

## Research Article

# Affect Detection from Text-Based Virtual Improvisation and Emotional Gesture Recognition

Li Zhang<sup>1</sup> and Bryan Yap<sup>2</sup>

<sup>1</sup> School of Computing, Engineering & Information Sciences, Northumbria University, Newcastle NE1 8ST, UK

<sup>2</sup> Department of Mechanical Engineering, University of Bristol, Bristol BS8 1TR, UK

Correspondence should be addressed to Li Zhang, l.zhang@cs.bham.ac.uk

Received 1 June 2012; Revised 15 October 2012; Accepted 2 November 2012

Academic Editor: Owen Noel Newton Fernando

Copyright © 2012 L. Zhang and B. Yap. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We have developed an intelligent agent to engage with users in virtual drama improvisation previously. The intelligent agent was able to perform sentence-level affect detection from user inputs with strong emotional indicators. However, we noticed that many inputs with weak or no affect indicators also contain emotional implication but were regarded as neutral expressions by the previous interpretation. In this paper, we employ latent semantic analysis to perform topic theme detection and identify target audiences for such inputs. We also discuss how such semantic interpretation of the dialog contexts is used to interpret affect more appropriately during virtual improvisation. Also, in order to build a reliable affect analyser, it is important to detect and combine weak affect indicators from other channels such as body language. Such emotional body language detection also provides a nonintrusive channel to detect users' experience without interfering with the primary task. Thus, we also make initial exploration on affect detection from several universally accepted emotional gestures.

## 1. Introduction

Human behavior in social interaction has been intensively studied. Intelligent agents are used as an effective channel to validate such studies. For example, mimicry agents are built to employ mimicry social behavior to improve human agent communication [1]. Intelligent conversational agents are also equipped to conduct personalised tutoring and generate small talk behaviors to enhance users' experience. However, the Turing test introduced in 1950 still poses big challenges to our intelligent agent development. Especially, the proposed question, "can machines think?", makes many of our developments shallow.

We believe it will make intelligent agents possess human-like behavior and narrow the communicative gap between machines and human beings if they are equipped to interpret human emotions during the interaction. Thus in our research, we equip our AI agent with emotion and social intelligence as the potential attempts to answer the above Turing question. According to Kappas [2], human emotions are psychological constructs with notoriously noisy, murky, and fuzzy boundaries that are compounded with contextual

influences in experience and expression and individual differences. These natural features of emotion also make it difficult for a single modal recognition, such as via acoustic-prosodic features of speech or facial expressions. Since human being's reasoning process has taken related contexts into consideration, in our research, we intend to make our agent take multichannels of subtle emotional expressions embedded in social interaction contexts into consideration to draw reliable affect interpretation. The research presented here focuses on the production of intelligent agents with the abilities of interpreting dialogue contexts semantically to support affect detection as the first step of building a "thinking" machine. This research also makes exploration of detecting users' emotional gestures in order to accompany the affect detected from the improvisational contexts to draw stronger affect interpretation. In the meantime, the emotional body language recognition also provides an effective channel to reveal users' experience on a moment-by-moment basis. Our research is conducted within a previously developed online multiuser role-play virtual drama framework, which allows school children aged 14–16 to talk about emotionally difficult issues and perform drama performance

training. In this platform young people could interact online in a 3D virtual drama stage with others under the guidance of a human director. In one session, up to five virtual characters are controlled on a virtual stage by human users (actors), with characters' (textual) "speeches" typed by the actors operating the characters. The actors are given a loose scenario around which to improvise, but are at liberty to be creative. An intelligent agent is also involved in improvisation. It included an affect detection component, which detected affect from human characters' each individual turn-taking input (an input contributed by an individual character at one time). This previous affect detection component was able to detect 15 emotions including basic and complex emotions and value judgments, but the detection processing has not taken any context into consideration. The intelligent agent made attempts to produce appropriate responses to help to stimulate the improvisation based on the detected affect. The detected emotions are also used to drive the animations of the avatars so that they react bodily in ways that is consistent with the affect that they are expressing [3].

Moreover, the previous affect detection processing was mainly based on pattern-matching rules that looked for simple grammatical patterns or templates partially involving specific words or sets of specific alternative words. A rule-based Java framework called Jess was used to implement the pattern/template-matching rules in the AI agent allowing the system to cope with more general wording and ungrammatical fragmented sentences. From the analysis of the previously collected transcripts, the original affect interpretation based on the analysis of individual turn-taking input itself without any contextual inference is proved to be effective enough for those inputs containing strong clear emotional indicators such as "yes/no," "haha," and "thanks." There are also situations that users' inputs do not have any obvious emotional indicators or contain very weak affect signals. thus contextual inference is needed to further derive the affect conveyed in such user inputs.

