

# Northumbria Research Link

Citation: Zhang, Li, Gillies, Marco, Dhaliwal, Kulwant, Gower, Amanda, Robertson, Dale and Crabtree, Barry (2009) eDrama: Facilitating online role-play using an AI actor and emotionally expressive characters. *International Journal of Artificial Intelligence in Education*, 19 (1). pp. 5-38. ISSN 1560-4292

Published by: IOS Press

URL: <http://ijaied.org/pub/990/> <<http://ijaied.org/pub/990/>>

This version was downloaded from Northumbria Research Link:  
<http://nrl.northumbria.ac.uk/3863/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

[www.northumbria.ac.uk/nrl](http://www.northumbria.ac.uk/nrl)



# **E-Drama: Facilitating Online Role-play using an AI Actor and Emotionally Expressive Characters**

**Li Zhang**, *School of Computing, University of Teesside, TS1 3BA, UK*

**Marco Gillies**, *Department of Computer Science, University College London, Malet Place, London WC1E 6BT, UK*

**Kulwant Dhaliwal**, *Hi8us Midlands Ltd., Unit F1, The Arch, 48-52 Floodgate Street, Birmingham, B5 5SL, UK*

**Amanda Gower, Dale Robertson, Barry Crabtree**, *BT Innovate, Adastral Park, Ipswich, Suffolk, IP5 3RE, UK*

**Abstract.** This paper describes a multi-user role-playing environment, referred to as “e-drama”, which enables groups of people to converse online, in scenario driven virtual environments. The starting point of this research, is an existing application known as “*edrama*”, a 2D graphical environment in which users are represented by static cartoon figures. Tools have been developed to enable integration of the existing *edrama* application with several new components to support avatars with emotionally expressive behaviours, rendered in a 3D environment. The functionality includes the extraction of affect from open-ended improvisational text. The results of the affective analysis are then used to: (a) control an automated improvisational AI actor – EMMA (emotion, metaphor and affect) that operates a bit-part character in the improvisation; (b) drive the animations of avatars using the Demeanour framework in the user interface so that they react bodily in ways that are consistent with the affect that they are expressing. Finally, we describe user trials that demonstrate that the changes made improve the quality of social interaction and users’ sense of presence. Moreover, our system has the potential to evolve normal classroom education for young people with or without learning disabilities by providing 24/7 efficient personalised social skill, language and career training via role-play and offering automatic monitoring.

**Keywords.** E-drama, affect detection, emotionally expressive behaviour and an educational improvisational interactive environment.

## **INTRODUCTION – OVERVIEW OF THE SPRINGBOARD *EDRAMA***

We started our research based on a system called *edrama*, developed by Hi8us Midlands Ltd (<http://www.edrama.co.uk>), a charitable company. It is online multi-user role-play software that could be used for education or entertainment. In *edrama*, young people could interact with others online in a 2D based interface under the guidance of a human director. The interface incorporates 2D static avatars and a text chat interface, with different photographic backgrounds as scenes to set the role-play. Up to five human actors and one human director are involved in any one session. Actors can choose the clothes and bodily appearance for their own characters. The actors are given a loose scenario within which to improvise. Since in *edrama*, only the characters’ names will be shown on the screen (not users’ real names), it allows users to remain anonymous. This is particularly useful to

young people who may be afraid of expressing their views in front of their peers. Moreover, facilitation of role-playing is a crucial aspect of *edrama*. This ensures that the “chat” is purposeful and assists users to respond to the given situations. The recorded improvisational scripts for *edrama* sessions are also very valuable for the production of films or short movie clips.

This system has been used as a tool for teaching in the formal education sector, not only in the teaching of drama, but in a wide range of subject areas, such as career advice (e.g. “Dream factory” system) and creative writing. Thus, *edrama* has the potential to be used to deliver almost any type of training, in an engaging and entertaining way. Due to the collaborative and multi-user nature of *edrama*, people are able to learn together remotely, with potential geographical and social barriers removed.

Although this 2D version of *edrama* (see Figure 1) has been successfully implemented in a range of situations and continues to be deployed as new opportunities arise, it has the potential to benefit from additional features, such as using animated 3D characters and an automated improvisational AI actor (a computer-controlled actor, who plays a minor character in a scenario, helps monitor and contributes to the progression of drama improvisation). This paper describes an alternative version of the *edrama* software which we refer to as “e-drama” – developed in collaboration with Hi8us, Maverick TV, Birmingham University and BT Innovate with the support of the PACCIT programme. Our collaboration aims to enrich the user-experience with emotionally responsive characters, including an additional non-human automated AI actor within a 3D application. The AI actor detects affective states from users’ open-ended text input and it also makes an appropriate response according to the detected affective states and its role in the drama. The addition of 3D capabilities includes character and background scene rendering and enables real-time processing of animation to visually update the current emotional state of every character on screen by adopting the detected affective states from the AI agent.

The affect detection functionality embedded in the AI agent is limited to the language phenomena we focused on for this current study (for detail, see affect detection module section). In our affect detection processing, we only detect affective states based on single user text input without any consideration of context-based information, although the work is also accompanied by basic research into how affect is conveyed linguistically. However, the affect detection function provides an efficient channel for the expressive animation engine to adopt real-time detected affective states implied in the users’ input into the production of the users’ avatars’ real-time expressive behaviour.



Fig.1. Children using Hi8us’ 2D version of *edrama* (previous version).

Thus, the new version of e-drama has the potential to be integrated into a school environment to provide a fun experience for training and learning in the following subjects: social skills, personal development, languages and the sciences. Our system creates a safe anonymous efficient learning environment and it could also be used by young people with a learning disability or language impairment to engage in learning and interaction without the fear of failure. In the long-run, our application provides a great possibility to help young people achieve their full potential in terms of both social and learning skills, which could increase their confidence and improve their self-esteem.

In the following sections we will describe relevant work and the three main components integrated into the original *edrama* system. We describe the prototype actor application supporting 3D backgrounds and avatars, and integration with existing *edrama* functionality; we then cover the two components that create the expressive characters: the improvisational AI actor and Demeanour. We go on to outline the user studies, including the scenarios used, the experiment setup and results.

As research objectives, we mainly intend to explore how to sense affect from open-ended improvisational text input. We also aim to create a computational model to automatically recognize metaphorical expression and the implied affective states to some extent. For our e-drama application, we have created the AI actor with an affect detection function. Therefore we intend to find out if the involvement of the AI actor will make any difference or not to users' enjoyment and engagement via user testing, although it could reduce some burden of a human director. In other words, we expect the AI actor to perform as well as any other 14-15 year old school pupil. We expect that the AI actor's involvement will not necessarily improve the users' experience but we don't expect the experience to be worse either. Generally, we also intend to find out how to employ the detected affective states in the 3D animation engine to provide expressive animation for user avatars without interfering with the improvisation. We aim to conduct user studies to evaluate the above research developments. Finally, although the addition of 3D capabilities probably does not always provide extra benefits for educational applications, we envisage it has the potential to enhance users' overall experience and improve social integration. Employing 3D expressive animation could also be very educational for applications which teach autistic young people to learn emotional expression in non-verbal communication during interaction.

## **RELATED WORK**

Much research has been done on creating affective virtual characters in interactive systems (see Vinayagamoorthy et al. (2006) for an overview). Picard's work (2000) makes great contributions to building affective virtual characters overall. Now we report relevant technologies and related work in the areas of emotion modelling, conversational agents, expressive animation and interactive narrative in the following subsections.

### **Emotion Modelling**

Emotion theories, particularly that of Ortony, Clore and Collins (1988) (OCC), have been used widely therein. Ortony et al. include a rule-based system for the generation of the 22 emotion categories, which is widely used for emotion generation for the development of intelligent virtual agents.

In Prendinger and Ishizuka's work (2001), animated agents are capable of performing affective communication. They have defined eliciting conditional rules for more frequently used OCC emotion

categories. Then the generated emotional states are filtered by two control states – personality and social role awareness – in order to achieve believable emotional expression. Wiltshko's eDrama Front Desk (2003) is an online emotional natural language dialogue simulator with a virtual reception interface for pedagogical purposes. A natural language parser is used to analyse users' text input which indexes into a list of phrases that are frequently used. Then the system dialogue manager selects an output phrase for the computer character. The emotion model of this system is derived from the OCC model based on two factors: respect and power. Mehdi et al. (2004) combined the widely accepted five-factor model of personality (McCrae & John, 1992) with mood and the OCC emotion model in order to generate emotional behaviour for a fireman training application. Personality and mood have also played important roles in emotion generation and emotion intensity adaptation. Gratch and Marsella (2004) presented an integrated model of appraisal and coping, in order to reason about emotions and to provide emotional responses, facial expressions and potential social intelligence for virtual agents. Their main contribution is the introduction of computational description of emotional coping for the first time. They have also extended the scope of discourse on appraisal theories by incorporating influence between cognition and appraisal to obtain emotional coping.

In our application, although we haven't used any emotional modelling and coping strategies, we focus intensively on how emotion is conveyed linguistically. Emotional labels from Ekman (1982) and the OCC model have been borrowed for our work.

## **Conversational Agents**

There is also well-known research work on the development of emotional conversational agents, for example, Egges et al. (2003) have provided virtual characters with conversational emotional responsiveness. Elliott et al. (1997) 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 have their own emotion appraisal and responsiveness, but also understand users' emotional states according to their learning progress. Aylett et al. (2006) also focused on the development of affective behaviour planning for the synthetic characters. Cavazza et al. (2008) reported a conversational agent embodied in a wireless robot to provide suggestions for users on a healthy living life-style. This work uses a Hierarchical Task Network (HTN) planner and semantic interpretation. 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 based on 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 fail when open-ended user input going beyond the planner's knowledge has been used intensively during interaction. However, our system has the potential to deal with this challenge.

## **Expressive Animation**

There has been extensive work on the expression of emotion for virtual characters (Vinayagamoorthy et al., 2006), for example work by Pelachaud and Poggi (2002) and Tanguy et al. (2003) on facial expression or Chi et al. (2000) and Hartmann et al. (2006) on bodily expression. There is more to

human expressive behaviour than emotion. For example, Cassell and her colleagues (Cassell & Thórisson, 1999; Cassell et al., 2002) have done extensive work on gesture and other forms of expression during conversation and Gillies et al. (2002) have worked on the expression of interpersonal relationships during interaction.

There are two basic classes of method for creating expressive body animation. Procedural methods (e.g. Kopp & Wachsmuth, 2004) generate animations algorithmically. They tend to be very flexible in terms of the range of animations they produce but tend to lack the individual personality of hand animated or motion captured data. The other approach is to use data, either created by hand and animator or motion captured from an actor. The main drawback of this method is that you are limited to the motion data that was originally created and it is difficult to create new animations at run time. A number of methods have been proposed to overcome this problem. The most relevant ones for emotional animation are interpolation methods (e.g. Rose et al., 1998), which blend a number of animations together to create new ones, and style learning methods (Brand & Hertzmann, 2000; Li et al., 2002) in which a dynamical system characterising a set of animations is learned, and then used to generate new animations. Both of these approaches work on data sets of similar motions. This limits the diversity of the animations that they can produce, essentially stylistically different versions of the same movement. This limitation means that many emotions are not easily expressed, as they would require very different types of movement. For this reason we take the approach of composing rather than interpolating animations (essentially a simplified, though more general version of the approach in (Heck et al., 2006)).

## **Interactive Narrative**

There are two basic approaches to interactive narrative. The first is a top-down story centred approach (e.g. Riedl & Young, 2006). These methods have a central story manager that has a branching or graph structured story, the details of which are instantiated based on user interaction, often by using a planning algorithm. This is well suited to structured interactions with a strong pre-authored narrative, but less suited to the more free form narratives that result from the role-plays in e-drama. In our case the direction of the narrative should primarily emerge from the actions of the participants. We therefore use an approach closer to the bottom-up, character-driven emergent narratives such as FearNot! (e.g. (Aylett et al., 2006)). In this approach the narrative emerges from the interaction between artificially intelligent characters and between the characters and players. There have also been attempts to make hybrid systems where the actions of characters are influenced by a drama manager or overall story goals (e.g. (Mateas, 2002) or (Magerko et al., 2004)). These systems are individual experiences in which a single person interacts with multiple characters, which means that the characters primarily drive the narrative. Our system inverts this set up with multiple participants whose interactions drive the narrative and a single AI character that acts as a facilitator, rather than the primary focus of the story.

