deal with such challenge.

Our work is distinctive in the following aspects: (1) affect detection in metaphorical expression; (2) real-time affect sensing for basic and complex affects, meta-emotions, value judgments etc (including 25 affective states) in improvisational role-play situations from open-ended user input; (3) and real-time simple facial and gesture emotional animation activated by the detected affective states.

3.   Affect detection from metaphor

Before we introduce the new developments on affect, food, and animal metaphor, we introduce our previous work on affect detection briefly. In our previous work [24], the affect detection component could deal with language with online chatting features such as abbreviations, acronyms etc. We have produced pre-processing procedures including spelling checking, abbreviation checking, and Metaphone algorithm dealing with letter repetitions in interjections and onomatopoeia in order to recover the standard user input. The recovered user input is sent to the Rasp parser to obtain syntactic information. We have particularly focused on affect detection from users' input with potential emotional implication, such as diverse imperatives ("Lisa, go away", "you leave me alone", "Dave bring me the menu", "do it or I'll kill u") and statements with a structure of 'first-person + present-tense verb' ("I like it",

"I hate u", "I enjoy the meal") [24]. In addition, we only used a simple approach for the detection of affect intensity by checking punctuation (e.g. repeated exclamation marks) and capitalization etc in users' input.

Overall, we have adopted rule-based reasoning, robust parsing, pattern matching, semantic and sentimental profiles (e.g. WordNet and a semantic profile [11]) in our approach. Jess, the rule engine for Java platform, has been used to implement the rule-based reasoning while Java has been used to implement other algorithms and processing with the integration of the off-the-shelf language processing tools, such as Rasp and WordNet.

We have made further developments on affect detection especially from four different types of metaphorical expressions – affect, food, animal and size metaphor – in order to provide some extent automatic metaphorical understanding in the interactive learning environment for autistic young people. Affect intensity has been further explored on size metaphor, size adjectives and degree adverbs. The implementation detail is presented in the following.

3.1 Affect detection on affect metaphor

Affect terms have been used intensively during online interaction. Besides they have been used literally to convey users' emotional states (e.g. "I am angry", "I get bored"), affect terms have been mentioned in affective metaphorical language. One category of such metaphorical expression is 'Ideas/Emotions as Physical Objects" [1, 15], e.g. "joy ran through me", "my anger returns in a rush", "fear is killing me" etc. In these examples, emotions and feelings have been regarded as external entities. The external entities are often, or usually, physical objects or events. Therefore, affects could be treated as physical objects outside the agent in such examples, which could be active in other ways [1]. Implementation has been carried out to provide the affect detection component the ability to deal with such affect metaphor.

WordNet-affect domain (part of WordNet-domain 3.2) [22] has been used in our application. It provides an additional hierarchy of 'affective domain labels', with which the synsets representing affective concepts are further annotated. Rasp has been used to detect statements with a structure of 'a singular common noun subject + present-tense lexical verb phrase' or 'a singular common noun subject + present-tense copular form + -ing form of lexical verb phrase'. Various user inputs could possess such syntactic forms, e.g. "the girl is crying", "the big bully runs through the grass" etc. We use WordNet-affect to refine the user inputs in order to obtain metaphorical affective expression. The singular common noun subject is sent to WordNet-affect in order to obtain the hierarchical affect information. If the

subject is an affective term such as 'panic', then the hierarchical affect information obtained from WordNet-affect is 'negative-fear -> negative-emotion -> emotion -> affective-state -> mental-state'. The system realizes that a mental state has been used as a subject which carries out an activity indicated by the verb phrase(s). Thus the system regards such expression as affective metaphor belonging to the category of 'affects as entities'. A further processing based on the hierarchical affect result leads to the exact affective state conveyed in user's input – fear (negative emotion). If such input has a first-person object, 'me' (such as "panic is dragging me down"), then it indicates the user currently experiences fear. Otherwise if such input has a third-person object, 'him/her' (such as "panic is sweeping over and over him"), it implies that it's not the user who currently experiences 'fear', but another character.

