
Affect detection and metaphor in e-drama

Li Zhang*, John A. Barnden,
Robert J. Hendley, Mark G. Lee,
Alan M. Wallington and Zhigang Wen

School of Computer Science,
University of Birmingham,
Birmingham, B15 2TT, UK
Fax: +0121 4144281
E-mail: l.zhang@cs.bham.ac.uk
*Corresponding author

Abstract: We report work on adding affect-detection to an existing e-drama programme, a text-based software system for (human) dramatic improvisation in simple virtual scenarios, for use primarily in learning contexts. The system allows a human director to monitor improvisations and make interventions, for instance in reaction to excessive, insufficient or inappropriate emotions in the characters' speeches. Within an endeavour to partially automate directors' functions, and to allow for automated affective bit-part characters, we have developed an affect-detection module. It is aimed at detecting affective aspects (concerning emotions, moods, rudeness, value judgments, etc.) of human-controlled characters' textual 'speeches'. The work also accompanies basic research into how affect is conveyed linguistically. A distinctive feature of the project is a focus on the metaphorical ways in which affect is conveyed. The project addresses the special issue themes such as making interactive narrative learning environments more usable, building them, and supporting reflection on narrative construction.

Keywords: affect detection; e-drama; intelligent virtual actor and metaphor.

Reference to this paper should be made as follows: Zhang, L., Barnden, J.A., Hendley, R.J., Lee, M.G., Wallington, A.M. and Wen, Z. (2008) 'Affect detection and metaphor in e-drama', *Int. J. Continuing Engineering Education and Life-Long Learning*, Vol. 18, No. 2, pp.234–252.

Biographical notes: Dr. Li Zhang has a PhD from the School of Computer Science, University of Birmingham. She works as a research Fellow on e-drama project in the School of Computer Science, University of Birmingham since February 2004. Her principle research interests are Intelligent User Interfaces, Natural Language Processing, Affect Detection in Text and Intelligent Conversational Virtual Agent. With a computational linguistic background, she is also interested in automatic processing of figurative language.

Professor John A. Barnden is leading the EPSRC/ESRC/DTI-funded project on 'e-drama', one aspect of which will draw on the metaphor processing research in the ATT-Meta project. His research interests mainly are Natural language Processing and Artificial Intelligence. He was a coproposer of a series of EURESCO conferences focusing on metaphor. He recently joined the organisational board of the Researching and Applying Metaphor (RAAM) series of international conferences. He was a keynote speaker at RAAM-IV in

2001, and on metaphor at two recent workshops. He coorganised workshops on figurative language at the Corpus'2001 and'2003 conferences (Lancaster University) and at ACL 2003 (Sapporo). He is an invited contributor to the Handbook on Metaphor being edited by Ray Gibbs for Cambridge University Press.

Robert J. Hendley works in the School of Computer Science at the University of Birmingham. He has worked for many years in the application of Artificial Intelligence techniques to the interaction between people and computers. This has included work on intelligent and adaptive interfaces, information visualisation, teaching and learning systems, and user modelling.

Dr. Mark Lee is a lecturer at the School of Computer Science at the University of Birmingham. His principle research interests are Natural Language Processing with particular interests in the pragmatics of natural language dialogue and the understanding of figurative language. His current research projects include work on multi-user communication systems, a computational treatment of metaphor understanding and an analysis of doctor patient dialogues. He has also published on more general topics of Artificial Intelligence and Human-Computer Interaction. He is also a committee member for the UK Special Interest Group on Computational Linguistics.

Dr. Alan Wallington has a PhD from the Centre for Computational Linguistics at UMIST, Manchester and is currently employed as a research fellow in the School of Computer Science at the University of Birmingham. His principle research interests lie in psychologically informed models of Natural Language Processing. This includes work in ambiguity resolution and more recently on the nature of figurative language. He is a coinvestigator on two current projects: 'Coping with the Open-Endedness of Everyday Metaphorical Language: theory, corpus study, and computational implementation' and on 'E-Drama: enhancement of people, technology, and their interaction'.

Dr. Zhigang Wen currently works as a research fellow in the e-drama project at the School of Computer Science, University of Birmingham. His main research interests include real time virtual human animation, gesture generation system, and natural language processing for 3D embodied conversational agent. He received an MSc in Computer Graphics and a PhD in real time virtual human animation.

1 Introduction and relationship to other work

Improvised drama and role-play are widely used in education, counselling, and conflict resolution. Various researchers have explored virtual, computer-based frameworks for such activity, leading to e-drama (virtual drama) systems in which virtual characters (avatars) interact under the partial control, at least, of human actors (e.g. Machado, Prada, and Paiva, 2000). The springboard for our own research was an already existing e-drama system (*e-drama*) created by Hi8us Midlands Ltd (<http://www.edrama.co.uk>), a charitable company. One main aspect of our project has been the addition of types of intelligent automation to the *e-drama* system. The original system had already been used in schools for creative writing, careers advice, and teaching in a range of subject areas such as history. Hi8us's experience with *e-drama* suggests that the use of e-drama helps

school children lose their usual inhibitions about drama improvisation, because they are not physically present on a stage and are anonymous. It permits a group of young people to jointly participate in live drama improvisation online. The participants can be in the same room or geographically separated.