## **E-DRAMA**

To take part in an e-drama session a user runs an “actor” client locally on their desktop. The user must first login, to select an available session and character to play. The login screen enables each actor to login with a unique ID. This ID can be anonymous – a numerical identifier for example or can be a

username. In all cases this login does not require any personal details and is not referred to in the role-play. The login interface presents a number of options, including key stage (if required), the role-play option and the characters available in the session. A total of five characters are available in each role-play.

Customisation of the character takes place in a virtual “dressing room”, available after login (see Figure 2). This includes scenario details and customisation tools. The 3D window includes an interactive text panel which displays background information about the scenario and selected character, and also renders a default avatar in the 3D dressing room, animated with general waiting poses. Actual customisation is through the e-fit tool in the flash interface, which provides click through image based selection of gender, head, torso and leg options. The results are displayed in real time in the 3D dressing room window.

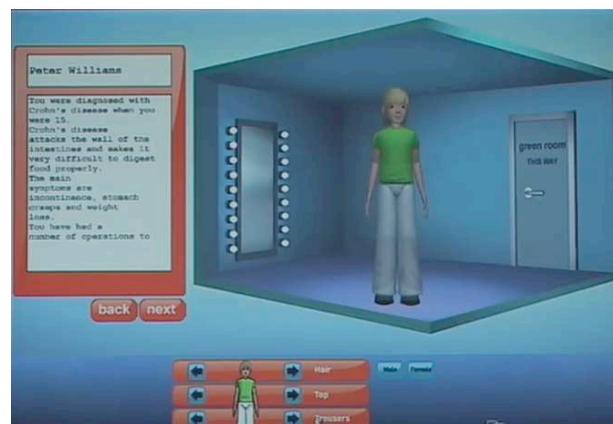


Fig.2. Avatar customisation.

From this point the user moves into the multi-user sections. The first of these is the “green room” (see Figure 3), which is a warm up space to meet other actors and the director. Usually, the characters are encouraged to get to know each other and discuss their roles in the green room to warm up. In e-drama, we provide loose scenarios (details for scenarios in user study section) and character profiles without any constrained scripts to facilitate creative improvisation. The 3D window displays the green room scenery and all logged in avatars. Text chat is displayed in speech bubbles above the avatars’ heads. Text is input via the flash panel.

The director activates the stage once all the characters have appeared in the green room and are warmed up. The director signals the start of the role-play using an “Action” command, which warns the actors of the scene change. A background scene from a library of options is displayed on all clients. This may be updated at anytime during the session. The role-play is ended using a “Cut” command from the director – at which point the application will close down all actor clients.

In both the green room and stage environments, each actor is given a set position on screen. This tableaux format provides the user with a view of the whole role-play and ensures avatar positions are consistent on each actor client.

When the director speaks to the group or individual avatars a 2D image representing the director overlays the client window and text appears in a speech bubble (see Figure 4). In this way the director can appear to the group or to single clients and give directions to assist the role-play. When an actor types in text in the chat panel, the text appears in bubbles above the avatar. Each character is animated according to its emotional profile and to the text input of users during the session.



Fig.3. Four actors in the green room.

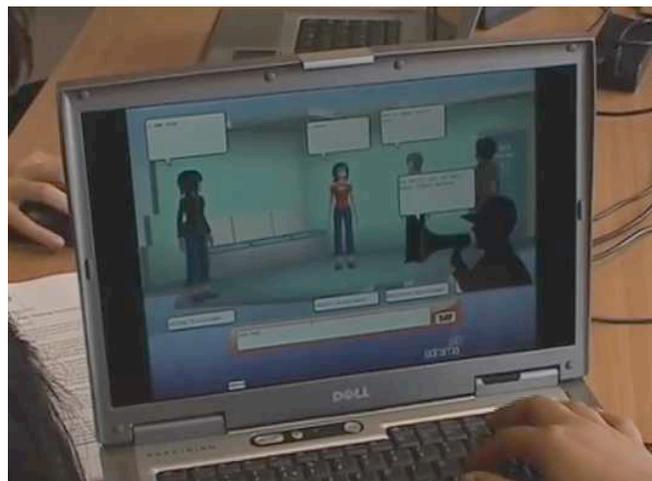


Fig.4. Director interacts with the actors.

## E-drama Application

This section describes, from a technical perspective, the architecture of the e-drama application. New features for e-drama include: the introduction of 3D animated gesturing avatars and 3D computer-generated backgrounds, an automated improvisational actor and the addition of documentary film material about the scenarios for the role-play, supplied by our industrial partner, Maverick TV Ltd. This is made available during an initial phase of an e-drama session, in which the actors familiarise themselves with the background to the scenario. Actors are given loose scenarios to read which provide a structure for creative improvisation. The scenarios chosen for our application are School Bullying, Crohn's Disease and Homophobic Bullying (see user testing sections for details).

The architecture of the new version of the e-drama software is shown in Figure 5. It consists of two main user interfaces, an “actor” client application that is used by the actors, and a “director” client application which is available solely to the director. The “actor” clients communicate with each other and the director through a central server.

The director interface remains largely unchanged compared to the 2D version. It is a web-based interface which incorporates a number of tools to start role-play sessions, view the scene and avatars, and monitor the conversation. The director can start, stop and change background scenes, and talk to one or all of the participants using text chat. In contrast, the actor client has undergone significant developments to support the real-time rendering of expressive characters and is the focus of the user study.

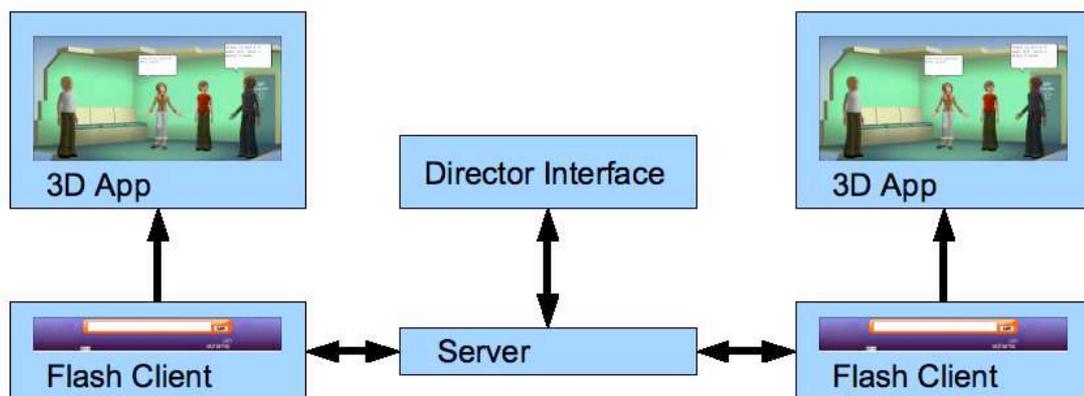


Fig.5. The architecture of the e-drama networked application.

The 3D version of the e-drama actor client is an MS Windows based application written using the Microsoft Foundation Classes user interface library. The application consists of two child windows; one houses the Flash Player ActiveX control to enable Flash interfaces to be displayed within that window; the other is a window that displays the 3D visuals. Hi8us’ original *edrama* is a web-browser hosted Flash application (Flash is a technology for developing interactive, graphical web applications). In order to maintain consistency the main user interface and networking software of the new client are still written in Flash. This flash component is embedded in a Microsoft Windows application, which also contains a 3D graphics component that displays the scene containing the avatars.

The 3D graphics are based around the TARA engine for creating real-time 3D enabled applications, developed by BT Innovate. The engine provides a set of extensible components that are used to render geometry and effects using Microsoft DirectX 3D graphics library. TARA allows its core components to be replaced by other functional components tailored to meet a specific need. This provides a mechanism in which to integrate new technologies into TARA enabled applications, without the components of the application interface needing to know about them; in this case the Demeanour framework (Gillies & Ballin, 2004). The creation of an alternative 3D system attempts to enrich the environment provided by e-drama to reinforce the emotional content of the role-play.

The flash component of the new e-drama application is the user interface that controls the flow of the application. It is a simplified version of the original 2D web-based interface so that a move from

one to another would require no learning on the part of the user; this also means that this prototype is compatible with the current Hi8us version therefore the two can be used in parallel. The flash component is the client to the external server passing state related messages for each e-drama client and capturing messages broadcasted by the server. The flash interface communicates with the 3D component through a local socket maintained by the application. Messages broadcasted by the server are passed through this socket to be processed. The communication between interface and 3D application is one way only, from interface to application as the application is responsible for reflecting the current state based upon users' interactions.

## Emotionally Expressive Characters

In our e-drama system, each scenario has a written description, or profile, of five characters who are able to participate. There is usually a main character or protagonist, who faces a conflict or issue. This character will have a counterpart who is the antagonist and takes an opposing view. The remaining characters will have specific relationships to these characters (parent, friend, enemy). In many scenarios the basic character profiles have a similar pattern to provide the basis of a productive role-play. In Hi8us' versions of *edrama* this information can only inform the performance of the actors engaged in the role-play. In the 3D version described here it becomes influential in how the avatars are animated on screen.

Using a combination of character profiles and detected affective states from user's text input it is possible to animate each character with expressive behaviour, without any direct user intervention via the e-drama interface. This employs a combination of two technologies, affect detection in open-ended improvisational text (Zhang et al., 2006) and the Demeanour framework (Gillies & Ballin, 2004). Figure 6 gives an overview of the control of the expressive characters. Users' text input is analysed by EMMA in order to detect affect in the text. The output is an emotion label with intensity derived from the text. This is then used in two ways. Firstly it is used by the bit-part character to generate a response. Secondly the label is sent to the emotional animation system (via an XML stream) where it is used to generate animation.

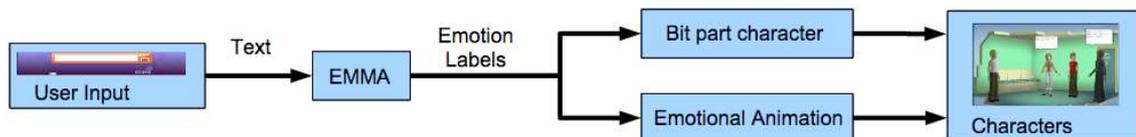


Fig.6. Affect detection and the control of characters.

## AFFECT DETECTION IN OPEN-ENDED IMPROVISATIONAL TEXT

In e-drama the actors (users) are given a scenario within which to improvise (improvisation in this context means to be creative in role-play and contribute to drama progression based on one's role). There is also a human director, who monitors the unfolding drama and can intervene by, for example, sending messages to actors, or by introducing and controlling a minor "bit-part" character to interact with the main characters. This character will not have a major role in the drama, but might, for example, try to interact with a character who is not participating much in the drama or who is being ignored by the other characters. Alternatively, it might make comments intended to "stir up" the

emotions of those involved, or, by intervening, diffuse any inappropriate exchange developing. But this places a heavy burden on directors, especially if they are, for example, teachers who are unpractised in the directorial role. One research aim is to partially automate the key directorial functions which mainly involve affect detection. For instance, a director may intervene when emotions expressed or discussed by characters are not as expected based on their role in a scenario (e.g. the bullied victim of the school bullying scenario doesn't seem bothered by the big bully and on the contrary feels very happy). Hence we have developed an affect-detection module. The module identifies affect in characters' text input, and makes appropriate responses to help stimulate the improvisation. Within affect we include: basic and complex *emotions* such as anger and embarrassment; *meta-emotions* such as desiring to overcome anxiety; *moods* such as hostility; and *value judgements* (of goodness, etc.). Although merely detecting affect is limited compared to extracting full meaning, this is often enough for stimulating improvisation. The results of this affective analysis are then used to: (a) control an automated improvisational AI actor, EMMA, that operates a bit-part character in the improvisation; (b) drive the animations of the avatars in the user interface so that they react bodily in ways that are consistent with the affect that they are expressing, for instance by changing posture. The response generation component of EMMA uses this interpretation to build its behaviour driven mainly by EMMA's role in the improvisation and the affect expressed in the statement to which it is responding. The intention of EMMA's response is to stimulate the improvisation.