If the subject of the user input is not an affect term (e.g. "the girl is crying", "the boy sweeps the floor"), other suitable processing methods (e.g. checking syntactic information and affect indicators etc) are adopted to extract affect. On the whole, such processing is indeed at a very initial stage. However, it provides a useful way to recognize both affect from user input and affect metaphor in which emotions are used as external entities.

3.2 Affect detection on food metaphor

Food has been used extensively as metaphor for social position, group identity, religion, etc. E.g. food could be used as a metaphor for national identity. British have been called 'roastbeefs' by the French, while French have been referred to as 'frogs' by the British. It has also been used to indicate social hierarchy. E.g. in certain Andean countries, potatoes have been used to represent poor rural farmers of native American descent and white flour and bread have been used mainly to refer to wealthy European descent. In our school bullying scenario, the big bully has called the bullied victim (Lisa) names, such as "u r a pizza", "Lisa has a pizza face" to exaggerate that fact that the victim has acne. Another most commonly used food metaphor is to use food to refer to a specific shape. E.g. body shape could be described as 'banana', 'pear' and 'apple' (http://jsgfood.blogspot.com/2008/02/food-metaphors.html). In our application, "Lisa has a pizza face" could also be interpreted as Lisa has a 'round (shape)' face. Therefore, insults could be conveyed in such food metaphorical expression. Inspection of the collected transcripts during the testing indicated that autistic young people have used food metaphor to imply potential insults towards other characters (e.g. "u pizza face", "u banana"), although some of them seemed struggling in understanding such usage at the beginning (because of

their autism condition).

We especially focus on the statement of 'second-person/a singular proper noun + present-tense copular form + food term' to extract affect. A special semantic dictionary has been created by providing semantic tags to normal English lexicon. The semantic tags have been created by using Wmatrix [19], which facilitates the user to obtain corpus annotation with semantic and part-of-speech tags to compose dictionary. The semantic dictionary created consists mainly of food terms, animal names, measureable adjectives (such as size) etc with their corresponding semantic tags due to the fact they have the potential to convey affect and feelings.

In our application, Rasp informs the system the user input with the desired structure - 'second-person/a singular proper noun + present-tense copular form + noun phrases' (e.g. "Lisa is a pizza", "u r a hard working man", "u r a peach"). The noun phrases are examined in order to recover the main noun term. Then its corresponding semantic tag is derived from the composed semantic dictionary if it is a food term, or an animal-name etc. E.g. "u r a peach" has been regarded as "second-person + present-tense copular form + [food-term]". WordNet has been employed in order to get the synset of the food term ('peach'). If among the synset, the food term has been explained as a certain type of human being, such as 'beauty', 'sweetheart' etc. Then another small slang-semantic dictionary collected in our previous study containing terms for special person types (such as 'freak', 'angle') and their corresponding evaluation values (negative or positive) has been adopted in order to obtain the evaluation values of such synonyms of the food term. If the synonyms are positive (e.g. 'beauty'), then we conclude that the input is an affectionate expression with a food metaphor (e.g. "u r a peach").

However, in most of the cases, WordNet doesn't provide any description of types of human beings when explaining a food term (e.g. 'pizza', 'meat' etc). According to the nature of the scenarios (e.g. bullying) we used, we simply conclude that the input implies insulting with a food metaphor when calling someone food terms ("u r walking meat", "Lisa is a pizza").

Another interesting phenomenon drawing our attention is food as shape metaphor. As mentioned earlier, food is often used as a metaphor to refer to body shapes (e.g. "you have a pear body shape", "Lisa has a garlic nose", "Lisa has a pizza face"). They might indicate literal truth, but most of which are potentially used to indicate very unpleasant truth. Thus they could be regarded as insulting. We extend our semantic dictionary created with the assistance of Wmatrix [19] by adding terms of physiological human body parts, such as face, nose, body etc. For the user's input with a structure of 'second-person/a singular proper noun + have/has +

noun phrases' informed by Rasp, the system provides a semantic tag for each word in the object noun phrase. If the semantic tag sequence of the noun phrase indicates that it consists of a food term followed by a physiological term ('pizza face'), the system interprets that the input implies insulting with a food metaphor.