The inspection of the collected transcripts also indicates that the improvisational dialogues are often multi-threaded. This refers to the situation that social conversational responses of different discussion themes to previous several speakers are mixed up due to the nature of the online chat setting. Therefore, the detection of the most related discussion theme context using semantic analysis is very crucial for the accurate interpretation of the emotions implied in those inputs with ambiguous target audiences and weak affect indicators.

In our previous study, we mainly focused on affect expressed by the human-controlled characters in their virtual improvisation and have not made attempts to find out users' experience by detecting users' emotions expressed in real world via body language and facial expressions while they are operating their characters in front of their computers. During the previous user testing, we also realized that although emotions expressed during their role-play in the virtual world may not be the same emotions experienced at that moment in the real world, the users' improvisation sometimes still reveals hints about their user experience of

using the system. And vice versa, the gestures showed in the real world also sometimes indicate users' feelings and emotions embedded in the virtual improvisation, such as boredom ("I'm getting bored about this" in the meantime showing a gesture such as checking time on their watches), excitement ("haha" in the meantime applauding), disagreement ("I do not like his attitude" by showing an arm-cross gesture), and confusion ("who is the bully? Aren't you the bully?" by showing a scratching head gesture). Thus, it is important to not only use gestures performed during the improvisation as an extra source of information to contribute to more reliable affect interpretation embedded in the virtual improvisation but also show a non-intrusive way to reveal users' experience of using the system on a moment-by-moment basis. In this research, besides the semantic interpretation of the virtual improvisation, we also conduct initial exploration of several universally accepted single upper body emotional gesture recognition in order to build an efficient intelligent user interface.

Moreover, context plays very important roles in the revealing of social goals that hide behind each social interaction. The cognitive research of Kappas [2] also discussed the diversity of affect embedded in "smile" facial expressions during social interaction and the importance of the understanding and employment of the related social contexts for the accurate interpretation of the implied affect in such facial expressions. For example, people tend to use smile facial expressions in order to show happiness and politeness, and to hide desperation and embarrassment. A broad social interaction context may include semantic interpretation of the conversation, discussion themes, tone of voice, and body language. Such a context will significantly help to interpret the most probable emotions and feelings implied in such smile facial expression. This is also one of the most challenging research topics in the affective computing field and the long-term research goal of the work presented here. But at this current stage, the context discussed in this paper indicates ones most recent personal or semantically most related social inputs during the drama improvisation. Such a context normally contains more than one user input and may provide a social communication or personal mood background to inform affect detection especially for those without strong affect indicators.

Thus, the novelty of the work presented here focuses on dealing with open-ended affect detection tasks during drama improvisation. We also employ latent semantic analysis to derive underlying semantic structures embedded in user inputs to go beyond the constraints of affective linguistic indicators to inform affect detection, especially for those inputs with weak or no linguistic emotional features. In order to reason affect from social interaction contexts, relationships between characters and emotions experienced by the target audiences informed by the semantic-based processing are employed. This neural network-based contextual affect reasoning with the consideration of relationships and emotional histories simulates how emotions are developed during social interaction. This is rarely modeled in real-time interactive cognitive computational systems. The work

also makes attempts to use multimodal affect detection from virtual improvisation and human users' body language (gestures) to interpret "social contexts" and to deal with the above-mentioned challenging issues in affect computing.

## 2. Related Work

Tremendous progress in emotion recognition has been witnessed by the last decade. Endrass et al. [4] carried out study on the culture-related differences in the domain of small talk behavior. Their agents were equipped with the capabilities of generating-culture specific dialogues. There is much other work in a similar vein. Recently textual affect sensing has also drawn researchers' attention. Ptaszynski et al. [5] employed context-sensitive affect detection with the integration of a web-mining technique to detect affect from users' input and verify the contextual appropriateness of the detected emotions. However, their system targeted interaction only between an AI agent and one human user in nonrole-playing situations, which greatly reduced the complexity of the modelling of the interaction context.

Scherer [6] explored a boarder category of affect concepts including emotion, mood, attitudes, personality traits, and interpersonal stances (affective stance showed in a specific interaction). Mower et al. [7] argued that it was very unlikely that each spoken utterance during natural human robot/computer interaction contained clear emotional content. Thus, dialog modeling techniques, such as emotional interpolation, have been developed in their work to interpret those emotionally ambiguous or nonprototypical utterances. Such development would benefit classification of emotions expressed within dialogue contexts.