There has been only a limited amount of work directly comparable to our own, especially given our concentration on improvisation and open-ended language. However, *Facade* (Mateas, 2002) included shallow natural language processing for characters' open-ended utterances, but the detection of major emotions, rudeness and value judgements is not mentioned. Zhe and Boucouvalas (2002) demonstrated an emotion extraction module embedded in an Internet chat environment. It uses a part-of-speech tagger and a syntactic chunker to detect the emotional words and to analyse emotion intensity for the first person cases (e.g. "I" or "we"). The emotion detection focuses only on emotional adjectives, and does not address deep issues such as figurative expression of emotion. Also, the concentration purely on first-person emotions is narrow. We might also mention work on general linguistic cues that could be used in practice for affect detection (Craggs & Wood, 2004).

Our work is distinctive in several respects. Our interest is not just in (a) the first-person, positive expression of affect case: the affective states or attitudes that a virtual character X implies that it itself has (or had or will have, etc.), but also in (b) affect that the character X implies it lacks, (c) affect that X implies that other characters have or lack, and (d) questions, commands, injunctions, etc. concerning affect. We also aim for the software to cope partially with the important case of communication of affect via metaphor (Fussell & Moss, 1998), and to push forward the theoretical study of such language, as part of our research on metaphor generally (see, e.g. Barnden et al., 2004).

## **Our Affect Detection Module**

Affect interpretation in open-ended text is very challenging and even one person's judgement could vary in different circumstances. In order to achieve a reliable affect interpretation, we have combined various weak affect indicators during the processing. Our work was initially inspired by Weizenbaum's Eliza (1966), the first interaction system based on natural language textual input. Thus, we have adopted rule-based reasoning, robust parsing and semantic and sentimental interpretation to suit our application. Since our current system is a rule-based system with the integration with other

language processing tools, we also intend to adopt other statistical (e.g. Bayes theorem) or biometric techniques for the processing of some particular language phenomena as further development. These aspects will be reported at a later stage. The detailed affect detection processing is reported in the following section.

### ***A. Pre-processing Modules & Affect Detection using Rasp, Pattern Matching & WordNet and Responding Regimes***

The language in the textual “speeches” created in e-drama sessions severely challenges existing language-analysis tools if accurate semantic information is sought, even in the limited domain of restricted affect-detection. The language includes abbreviations, misspellings, slang, use of upper case and special punctuation (such as repeated exclamation marks) for affective emphasis, repetition of letters, syllables or words for emphasis, and open-ended interjective and onomatopoeic elements such as “hm”, “ow” and “grrr”. To deal with the misspellings, abbreviations, letter repetitions, interjections and onomatopoeia, several types of pre-processing occur before the main aspects of detection of affect, while the affective states or intensities implied in the repetition of letters, syllables, words and interjections are also recorded by the system for further processing. If there is affect derived from the text input at a later stage, then the affect detected by this stage is not counted (e.g. “*hahaha*, I’m going to kill u” – “threatening” rather than “happy”). If there is no affect recovered from the text, then the affect recorded at this pre-processing stage is counted (e.g. “Wowwww, he’s here” – “surprised” rather than “neutral”). We have also employed some contents of an online slang dictionary into our system in order to derive affect from such user input. We have reported our work on pre-processing modules to deal with these language phenomena in detail in Zhang et al. (2006).

Now we briefly introduce our work on the core aspects of affect detection. One useful pointer to affect is the use of imperative mood, especially when used without softeners such as “please” or “would you”. Strong emotions and/or rude attitudes are often expressed in this case. There are common imperative phrases we deal with explicitly, such as “shut up” and “mind your own business”. They usually indicate strong negative emotions. But the phenomenon is more general. Detecting imperatives accurately in general is by itself an example of the non-trivial problems we face. Expression of the imperative mood in English is surprisingly various and ambiguity-prone, as illustrated below. We have used the syntactic output from the *Rasp* parser (Briscoe & Carroll, 2002) and semantic information in the form of the semantic profiles for the 1,000 most frequently used English words (Heise, 1965) to deal with certain types of imperatives. Briefly, the grammar of the 2006 version of the *Rasp* parser that we have used incorrectly recognised certain imperatives (such as “you shut up”, “Dave bring me the menu” etc.) as declaratives. We have made further analysis of the syntactic trees produced by *Rasp* by considering the nature of the sentence subject, the form of the verb used, etc., in order to detect imperatives. We have also made an effort to deal with one special case of ambiguities: a subject + a verb (for which there is no difference at all between the base form and the past tense form) + “me” (e.g. “Lisa hit/hurt me”). The semantic information of the verb obtained by using Heise’s (1965) semantic profiles, the conversation logs and other indicators implying imperatives help to find out if the input is an imperative or not.

In an initial stage of our work, affect detection was based purely on textual pattern-matching rules that looked for simple grammatical patterns or templates partially involving specific words or sets of specific alternative words. This continues to be a core aspect of our system but we have now added robust parsing and some semantic analysis, including but going beyond the handling of imperatives discussed above.

A rule-based Java framework called Jess is used to implement the pattern/template-matching rules in EMMA allowing the system to cope with more general wording, 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 (Fellbaum, 1998). Since Jess is the rule engine for the Java platform and Java has APIs available for its integration with Jess, their integration would create a much more stable system than those created by the integrations between Java and any other expert programming language (such as Lisp or Prolog). Therefore, we chose to produce rule-based reasoning in our application using Jess. In the textual pattern-matching, particular keywords, phrases and fragmented sentences are found, but also certain partial sentence structures are extracted. This procedure possesses the robustness and flexibility to accept many ungrammatical fragmented sentences and to deal with the varied positions of sought-after phraseology in characters' utterances. The rules conjecture the character's emotions, evaluation dimension (negative or positive), politeness (rude or polite) and what response EMMA should make. The rule sets created for one scenario have a useful degree of applicability to other scenarios, though there will be a few changes in the related knowledge base according to the nature of specific scenarios.

However, it lacks other types of generality and can be fooled when the phrases are suitably embedded as subcomponents of other grammatical structures. In order to go beyond certain such limitations, sentence type information obtained from the *Rasp* parser has also been adopted in the pattern-matching rules. This information not only helps EMMA to detect affective states in the user's input (see the above discussion of imperatives), and to decide if the detected affective states should be counted (e.g. affects detected from conditional sentences won't be valued), but also helps EMMA to make appropriate responses. Additionally, the sentence type information can also help to avoid the activation of multiple rules, which could lead to multiple detected affect results for one user's input. Mostly, it will help to activate only the most suitable rule to obtain the speaker's affective state and EMMA's response to the human character. The following is the pseudo-code of one example rule for the user's input such as "Peter, don't argue with me".

```
(defrule example_rule
  (any string containing negation and the sentence type is "imperative")
  =>
  (obtain affect and response from knowledge base))
```

Thus, the declarative input such as "I don't like the operation" will not be able to activate the example rule due to different sentence type information.

Additionally, a reasonably good indicator that an inner state is being described is the use of "I", especially in combination with the present or future tense (e.g. "I'll scream", "I hate/like you", and "I need your help"). We especially process "the first-person with a present-tense verb" statements using WordNet. We use WordNet to find the synonyms of the original verb in the user's input. These synonyms are then refined by using Heise's (1965) semantic profiles in order to obtain a subset of close synonyms. The newly composed sentences with the verbs in the subset respectively replacing the original verb have extended the matching possibilities in the pattern-matching rules to obtain the user's affective state in the current input.

For example, if the user's input is "I enjoy the movie very much", we use WordNet to obtain the synonyms of the verb "enjoy". The set of synonyms is refined by using semantic profiles from Heise's dictionary and we obtain rough synonyms "love" and "like". Then we use "love" to replace the verb "enjoy", and send the newly built sentence "I love the movie very much" to the pattern-matching rules

in order to obtain the speaker's affective state and EMMA's response. If we cannot successfully obtain such information, we will build another input sentence using the other synonym "like" and send the sentence "I like the movie very much" to the pattern-matching rules. In general, using WordNet provides us with the benefit of making our affect detection approach more generalised.

After the automatic detection of users' affective states, EMMA needs to make responses to the human characters during the improvisation. We have also created responding regimes for the EMMA character. Most importantly, EMMA can adjust its response likelihood according to how confident EMMA is about what it has discerned in the utterance at hand. Some example transcripts collected during user testing are displayed in section "The user study – AI branch testing".

### ***B. Metaphorical Language Processing in EMMA***

The metaphorical description of emotional states is common and has been extensively studied (Fussell & Moss, 1998) – e.g. "He nearly exploded" and "Joy ran through me," where anger and joy are being viewed in vivid physical terms. Such examples describe emotional states in a relatively explicit if metaphorical way. But affect is also often conveyed more implicitly via metaphor, as in "His room is a cess-pit": affect (such as "disgust") associated with a source item (cess-pit) gets carried over to the corresponding target item (the room). We use EMMA as a useful application of theoretical inspiration for metaphor processing.

Our approach to metaphor handling in the EMMA affect-detection module is partly to look for stock metaphorical phraseology and straightforward variants of it. As an example of stock phrase handling, insults in e-drama are often metaphorical, especially the case of animal insults ("you stupid cow", "you dog"). Particularly the "second-person/a singular proper noun + present-tense copular form" statements (such as "you're a rat", "Lisa is a pig") and the second-person phrases (such as "you dog") are often used to express insults. In the EMMA affect-detection module, we use *Rasp* to locate input of this kind. We have also employed an on-line animal-name dictionary (<http://dictionary.reference.com/writing/styleguide/animal.html>), including names of animals, animal groups, young animals, etc., since usually calling someone a baby animal name (e.g. "puppy") may imply affection while calling someone an adult animal name could convey an insult. Then we use this animal-name dictionary to find out if there is any potentially insulting/affectionate animal name present in the "second-person/a singular proper noun + present-tense copular form" statements or in the second-person phrases. If there is, then we will use WordNet to analyse the animal name. If WordNet provides the semantic information of the animal name containing the description of the characteristics of a person/woman/man, such as "an adjective + person/woman/man", then we will classify input of this kind as metaphorical language. Additionally, we also use another semantic profile developed by Esuli and Sebastiani (2006) to obtain the evaluation value of the adjective preceding the word "person/woman/man". If it is negative (e.g. "a disagreeable person", "a disgraceful person"), then we classify the user's input as metaphorical insulting language. If it is positive (e.g. "a lovely person", "a famous man"), then the user's input will be considered as metaphorical affectionate language. Otherwise, the user's input will be regarded as metaphorical objective language. We list a run-time example in the following:

<the bully (Mayid)> u r a pig, Lisa!!

*Reasoning:*

1. *Rasp* detects such input: "second-person + present-tense copular form";
2. Using animal-name dictionary to find out an animal name in the input: "pig";
3. WordNet: pig -> "a disagreeable woman";

4. *Semantic dict*<sup>1</sup>: “disagreeable”-> negative;
  5. => *metaphorical insulting language + strong intensity (exclamation marks)*;
  6. *The bully seems very rude.*
- <the AI character>Mayid, that’s enough.

However, in some cases, WordNet is not able to provide any characteristics of human beings as one potential explanation for an animal name. In such cases, our system is not capable of sensing the metaphorical expression and the affect implied in it (see the following example).