In the *e-drama* system, ‘actors’ (human users) control virtual characters on a virtual stage, with textual ‘speeches’ displayed as text bubbles typed by the actor operating the character. One director and up to five actors are involved in an e-drama session. The actors and director may be located distantly from each other, as all communication is via the internet, through the intermediary of a server operated by Hi8us. A graphical interface shows the characters and virtual stage on each actor’s terminal and the director’s. Actors can choose the clothes and bodily appearance for their own characters. A possible state of the graphical interface is shown in Figure 1. The characters’ visual forms were originally 2D static cartoon figures, with backdrops that could include real-life photographic images. However, we now brought in the option of 3D animated gesturing avatars (though still somewhat cartoon-like) and 3D computer-generated settings using technology from one of our industrial partners, British Telecom (BT). Another useful feature recently introduced is that during an initial phase of an e-drama session, in which the actors familiarise themselves with the background to the scenario, actors can see documentary film material about the background, supplied by our third industrial partner, Maverick TV Ltd.

Figure 1 One example of the *e-drama* virtual stage



The actors are given a loose scenario around which to improvise, but are at liberty to be creative. For example, one scenario we have used is a school-bullying in which one involving a schoolgirl Lisa, who is being bullied by her classmate Mayid. Lisa is a shy child and afraid of Mayid. Other characters are Lisa's mother, a friend, and a school teacher. Actors are expected to improvise interesting interchanges within these parameters. It is expected that normally the Mayid character will express hostility to Lisa and that she will express fear, but actors can be creative, so that for example, the Mayid actor might play him as repenting of his bullying. The other main scenario we have concentrated on is about in a situation in which someone is afflicted with Crohn's disease, an embarrassing bowel illness.

The human director has a number of roles. He/she must constantly monitor the unfolding drama and the actors' interactions, or lack of them, to intervene if they are not keeping to the general spirit of the scenario. For example, a director may intervene when the emotions being expressed or discussed are not as expected (or are not leading usefully in a new interesting direction). The director may also intervene if, for example, one character is not getting involved, or is unduly dominating the improvisation.

One type of director intervention is a message sent either to all the actors or to one actor privately (the latter is a 'whisper'). For example, the director might suggest to an actor that he/she interacts more (or less) with another. However, another important form of intervention is for the director is to introduce and control a 'bit-part' character. 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 an inappropriate exchange developing between those two characters.

Clearly, all these imposes a heavy burden on the director. Playing the role of the bit-part character and interacting with other characters whilst keeping interventions limited so as to maintain the main improvisatory drama amongst the actors, makes it difficult to fully monitor the behaviour of all the other actors and send appropriate messages to them should they stray off topic or exhibit inappropriate emotions. The difficulty is particularly acute if the directors are novices, such as teachers trying to use e-drama in their lessons.

The aim of one major research is accordingly to automate some directorial functions, either to take some of the burden away from a human director, or to provide a fully automated (though necessarily very restricted) director. With a fully-automated director, even if highly restricted in what it could do, little or no human supervision might be required for at least minimally adequate improvisations, and *e-drama* could, for example, be added to websites about certain topics allowing visitors to engage in online role-play germane to the topic. However currently, our main work is on merely assisting a human director. The assistance is by

- 1 fully-automated control of an optionally-included bit-part character
- 2 sending of automated suggestions to the human director about the progress of the improvisation or about the messages to send to the human actors.

Point (2) is addressed briefly below. Our main focus in this paper is on (1).

For purpose (1), we have created a simple automated actor, Emotion, Metaphor and Affect (EMMA), which controls a bit-part character who is an acquaintance of the other

characters. EMMA contains an affect-detection module, which tries to identify affect in the characters' speeches, allowing the EMMA character to make responses that, it is hoped, will stimulate the improvisation, thus leading to less need for intervention by the human director. Within affect we include: basic and complex *emotions* such as anger and embarrassment, respectively; *meta-emotions* (emotions about emotions) such as desiring to overcome anxiety; *moods* such as hostility; and *value judgments* (judgments of goodness, importance, etc.). Although merely detecting affect is limited compared to the extracting full meaning of the characters' utterances, we have found that in many cases this is sufficient for the purposes of stimulating the improvisation.

Even limited types of affect detection can be useful. We do not purport to make EMMA to detect all types of affect under all ways where affect can be expressed or implied, or to do it with a high degree of reliability. The spirit of the project is to see how far we can get with practical processing techniques, while at the same time, investigating theoretically the nature of, and potential computational ways of dealing with, forms of affective expression that may be too difficult to handle currently in a usable implemented system.

Much research has been done on creating affective virtual characters in interactive systems. Emotion theories, particularly that of Ortony, Clore and Collins (1988; OCC), have been used widely therein. Prendinger and Ishizuka (2001) used the OCC model in part to reason about emotions and to produce believable emotional expressions. The *e-Drama Front Desk* (Wiltschko, 2003), which is used for pedagogical purposes, uses an online emotional natural language dialogue simulator with a virtual reception interface. Mehdi et al. (2004) combined a widely accepted five-factor model of personality (McCrae and John, 1992), mood and OCC in their approach for the generation of emotional behaviour for a fireman training application. Gratch and Marsella (2004) presented an integrated model of appraisal and coping, to reason about emotions and to provide emotional responses, facial expressions, and potential social intelligence for virtual agents. Egges, Kshirsagar and Magnenat-Thalmann (2003) provided the virtual characters with conversational emotional responsiveness. Elliott, Rickel and Lester (1997) demonstrated tutoring systems that reason about the users' emotions. There is much other work in a similar vein.

However, there has been only a limited amount of work directly comparable to our own, especially given our concentration on improvisation and open-ended language. Although Façade (Mateas, 2002) included shallow natural language processing for characters' open-ended utterances, 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 chatting environment (see also, Boucouvalas (2002)). 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 (e.g. 'I' or 'we'). Unfortunately, the emotion detection focuses only on emotional adjectives, and does not address deep issues such as figurative expression of emotion (discussed below). Also, the concentration purely on first-person emotions is narrow. There has been a relevant work on general linguistic clues that could be used in practice for affect detection (e.g. Craggs and Wood, 2004).