As mentioned earlier, emotion can also be manifested through body language such as facial expressions and gestures. Research of the HUMAINE society has studied emotion recognition from multimodal channels extensively. For example, Castellano et al. [8] focused on drawing affect detection results from diverse channels. They employed facial expressions, body language, and speech for the recognition of eight emotional states. Bayesian classifiers were used for the recognition tasks. The classifier trained with the integration of multimodal data outperformed those with training features extracted purely from one communication channel. Billon et al. [9] presented a continuous gesture recognition system based on principal component analysis (PCA). In their system, movements from any motion capture system can be reduced to single artificial signatures by using properties from PCA. This artificial gesture representation was used in real-time to simultaneously perform gesture segmentation and recognition.

Moreover, as discussed earlier, naturalistic emotion expressions usually consist of a complex and continuously changed symphony of multimodal expressions, rather than rarely unimodal expressions. However, most existing systems consider these expressions in isolation. This limitation may cause inaccuracy or even lead to a contradictory result in practice. For instance, currently many systems can accurately

recognize smile from facial expressions, but it is inappropriate to conclude that a smiling user is really happy. In fact, the same expression can be interpreted completely differently depending on the context that is given [2]. It also motivates this research to use semantic interpretation of text-based social contexts to inform affect detection and to detect affect embedded in upper body language as another channel to understand human behavior better during the improvisation.

Comparing with the above related work, the work presented in this paper focuses on the following aspects: (1) real-time affect sensing for basic and complex emotions in improvisational role-play situations from open-ended individual turn-taking inputs; (2) reasoning affect from social interaction contexts with the consideration of interpersonal relationships between characters and target audiences' most recent emotional indications; (3) employing latent semantic analysis to go beyond linguistic constraints and derive underlying semantic structures embedded in emotional expressions, especially inputs with weak emotional linguistic indicators and ambiguous target audiences; (4) employing multimodal affect detection from virtual improvisational dialogue and gestures to draw a more reliable affect detection conclusion.

## 3. Semantic Interpretation of Social Contexts

From the inspection of the collected transcripts, we noticed that the language used in our application domain is often complex, idiosyncratic, and invariably ungrammatical. It contains abbreviations and borrows heavily from the language of chatrooms. Compared to the language normally analysed in computational linguistics it provides significant additional challenges. We also implemented preprocessing components previously to deal with misspellings, abbreviations, and so forth. Most importantly, the language also 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. It thus guides not only our previous research but also our current developments. For example, in our previous work, we undertook several analyses of any given utterance. These would each build representations which may be used by other components (e.g., syntactic structures) and would construct (possibly weak) hypotheses about the affective state conveyed in the input. Previously we adopted rule-based reasoning, robust parsing, pattern matching, semantic, and sentimental profiles for affect detection analysis. In our current study, we also integrate contextual information to further derive the affect embedded in the interaction context and to provide affect interpretation for those without strong affect indicators.

In order to detect affect accurately from the improvisational inputs without strong affect indicators and clear target audiences, we employ the semantic meaning of the social

interaction context to inform the affect detection processing. In this section, we discuss our approaches of using latent semantic analysis (LSA) [10] and its related packages for terms and documents comparison to recover the most related discussion themes and potential target audiences to benefit affect detection.

In our previous rule-based-driven affect detection implementation, we mainly relied on keywords and partial phrases matching with simple semantic analysis using WordNet, and so forth. However, we notice that many terms, concepts, and emotional expressions can be described in various ways. Especially if the inputs contain no strong affect indicators, other approaches focusing on underlying semantic structures in the data should be considered. Thus, latent semantic analysis is employed to calculate semantic similarities between sentences to derive discussion themes for such inputs.

Latent semantic analysis generally identifies relationships between a set of documents and the terms they contain by producing a set of concepts related to the documents and terms. In order to compare the meanings or concepts behind the words, LSA maps both words and documents into a “concept” space and performs comparison in this space.

In detail, LSA assumes that there are some underlying latent semantic structures in the data which are partially obscured by the randomness of the word choice. This random choice of words also introduces noise into the word-concept relationship. LSA aims to find the smallest set of concepts that spans all the documents. It uses a statistical technique, called singular value decomposition, to estimate the hidden concept space and to remove the noise. This concept space associates syntactically different but semantically similar terms and documents. We use these transformed terms and documents in the concept space for retrieval rather than the original terms and documents.

In our work, we employ the semantic vectors package [11] to perform LSA, analyze underlying relationships between documents, and calculate their similarities. This package provides APIs for concept space creation. It applies concept mapping algorithms to term-document matrices using Apache Lucene, a high-performance, full-featured text search engine library implemented in Java [11]. We integrate this package with our intelligent agent’s affect detection component to calculate the semantic similarities between improvisational inputs without strong affect signals and training documents with clear discussion themes. In this paper, we target the transcripts of the school bullying scenario for context-based affect analysis. In this scenario, it is mainly about the bully, Mayid, which is picking on a new schoolmate, Lisa. Elise and Dave (Lisa’s friends), and Mrs Parton (the school teacher) are trying to stop the bullying.