<the bully (Mayid)> u r a donkey, Lisa!!

*Reasoning:*

1. *Rasp detects such input: “second-person + present-tense copular form”;*
2. *Using animal-name dict to find out an animal name in the input: “donkey”;*
3. *WordNet: donkey -> “ass” (no explanation such as any characteristics of a human being);*
4. => *not metaphorical language + strong intensity (exclamation marks)*;
5. *No affect detected and affect intensity discounted.*

<the AI character> r u an animal lover?

In order to solve this problem, we intend to create a semantic dictionary with different semantic tags for young and adult animal names. As we mentioned earlier, calling someone an adult animal name may indicate an insult, while we may infer affection if calling someone a baby animal name. When WordNet fails, we may resort to the semantic dictionary to obtain the tags for the animal names mentioned in the user input. Then we could at least infer if the user input implies an insult or not.

Not only might animal names present in the above statements convey affective states, but also the use of special person-types (e.g. “freak”) or mythical beings (e.g. “devil”, “angel”) could imply insults or approbations. Thus if there is a noun, but not an animal name, present in the “second-person/a singular proper noun + present-tense copular form” statements or in the second-person phrases, then we collect all of its synonyms using WordNet. If any of the special person-types or mythical beings collected in previous e-drama transcripts is shown among the synonyms, then we conclude that it has an insulting/affectionate flavour.

Sometimes, adjectives instead of nouns are used to directly convey affective states, such as “Lisa is a stupid girl”, “you’re a good mum” and “you stupid boy”. If there is no noun present in the above statements or we are not able to obtain any semantic information by only analysing nouns in the above sentence structures, adjectives become very helpful. We could find out the evaluation values of these adjectives, again using the semantic profile developed by Esuli and Sebastiani (2006). In this way, we may at least find a positive or negative flavour in the user’s input.

One particular phenomenon of theoretical and practical interest is that physical size is often metaphorically used to emphasize evaluations, as in “you are a big bully”, “you’re a big idiot”, and “you’re just a little bully.” The bigness is sometimes literal as well. “Big bully” expresses strong disapproval (Sharoff, 2005) and “little bully” can express contempt, although “little” can also convey sympathy or be used as an endearment. Such examples are not only important in practice but also theoretically challenging. Our work on metaphor outside the e-drama research is focused on an approach and system called ATT-Meta (Barnden et al., 2004). This approach is heavily dependent on detailed utterance-meaning analysis and on rich knowledge bases and reasoning processes, and is

---

<sup>1</sup> Throughout, “dict” is to be understood as shorthand for “dictionary”.

currently unsuitable for direct use in the e-drama system. However, examples arising in e-drama transcripts provide useful data guiding the further development of ATT-Meta and can pose useful challenges to current metaphor theory generally.

The functionality we have achieved so far on automatic affect interpretation of metaphorical language in EMMA is still limited. Other complex and challenging metaphorical language phenomena are far beyond the capabilities of our current system. There are also other important factors, such as irony and lies, which could severely challenge the performance of our current system. However, they also indicate where our future strength needs to lie.

### ***C. One Special Case of Imperative Mood Processing in EMMA***

In e-drama transcripts, imperatives have been used intensively to express affective states, such as mentioned above in section A. There is one special case of imperatives that is particularly interesting: “imperative + conjunction + future tense”, such as “do it and I’ll like it (encouraging)”, “eat it and he’ll buy more (encouraging)”, “do it or I’ll kill you (threatening)”, etc. Although people may argue that they tend to be more like conditional sentences instead of imperatives, in our case we currently simply regard them as one special case of imperatives since such type of sentences may imply affective states that should be considered during improvisation, while affects detected in conditional sentences usually will not be valued, such as “I like the film if it ends in this way”. *Rasp* regards such special imperative input as question sentences. The further processing of *Rasp*’s output has changed the sentence type to imperatives. In this special imperative processing, we believe that the user implies encouraging if the second part of the input – “future tense” – is followed by a positive verb (e.g. “be brave and I’ll support/help you” and “try it and you’ll like it”), while the input has a threatening flavour if “future tense” is followed by a negative verb (e.g. “do it or he/she/I will kick/kill/bang you”). First, we use *Rasp* to detect such input and then we will locate the verb in the “future tense”. Evaluation value of this verb will be obtained by using the semantic dictionary developed by Esuli and Sebastiani (2006). In this way, we may obtain encouraging or threatening affective states from the user’s input, which has been regarded as no emotions at all in the previous processing.

### ***D. The Affect Detection Model***

After the description of various affect processing components, we have summarised these components into one overall affect detection model presented in Figure 7. In order to remind readers, the pre-processing modules are presented in detail in section A. The language phenomena, “second-person/a singular proper noun + present-tense copular form” statements, are described in detail in section B. Special imperatives, “imperative + conjunction + future tense”, are demonstrated in section C.

EMMA has adopted these affect processing procedures to detect affective states in users’ text input. When EMMA is involved in the improvisation, it plays a minor character in the pre-defined loose scenarios. This character is a close friend of the victim character in the school bullying scenario and a close friend of the leading character who has a terrible disease in the Crohn’s Disease scenario. EMMA’s responses are mainly determined by its roles in both scenarios and the detected affective states of other characters’ text input. E.g. if one character is outrageous, then EMMA will try to mediate the discussion (e.g. “Hey guys, let’s all calm down”). If the big bully insults the victim, then EMMA will stand up for its friend and stop the bullying (e.g. “Mayid, is that the way you talk to people”). EMMA will show its sympathy and support if the sick character in the Crohn’s Disease scenario worries about his disease and side-effects of the operation (e.g. “Peter, you’ll get better and I’ll sleep in the hospital with you”).

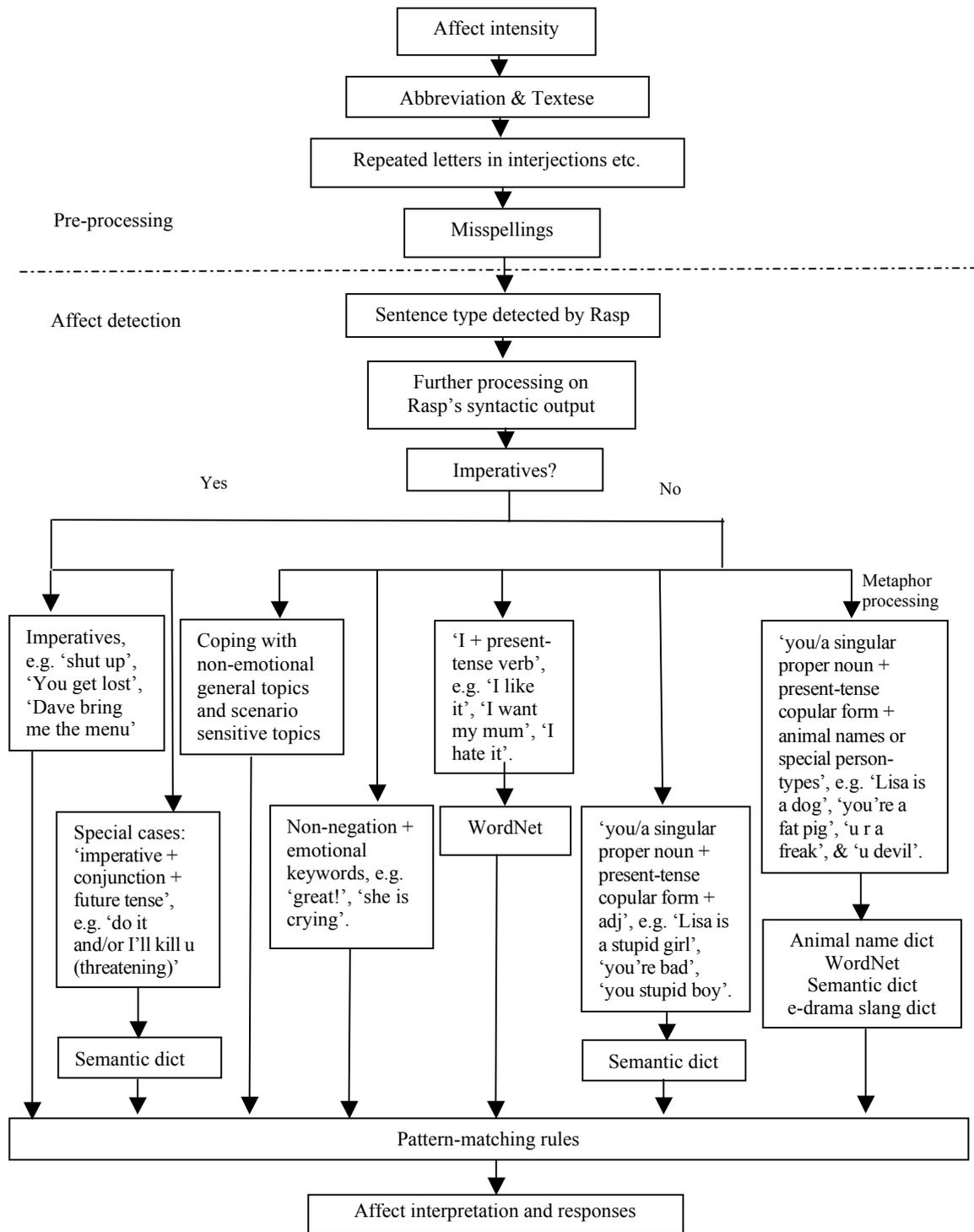


Fig.7. The affect detection model.

The detected affective states in the user's text input and EMMA's responses to other characters have been encoded in an XML stream, sent to the server by EMMA. Then the server broadcasts the XML stream to all the clients so that the detected affective states can be picked up by the animation engine to contribute to the production of 3D gestures and postures for the avatars. In the following section we will discuss the generation of emotional believable animation.

## **THE USERS' AVATARS AND EMOTIONAL ANIMATION**

The topics discussed in the e-drama scenarios are often highly emotionally charged and this is reflected in the animation of the characters. Each participant in e-drama has their own animated graphical character (avatar). In order for the characters to enhance the interaction the characters all have emotionally expressive animations. Garau et al. (2001) point out that avatars that do not exhibit appropriate emotional expression during emotionally charged conversation can be detrimental to an interaction. The problem with animated avatars is that they can be very complex to use if users have to directly control the avatars' animation. Vilhjálmsón and Cassell (1998) have shown that users find controlling the expressive behaviour of animated avatars difficult and their experience and interaction are improved if they use an avatar whose behaviour is controlled autonomously. This is due partly to the complexity of controlling expressive behaviour, but also due to the sub-conscious nature of many types of expression. As people produce expressive behaviour sub-consciously they are not explicitly aware of which behaviours are produced and therefore find it difficult to consciously control the character's behaviour. We therefore have an autonomous model of affective animation for our avatars based on the affective states detected in users' text input. These detected affective states control the animation of the user avatars using the Demeanour expressive animation framework (Gillies & Ballin, 2004). This system makes it possible to express the wide range of emotions detected by EMMA and also takes contextual information into account. The expressive animation displayed by the characters takes two forms. The first concerns attention in group social interaction while the second is the expression of emotion through animation. The overall architecture used is shown in Figure 8.

### **Social Attention**

E-drama deals with social interactions in groups of three to five people. This adds some complexity over and above the case of a two-person interaction. One particular area of complexity is *social attention*, which member of the group does an avatar concentrate on at a given time? This is an important aspect of animation for two reasons. Firstly it is important for animating the avatars' direction of gaze. Avatars will look at the other avatars they are attending to. Secondly, and perhaps most importantly, the avatars should respond to the behaviour of the people they are attending to. In the next section we discuss how the avatars respond to each other.