Our work is distinctive in several respects. Our interest is not just in

- 1 the first-person, positive expression of affect, i.e. the affective states or attitudes that a virtual character X implies that it itself has (had or will have, etc.)

- 2 affect that a character X implies it lacks
- 3 affect that X implies other characters have or lack
- 4 questions, injunctions, commands, and so on. concerning affect.

Also, we aim for the software to cope partially with the important case of communication of affect via metaphor (Fussell and Moss, 1998; Kövecses, 1998), and to push forward the theoretical study of such language, as part of our general research on metaphor (see, e.g. Barnden et al., 2004).

Our project does not involve using or developing deep and scientific models of how emotional states, etc. function in cognition. Instead, the deep questions investigated are on linguistic matters such as the metaphorical expression of the affect. The prime importance in studying how ordinary people understand and talk about affect in ordinary life is their common-sense views of how affect works, irrespective of how scientifically accurate those views are. Metaphor is strongly involved in such views.

It should also be appreciated that this paper does not address the emotional, etc. states of the actors (or director, or any audience). Our focus is on the affect that the actors make their characters express or mention. While an actor may work him/herself up into, or be put into, a state similar to or affected by those in his/her own characters' speeches or those of other characters, such interesting effects, which go to the heart of the dramatic experience, are beyond the scope of this paper, and so is the possibility of using information one might be able to get about actors' own affective states as a hint about the affective states of their characters or vice-versa.

Various characterisations of emotion are used in emotion theories. The OCC model uses emotion labels (anger, etc.) and intensity, while Watson and Tellegen (1985) use positivity and negativity of affect as the major dimensions. We have drawn ideas from several such sources. We use an evaluation dimension (negative–positive), affect labels, and intensity. The basic emotion labels (such as 'angry') we use are taken from Ekman (1992), while other comparatively complex affect labels (such as 'approving') are taken from the OCC model. There are 25 affect labels used in our system currently. Affect labels plus intensity are used when strong text clues signalling affect are detected, while the evaluation dimension plus intensity is used when only weak text clues are detected.

2 Our current affect detection

Before any automatic recognition and response components could be built for use in our automated actor EMMA, a detailed analysis of the language used in *e-drama* sessions was necessary. A small corpus of sessions was analysed by hand to identify the range of linguistic forms used and to provide insight for the automatic processing. In fact, this analysis is often very difficult and unreliable, but it does reveal some important observations.

- The language used is often complex and idiosyncratic. It is almost invariably ungrammatical, it uses abbreviations, it contains misspellings, and it borrows heavily from the language of text-messaging (*textese*) and chat-rooms. Compared to the language normally analysed in computational linguistics, it provides significant additional challenges.

- The literal meaning of the statements is often less important to its interpretation than the affect that it is expressing. The content of a statement is still important in building an understanding and in responding appropriately, but understanding the affective state being expressed is critical.
- The language contains a large number of weak cues to the affect that is being expressed. These cues may be contradictory or they may work together to enable a stronger interpretation of the affective state. In order to build a reliable and robust analyser of affect, it is necessary to undertake several diverse forms of analysis and to enable these to work together to build stronger interpretations.

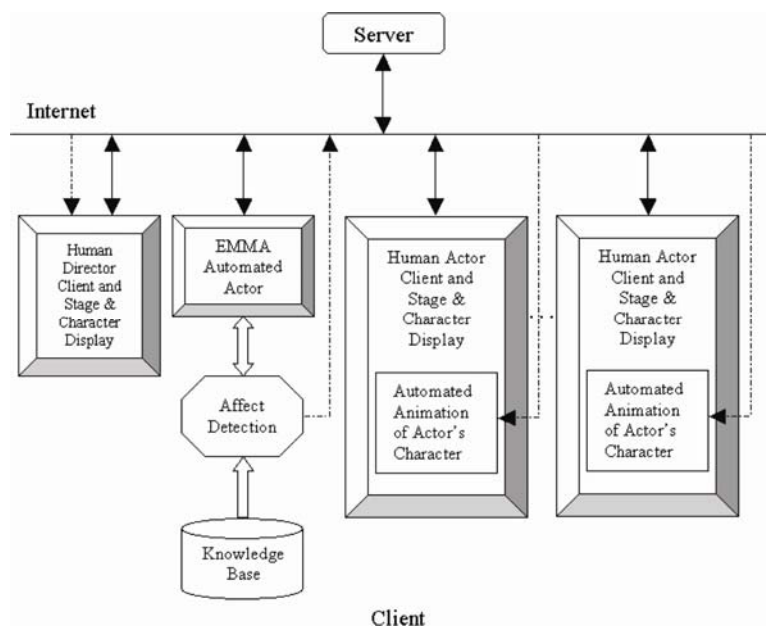
This leads to a system where the emphasis is moved away from building a representation of the meaning of the statement to one where more weight is given to building a robust representation of affective connotations.

The results of this affective analysis are then used to:

- Control an automated actor (EMMA) that operates a character in the improvisation: see (1) in Section 1.
- Independently of this, help to create the directorial suggestions mentioned in Section 1, in point (2).
- Additionally, drive the animations of the avatars in the user interface so that they react bodily in ways that is consistent with the affect that they are expressing, for instance by changing posture or facial expressions.

The overall architecture is shown in Figure 2.