In order to compare the improvisational inputs with documents belonging to different topic categories, we have to collect some sample documents with strong topic themes. Personal articles from the Experience Project (<http://www.experienceproject.com/>) are used for this purpose. These articles belong to 12 discussion categories including education, family and friends, health and wellness, lifestyle and style, and pets and animals. Since we intend to

perform discussion theme detection for the transcripts of those employed testing scenarios (including school bullying and Crohn’s disease), we have extracted sample articles close enough to these scenarios including articles of Crohn’s disease (five articles), school bullying (five articles), family care for children (five articles), food choice (three articles), school life including school uniform (10 short articles), and school lunch (10 short articles). Phrase and sentence level expressions implying “disagreement” and “suggestion” have also been gathered from several other articles published on the Experience website. Thus, we have training documents with eight discussion themes including “Crohns disease,” “bullying,” “family care,” “food choice,” “school lunch,” “school uniform,” “suggestions,” and “disagreement.” The first six themes are sensitive and crucial discussion topics to the above scenarios, while the last two themes are intended to capture arguments expressed in multiple ways. Affect detection from metaphorical expressions often poses great challenges to automatic linguistic processing systems. In order to detect a few frequently used basic metaphorical phenomena, we include four types of metaphorical examples published on the following website: <http://knowgramming.com/>, in our training corpus. These include cooking, family, weather, and farm metaphors. We have also borrowed a group of “Ideas as External Entities” metaphor examples from the ATT-Meta Project databank (<http://www.cs.bham.ac.uk/~jab/ATT-Meta/Databank/>) to enrich the metaphor categories. Individual files are used to store each type of the metaphorical expressions, such as `cooking_metaphor.txt`, `family_metaphor.txt`, and `ideas_metaphor.txt`. All the sample documents of the above 13 categories are regarded as training files and have been put under one directory for further analysis.

We have taken one example interaction of the school bullying scenario produced by testing subjects during our previous user testing in the following to demonstrate how we detect the discussion themes for those inputs with weak or no affect indicators and ambiguous target audiences:

- (1) Mayid: *ugh! Ur such a wimp* Lisa. [angry]
- (2) Lisa: ☹☹☹ [sad]
- (3) Mayid: Lisa is just *an attention seeker*. [angry]
- (4) Lisa: I’ve got something in my eye. [*Topic themes: “bullying” and “disease”; Target audience: Mayid; Emotion: sad*]
- (5) Mayid: *stop crying*. [disapproval]
- (6) Elise: lisa, what’s up. *r u ok?* [caring]
- (7) Mrs Parton: *detection*, Mayid. [angry]
- (8) Lisa: *leave me alone*. [angry]
- (9) Mayid: I *aint* going to leave you alone. [disapproval]
- (10) Mayid: I’m born to make ur life a misery. [*Topic theme: “bullying”; Target audiences: Mrs Parton and Lisa; Emotion: angry*]
- (11) Lisa: I bet his family hates him because he is so mean. [*Topic themes: “bullying” and “family care”; Emotion: angry*].

Since our previous affect detection focuses on affect interpretation from inputs with strong emotion signals, it provides affect annotation for such inputs in the above example. The emotion indicators are also illustrated in italics in the above interaction. The inputs without an affect label followed straightaway are those with weak or no strong affect indicators (4th, 10th, and 11th inputs). Therefore, further processing is needed to recover their most related discussion themes and identify their most likely audiences in order to identify implied emotions more accurately. Our general idea for the detection of discussion themes is to use LSA to calculate semantic distances between each test input and all the training files with clear topic themes. Semantic distances between the test input and the 13 topic terms (such as “disease”) are also calculated. The detected topics are derived from the integration of these semantic similarity outputs. We start with the 4th input to demonstrate the theme detection.

First of all, in order to produce a concept space, the corresponding semantic vector APIs are used to create a Lucene index for all the training samples and the test file (“test\_corpus1.txt” contains the 4th input). This generated index is also used to create term and document vectors, that is, the concept space. Various search options could be used to test the generated concept model. In order to find the most effective approach to extract the topic themes, we provide rankings for all the training files and the test input based on their semantic distances to a topic theme as the first step. We achieve this by searching for document vectors closest to the vector for a specific term (e.g., “bullying”). The 4th input thus obtains the highest ranking for the topic theme, “bullying,” among all the rankings for the 13 topics. Figure 1 shows the partial output for the searching for document vectors closest to the vector for the topic term, “bullying.” But there are multiple ways to describe a topic theme (e.g., “disagreement”). It affects the file ranking results more or less if different terms indicating the same themes are used. Thus, we need to use other more effective search methods to accompany the above findings.