There have been numerous models of gaze behaviour developed for animated virtual characters. Some such as those of Chopra-Khullar and Badler (1999) or Gillies and Dodgson (2002) deal with gaze directed to inanimate objects in the environment. Others deal with social interaction. The model used by BodyChat (Vilhjálmsón & Cassell, 1998) uses gaze as an important factor in determining turn taking in conversation. Lee et al. (2002) and Vinayagamoorthy et al. (2004) extend this type of model with a statistical model of saccade behaviour based on data from real conversations. The Rickel gaze model (Lee et al., 2007) takes into account a very wide range of factors that can influence gaze

including social, cognitive and emotional factors. These existing models are primarily aimed at two party conversations, so we have developed a model of gaze, that is based on the same principles as the models mentioned but extends them to handle the type of multi-party conversations found in e-drama.

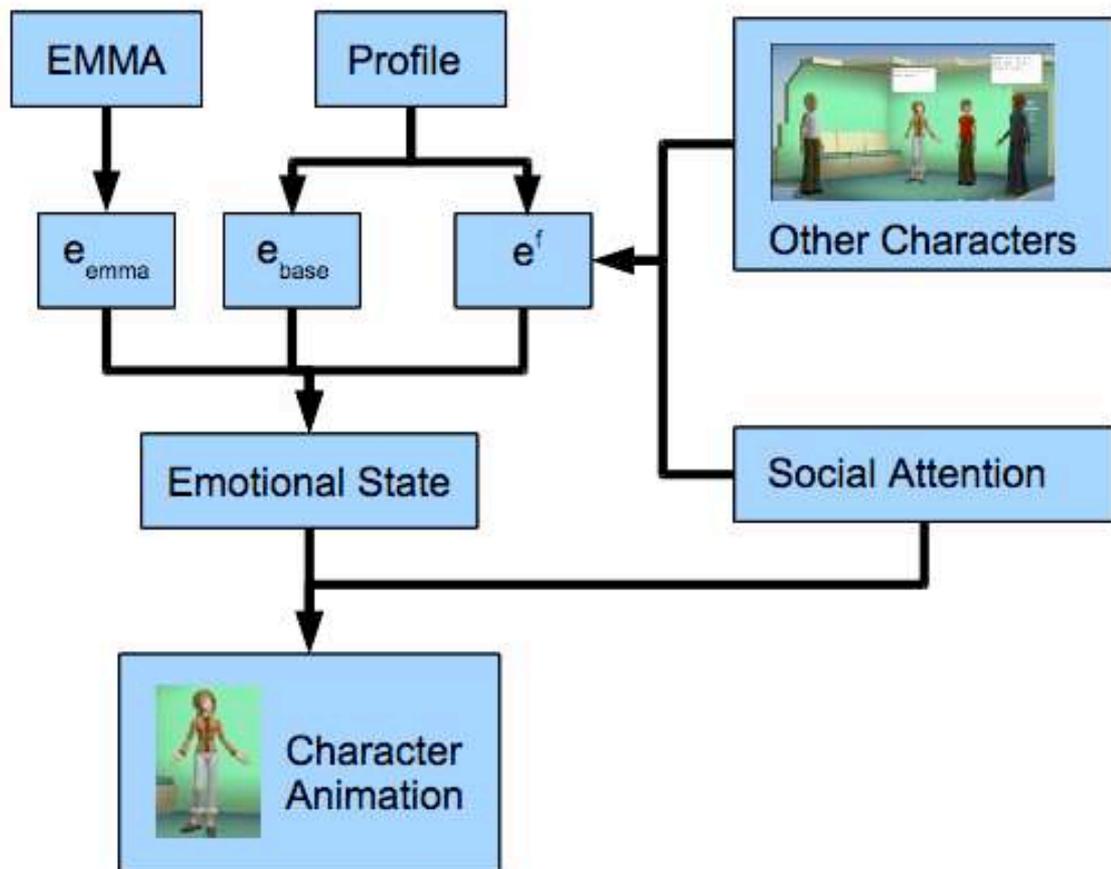


Fig.8. The character animation system.

We have developed a novel model for creating social attention behaviour for avatars. The model assumes that the avatar will have a single focus of attention at a given time (though there may be some peripheral attention, we assume the effect on behaviour is minimal). This focus of attention will either be another avatar, some aspect of the scene or an “attending to nothing” focus. As the non-avatar foci are less behaviourally important in this context (where the scene does not contain objects salient to the conversation) we model them simply as looking at random locations. Our model therefore has to choose between attending to the different avatars in the scene or a random location, and also decide how long to spend on a given focal point.

The model is based on statistics, presented by Argyle (1976), on the proportion of time spent looking at the conversational partner in two person conversation. In order to determine an avatar’s attention the model uses two statistics, the maximum proportion of time spent attending to (looking at) another avatar (attention proportion) and the maximum length of time spent continuously looking at

another avatar (maximum gaze length). These statistics are highly affected by the state of a conversation, and in particular who is talking. Argyle gives separate statistics for gaze while talking and while listening. These statistics can readily be translated into attention proportions and maximum gaze lengths for two situations. The first is the statistics for an avatar attending to other avatars while talking. The second is for attending to another avatar that is talking, while not talking. These are sufficient for a two-person conversation in which there is normally one listener and one speaker (or possibly two speakers), but in groups there is also the possibility of listeners attending to each other as well as the speaker. What is needed are attention proportions and maximum gaze lengths for avatars that are not talking and attending to other avatars that are not talking. As we lack empirical data for this condition we assume that the values in this case are much smaller than in the other case. Table 1 shows the values used in our model for the three conditions, *talking*, *listening* and *neither*.

Table 1  
The attention statistics used in our model

	<b>Attention proportion</b>	<b>Maximum gaze length (seconds)</b>
Talking	0.35	4.0
Listening	0.75	5.0
Neither	0.3	3.0

These statistics are used by our algorithm to determine an avatar’s focus of attention. A new focus is chosen based on the difference between the current proportion of time attended to and the target proportion given in Table 1. This ensures that avatars that have not been sufficiently attended to have a high probability of being chosen, and avatars that have exceeded the target proportion are not chosen. The current focus is maintained until the time spent looking at it exceeds the maximum gaze length. The algorithm is summarized as follows:

1. For each candidate avatar,  $i$ :
  - 1.1. update  $state(i) \in \{talking, listening, neither\}$
  - 1.2. update attention proportion target  $AP\_target(i)$
  - 1.3. calculate current attention proportion  $AP(i)$
  - 1.4. calculate probability of choosing candidate
    - a)  $P(i) = AP(i) - AP\_target(i)$
    - b) if  $P(i) < 0$  then set  $P(i) = 0$
2. normalise all probabilities  $P(i)$
3. select focus =  $i$  with probability  $P(i)$ 
  - 3.1. if all candidates have  $P(i) = 0$ , then select a random location
4. when the current focus has exceeded its maximum gaze length repeat from step 1

The avatar is animated as looking at its current focus of attention by turning its head towards the focus. However, the use of the focus of social attention is not merely used as a cosmetic gaze model for animation purposes. As pointed out by Gillies et al. (2002), it is important the external animation of an avatar’s gaze is linked to the internal perceptual processes that are used to determine its behaviour in such a way that they combine to reinforce the “readability” of the behaviour. We therefore use the focus of attention in a fundamental way to determine the behaviour of our avatars, as described in the next section.

## Emotional Expression

The expressive animation engine, Demeanour, makes it possible for our characters to express the affective states detected by EMMA. When EMMA detects an affective state in a user's text input, this is passed to the Demeanour system attached to this user's character and a suitable emotional animation is produced. However, the emotions detected for an avatar by EMMA are not the only form of affective information available for animating the avatars. There is considerable contextual information that can be used to augment the expressive animation. Firstly, as well as expressing their own emotions avatars should respond to the strong emotional expressions of other participants. If an avatar produces a strong affective state then other avatars will also produce a milder response. Secondly, interaction with e-drama involves taking part in a well-defined scenario and each participant plays a well-defined character with particular personality traits and relationships with other characters. Information about the characters in the scenario can be used to shape the affective responses of the avatars. Certain characters may have innate predispositions to certain emotions, which might be interpreted as a personality feature or a mood lasting for the whole role-play. For example, a character suffering from Crohn's disease might have a tendency towards sadness. This type of personality trait may be expressed by a default, low-level emotional state. Another influence of the characters being played is that the relationships between characters should influence how characters respond to each other. The response an avatar makes to another participant's emotional state should depend critically on the relationship between them. For example, in the Homophobic Bullying scenario we use, the relationship of the bullied character to his friend is very different from his relationship to the bully, and this should be reflected in the affective responses. For example, when the bullied character is sad, the friend responds with an empathic response of sadness, whereas the bully responds with a gloating response of happiness. Adding this information from the scenario to the animation system not only improves the quality of animation, but it also reinforces the personalities of the characters through their animations, and thus helps the participants to better role play their assigned parts.

EMMA is able to detect a wide range of complex emotions and mental states. This means the emotion generation system we use cannot be limited to a small set of emotions such as the basic emotions (Ekman, 1982) that are often used by emotional animation systems. The range of different mental states means that the types of movement used to express those mental states are very diverse. Many emotional animation techniques such as interpolation based methods (Rose et al., 1998) are not suitable because they animate a single type of movement (or a small set) in different emotional styles. With many of the emotions that e-drama handle it is as much the choice of movement that expresses the emotion as the style in which it is performed. For this reason we animate the characters using sets of emotion specific movements that are localized to specific areas of the body. New animations are generated by applying different animations to different parts of the body.

Demeanour is able to animate the wide range of heterogeneous mental states that EMMA can detect, and does so using both the output of EMMA and other contextual information. Each avatar is capable of a number of different emotions, each with a numerical value. The values are calculated based on three factors, a term representing the character's innate emotional tendencies, a term representing the output from EMMA, and a term representing responses to other avatar's emotions. The following equation is used:

$$e^i = e_{base}^i + e_{emma}^i + \sum_{\hat{e}} w_{e\hat{e}}^{if} \hat{e}^f \quad (1)$$

Where  $e^i$  is the value of an emotion  $e$  for character  $i$ ,  $e_{base}^i$  represents the innate emotion tendency and

$e_{emma}^i$  is the value calculated by EMMA. When determining responses to other avatars, the social attention model described in the previous section is used to determine what to respond to. The avatar will only respond to emotions expressed by its current focus of attention, thus making responses realistic, by following human attention patterns, and easy to understand, as the avatar's gaze should make it clear what a given response is to. Thus in the final term we sum over all possible other emotions  $\hat{e}^f$  of the focus of attention, and weight it by a value  $w_{e\hat{e}}^{if}$  which is used to represent the relationship between the avatar and the focus of attention. It is the weighting of the effect of character  $f$ 's emotion  $\hat{e}$  on the emotion  $e$  of character  $i$ . For example, a positive value for  $w_{e\hat{e}}^{if}$  can represent empathy while a negative value can represent hostile relationships. Each character has a separate value of  $e^i$  for each emotion  $e$ , and a value  $w_{e\hat{e}}^{if}$  for each emotion and each other character  $j$  in the scenario. This provides a rich set of parameters with which to describe the relationships and personalities required by an emotionally intense scenario such as the ones used in e-drama.



Fig.9. Examples of emotional animation: uncertain, negative, threatening, reproachful, sad, approving, grateful.

Once values for  $e^i$  have been calculated, they are used to generate affective animation. The emotion of the character is chosen as the emotion with the highest value of  $e^i$ . The animation system is based around a set of short animation clips. These clips are grouped into sets of clips, each group representing one of the possible emotions. The current emotion is animated based on the clips in its group. A subset of the clips is chosen at random and they are combined together to produce a new animation. Each clip only affects an individual part of the body (torso, legs, arms) and thus several clips can be easily combined at the same time, as shown in Figure 10. In order to produce varied behaviour over long time periods, every few seconds a new set of clips is chosen from the group of the current affective state, and these new clips are used to generate a new animation. Examples of affective animations are shown in Figure 9. The animation system also implements affective decay. Any emotion will eventually revert to a neutral state if it is not replaced by a new one. The time taken to revert to a neutral state is chosen randomly for each event, with a constant mean and variance for all emotions.



Fig.10. Combining animations from different body parts to create a complete animation.