Figure 2 Application architecture: fat arrows show flow of information inside a programme; solid thin arrows show flow of character utterances and director messages formatted in XML; dashed lines show flow of information about detected affect and flow of automated suggestions to the director



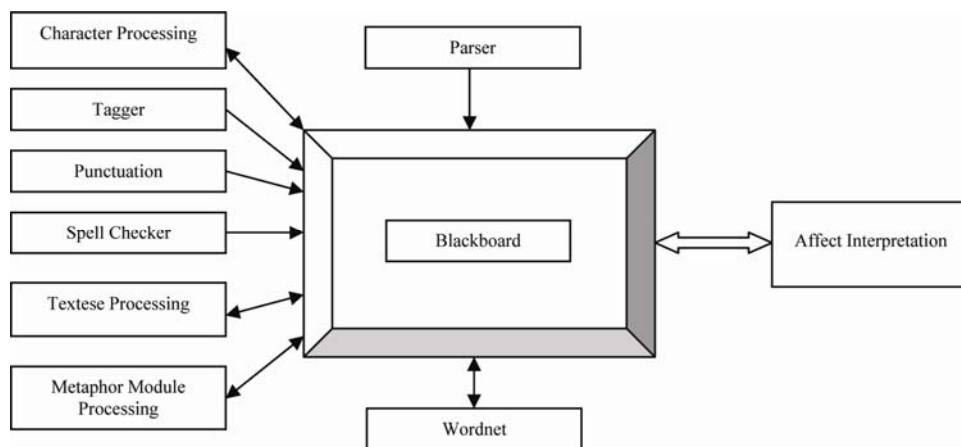
Within the affect detection component, we need to undertake several analyses of any given utterance by another character. These will each build representations which may be used by other components (e.g. syntactic structure) and will construct (possibly weak) hypotheses about the affective state. The architecture adopted is a blackboard based one. Each knowledge source undertakes its processing and writes its results to a central data structure (the blackboard) where they can be used by other knowledge sources and where hypotheses can be supported by multiple knowledge sources.

A rule-based component takes these hypotheses and builds a single interpretation of the affective state being expressed by the utterance being analysed. This interpretation is then transmitted to the other components of the system: EMMA (the automated actor), the automated directorial-suggestion generator and the animation component in the software client that handles the actor whose utterance is being processed.

The response generation component of EMMA uses this interpretation to build its behaviour driven mainly by its role in the improvisation and the affect expressed in the statement to which it is responding.

Figure 3 illustrates the overall structure of the language processing.

Figure 3 Blackboard architecture



2.1 Pre-processing modules

The language in the 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 misspellings, ungrammaticality, abbreviations (often as in text messaging), 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'. In the examples we have studied, which so far involve teenage children improvising around topics such as school bullying, the genre is similar to internet chat (Werry, 1996). 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.

A lookup table has been created containing expansions of abbreviations commonly used in internet chat-rooms and texting and abbreviations that we have found by analysing previous e-drama sessions (e.g. ‘im (I am)’ and ‘c u (see you)’). In particular, the current module deals with the abbreviations such as numbers embedded within words (e.g. ‘l8r’ for ‘later’), using the lookup table. The abbreviation module can handle most of the abbreviation found in users’ input, but in later work, we hope to incorporate a more general treatment that will be able to deal with some novel instances that follow standard abbreviatory conventions.

Unfortunately, certain abbreviations can be ambiguous, e.g. ‘2’ may stand for ‘to’, ‘too’, or ‘two’ (although the last is rare in our genre), as exemplified by ‘I’m 2 hungry 2 walk’. A lookup table on its own cannot solve such context-sensitive ambiguity. In order to solve this problem, part-of-speech information has been assigned to the surrounding words using the lexicon from Brill’s tagger (Brill, 1994). Then, simple strategies are used to find the appropriate words for the ambiguous items. For example with the textese use of the number ‘2’, two strategies are used. If there is a word following ‘2’ the preposition or infinitival ‘to’ will be chosen, if the tagging information of its next word belongs to the following categories: verb base forms, pronouns, common nouns (singular and plural), articles, possessive pronouns or proper nouns (e.g. ‘2 do it and ‘talk 2 you’). If there is no word following ‘2’, then ‘to’ will be chosen if the tagging information of its previous word is a base form verb, a present tense verb or a past tense verb (e.g. ‘I’d love 2’). Otherwise, ‘too’ will be chosen (e.g. ‘me 2’). Since these two simple strategies will inevitably lead to errors in some special cases, we have evaluated those using examples from previous e-drama transcripts and obtained an 85.7% accuracy rate, which is currently adequate.

Letter repetition comes in two flavours that can usefully be distinguished. One is the repetition added to ordinary words (e.g. ‘yessss’, ‘seeee’, and ‘loook’) and the other is repetition added to interjections or onomatopoeic elements (e.g. ‘grrrrrrrr’, ‘agggghhhh’). The iconic use of word length here (i.e. written word length corresponding roughly to an imagined sound length) normally implies strong affective states in the characters’ input. Usefully, adding letters does not change the pronunciation in a great deal. We use two simple strategies to deal with such words in users’ input. If there are three consecutive occurrences of the same letter (no matter whether they are vowels or consonants) in one word, or if the first letter is equal to the third and the second to the fourth (e.g. ‘haha’), then they are the ones we are looking for. A dictionary was created containing base forms of various interjections and onomatopoeic elements together with some ordinary words that are often subject to letter-repetition in e-drama sessions. Then the Metaphone spelling-correction algorithm (<http://aspell.net/metaphone/>), whose working strategy is based on pronunciations, works with the dictionary to locate the base forms of such input words. We also aim to develop a detector of onomatopoeic elements that does not rely on particular base forms. We must stress that the added letter-repetition is not simply eliminated, but the fact of its occurrence is recorded for the purposes of affect-detection.

Finally, the Levenshtein distance algorithm (<http://www.merriampark.com/ld.htm>) with a contemporary English dictionary deals with spelling mistakes in users’ input. Having described the necessary pre-processing, we now turn to the core detection of affect in users’ input.

2.2 Processing of imperative moods

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 special, 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 and 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.