Another effective approach is to find the semantic similarity between documents. All the training documents contain clear discussion themes indicated by their file names. If the semantic distances between training files and the test file are calculated, then it provides another source of information for topic theme detection. Therefore we use the CompareTerms semantic vector API to find out semantic similarities between all the training corpus and the test document. We provide the top five rankings for semantic similarities between the training documents and the 4th input in Figure 2.

The similarity results listed in Figure 2 show that there are three training files (bullied3.txt, bullied2.txt, and crohn2.txt) semantically most similar to the test file. These three training files, respectively, recommend the following two most related discussion themes: “bullying” and “disease.” In the first step processing mentioned earlier (see Figure 1), to find document vectors closest to that of a topic theme, the test sentence also achieves the best ranking for the “bullying” topic theme. With the integration of the semantic similarity

```
Found vector for "bullying"
Search output follows ...
0.7441609525156784: F:\bullied1.txt
0.4718913937460364: F:\bullied2.txt
0.27654129794720456: F:\school_uniform.txt
0.2309271313719645: F:\bullied3.txt
0.22194621557020616: F:\test_corpus1.txt
0.213115076224951: F:\crohn4.txt
0.16244469025823494: F:\crohn1.txt
0.16220792456248428: F:\food2.txt
0.15936997902429156: F:\family_care3.txt
```

FIGURE 1: Partial example output for searching for document vectors closest to the vector for the topic theme, “bullying” (test\_corpus1.txt containing the 4th input).

```
Similarities between training files and the 4th input:
Similarity of "bullied3.txt" with "test_corpus1.txt": 0.8513
Similarity of "bullied2.txt" with "test_corpus1.txt": 0.7358
Similarity of "crohn2.txt" with "test_corpus1.txt": 0.7202
Similarity of "crohn3.txt" with "test_corpus1.txt": 0.7131
Similarity of "bullied1.txt" with "test_corpus1.txt": 0.6811
Similarity of "crohn4.txt" with "test_corpus1.txt": 0.6563
```

FIGURE 2: Part of the output for the semantic similarities between training documents and the test file (the 4th input).

results between document vectors, our processing concludes that the 4th input from Lisa relates most closely to topics of “bullying” and “disease.” In order to identify the target audiences of the 4th input, we start from the 3rd input to derive its topic themes until retrieving the input with at least partially the same topic themes as those of the 4th input. The processing detecting the 3rd input is most likely to indicate “bullying,” which is also part of the themes embedded in the 4th input. The backtracking for target audience detection stops at the 2nd input, the last round of the input contributed by the same speaker (Lisa). Thus, the target audience of the 4th input is Mayid, who showed bullying behavior in the 3rd input.

In a similar way, the conversation theme processing has identified the following two semantically most similar training documents (bullied3.txt and bullied2.txt) to the 10th input from Mayid shown in Figure 3. These two training files recommend the same discussion theme: “bullying.” The 10th input also achieves the highest ranking for the enquiry of search for document vectors closest to the vector for the topic theme “bullying.” Since the 9th input is also contributed by Mayid, in order to find the target audience of the 10th input, topic theme detection starts from the 8th input from Lisa. The previous version of the affect detection processing that identified this input showed an “angry” emotion using a phrase with a strong affect indicator. Moreover, the 7th input from the school teacher also showed an “angry” emotion with a strong affective phrase “detection” and indicated its target audience, Mayid, by mentioning this character’s name. The original affect detection processing also regarded these two phrasal inputs (the 7th and 8th inputs) as “bullying-” related improvisation. The 6th input mentioned clearly its target audience, Lisa, in its input as well with an identified “caring” theme by the original affect detection. Thus, based

Similarities between training files and the 10th input:  
 Similarity of “bullied3.txt” with “test\_corpus2.txt”: 0.6791  
 Similarity of “bullied2.txt” with “test\_corpus2.txt”: 0.6129  
 Similarity of “bullied1.txt” with “test\_corpus2.txt”: 0.5780  
 Similarity of “crohn3.txt” with “test\_corpus2.txt”: 0.5757

FIGURE 3: Part of the results showing the semantic similarities between training document vectors and the 10th input.

on the above reasoning, Lisa and Mrs Parton are the two target audiences of the 10th input from Mayid, who have aroused the interaction in the first place.

By searching for document vectors closest to the vector for the discussion themes “bullying” and “family care,” the last input (11th input) from Lisa shows high semantic closeness to these two topics, respectively, with semantic distance scores over 0.65 and 0.76. The similarity processing indicates that it is most similar to “bullied3.txt (0.78),” “bullied1.txt (0.76),” and “family\_care2.txt (0.75)” in the semantic domain. Thus, this input is most likely aroused by the 10th input containing a similar “bullying” theme. Thus, its most likely target audience of the 11th input is Mayid, who started the bullying topic in the 10th input.