However, examples, such as "you have a banana body shape" and "you are a meat and potatoes man", haven't been used to express insults, but instead the former used to indicate a slim body and the latter to indicate a hearty appetite and robust character. Other examples such as "you are what you eat" could be very challenging theoretically and practically. They also indicate the direction of the future extension of our current system.

3.3 Affect detection on animal metaphor

In order to provide an interactive improvisational environment with some extent of metaphorical understanding, we have integrated our previous work on animal metaphor processing [25] using WordNet and a semantic profile developed by Esuli and Sebastiani [11] into our current framework. Briefly we mainly intend to provide automatic processing on the user input with a structure of "second-person/a singular proper noun + present-tense copular form + [animal-name]", which could convey affectionate ("u r a lion") or insults ("Lisa is a pig"). WordNet has been used to analyse the animal name. If WordNet provides an 'adjective + noun' description of the characteristics of a person/woman/man (e.g. "a famous man" and "a disgraceful woman") as one interpretation of the animal name, then Esuli and Sebastiani's semantic profile is used to obtain the evaluation value of the adjective. If it's positive ("a famous man"), then the user input is metaphorical affectionate expression, otherwise ("a disgraceful woman") it is a metaphorical insulting expression. However, as stated in the food metaphor, WordNet has rarely provided descriptions of the characteristics of a human being as interpretation of an animal name, but only for those animal names with strong affect implication in culture background. In common sense, calling someone a baby animal name may indicate affection, while calling someone an adult animal name may convey insults. Thus we have put animal names in our own newly created semantic dictionary with the semantic tag – 'L2' to indicate living creatures generally for adult animal names [19] and the semantic tag – 'L2y' to indicate young living creatures for baby animal names (e.g. 'puppy', 'bunny', 'kitten' etc). When WordNet is not able to provide any characteristic description of a human being as a further explanation for an animal name, we look up the semantic dictionary to recover a semantic tag for this animal name. If the semantic tag indicates that it is a baby animal name, then the user input is

regarded as affectionate. If the tag implies that an adult animal name has been used in the user input, we simply interpret that the input is animal metaphorical expression currently. Although autistic young people generally have difficulty with understanding and usage of metaphorical expression, our user study indicated that animal insulting or affectionate metaphor has been used extensively during the improvisation by the testing subjects. They also have used special person types, such as "u nazi", "you soviet", "u git" etc to imply insulting.

Overall, we make some initial attempts on understanding of several different types of metaphor. Although the processing we report is at a very initial stage and there are other figurative language phenomena such as irony, humor, simile etc that we haven't even touched, our work could point out a potential positive direction for affect detection in metaphorical figurative language. It also provides an interesting framework with automatic metaphor interpretation to enable autistic children to learn emotional expressions in metaphor.

4. System architecture and emotional animation

As mentioned above, our system adopts client/server architecture. The conversation AI agent and other human-controlled characters consist of clients. The server broadcasts messages sent by one client to all other clients. Thus user's text input from normal user client is sent to the AI agent client via the server. Then the AI agent, who plays a minor role in the improvisation with other human-controlled characters, analyzes the user's text input and derives the affective implication out of the text. Then the AI agent also searches its knowledge base to provide a suitable response to the human players using the detected affective states. We have particularly created the AI agent's responses in a way which could stimulate the improvisation by generating sensitive topics of the storyline. Then an XML stream composed of the detected affective state from one user input and the AI agent's response is dynamically created and broadcasted to all other clients by the server. The users' clients parse the XML stream to obtain the information of the previous "speaker's" emotional state and the current AI character's response. An animation engine has embedded in each user client which updates the user avatars' emotional facial and gesture animation on each user's terminal. Therefore, if the previous human-controlled character expresses 'anger' affective state by saying "r u messing with me!!!", the animation engine in each user client updates emotional animation of that character on each terminal using cross behavior via simple facial and gesture animation (see Figure 1). In each session, up to four characters are engaged in.