The *Rasp* parser recognises some types of imperatives directly. Unfortunately, the grammar of the 2002 version of the *Rasp* parser that we have used does not deal properly with certain imperatives, which means that examples like ‘you shut up’, ‘Dave bring me the menu’, ‘Matt don’t be so blunt’ and ‘please leave me alone’, are not recognised as imperatives, but as normal declarative sentences. Therefore, further analysis is needed to detect imperatives, by additional processing applied to the possibly-incorrect syntactic trees produced by *Rasp* (we are currently investigating whether a new version of *Rasp* announced in 2006 would work better for us).

If *Rasp* outputs a subject, ‘you’, followed by certain verbs (e.g. ‘shut’, ‘calm’, etc.) or certain verb phrases (e.g. ‘get lost’, ‘go away’, etc.), the sentence type will be changed to imperative (note: in ‘you get out’ the ‘you’ could be a vocative rather than the subject of ‘get’, especially as punctuation such as commas is often omitted in our genre; however, these cases are not worth distinguishing and we assume that the ‘you’ is a subject). If a softener ‘please’ is followed by the base forms of a verb, then the input is taken to be imperative. If a singular proper noun is followed by a base form of the verb, then this sentence is taken to be an imperative as well (e.g. ‘Dave get lost’). However, when a subject is followed by a verb for which there is no difference at all between the base form and the past tense form, then ambiguity arises between imperative and declarative (e.g. ‘Lisa hit me’), and cleverer processing is needed.

There is an important special case of this ambiguity. If the object of the verb is ‘me’, then in order to solve the ambiguity, we have adopted the evaluation value of the verb from Heise’s compilation of semantic differential profiles (Heise, 1965). In these profiles, Heise listed values of evaluation, activation, potency, distance from neutrality, etc. for the 1,000 most frequently used English words. In the evaluation dimension, positive values imply goodness. Because normally people tend to use ‘a negative verb + me’ to complain about an unfair situation, but if the evaluation value is negative for such a verb, then this sentence is probably not imperative but declarative (e.g. ‘Mayid hurt me’). Otherwise, other factors implying imperative are checked in this sentence, such as exclamation marks and capitalisations. If these factors occur, then the input is probably an imperative. Otherwise, the conversation logs are checked to see if there is any question sentence directed towards this speaker recently. If there is, then the input is conjectured to be declarative.

There is another type of sentence: ‘don’t you + base form of verb’ that we have started to address. Though such a sentence is often interrogative, it is also often a negative version of an imperative with a ‘you’ subject (e.g. ‘Don’t you dare call me a

dog,’ ‘Don’t you call me a dog’). Normally, *Rasp* regards it as a question sentence (even if there is no question mark, but in any case in our genre punctuation is not used reliably). Thus, further analysis has also been implemented for such a sentence structure to change its sentence type to imperative. Although currently this has limited effect, as we only infer a (negative) affective quality when the verb is ‘dare’, we plan to add semantic processing in an attempt to glean affect more generally from ‘Don’t you ...’ imperatives.

In general, the imperative-mood detection is one useful tool for extracting potential affective flavour from users’ input.

2.3 Affect detection by pattern matching

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 (<http://herzberg.ca.sandia.gov/jess/>) is used to implement the pattern/template-matching rules in EMMA allowing the system to cope with more general syntax and wording.

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. However, it lacks other types of generality and can be fooled when the phrases are suitably embedded as subcomponents of other grammatical structures. For example, if the input is ‘I doubt she’s really angry’, pattern-matching that looks for anger in a simple way will fail to provide the expected results. Below, we indicate our path beyond these limitations.

The transcripts analysed to inspire our initial knowledge base and pattern-matching rules had independently been produced earlier from Hi8us *e-drama* improvisations based on a school bullying scenario. The actors were school children aged from 8 to 12. We have also worked on another, distinctly different scenario – Crohn’s disease, based on a TV programme by Maverick Television Ltd. (another of our industrial partners). One interesting feature in this scenario is meta-emotion (and cognition about emotion) because of the need for people to cope with emotions about their illnesses. 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 database according to the nature of specific scenarios.

When the character Mayid says “Lisa, you Pizza Face! You smell”, EMMA detects that he is insulting Lisa. Thus, he is very *rude*. Patterns such as ‘you smell’ have been used for rule implementation. The rules conjecture the character’s emotions, evaluation dimension (negative or positive), politeness (rude or polite), and what response EMMA should make. Some more examples of user’s input that indicate the speaker is in a rude emotional state are ‘no one likes u’, ‘keep your mouth shut’, ‘who do you think you are’, ‘losers’, and ‘you stink’.

Multiple exclamation marks and capitalisation are frequently employed to express emphasis in e-drama sessions. If emotion and exclamation marks or capitalisation are

detected in a character's utterance, then the emotion intensity is deemed to be comparatively high (and emotion is suggested even in the absence of other indicators).

A reasonably good indicator that an inner state is being described is the use of 'I' (see also, Craggs and Wood (2004)), especially in combination with the present or future tense. In the school-bullying scenario, for example, when 'I' is followed by a future-tense verb the affective state 'threatening' is normally being expressed; and the utterance is usually the shortened version of an implied conditional, e.g. "I'll scream (if you stay here)". Note that when 'I' is followed by a present-tense verb, a variety of other emotional states tend to be expressed, e.g. 'I want my mum', 'I hate you', 'I like you'. Further analysis of first-person, present-tense cases is discussed in Section 2.4.