In general, the conversation theme detection using semantic vectors analysis is able to help the AI agent to detect the most related discussion themes and therefore to identify the most likely target audiences. We believe these are very important aspects for the accurate interpretation of the emotion context. We also envisage that the above processing would be really helpful to distinguish small talk (task unrelated discussion) behaviors from task-driven talk during human agent/robot interaction. Thus, it may enable the AI agent to respond more appropriately during the social interaction. In the following section, we discuss how cognitive cues such as relationships and emotion contexts of target audiences are used to inform context-based affect interpretation.

#### 4. Context-Based Affect Detection

The research of Wang et al. [12] discussed that feedback of artificial listeners can be influenced by interpersonal relationships, personalities, and-culture related aspects. The cognitive emotion research of Hareli and Rafaeli [13] also pointed out that “one person’s emotion is a factor that can shape the behaviors, thoughts, and emotions of other people.” They also believed that “emotions may affect not only the person at whom the emotion was directed but also third parties who observe an agent’s emotion.” In our application domain, one character’s manifestations of emotion can also thus influence others. For example, if two characters share positive relationships and one of them experiences “sad” emotion, then it is more likely the other character responds with an empathic response of “sadness.” Otherwise if they have a negative relationship, then the other character is more inclined to show a gloating response of “happiness.” Thus, such interpersonal relationships (such as positive (friendly)

or negative (hostile or tense) relationships) are also employed to advise the affect detection in the social contexts.

In the example interaction mentioned in Section 3, the topic theme processing has identified that the most likely audience of the 4th input from Lisa is Mayid. That is, the most related social context of the 4th input is the 3rd input indicating a “bullying” negative theme contributed by Mayid. Especially Lisa (the bullied victim) and Mayid (the bully) have a negative relationship and Mayid expressed “anger” via a “bullying” input; thus, the 4th input from Lisa with identified topic themes of “bullying” and “disease” will be most likely to show “sad” or “scared” emotional implications.

Moreover, the topic theme detection also reveals that the 10th input from Mayid is mainly related to the “bullying” topic too and its target audiences are Mrs Parton and Lisa. Since Mayid is the bully and Mrs Parton tries to find out what is going on and stops the bullying, this character shares tense relationships with both Mrs Parton and Lisa. Also, the 7th and 8th inputs contributed by Mrs Parton and Lisa consist of the most related social context of the 10th input and these two inputs both implied “angry” emotions. The 9th input from Mayid with a strong affect indicator “aint” is detected to show “disapproval” by the original version of the system. Thus, the 10th input with a “bullying” theme from the same speaker is most likely built on from his previous input. Embedded in a negative emotion context, it most probably indicates “anger.” In a similar way, the 11th input from Lisa has the topic themes of “bullying” and “family care.” It is also embedded in a negative context, that is, the 10th input, indicating an “angry” emotion with a bullying theme. The speaker also has a tense relationship with the target audience, Mayid. Thus, the 11th input is more inclined to imply the “bullying” theme rather than the “family care” topic and to indicate “anger.”

We implement the above reasoning of emotional influences between characters with the consideration of their interpersonal relationships and recent emotions of target audiences using Backpropagation, a supervised neural network algorithm. Neural networks are generally well known for classification tasks and pattern recognition. Backpropagation is also one of the most classic supervised neural network algorithms. It is chosen due to its promising performances and robustness for the modeling of the problem domain.

We use this neural network implementation to accept most recent emotions of the current input’s potential target audiences and an averaged relationship value between the target audiences and the speaking character as inputs. The number of target audiences could range from one to four for one social input in one drama improvisation session with altogether five characters. The output of the neural network will be the most probable emotion implied in the current input expressed by the speaking character. In this context-based affect detection application, we consider the most frequently used 10 emotions (“neutral,” “approval,” “disapproval,” “angry,” “grateful,” “regretful,” “happy,” “sad,” “worried,” and “caring”) in the bullying scenario as the output detected affective states.

Moreover, since neural networks with one hidden layer are capable enough for the target problem domain, a model with one single hidden layer is chosen in our application. The three-layer topology of the neural network includes one input, one hidden and one output layer, with five nodes in the input layer and 10 nodes, respectively, in the hidden and output layers. The five nodes in the input layer indicate the most recent emotional implications expressed by potential up to four target audiences and an averaged interpersonal relationship value. We use three values to define relationships: 1 for a positive relationship, 0 for a neutral relationship, and  $-1$  for a negative relationship. An average relationship value will be calculated and used as one input to the neural network if the user input has more than one potential target audience. The input emotions are represented in the following way.