Emotional expressive animation using facial, gesture and social attention (such as eye gazing) has been widely studied. There has been extensive work on

the expression of emotion for virtual characters (Vinayagamoorthy et al. [23]), for example work by Pelachaud and Poggi [18] on facial expression and Hartmann et al. [13] on bodily expression. There is more to human expressive behaviour than emotion. For example, Cassell et al. [5] have done extensive work on gesture and other forms of expression during conversation and Gillies et al. [12] have worked on the expression of interpersonal relationship during interaction.
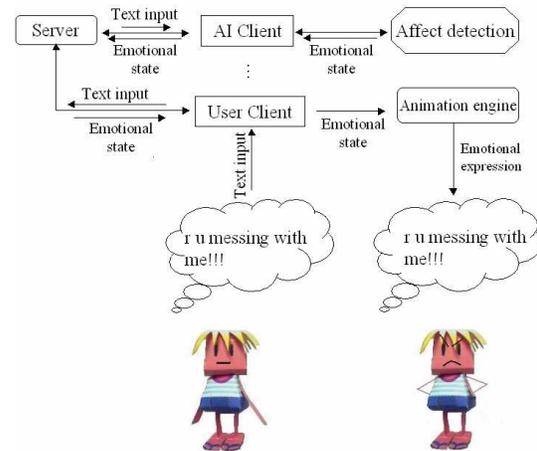

Figure 1. An interactive real-time example

We have adopted an approach of generating simple facial and gesture animation dynamically. We have assigned different lip shapes, eye brow shapes and arm positions dynamically to different emotional expression. Expressive animation has been considered for eight emotional states including 'neutral', five of Ekman's basic emotions – 'happy', 'sad', 'fear', 'angry', 'surprise' – and another two complex emotions, 'threatening', 'caring', and an non-emotional state 'greeting'. If the AI character derives an emotional state from a human-controlled character's text input, emotional animation engine in each client updates the emotional expression of that user' avatar on each client terminal. Since the detection of affective states only takes less than 1 second, the emotional animation updating happens right after the server broadcasts the user's new input. The overall system could provide the effects that the users' avatars move in a way which is consistent with their emotional states implied in their text input in real-time application. Although user avatars' emotional animation is truly basic and simple, we obtained very positive feedback from the testing subjects based on the analysis of the post questionnaires and the discussion in the debriefing session. More results are reported on this aspect in user testing section. Therefore it provides a potential platform for autistic children to understand emotional expression in verbal and non-verbal communication during interaction.

Relationships between characters play an important

role in how characters respond to one another. We have implemented a simple emotion appraisal model for the AI character. In School Bullying scenario, there are four characters: the bully, the victim, and two close friends of the victim who try to stop the bullying. The AI agent plays a minor role – one male close friend of the bullied victim. If the AI character realizes that the bully seems being aggressive ('rude', 'angry' or 'threatening') implied in his text input, the AI character becomes 'angry' due to the fact that they have a negative relationship. Similarly, the AI character would become 'caring', if the bullied victim indicates 'sad' or 'fear' in the text input during the interaction. It also indicates that the AI character and the bullied victim have a positive relationship. In the meantime, the AI character's responses also reflect its current emotional states.

5.   User testing and conclusion

We have conducted a user testing with 24 school children with autism spectrum disorders age from 11 to 14 in Education Village, Darlington. The main intention of the user testing is to evaluate if such improvisational interactive system with the AI character and emotional animation for users' avatars would be useful or not to improve affective communication and social integration for autistic pupils. We also intend to find out if the AI character's involvement and the simple emotional expressive animation for users' avatars will enhance users' experience on learning emotional expression in verbal and non-verbal communication. The testing subjects have different levels of autism condition, although they all have the ability to engage in an online activity and reasonable level of verbal communication skills. We have classified them based on their condition into three groups. Group 1 has the children age 11 – 13 with a bit more serious autism condition and low-level ability in learning, and they are from a special need school in Education Village. Group 2 is composed of children age 13 – 14 with medium level of learning ability and autism condition, which enables them to go to mainstream school. Group 3 consists of age 13 – 14 autistic young people with good ability in learning and the slightest autism condition of these three groups. They are also from a mainstream school in Education Village.