## THE USER STUDY – 3D BRANCH TESTING

Our first user study involved two trials of the prototype 3D e-drama application. These were completed in July and October 2006, as new animation capabilities were added to the prototype system (see Figure 11). This section describes the scenarios, the user study and results.



Fig.11. Children using the 3D version of e-drama (new version).

### Objectives

The main objective of the study was to test whether a 3D interface, and affective character animation improved the role-playing experience of the users of e-drama. In detail we had 3 hypotheses:

1. The participants would prefer 3D avatars and avatars with affective animation, that is their subjective appreciation of the avatars would improve with a 3D system and further improve with affective animation.
2. The quality of social interaction would improve slightly with a 3D environment and greatly with affective animation.
3. Other, general, measures of the quality of the experience (enjoyment, presence) would improve with a 3D environment and further with affective animation.

## **The Scenarios**

Three scenarios were used in the user testing; the first was entitled “Big Night Out” which was delivered in the 2D version and served as a warm up; the other two delivered with AI characters in the 3D version were homophobic bullying and Crohn’s disease. In each case, introductory videos produced by Maverick TV were shown to the trialists. This additional video material helps participants identify with the sensitive issues being explored in the scenarios. In these scenarios, Mr Dhanda (Homophobic Bullying) and Dave (Crohn’s Disease) are AI characters driven by EMMA.

### *Big Night Out*

In this scenario, five teenagers – Jenny, Pip, Paul, Imran and Leigh – plan to meet after school to go to the Megabowl club, which opens on that night. Since they are too young to be allowed to enter, Jenny, who lives with her single mum, has got some fake IDs. She has also stolen her mum’s credit card. Pip is worried that they will be found out and his family will be disappointed with him. The other three are also engaged in discussing whether or not they will use the fake IDs to enter the club.

### *Homophobic Bullying*

In this scenario the character Dean (16 years old), captain of the football team, is confused about his sexuality. He has ended a relationship with a girlfriend because he thinks he may be gay and has told her this in confidence. Tiffany (ex-girlfriend) has told the whole school and now Dean is being bullied and is concerned that his team mates on the football team will react badly. He thinks he may have to leave the team. The other characters are: Tiffany who is the ring leader of the bullying, and wants Dean to leave the football team, Rob (Dean’s younger brother) who wants Dean to say he is not gay in order to stop the bullying, Lea (Dean’s older sister) who wants Dean to be proud of who he is and ignore the bullying, and Mr Dhanda (PE Teacher) who needs to confront Tiffany and stop the bullying.

### *Crohn’s Disease*

In this scenario the character Peter has had Crohn’s disease since the age of 15. Crohn’s disease attacks the wall of the intestines and makes it very difficult to digest food properly. The character has the option to undergo surgery (ileostomy) which will have a major impact on his life. The task of the role-play is to discuss the pros and cons with friends and family and decide whether he should have the operation. The other characters are: Mum, Janet, who wants Peter to have the operation, Matthew (older brother) who is against the operation, Dad, Arnold, who is not able to face the situation, and David (the best friend) who mediates the discussion. The setting is a night out for an evening meal.

## Procedures

There were 3 conditions in the user study:

1. Hi8us' 2D version with no animation or affect detection
2. The 3D version with the automated AI characters but limited animation
3. The 3D version with the automated AI characters and full animation

In the version with limited animation the animations only occurred when an emotion was detected by the affect detection component and there was only one animation per emotion. In the full animation condition the full behaviour model described in this paper was used, including social attention, emotional animation and responsive behaviour.

The intention of using the 2D version without affect detection instead of using the 2D version with this function is simply because in the 2D version the users' avatars do not move at all whether the affect detection is activated or not – and visually there is no difference between these two systems. The only functional difference between these two systems is that the system with affect detection involves the AI actor playing the minor characters, while the one without affect detection has all human-controlled characters. In order not to introduce any extra effects to the users of the 2D version, we chose to work with the 2D version without affect detection and the AI actor. However, in order to activate the animation engine using the detected affective states for the 3D versions, we have to include the AI actor and affect detection for both 3D versions to carry out the user study.

The comparison between the 2D version and the 3D versions was performed within subjects while the comparison between the two 3D conditions was performed between subjects. (Such arrangements simply gave us the opportunities to discover directions for further improvements from the testing subjects gradually. There is no deep rationale of such arrangements on testing methodology aspects.) The participants were therefore divided into two groups as show in Table 2.

Table 2  
Experimental conditions

Group	Condition 1	Condition 2
A	2D	3D with limited animation
B	2D	3D with full animation

The two groups were tested in different sessions. Group A used the 2D and 3D versions on different days while group B used them on the same day.

The participants were then asked to role play using the e-drama system for 10-15 minutes per session. They undertook 3 sessions in the first trial but only 2 sessions in the second one. They had less time to undertake the 2D session in the second trial, due to the technical difficulties. We believe that this did not greatly affect the results as there was no significant difference between the scores of the two groups in the 2D condition (other than a low level of significance for the difficulty measure, which as we shall see was not significant in any of the conditions we were actually measuring).

## Participants

There were 10 participants per group. The participants were all female aged between 13 and 14 and pupils at Swanshurst School, a Specialist Science College in Billesley, Birmingham. Using

participants of the same gender allowed us to control for gender effects in social interaction. An accusation often levelled at 3D interactive environments is that they are mostly aimed at male users. As we wanted to ensure that our software is not solely effective with male users we chose female participants.

The participants were randomly assigned into groups of 4 and given a scenario and character. None of the participants knew who the other members of their group were. However, due to the proximity of the terminals, sometimes they were able to establish the identities of fellow participants.

## Measures

We have used post-experiment questionnaires for measurement. Though, these may be less objective than behavioural (e.g. Friedman et al., 2007) or physiological (e.g. Slater et al., 2006) measures, they make it possible to investigate a wider range of complex factors within the user experience. This was important to understand users' reactions to e-drama as a whole. Participants were asked to complete a questionnaire about their experience with the 2D *edrama* before using the 3D version and a second one after using the 3D version. The two questionnaires were mostly identical, but some minor changes were made to the questions to make them applicable, and 7 questions were added that were not applicable to the 2D version. The first questionnaire had 71 questions and the second had 78. All of the questions were 7 point Likert like scales.

The questions were divided into 7 categories, shown in Table 3. For each participant the mean was taken of their answers to the questions in each category and this was used as their score for that category. The mean was then taken for each condition for each category.

Table 3  
The categories of questions used for the questionnaires

Category	Number of questions	Example Questions
Enjoyment	12	"How much did you enjoy the roleplays?"
Difficulty	14	"I needed help to use e-drama"
Presence	2	"I forgot I was at school when I was doing the role-play"
Co-Presence	8	"Did you feel close to the group online?"
Quality of Social Interaction	6	"Did you get to have your say?"
Own Avatar Appearance	3	"I wanted the Avatar to look more like me"
Own Avatar Behaviour	5	"My Avatar was expressive"
Other Avatars	10	"I was paying attention to other people's Avatars"

## Results

The first comparison was between the 2D condition and the two 3D conditions, as shown in Table 4. The main significant result that was obtained consistently between the groups was that the quality of social interaction improved with the 3D condition. Interestingly the participants reported evaluation of the avatars was not significantly different between conditions, except for weak significance in the case

of the other avatars in the full animation condition. This might be because the participants did the first questionnaire before doing the 3D version and so were not directly comparing the avatars in the two systems.

Table 4  
Comparison of 2D and 3D conditions  
(A single \* indicates significance to 90% probability, a double \* indicates significance to 95%, and a triple \* indicates 99%)

Category	2D Mean	3D Mean	t-value	p-value
<b>Group A</b>				
enjoyment	5.075	4.875	-0.418	0.680891
difficulty	2.042	2.07	0.087	0.931632
presence	4.9	3.6	-3.00	0.007685 ***
co-presence	3.5	3.5	0	1.0
Social Dynamics	4.05	4.95	2.347	0.030567 **
Avatar Appearance	3.266	3.7	0.979	0.340556
Avatar Behaviour	3.3	2.9	-0.669	0.511985
Other Avatars	4.483	4.03	-1.33	0.200126
<b>Group B</b>				
enjoyment	4.5	5.516	1.53	0.143401
difficulty	3.278	2.528	-1.494	0.152499
presence	4.8	5.35	0.992	0.334346
co-presence	3.614	4.137	1.016	0.323092
Social Dynamics	3.633	5	2.306	0.033222 **
Avatar Appearance	3.76	3.033	-1.568	0.134293
Avatar Behaviour	3.85	4.96	1.863	0.078864 *
Other Avatars	4.5	5.35	2.072	2.072 *

The second comparison was between the 3D conditions with limited and full animation, as shown in Table 5.

Table 5  
Comparison of limited animation and full animation conditions  
(A single \* indicates significance to 90% probability, and a triple \* indicates 99%)

Category	Group A	Group B	t-value	p-value
Enjoyment	4.875	5.516	1.324	0.202072
Difficulty	2.071	2.528	1.186	0.251042
Presence	3.6	5.35	3.719	0.001571 ***
co-presence	3.5	4.137	1.481	0.155899
Social Dynamics	4.95	5	0.092	0.927714
Avatar Appearance	3.7	3.033	-1.862	0.079011 *
Avatar Behaviour	2.9	4.96	4.616	0.000215 ***
Other Avatars	4.03	5.35	3.660	0.001791 ***

These results show a significant improvement of the evaluation of the behaviour of the participants own avatar and of the other avatars, demonstrating that realistic animation and emotionally expressive behaviour have a strong effect on people's evaluations of avatars. There was also a significant improvement in presence and a notable but non-significant improvement in co-presence showing that this has a real effect on the participant's experience. Interestingly there was a

weakly significant result that the avatars' appearance was considered worse in the full animation condition. This may be because participants concentrated less on the appearance when the characters' behaviour was more lively.

Following the testing the participants were invited to make comments about their experience in an open interview. Regarding the 3D developments feedback included, "I think it's good that it's 3D because you can sort of.. it gives you more of a vision on how everyone's acting.. sort of thing and how people can react.. because you can sort of see all the shadows and stuff it's more realistic and it gets you.. into it a bit more". Regarding the animation one pupil observed, "They did move differently, like if I said something, erm, like lovingly towards someone they did an action kind of expressing that as well".

## **Discussion**

These results are interesting when compared to our original hypotheses. Hypotheses 1 and 3 seem to be partially confirmed. Simply moving from 2D to 3D does not seem to improve the participant's subjective evaluation of the avatars or any general measures of quality of experience such as enjoyment and presence. However, moving to affective animation does produce strongly significant improvements of evaluation of the avatars' behaviour and of presence. This result seems to indicate that when going to 3D it is important not only to improve the appearance of the avatar but also their behaviour. This result seems consistent with the study by Garau et al. (2003). It also supports the idea that simply moving to 3D does not improve learning or interaction. However, moving to 3D can support new behaviours of styles of interaction, in this case expressive behaviour, which can improve the experience of using an environment. However, the results get more complex when we look at hypothesis 2. Our initial belief was that simply moving to 3D would have a minimal effect on social dynamics, and that affective animation would greatly improve it. This seems intuitive as it is not immediately obvious what moving to 3D contributes to the quality of social interaction, but improving the body language of the avatars would have a strong effect. However, our results show the exact opposite. The only significant improvement in the 3D version was in the quality of social interaction, while it did not improve with affective animation. One possible explanation of the first result is that the ability to visualise the spatial relationships between characters can improve the sense of social interaction. Alternatively, this result might be due to the influence of the AI character, a more detailed study of this aspect is described in the next section. The second result may be explained by the result of Cassell and Thórisson (1999), that in fact simply using emotional expression does little to improve social interaction if basic non-verbal social cues are lacking.