According to their distances to “neutral,” happy = 0.1, grateful = 0.2, caring = 0.3, approval = 0.4, neutral = 0.5, regretful = 0.6, disapproval = 0.7, sad = 0.8, worried = 0.85, and angry = 0.9. Other ways of assigning values for emotion inputs (e.g., all values for emotions are distributed between  $-1$  and 1 with positive values assigned for positive emotions and negative values assigned for negative emotions) were also attempted, which produced exactly the same results, that is, recommending the same emotional indication. If there are less than four target audiences available for a conversational input, the value for the neural emotion (0.5) is used to represent the emotion of an unintended audience with the intention of not providing any emotional influence to the speaking character. Finally, the 10 nodes in the output layer represent the 10 output emotions.

The 500 example inputs with agreed annotations extracted from the selected five example transcripts of the bullying scenario are also used for the training of the neural network. A sequence consisting of up to four emotion values, a score for relationship interpretation, and a subsequent speaker’s emotion are regarded as one element of training data. In this way, 500 training data elements are used to train the Backpropagation algorithm. Standard error functions of Backpropagation are used to calculate errors in the output and hidden layers. Then they are, respectively, used to adjust the weights from the hidden to output layer and the weights from the input to hidden layer.

In order to maintain the algorithm’s generalization capabilities, the training algorithm minimizes the changes made to the network at each step. This can be achieved by reducing the learning rate. Thus, by reducing the changes over time, the training algorithm reduces the possibility that the network will become overtrained and too focused on the training data. After the neural network has been trained to reach a reasonable average error rate (less than 0.05), it is used for testing to predict emotional influence of other participant characters towards the speaking character in the test interaction contexts.

In the above example interaction we discussed in Section 3, for the emotion detection of the 4th input, we have the following sequence used as the inputs to the Backpropagation algorithm:

- (1) the most related emotion context: “angry (implied in the 3rd input from the audience, Mayid)-0.9, null-0.5, null-0.5, and null-0.5”. “Null” is used to represent the absence of other audiences;
- (2) relationship: “ $-1$ ,” Lisa and Mayid share a negative relationship.

The neural network uses the above as inputs and outputs “sad” as the implied emotion in the 4th input as discussed earlier. Similarly, for the 10th input from Mayid, the Backpropagation algorithm outputs that he is most likely to be “angry.” The neural network-based reasoning also detects the 11th input from Lisa containing an “angry” emotion with the inputs of an “angry” emotion context and a tense relationship with the target audience, Mayid. Another three transcripts of the bullying scenario are also used for the testing of the neural network. Two human judges are also used to provide affect annotation of the test example inputs. 230 emotional contexts with agreed affect annotation are extracted to evaluate the performance of Backpropagation. In this way, we can provide a channel for context-based affect interpretation as emotion shifters in the social contexts. In the following section, we discuss initial exploration on several emotional gestures recognition so that it provides a non-intrusive way to identify users’ experience and another source for more reliable affect interpretation.

## 5. Initial Developments in Emotional Gesture Recognition

As discussed earlier, since human emotions are psychological constructs with notoriously noisy and vague boundaries, affect detection from a single-isolated channel sometimes may not be sufficient enough. Thus, our research goal is to combine the affect detection results obtained, respectively, from the above semantic interpretation of the dialogue contexts and emotional body language recognition in order to draw a stronger conclusion on affect detection. As mentioned earlier, the gestures showed in the real world also sometimes indicate users’ feelings and emotions embedded in the virtual improvisation. For example, we have observed the following body language of some of the human players during the improvisation of the example interaction mentioned in Section 3:

- (1) Mayid: *ugh! Ur such a wimp* Lisa. [The human participant showing a smile facial expression]
- (2) Lisa: ☹☹☹
- (3) Mayid: Lisa is just *an attention seeker*.
- (4) Lisa: I’ve got something in my eye.
- (5) Mayid: *stop crying*.
- (6) Elise: lisa, what’s up. *r u ok?*
- (7) Mrs Parton: *detection*, Mayid. [The human participant showing an arm-cross gesture]
- (8) Lisa: *leave me alone*.
- (9) Mayid: I *aint* going to leave you alone.

- (10) Mayid: I'm born to make ur life a misery. [The human participant showing an arm-cross gesture]
- (11) Lisa: I bet his family hates him because he is so mean. [The human participant showing one hand holding hip gesture].

Emotional facial expressions were used to indicate the emotions users were experiencing during the improvisation. For example, the human player that controlled the Mayid character showed a smile facial expression to indicate a gloating response of "happiness." Moreover, the testing subjects involved in the above improvisational session also carried their emotional experience in the virtual improvisation to the real world via emotional gesture display. For example, the participants who played the school teacher, the bully (Mayid), and the bullied victim (Lisa), respectively, indicated the heat of the discussion using arm cross (skepticism) and hand-on-hip gestures (anger).