The methodology for the user testing is presented in the following. We made each participant have an experience of the two scenarios, School Bullying (SB) and Skin Cancer (SC). In each scenario, there are four characters: 1 leading character (the bullied victim in SB and the character with cancer in SC), 2 supporting characters (the bully and the close friend of the bullied victim in SB and the Mum and Dad of the character with cancer in SC) and 1 minor character (another close friend of the bullied victim in SB and the close friend of the character with cancer in SC). The AI agent plays a minor

character in School Bullying scenario only with other characters all human-controlled, while in Skin Cancer scenario, we had all the characters played by human users. In this way, we could not only find out the contribution of the affect detection component and the expressive animation, but also compare the performance of the AI minor character in SB with that of the human-controlled minor character in SC.

During the role-play, the AI character only accepted user input for affect sensing and provided an appropriate response based on the detected affective states every two turn-takings of users' input, but it carried out affect analysis and made a response immediately whenever its confidence score of affect interpretation reached a reasonably high level or its character's name was mentioned by the human-controlled characters. We concealed the fact an AI character was involved in one scenario in order to gain more feedback and fair evaluation of the AI agent's performance. In general, the participants experienced the AI character's contribution to the drama improvisation and the simple facial and gesture expressive animation of their avatars in the School Bullying scenario, while the participants didn't have such features, but only worked with still cartoon avatars and all human-controlled characters in Skin Cancer scenario. Post questionnaires have been filled up by the testing subjects. We have questions regarding to the overall enjoyment and engagement of the role-play (18 questions), the appearance of the user avatars and quality of social interaction (12 questions), the expressive emotional behavior of the user avatars and the movement of the avatars in corresponding with users' text input (6 questions) and the performance of each character in each scenario (8 for School Bullying and 7 for Skin Cancer). In total, we have 51 questions.

We especially report a few results on how affect detection component contributed to the expressive animation. In post questionnaires, we have questions regarding to expressive animation and avatars' movement in corresponding to users' text input, such as "my avatar seemed to show emotions in consistent to my text input (question 1)", "my avatar moved in the way I thought it should (question 2)", "I could tell what the characters felt by the way their avatars moved (question 3)" and "what I said made my avatar move differently (question 4)". Users are given a scale from 1 (not at all) to 7 (very much) to indicate their extent of agreement for each question. Figure 2 illustrates the average values of users' feedback for these four particular questions. Users' average scores for questions 1, 2, 3 and 4 are respectively 4.3, 4.4, 5.2 and 4.3. Questions 1, 2, and 4 are mainly referring to testing subjects' experience on their own avatars' expressive animation via automatic affect detection. Considering the fact that the AI character only carried out affect sensing every 2 turn-takings of users' input, it is acceptable that testing

subjects had the impression that their own avatars had expressed their intended emotional or sometimes non-emotional expressions appropriately roughly in half of the times in general. While they observed other participants' avatars during the interaction as question 3 stated, the testing subjects believed that they witnessed more chances that other characters' avatars seemed have showed what their characters felt during the role-play. We report Cohen's Kappa statistics for affect detection performance later in this session in order to provide a suitable measure of how accurate the affect sensing processing is.

We have also compared the performance of the AI minor character in SB with that of the human-controlled minor character in the same scenario, respectively based on to what extent the AI and human-controlled minor characters have said strange things and to what extent they have contributed to the improvisation usefully based on the analysis of the questionnaires filled up by the three groups with low-level (Group 1), medium-level (Group 2) and good-level (Group 3) learning abilities.