As we mentioned earlier, pure pattern-matching rules only using keywords and simple templates can be fooled when the phrases are suitably embedded as subcomponents in 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. The general sentence structure 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, but also helps EMMA to make appropriate responses. For example, we will not count the affective states presented in conditional or question sentences as the *speakers'* current emotional state. *Rasp* will inform the pattern-matching rule with sentence type information. If the current input is a conditional or question sentence with affective keywords or structures in, then the affective states will not be valued. For example, if the input is "I like the place when it is quiet", *Rasp* works out its sentence type: a conditional sentence and the rule for structures containing 'like' with a normal declarative sentence label will not be activated. Instead, the rule for the keyword 'when' with a conditional sentence type label will be fired. Thus, an appropriate response will be obtained.

In addition, as discussed in Section 2.2, we use *Rasp* to indicate imperative sentences, such as when Mayid (the bully) said "Lisa, don't tell Miss about it". The pseudo-code example rule for such input is as follows:

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

Thus, the declarative input such as "I won't tell Miss about it" will not be able to activate the example rule due to different sentence type information. Especially, we have assigned a special sentence type label ('imp + please') for imperatives with softener 'please'. Only using this special sentence type label itself in the pattern-matching rule helps us effortlessly to obtain the user's linguistic style ('polite') and probably a polite response from EMMA as well according to different roles in specific scenarios. 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. In general, using sentence structure information obtained from *Rasp* in the pattern-matching rules helps our approach to go beyond certain limitations and enables it to work more efficiently and accurately.

Aside from using the *Rasp* parser, we have also worked on implementing simple types of semantic extraction of affect using affect dictionaries and electronic thesauri, such as WordNet (2006). The way we are currently using WordNet is briefly as follows.

2.4 Using WordNet for a first person case

As we mentioned earlier, use of the first-person with a present-tense verb tends to express an affective state in the speaker, especially in discourse in which affect is salient, as is the case in scenarios such as School Bullying and Crohn's Disease. We have used the *Rasp* parser to detect such a sentence. First of all, such user's input is sent to the pattern-matching rules in order to obtain the speaker's current affective state and EMMA's response to the user. If there is no rule fired (i.e. we do not obtain any information of the speaker's affective state and EMMA's response from the pattern-matching rules), further processing is applied. We use WordNet to track down the rough synonyms of the verb (possibly from different WordNet 'synsets') in the verb phrase of the input sentence, in order to allow a higher degree of generality than would be achieved just with the use of our pattern-matching rules. In order to find the closest synonyms to the verb in different synsets, the semantic profiles of the 1,000 most frequently used English words (Heise, 1965) have been employed, especially to find the evaluation values of every synonym of the original verb. We transform positive and negative evaluation values in Heise's dictionary into binary 'positive' and 'negative' only. For example, if a word has a positive evaluation value 1.90, then we will simply label the value as 'positive'. Then if any synonym has the same evaluation value ('positive' or 'negative') as that of the original verb, then it will be selected as a member of the set of closest synonyms. We use one closest synonym to replace the original verb in the user's input. This newly built sentence will be sent to the pattern-matching rules in order to obtain the user's affective state and EMMA's response. Such processing (using a closest synonym to replace the original verb and sending the newly built sentence to the pattern-matching rules) continues until we obtain the speaker's affective state and appropriate response.

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.

2.5 Responding regimes

EMMA normally responds to, on average, every N th speech by another character in the e-drama session, where N is a changeable parameter (currently usually set to 3). However, it also responds when EMMA's character's name is mentioned, and makes no response if it cannot detect anything useful in the utterance it is responding to. The one-in- N average is achieved by sampling a random variable every time another character says something. As a future development, we plan to have N dynamically adjustable

according to how confident EMMA is about what it has discerned in the utterance at hand so that it is less likely to respond if it has less confidence. EMMA makes a random response from several stored response candidates that are suitable for the affective quality it has discerned in the utterance it is responding to.

In addition, EMMA sometimes takes a portion of the user's input string as a sub-component of its response. But the pre-processing module, reported in Section 2.1, replaces the abbreviations and misspellings with surmised expansions and corrected versions. This helps to obscure the fact that part of EMMA's response is only a reflection. For example:

<Another character> I'm ur (*your*) best m8 (*mate*).

<EMMA character> Why are you my best mate?

On the other hand, we also purposely put abbreviations and other phenomena used in the internet chat-rooms in EMMA's responses in order to make EMMA 'behave' like a school kid. For example:

<Another character> You get lost!

<EMMA character> Hey, coool it folks (*notice the lengthening of 'cool'*).

<Another character> He didn't do it.

<EMMA character> No1 believes you, m8. (*where 'no1' stands for 'no one'*)

3 Metaphorical expression of affect

The metaphorical description of emotional states is common and has been extensively studied (Fussell and 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). In other work, we have conducted research on metaphor in general (see, e.g. Barnden et al., 2004; Barnden, 2006), and are now applying it to the e-drama application, and conversely using the application as a useful source of theoretical inspiration.

In most discourse genres, metaphorical phraseology tends to be of conventional form, the extreme being stock phrases such as 'sit on the fence'. Such phrases can be stored in a lexicon and directly recognised. Our intended approach to affective metaphor handling in the EMMA module is partly to look for stock phraseology and simple variants of it. As an example of stock phrase handling, insults in e-drama often use stock metaphorical phraseology, especially the case of animal insults ('you cow', 'you dog'). Simple pattern-matching rules are currently used in EMMA to deal with some animal insults. We now aim to use WordNet, semantic profiles, etc. in a more general treatment.