## **THE USER STUDY – AI BRANCH TESTING**

We have also conducted another branch of user testing with 120 secondary school students during February - September 2006 to evaluate the improvisational AI actor, EMMA. The aim of the testing was primarily to measure the extent to which having EMMA as opposed to a person play a character affects users' level of enjoyment, sense of engagement, etc. (The main intention of involving the AI actor is to reduce the human director's burden, but not to improve the users' experience. We expect the AI actor will not worsen the users' experience and perform as good as another human being if it's involved in improvisation.) We concealed the fact that EMMA was involved in some sessions in order

to have a fair test of the difference that is made. The scenarios we used for this AI branch testing are Crohn's Disease (the same as the one used for 3D branch testing) and School Bullying (different from the Homophobic Bullying scenario mentioned above). EMMA played "Dave" in both scenarios (in school bullying, Dave is a close friend to the bullied victim and helps to stop the bullying). We obtained surprisingly good results. Having a minor bit-part character called "Dave" played by EMMA as opposed to a person made no statistically significant difference to measures of user engagement and enjoyment, or indeed to user perceptions of the worth of the contributions made by the character "Dave". Users did comment in debriefing sessions on some utterances of Dave's, so it was not that there was a lack of effect simply because users did not notice Dave at all. Furthermore, it surprised us that few users appeared to realise that sometimes Dave was computer-controlled. We stress, however, that it is not an aim of our work to ensure that human actors do not realise this.

The outline experimental methodology used in the 2006 testing was as follows. Subjects were 14–16 year old students at local Birmingham schools. Forty students were chosen by each school for the testing. There was no control of gender. Four 2-hour sessions took place at each school, each session involving a different set of ten students.

In a session, the main phases were as follows: an introduction to the software; a First Improvisation Phase, where five students are involved in a school bullying improvisation and the remaining five in a Crohn's disease improvisation; a Second Improvisation Phase in which this assignment is reversed; filling out of a questionnaire by the students; and finally, a group discussion acting as a debrief phase. For each improvisation, characters were pre-assigned to specific students. Each Improvisation Phase involved some preliminaries (background familiarisation, appearance choosing, etc.) followed by 10 minutes of improvisation proper.

In half of the school bullying improvisations and half of the Crohn's disease improvisations, a minor character called Dave in each case was played by one of the students and by EMMA in the remaining improvisations. When EMMA played Dave, the student who would otherwise have played him was instructed to sit at another student's terminal and thereby serve as an audience member. Students were told that we were interested in the experiences of audience members as well as of actors. Almost without exception, students appeared not to have suspected that having an audience member resulted from not having Dave played by another student. At the end of one exceptional session, some students asked whether one of the directors was playing Dave. Of the two improvisations a given student was involved in, exactly one involved EMMA playing Dave. This was either the first session or the second. This EMMA involvement order and the order in which the student encounters school bullying and Crohn's disease were independently counterbalanced across students.

Inspection of the transcripts collected also indicates that EMMA usefully pushed the improvisation forward on various occasions. Some example transcripts of the Crohn's Disease scenario created during the testing are displayed in the following. As mentioned previously, Peter has Crohn's disease and Dave is his best friend. Janet and Arnold are Peter's parents. Peter invites everyone out to a restaurant to discuss if he will go through another life-changing operation or not. Dave was played by EMMA, which successfully led the improvisation on the desired track.

Arnold Williams: no its very embarrassing  
Peter Williams: EXCUSE ME!!!  
Janet Williams: i want a divorce right now  
Peter Williams: I'M THE ILL ONE

Dave Atkins: Arnold, Peter needs ur support and u can't just ignore it. [*detecting Arnold was embarrassed about talking about Peter's disease in public.*]

Janet Williams: nobody cares peter

Arnold Williams: we know peter now stop talking about it

Peter Williams: yes i need your support [*Peter echoed Dave's suggestion.*]

Dave Atkins: Hey guys, lets all calm down a bit. [*detecting Arnold was rude by saying "stop talking about it"*]

Arnold Williams: we will discuss it later

Janet Williams: well you have got mien

Peter Williams: help me daddy

Arnold Williams: not now son

Peter Williams: well when

Janet Williams: he is not your daddy

Arnold Williams: another time

Peter Williams: i dont like u

Arnold Williams: wife i need your support on this

Dave Atkins: I think we all should respect Peter's decision. [*detecting Peter was expressing his opinion, although Dave's response was not very appropriate.*]

Peter Williams: u should love me

Janet Williams: peter lets go home u have spolit it dad

Arnold Williams: i do son

Janet Williams: i hate u dad

Arnold Williams: lets talk about it another time

Dave Atkins: Guys, let's try 2 sort this out calmly. [*detecting Janet was angry by saying "I hate u dad".*]

Arnold Williams: thats not very nice mum

Peter Williams: yes calmly [*Peter echoed Dave's suggestion.*]

We have conducted an initial evaluation of the quality of EMMA's determinations about emotion during these testing sessions, by comparing EMMA's determinations during one of the School Bullying improvisations with emotion labels later assigned offline by two members of our team (not involved in the development of EMMA's algorithms). We used the kappa statistic of Carletta (1996). It is a measure of the pairwise agreement among a set of coders making category judgements, correcting for expected chance agreement. A value of at least 0.6 – 0.8 is generally required by researchers looking for good inter-annotator agreement. We calculated the statistic for each pair among the three labellers (EMMA and two humans). The inter-human agreement was only 0.32, and so it is not surprising that the EMMA/human values were only 0.32 again and 0.23. Although they are not ideal, at least these results give grounds for hope that our affect detection with further refinement can come near the rather low human/human level of agreement.

Figure 12 also shows some evaluation results from a "within subjects" analysis based on the collected questionnaires looking at the difference made PER SUBJECT by having EMMA IN (= playing Dave, in either scenario) or OUT. When EMMA is out, the overall boredom is 31%. When EMMA is in, it changes to 34%. The results of "human Dave and EMMA Dave said strange things" respectively are 40% and 44%. When EMMA changes from in to out, the results of "improvisation kept moving" are respectively 54% to 58% and the results of "the eagerness to make own character

speak” are respectively 71% to 72%. Although the measures were “worsened” by having EMMA in, in all cases the worsening was numerically fairly small and not statistically significant. Other statistical analysis results also indicate that when EMMA is in, users’ abilities to concentrate on the improvisation in the Crohn’s Disease scenario are somewhat higher than those when EMMA is out. It seems to be showing that EMMA can make a real positive difference to an aspect of user engagement when the improvisation is comparatively uninteresting.

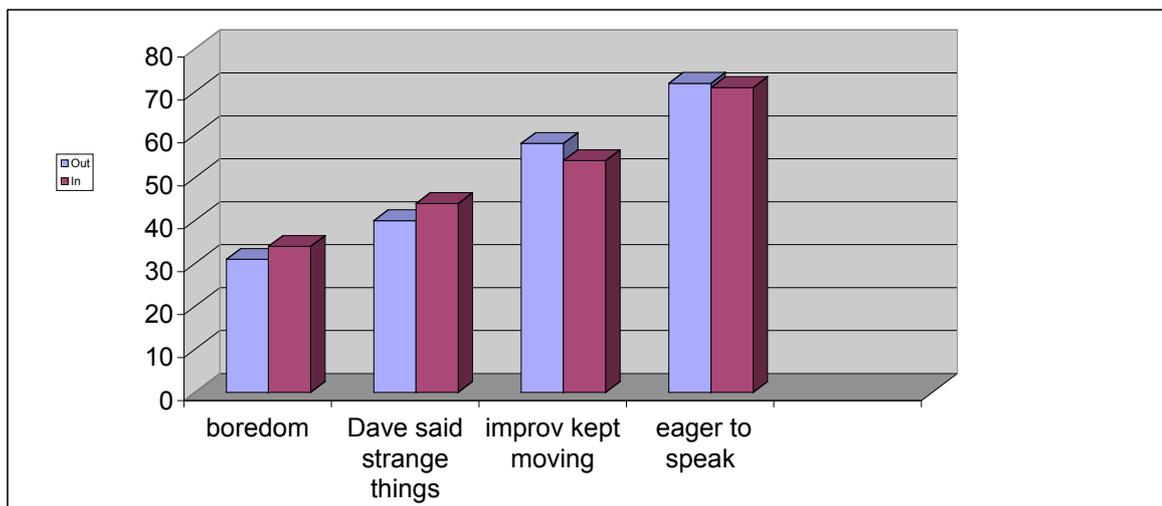


Fig.12. Statistical results for “boredom”, “Dave said strange things”, “improvisation kept moving” and “eager to make own character speak” when EMMA is OUT or IN an “improv”.

## CONCLUSIONS

E-drama provides a platform for participants to engage in focused discussion around emotionally charged issues. This new prototype provides an opportunity for the developers to explore how emotional issues embedded in the scenarios, characters and dialogue can be represented visually without detracting from the learning situation.

During the evaluation of the e-drama systems, the testing subjects enjoyed the experience thoroughly no matter if they used 2D or 3D versions, although testing results showed some preferences for the 3D animated version. It indicates that our e-drama system creates a new channel for young people’s classroom communication. It could play a pioneer role in introducing a new innovative interactive learning and teaching approach to normal classroom situations. To young people with learning difficulties or communication impairment, our e-drama system has the potential to provide automatic monitoring and 24/7 virtual learning companions for them to engage in learning and interaction in an anonymous trustworthy virtual learning environment.

The user trials demonstrate that the creation of a 3D animated version of e-drama indicates a marked improvement on the role-playing experience using the e-drama system. The 3D version of the system, with the automated AI characters, may contribute to improving the perceived quality of social interaction over and above the original 2D version. In addition to this, adding emotionally appropriate animations to the user avatars improves both the participants’ evaluation of those characters and their sense of presence. There is great potential for the use of e-drama in education in areas such as

citizenship, PHSE and drama. Beyond the classroom e-drama can be easily customised for use in professional training, where face to face training can be difficult or expensive, such as customer services training and e-learning in the workplace.

Our research shows that the application of expressive characters to online role-play contributes positively to an already engaging user experience. Future work could include the exploration of automated bit-part characters to fully develop a non-human director. Additionally tools to enable participants to replay the role-plays have been considered. These could enable further reflection and group discussion, allowing for comparisons of sessions between different groups of learners. Replays could even be altered to adjust the emotional states of each character and generate different online “performances”, which could create emotionally rich experiences for audiences as well as participants.

## ACKNOWLEDGEMENTS

This work is a result of the LINK project E-Drama - Enhancement of People, Technology and their Interaction: supported by PACCIT (People @ The Centre of Communication and Information Technologies).