From the inspection of the results, the AI agent seemed haven't impressed the first group users with the scores for 'useful contribution' AI vs Human, 47% - 63% and the scores for 'saying strange things' AI vs Human, 63% - 47%. In fact during the testing, Group 1 users with serious autism condition seemed to need more time to get into their characters and they were inclined not to stay in their characters during the role-play. Within a restricted period of time, it seemed hard for them to be creative and professional on their improvisation as good as other groups of users. Most of them also needed assistance to fill up their questionnaires. However, once they began to enjoy the role-play, it was also really hard for them to stop the session. The pupils have asked "why do we stop?". Some of them also gave comments such as "interesting, I want to do more" etc. Inspection of the transcripts indicated that their improvisation was reasonably acceptable. In the meantime, they also believed that the AI character contributed less than the human-controlled minor character and said more strange things as well, probably because they realized the AI character wasn't involved in that much in their scenario unrelated topic discussion. For the other two groups of users with medium and good levels of learning abilities and less serious autism condition, the AI character has scored slightly better than the human-controlled minor character with the scores for 'useful contribution' AI vs Human respectively, 56% - 47% (Group 2) and 60% - 53% (Group 3) and the scores for 'saying strange things' AI vs Human respectively, 42% - 60% (Group 2) and 53% - 67% (Group 3). Some pupils from Group 3 expressed they felt more relaxed when doing improvisation online than in real-life situations. The overall results could probably also imply that the AI character's verbal communication style (designed in a normal teenager user way) is closer to that of the Group 2 & 3 users than that of the Group 1 users.
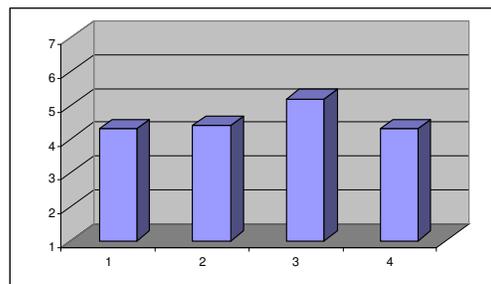

Figure 2. Average scores for sample questions in post questionnaires

Moreover, we have provided Cohen's Kappa in order to evaluate the efficiency of the affect detection processing for the detection of 25 affective states, although simple emotional facial and gesture animation could only deal with 8 particular emotional states. Two human judges (not involved in any development stage) have been employed to annotate part of the recorded transcripts of the SB scenario (52 turn-taking user input) filed from the testing. The inter agreement between human judge A and B is 0.896. The agreements for human judge A/the AI agent and human judge B/the AI agent are respectively 0.662 and 0.729. Usually a good agreement is between 0.6 and 0.8. Thus the AI agent's affect detection performance is acceptable and could achieve satisfactory level in good cases. We will provide more statistical analysis results on the recorded transcripts at a later stage. Finally, the testing subjects gave an average score of 5.79 from range 1 (not at all) to 7 (very much) to indicate to what extent such experience would be useful for real-life situations. No matter if there are simple emotional animations for their avatars or not, users gave an average score of 5.71 from range 1 (not at all) to 7 (very much) to claim to what extent they enjoyed their avatars.

In conclusion, we have reported our new developments on an interactive improvisational environment for autistic young people to improve social affective communication, language learning and enhance emotional learning experience. Our work also makes an attempt in encouraging autistic children to use and understand metaphorical emotional expression during role-play. Finally, our system has the potential to evolve home care and normal classroom education for young people with or without learning disabilities by providing 24/7 efficient personalised social skill and language training and offering automatic monitoring to their progress.

References

[1] ATT-Meta Project Databank: Examples of Usage of Metaphors of Mind. http://www.cs.bham.ac.uk/~jab/ATT-Meta/Databank/. 2008.

[2] R. Aylett, S. Louchart, J. Dias, A. Paiva, M. Vala, S. Woods, and L.E. Hall. "Unscripted Narrative for Affectively Driven Characters". IEEE Computer Graphics and Applications. 26(3), 2006, 42-52.

[3] S. Baron-Cohen. "Two new theories of autism: Hyper-systemizing and assortative mating". Archives of Diseases in Childhood, 91, 2006, 2-5.

[4] A.C. Boucouvalas. "Real Time Text-to-Emotion Engine for Expressive Internet Communications". In Being There: Concepts, Effects and Measurement of User Presence in Synthetic Environments. G. Riva, F. Davide and W. IJsselsteijn (eds.), 2002.

[5] J. Cassell & K.R. Thórisson. "The Power of a Nod and a Glance: Envelope vs. Emotional Feedback in Animated Conversational Agents". International Journal of Applied Artificial Intelligence, 1999.

[6] M. Cavazza, C. Smith, D. Charlton, L. Zhang, M. Turunen and J. Hakulinen. "A 'Companion' ECA with Planning and Activity Modelling". In Proceedings of the 7th International Conference on AAMAS. Portugal, 2008.