It is common also for stock phraseology to be modified in a way that defeats a lexicon-based approach and raises the need for some knowledge-based reasoning. For example, a Google™ search found the following variant of the 'on the fence' stock phrase: "It will put them on a shaky fence between protecting their copyrights and

technology terrorists". The phenomenon is discussed in more detail in (Barnden, 2006). Similarly, some *e-drama* transcript examples of metaphor have gone in an open-ended way beyond conventional wording even though based on familiar metaphorical conceptions. Such examples include (after conversion of textese, removal of spelling errors, etc.): "I think the mirror breaks all the time you look in it", "you're looking in the mirror right now, but you probably cannot see yourself with all the cracks", "do you even have any brains to think about that one!" and "I'm trying very hard but there's too much stuff blocking my head up" (to illustrate the problems of textese, etc. in our genre, the original form of the second of these examples was: "ur lukin in da mirror rite now, but u probz cnt c yaself wid all da craks").

One particular phenomenon of theoretical and practical interest is that physical size is often metaphorically used to emphasise 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.

We have encountered surprisingly creative uses of metaphor in *e-drama*. For example, in the school-bullying scenario, Mayid is portrayed as having already insulted Lisa by calling her a 'pizza' (short for 'pizza-face'). This figurative insult was given a theoretically intriguing, creatively metaphorical elaboration in one improvisation, where Mayid said "I'll knock your topping off, Lisa".

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 upon detailed utterance-meaning analysis and on rich knowledge bases and reasoning processes, and is currently unsuitable for direct use in the *e-drama* system. However, examples arising in *e-drama* transcripts such as the more creative ones above provide useful data guiding the further development of ATT-Meta and can pose useful challenges to current metaphor theory generally.

4 User testing

We conducted a 2-day pilot user test with 39 secondary school students in May 2005, in order to try out and refine a testing methodology. 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. We concealed the fact that EMMA was involved in some sessions in order to have a fair test of the difference that is made. 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. Also, the frequencies of human 'Dave' and EMMA 'Dave' being responded to during the improvisation (sentences of Dave's causing a response divided by all sentences said by 'Dave') are both roughly around 30%, again suggesting that users notice Dave. Additionally, the frequencies of other side-characters being responded to be roughly the same as the 'Dave' character – 'Matthew': around 30% and 'Elise': around 35%.

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.

More extensive user testing at several Birmingham secondary schools has been conducted recently (up to September 2006). We have conducted an initial evaluation of the quality of EMMA's determinations about emotion during these school 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 centrally involved in the development of EMMA's algorithms). The humans were constrained to use the set of emotion labels that EMMA uses. It is still unclear how to conduct such an evaluation, because the conscious thoughts of a human annotator (labeller) about emotions revealed by an utterance in an offline labelling task may be different from the emotions the annotator would have unconsciously understood during an online e-drama session, and EMMA only surmises an emotion when there is a strong evidence whereas a human labeller may proceed on a different level of evidence. Also, it transpires that there was mediocre agreement between the two human labellers, and the task is artificial for them because they might normally have assigned an emotion outside the prescribed set of 25. To compare the labelling of the two human labellers to each other and to the labelling by EMMA, we used the often-used *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. The statistic, *K*, is calculated as $K = (P(A) - P(E)) / (1 - P(E))$ where $P(A)$ is the proportion of times two coders agree and $P(E)$ is the proportion of times we would expect them to agree if they categorised randomly. A value of at least 0.6–0.8 is generally required by the researchers looking for a good inter-annotator agreement. We calculated *K* for each pair among the three labellers (EMMA and two humans). The inter-human *K* was only 0.32, and so it is not surprising that the EMMA/human values were only 0.32 again and 0.23. However, we also performed a modified comparison in which the emotion labels were conflated to three (positive, negative, and neutral) by grossly lumping together, e.g. the labels deemed positive by our team. We then got a human/human *K* of 0.65, and EMMA/human values of 0.55 and 0.42. The latter are not good values, but they at least give grounds for hope that with further refinement of our affect detection we can come near the rather low human/human level of agreement.

The experimental methodology used in the 2006 testing was as follows, in outline. Subjects were 14–16 years 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 min of improvisation proper.

In half of the School Bullying (SB) improvisations and half of the Crohn's Disease (CD) 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 SB and CD were independently counterbalanced across students.

The questionnaire was largely composed of questions that were explicitly about students' feelings about the experience (notably enjoyment, nervousness, and opinions about the worth of the dramatic contributions of the various characters), with essentially the same set of questions being asked separately about the SB and the CD improvisations. The other data collected were: for each debrief phase, written minutes and an audio and video record; notes taken by two observers present during each Improvisation Phase; and automatically stored transcripts of the sessions themselves, allowing analysis of linguistic forms used and types of interactivity. Inspection of the transcripts collected indicates that EMMA usefully push the improvisation forward on various occasions. The questionnaire answers remain to be subjected to statistical analysis, with the main independent variable of interest being the involvement or otherwise of EMMA in improvisations.

5 Conclusion and ongoing work

We have implemented a limited degree of affect-detection in an automated bit-part character in an e-drama application, and fielded the character successfully in pilot user-testing. Although there is a considerable distance to go in terms of the practical affect-detection that we plan eventually to implement, the already implemented detection is able to cause contributions by the automated character that are reasonably appropriate to the discourse.

Our project makes a contribution to the issue of what types of automation should be included in interactive narrative environments, and as a part of that the issue of what types of affect should be detected (by directors, etc.) and how.

Additionally, a record of an improvisation is automatically filed by our system, and these e-drama session records will be used in a varied playback function we are developing. This function not only enables us to replay the original drama improvised by the actors, but also provides us with the possibility of changing certain character features for the replay, such as clothing and skin colour. Thus, the system also supports reflection on narrative coconstruction and its affective dimension.

Our remaining work on affect detection within the project will be on the metaphorical aspects of it, both in the sense of developing our theoretical ideas about metaphor further in the light of the e-drama data and in the sense of implementing certain limited forms of metaphor processing that can reveal useful extra hints about the affect that is present.