## REFERENCES

- Argyle, M., & Cook, M. (1976). *Gaze and Mutual Gaze*. Cambridge University Press.
- Aylett, R. S., Dias, J., & Paiva, A. (2006). An affectively-driven planner for synthetic characters. In D. Long, S. F. Smith, D. Borrajo & L. McCluskey (Eds.) *Proceedings of the Sixteenth International Conference on Automated Planning and Scheduling (ICAPS 2006)* (pp. 2-10). AAAI Press.
- Aylett, R., Louchart, S., Dias, J., Paiva, A., Vala, M., Woods, S., & Hall, L. E. (2006). Unscripted Narrative for Affectively Driven Characters. *IEEE Computer Graphics and Applications*, 26(3), 42-52.
- Barnden, J. A., Glasbey, S. R., Lee, M. G., & Wallington, A. M. (2004). Varieties and Directions of Inter-domain Influence in Metaphor. *Metaphor and Symbol*, 19(1), 1-30.
- Brand, M., & Hertzmann, A. (2000). Style Machines. In *Proceedings of the 27th Annual Conference on Computer Graphics SIGGRAPH 2000* (pp. 183-192). New Orleans, Louisiana, USA, July 23-28. ACM.
- Briscoe, E., & Carroll, J. (2002). Robust Accurate Statistical Annotation of General Text. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation* (pp. 1499-1504). Las Palmas, Gran Canaria.
- Carletta, J. (1996). Assessing Agreement on Classification Tasks: The Kappa statistic. *Computational Linguistics*, 22(2), 249-254.
- Cassell, J., & Thórisson, K. R. (1999). The Power of a Nod and a Glance: Envelope vs. Emotional Feedback in Animated Conversational Agents. *International Journal of Applied Artificial Intelligence*, 13(4-5), 519-538.
- Cassell, J., & Vilhjálmsón, H. (1999). Fully Embodied Conversational Avatars: Making Communicative Behaviors Autonomous. *Autonomous Agents and Multi-Agent Systems*, 2(1), 45-64.
- Cassell, J., Stocky, T., Bickmore, T., Gao, Y., Nakano, Y., Ryokai, K., Tversky, D., Vaucelle, C., & Vilhjálmsón, H. (2002). MACK: Media lab Autonomous Conversational Kiosk. In *Proceedings of Imagina '02*. February 12-15, Monte Carlo.
- Cavazza, M., Smith, C., Charlton, D., Zhang, L., Turunen, M., & Hakulinen, J. (2008). A ‘Companion’ ECA with Planning and Activity Modeling. In L. Padgham, D. Parkes, J. Müller & S. Parsons (Eds.) *Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems* (pp. 1281-1284). May 2008. Portugal.
- Chi, D., Costa, M., Zhao, L., & Badler, N. I. (2000) The EMOTE Model for Effort and Shape. In *Proceedings of*

- the 27th Annual Conference on Computer Graphics SIGGRAPH 2000* (pp. 173-182). New York, NY: ACM Press.
- Chopra-Khullar, S., & Badler, N. (1999). Where to look? Automating visual attending behaviors of virtual human characters. In O. Etzioni, J. P. Müller & J. M. Bradshaw (Eds.) *Proceedings of the Third Annual Conference on Autonomous Agents* (pp. 16-23). Seattle, Washington, United States.
- Craggs, R., & Wood, M. (2004). A Two Dimensional Annotation Scheme for Emotion in Dialogue. In Y. Qu, J. Shanahan & J. Wiebe (Eds.) *Exploring Attitude and Affect in Text: Theories and Applications. Papers from the AAAI Spring Symposium* (pp. 44-49). Technical Report SS-04-07. Menlo Park, CA: AAAI Press.
- Egges, A., Kshirsagar, S., & Magnenat-Thalmann, N. (2003). A Model for Personality and Emotion Simulation. In V. Palade, R. J. Howlett & L. C. Jain (Eds.) *Proceedings of 7th International Conference of Knowledge-Based Intelligent Information and Engineering Systems (KES 2003)* (pp. 453-461). LNAI 2773. Berlin: Springer-Verlag.
- Ekman, P. (1982) *Emotion in the Human Face*. Cambridge University Press.
- Elliott, C., Rickel, J., & Lester, J. (1997). Integrating Affective Computing into Animated Tutoring Agents. In *Proceedings of the IJCAI Workshop on Animated Interface Agents: Making Them Intelligent* (pp. 113-121). Nagoya, Japan.
- Esuli, A., & Sebastiani, F. (2006). Determining Term Subjectivity and Term Orientation for Opinion Mining. In *Proceedings of EACL-06, 11th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 193-200). Trento, Italy. ACL.
- Fellbaum, C. (1998). *WordNet, an Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Friedman, D., Steed, A., & Slater, M. (2007). Spatial Social Behavior in Second Life. In C. Pelachaud, J.-C. Martin, E. Andre, G. Chollet, K. Karpouzis & D. Pelè (Eds.) *Intelligent Virtual Agents: Proceedings of the 7th International Workshop on Intelligent Virtual Agents* (pp. 252-263). LNAI 4722. Berlin: Springer.
- Fussell, S., & Moss, M. (1998). Figurative Language in Emotional Communication. In S. R. Fussell & R. J. Kreuz (Eds.) *Social and Cognitive Approaches to Interpersonal Communication* (pp.113-142). Lawrence Erlbaum.
- Garau, M., Slater, M., Bee, S., & Sasse, M.A. (2001). The impact of eye gaze on communication using humanoid avatars. In M. Beaudouin-Lafon & R. J. K. Jacob (Eds.) *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 309-316). March 31 - April 5, 2001, Seattle, WA USA. ACM Press.
- Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., & Sasse, M.A. (2003). The Impact of Avatar Realism and Eye Gaze Control on Perceived Quality of Communication in a Shared Immersive Virtual Environment. In G. Cockton & P. Korhonen (Eds.) *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 529-536). New York, NY: ACM Press.
- Gillies, M., & Ballin, D. (2004). Integrating autonomous behavior and user control for believable agents. In G. Cockton & P. Korhonen (Eds.) *Proceedings of the Third International Joint Conference on Autonomous Agents and Multi-Agent Systems* (pp. 336-343). Columbia University, New York. IEEE Computer Society.
- Gillies, M., & Dodgson, N. A. (2002). Eye Movements and Attention for Behavioural Animation. *Journal of Visualization and Computer Animation*, 13, 287-300.
- Gillies, M., Ballin, D., Dodgson, N.A., & Slater, M. (2002). Integrating internal behavioural models with external expressions. In H. Prendinger (Ed.) *Proceedings of the International Workshop on Lifelike Animated Agents: Tools, Affective Functions, and Applications* (pp. 34-39). Held at 7th Pacific Rim International Conference on AI (PRICAI-02). August 19, Tokyo, Japan.
- Gratch, J., & Marsella, S. (2004). A Domain-Independent Framework for Modeling Emotion. *Journal of Cognitive Systems Research*, 5(4), 269-306.
- Hartmann, B., Mancini, M., & Pelachaud, C. (2006). Implementing Expressive Gesture Synthesis for Embodied Conversational Agents. In S. Gibet, N. Courty, J.-H. Kamp, (Eds.) *Gesture in Human-Computer Interaction and Simulation* (pp. 188-199). LNAI 3881. Berlin: Springer-Verlag.
- Heck, R., Kovar, L., & Gleicher, M. (2006). Splicing Upper-Body Actions with Locomotion. *Computer Graphics Forum*, 25(3), 459-466.
- Heise, D. R. (1965). Semantic Differential Profiles for 1,000 Most Frequent English Words. *Psychological Monographs*. 79(8), Whole of 601.

- Kopp, S., & Wachsmuth, I. (2004). Synthesizing Multimodal Utterances for Conversational Agents. *Journal of Computer Animation and Virtual Worlds*, 15(1), 39-52.
- Lee S. P., Badler, J. B., & Badler, N. (2002). Eyes Alive. *ACM Transactions on Graphics*, 21(3), 637-644.
- Lee, J., Marsella, S., Traum, D. R., Gratch, J., & Lance, B. (2007). The Rickel Gaze Model: A Window on the Mind of a Virtual Human. In C. Pelachaud, J.-C. Martin, E. Andre, G. Chollet, K. Karpouzis & D. Pelè (Eds.) *Proceedings of Intelligent Virtual Agents, 7th International Conference, (IVA 2007)* (pp 296-303). LNCS 4722. Berlin: Springer.
- Li, Y., Wang, T., & Shum, H-Y. (2002). Motion texture: a two-level statistical model for character motion synthesis. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques* (pp 465-472). ACM Press.
- Magerko, B., Laird, J. E., Assanie, M., Kerfoot, A., & Stokes, D. (2004). AI Characters and Directors for Interactive Computer Games. In *Proceedings of the Nineteenth National Conference on Artificial Intelligence, Sixteenth Conference on Innovative Applications of Artificial Intelligence* (pp 877-883). AAAI Press / MIT Press.
- Mateas, M. (2002). Interactive Drama, Art and Artificial Intelligence. Ph.D. Thesis. School of Computer Science, Carnegie Mellon University.
- McCrae, R. R., & John, O. P. (1992). An Introduction to the Five Factor Model and Its Application. *Journal of Personality*, 60, 175-215.
- Mehdi, E. J., Nico, P., Julie, D., & Bernard, P. (2004). Modeling Character Emotion in an Interactive Virtual Environment. In *Proceedings of AISB 2004 Symposium Language, Speech & Gesture for Expressive Characters*. Leeds, UK.
- Ortony, A., Clore, G.L., & Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge University Press.
- Pelachaud, C., & Poggi, I. (2002). Subtleties of Facial Expressions in Embodied Agents. *Journal of Visualization and Computer Animation*, 13, 301-312.
- Picard, R.W. (2000). *Affective Computing*. Cambridge, MA: MIT Press.
- Prendinger, H., & Ishizuka, M. (2001). Simulating Affective Communication with Animated Agents. In *Proceedings of Eighth IFIP TC.13 Conference on Human-Computer Interaction* (pp.182-189). Tokyo, Japan.
- Riedl, M. O., & Young, R. M. (2006). From Linear Story Generation to Branching Story Graphs. *IEEE Computer Graphics and Applications*, 26(3), 23-31.
- Rose, C., Cohen, M. F., & Bodenheimer, B. (1998). Verbs and Adverbs: Multidimensional Motion Interpolation. *IEEE Computer Graphics & Applications*, 18(5), 32-40.
- Sharoff, S. (2005). How to Handle Lexical Semantics in SFL: a Corpus Study of Purposes for Using Size Adjectives. *Systemic Linguistics and Corpus*. London: Continuum.
- Slater, M., Guger, C., Edlinger, G., Leeb, R., Pfurtscheller, G., Antley, A., Garau, M., Brogni, A., Steed, A., & Friedman, D. (2006). Analysis of Physiological Responses to a Social Situation in an Immersive Virtual Environment, Presence. *Teleoperators and Virtual Environments*, 15(5), 553-569.
- Tanguy, E. A. R., Willis, P. J., & Bryson, J. J. (2003). A Layered Dynamic Emotion Representation for the Creation of Complex Facial Animation. In T. Rist, R. Aylett, D. Ballin & J. Rickel (Eds.) *Proceedings of Intelligent Virtual Agents* (pp. 101-105). LNCS 2792. Berlin: Springer-Verlag.
- Vilhjálmsón, H., & Cassell, J. (1998). BodyChat: Autonomous Communicative Behaviors in Avatars. In K. P. Sycara & M. Wooldridge (Eds.) *Proceedings of Second International Conference on Autonomous Agents* (pp. 269-276). New York, NY: ACM Press.
- Vinayagamoorthy, V., Garau, M., Steed, A., & Slater, M. (2004). An Eye Gaze Model for Dyadic Interaction in an Immersive Virtual Environment: Practice and Experience. *Computer Graphics Forum*, 23(1), 1-12.
- Vinayagamoorthy, V., Gillies, M., Steed, A., Tanguy, E., Pan, X., Loscos, C., & Slater, M. (2006). Building Expression into Virtual Characters. In *Proceedings of Eurographics 2006*, 04-08 Sept 2006, Vienna, Austria.
- Weizenbaum, J. (1966). ELIZA - A Computer Program For the Study of Natural Language Communication Between Man and Machine. *Communications of the ACM*, 9(1), 36-45.
- Wiltschko, W. R. (2003). Emotion Dialogue Simulator. eDrama learning, Inc. eDrama Front Desk.

- Zhang, L., Barnden, J. A., Hendley, R. J., & Wallington, A. M. (2006). Exploitation in Affect Detection in Open-ended Improvisational Text. In *Proceedings of Workshop on Sentiment and Subjectivity* (pp. 47-55). Held at COLING-ACL 2006, Sydney, July 2006.
- Zhang, L., Gillies, M., Barnden, J. A., & Hendley, R. J. (2007). An Improvisational AI Agent and Emotionally Expressive Characters. In *Proceedings of the Workshop on Narrative Learning Environments* (pp. 60-69). *Supplementary Proceedings of the 13th International Conference of Artificial Intelligence in Education (AIED 2007)*, July 9-13, 2007, Los Angeles, USA. Amsterdam: IOS Press.
- Zhe, X., & Boucouvalas, A. C. (2002). Text-to-Emotion Engine for Real Time Internet Communication. In *Proceedings of International Symposium on Communication Systems, Networks and DSPs* (pp.164-168). Staffordshire University, UK.