[7] S. Cobb, L. Beardon, R. Eastgate, T. Glover, S. Kerr & H. Neale. "Applied virtual environments to support learning of social interaction skills in users with Asperger's Syndrome". Digital Creativity, 13, 2002.

[8] R. Craggs & M. Wood. "A Two Dimensional Annotation Scheme for Emotion in Dialogue". In Proceedings of AAAI Spring Symposium: Exploring Attitude and Affect in Text. 2004.

[9] A. Egges, S. Kshirsagar & N. Magnenat-Thalmann. "A Model for Personality and Emotion Simulation", In Proceedings of Knowledge-Based Intelligent Information & Engineering Systems, Lecture Notes in AI. 2003.

[10] C. Elliott, J. Rickel & J. Lester. "Integrating Affective Computing into Animated Tutoring Agents". In Proceedings of IJCAI'97 Workshop on Intelligent Interface Agents, 1997.

[11] A. Esuli and F. Sebastiani. "Determining Term Subjectivity and Term Orientation for Opinion Mining". In Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics, Trento, IT. 2006.

[12] M. Gillies & D. Ballin. "Integrating autonomous behavior and user control for believable agents". In Proceedings of the Third International Joint Conference on Autonomous Agents and Multi-Agent Systems. New York, 2004.

[13] B. Hartmann, M. Mancini. & C. Pelachaud. "Implementing Expressive Gesture Synthesis for Embodied Conversational Agents". Gesture in Human-Computer Interaction and Simulation, 6th International Gesture Workshop, France, 2005.

[14] H. Kose-Bagci, K. Dautenhahn & C.L. Nehaniv. "Emergent Dynamics of Turn-Taking Interaction in Drumming Games with a Humanoid Robot". In Proc. IEEE RO-MAN 2008, Munich, Germany, 2008.

[15] Z. Kövecses. "Are There Any Emotion-Specific Metaphors?" In Speaking of Emotions: Conceptualization and Expression. Athanasiadou, A. and Tabakowska, E. (eds.), Berlin and New York: Mouton de Gruyter, 1998, 127-151.

[16] H. Liu & P. Singh. "ConceptNet: A practical commonsense reasoning toolkit". BT Technology Journal, Vol22, 2004, Kluwer Academic Publishers.

[17] M. Mateas. Ph.D. Thesis. "Interactive Drama, Art and Artificial Intelligence". School of Computer Science, Carnegie Mellon University, 2002.

[18] C. Pelachaud & I. Poggi. "Subtleties of Facial Expressions in Embodied Agents", Journal of Visualization and Computer Animation, vol. 13, 2002.

[19] P. Rayson. "Matrix: A statistical method and software tool for linguistic analysis through corpus comparison". Ph.D. thesis, Lancaster University., 2003.

[20] M.A.M. Shaikh, H. Prendinger & I. Mitsuru. "Assessing sentiment of text by semantic dependency and contextual valence analysis". In Proceeding of ACII. 2007, 191-202.

[21] S. Sharoff. "How to Handle Lexical Semantics in SFL: a Corpus Study of Purposes for Using Size Adjectives". Systemic Linguistics and Corpus. London: Continuum, 2005.

[22] C. Strapparava and A. Valitutti. "WordNet-Affect: An Affective Extension of WordNet", 1083–1086, In Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004), Lisbon, Portugal, 2004.

[23] V. Vinayagamoorthy, M. Gillies, A. Steed, E. Tanguy, X. Pan, C. Loscos & M. Slater. "Building Expression into Virtual Characters". In Eurographics Conference State of the Art Reports, 2006.

[24] L. Zhang, J.A. Barnden, R.J. Hendley, M.G. Lee, A.W. Wallington & Z. Wen. "Affect Detection and Metaphor in E-drama". Int. J. Continuing Engineering Education and Life-Long Learning, Vol. 18, No. 2, 2008, 234–252.

[25] L. Zhang, M. Gillies & J.A. Barnden. "EMMA: an Automated Intelligent Actor in E-drama". In Proceedings of International Conference on Intelligent User Interfaces, Spain, 2008.