Acknowledgements

This work is supported by grant RES-328-25-0009 from the ESRC under the ESRC/EPSRC/DTI 'PACCIT' programme. We are grateful to Hi8us Midlands Ltd, Maverick Television Ltd, BT, and our colleagues R. Agerri, W.H. Edmondson and S.R. Glasbey. The work is also partially supported by EPSRC grant EP/C538943/1.

References

- Barnden, J., Glasbey, S., Lee, M. and Wallington, A. (2004) 'Varieties and directions of inter-domain influence in metaphor', *Metaphor and Symbol*, Vol. 19, pp.1–30.
- Barnden, J.A. (2006) 'Metaphor, semantic preferences and context-sensitivity', Invited chapter for a Festschrift volume. Kluwer.
- Boucouvalas, A.C. (2002) 'Real time text-to-emotion engine for expressive internet communications', in G. Riva, F. Davide and W. IJsselsteijn (Eds), *Being There: Concepts, Effects and Measurement of User Presence in Synthetic Environments* (pp.305–318).
- Brill, E. (1994) 'Some Advances in Rule-Based Part of Speech Tagging', Paper presented in the Proceedings of *AAAI*.
- Briscoe, E. and Carroll, J. (2002) 'Robust accurate statistical annotation of general text', Paper presented in the Proceedings of the *3rd International Conference on Language Resources and Evaluation*, Las Palmas, Gran Canaria, pp.1499–1504.
- Carletta, J. (1996) 'Assessing agreement on classification tasks: the kappa statistic', *Computational Linguistics*, Vol. 22, pp.249–254.
- Craggs, R. and Wood, M. (2004) 'A two dimensional annotation scheme for emotion in dialogue', Paper presented in the Proceedings of *AAAI Spring Symposium: Exploring Attitude and Affect in Text*.
- Egges, A., Kshirsagar, S. and Magnenat-Thalmann, N. (2003) 'A model for personality and emotion simulation', Paper presented in the Proceedings of *Knowledge-Based Intelligent Information and Engineering Systems (KES2003)*, Lecture Notes in AI. Berlin: Springer-Verlag.
- Ekman, P. (1992) 'An argument for basic emotions', *Cognition and Emotion*, Vol. 6, pp.169–200.
- Elliott, C., Rickel, J. and Lester, J. (1997) 'Integrating affective computing into animated tutoring agents', Paper presented in the Proceedings of *IJCAI'97 Workshop on Intelligent Interface Agents*.
- Fussell, S. and Moss, M. (1998) 'Figurative language in descriptions of emotional states', in S.R. Fussell and R.J. Kreuz (Eds), *Social and cognitive approaches to interpersonal communication*. Lawrence Erlbaum.
- Gratch, J. and Marsella, S. (2004) 'A domain-independent framework for modeling emotion', *Journal of Cognitive Systems Research*, Vol. 5, pp.269–306.
- Heise, D.R. (1965) 'Semantic differential profiles for 1,000 most frequent english words', *Psychological Monographs*. 70 8:(Whole 601).
- Kövecses, Z. (1998) 'Are there any emotion-specific metaphors?', in A. Athanasiadou and E. Tabakowska (Eds), *Speaking of Emotions: Conceptualization and Expression* (pp.127–151). Berlin and New York, NY: Mouton de Gruyter.
- Machado, I., Prada, R. and Paiva, A. (2000) 'Bringing drama into a virtual stage', Paper presented in the Proceedings of *ACM Conference on Collaborative Virtual Environments*. New York, NY: ACM Press.
- Mateas, M. (2002) *Ph.D. Thesis. Interactive Drama, Art and Artificial Intelligence*. School of Computer Science, Carnegie Mellon University.

- McCrae, R.R. and John, O.P. (1992) 'An introduction to the five factor model and its application', *Journal of Personality*, Vol. 60, pp.175–215.
- Mehdi, E.J., Nico, P., Julie, D. and Bernard, P. (2004) 'Modeling character emotion in an interactive virtual environment', Paper presented in the Proceedings of *AISB 2004 Symposium: Motion, Emotion and Cognition*. Leeds, UK.
- Ortony, A., Clore, G.L. and Collins, A. (1988) *The Cognitive Structure of Emotions*. Cambridge, UK: Cambridge University Press.
- Prendinger, H. and Ishizuka, M. (2001) 'Simulating affective communication with animated agents', Paper presented in the Proceedings of *Eighth IFIP TC.13 Conference on Human-Computer Interaction*. Tokyo, Japan, pp.182–189.
- Sharoff, S. (2005) 'How to handle Lexical semantics in SFL: a corpus study of purposes for using size adjectives', *Systemic Linguistics and Corpus*, London, UK: Continuum.
- Watson, D. and Tellegen, A. (1985) 'Toward a consensual structure of mood', *Psychological Bulletin*, Vol. 98, pp.219–235.
- Weizenbaum, J. (1966) 'ELIZA – a computer program for the study of natural language communication between man and machine', *Communications of the ACM*, Vol. 9, pp.36–45.
- Werry, C. (1996) 'Linguistic and interactional features of internet relay chat', *Computer-Mediated Communication: Linguistic: Social and Cross-Cultural Perspectives* (pp.47–64). Pragmatics and Beyond New Series 39. Amsterdam: John Benjamins.
- Wiltchko, W.R. (2003) 'Emotion dialogue simulator. edDrama learning, Inc. eDrama Front Desk.
- WordNet (2006) 'A Lexical database for the english language. Version 2.1. Cognitive Science Laboratory. Princeton University.
- Zhe, X. and Boucouvalas, A.C. (2002) 'Text-to-emotion engine for real time internet communication', Paper presented in the Proceedings of *International Symposium on Communication Systems, Networks and DSPs*, Staffordshire University, UK, pp.164–168.