Moreover, we also have the following recorded example interaction of the Crohn's disease scenario accompanied observed users' emotional gestures during the improvisation. In this scenario, Peter is the one with Crohn's disease since he was 15. He needs to go through another life-changing operation and wants to discuss pros and cons with his family and friends, including Mum (Janet), Dad (Arnold), his brother (Matthew), and his best friend (Dave):

- (1) Janet: Matthew..arent u my husband..lol
- (2) Arnold: wat u bin chattin while I was gone.
- (3) Matthew: no son
- (4) Peter: dad we are wearing the same tops!
- (5) Janet: haha. .like father like son
- (6) Arnold: Peter, Matthew I AM YOUR FATHER!
- (7) Arnold: y my lovely wif dressed like a detective.
- (8) Janet: wait..I'm confused..//who is my husband? [The human participant showing a scratching head gesture]
- (9) Peter: Oh, I forgot about what disease do I have. [The human participant showing a hand touching neck gesture].

As shown in the above example, the emotion "confusion" embedded in the virtual improvisation has been carried through via an emotionally consistent scratching head gesture. The regret emotion embedded in the virtual improvisation was also indicated by a hand touching neck annoying gesture.

Such emotional gesture study and developments may also help to identify ironic social interactions in daily life situations, such as showing an arm-cross gesture indicating potential disagreement and in the meantime saying "oh, this is great!" Or someone may applause and in the meantime saying "this is simply waste of time." In the long-term research goal, we also aim to extend our application to a broader daily life context to identify such complex phenomena of human behaviors during social interactions and to help users better in learning situations.

In this section, we made initial exploration on universally accepted single emotional gesture recognition. An unsupervised learning neural network algorithm, Adaptive Resonance Theory (ART-2), is used to perform the recognition task. The emotional gesture recognition results are also combined with the outputs of the semantic-based affect detection of social interaction contexts. The gesture recognition processing is thus able to contribute to the understanding of users' experience throughout the improvisation. Now we discuss the developments of five universally accepted emotional gesture recognition in detail in the following.

As known, body language is another effective channel to express emotions and feelings. It is also one of the effective indicators for mood, meaning, and motive. A single body language signal such as a gesture represents a word. Similar to words, without a sentence context, the meaning of a single gesture could be ambiguous sometimes. Thus, the interpretation of a single-isolated body language may not be reliable. Pease [14] indicated that clusters of body language signals provide a much more reliable context for the indication of meaning and emotions that hide behind the social interactions. Moreover, people from different culture or ethic groups may express emotions and mood using different body language. For example, people from India may shake their heads from side to side to indicate agreement while people from other cultures usually use nodding behavior to express active listening and agreement. In Japan, the depth of the bow indicates the amount of respect shown and also implies the relative status between two people. There are also gestures that can be recognized universally and across culture. In this research, at this current stage, we specifically focus on several universally accepted single upper body emotional gesture recognition. In future work, cluster of gestures will be considered to interpret affect much more appropriately.

Pease claimed that standard arm-cross is a universal gesture signifying defensive or negative attitude. Hands-on-hips pose especially when standing is regarded as one of the common gestures used to communicate an aggressive attitude. He also pointed out that when two men are standing in this pose, then a fight is about to occur. Scratching one's head is normally seen during exams and tends to indicate confusion. If someone is spotting with both hands holding forehead, it probably indicates frustration or a disaster situation. If a person shows one hand holding one elbow, it probably indicates shyness. Therefore, in this research, we make initial exploration to recognize the above gestures including folded arms (aggressive), one hand on elbow (shy), one hand scratching head (confused), both hands holding hips (angry), and both hands holding forehead (frustrated). These gestures mainly refer to upper body language in order to represent typical emotional behaviors expressed while the users are in a sitting position such as in our application. Some of these gestures are presented in Figure 4.

In order to recognize the target emotional gestures, Kinect [15], a motion sensing device produced by Microsoft, is used in this research. The device has an embedded standard



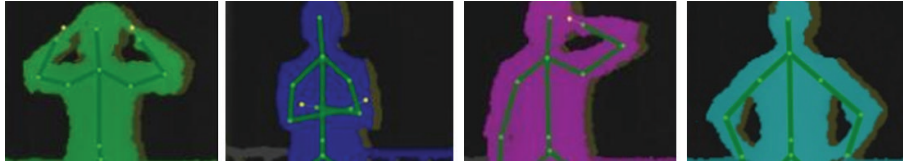


FIGURE 4: Emotional gesture recognition using skeleton tracking (from left to right, gestures indicating frustration, disagreement, confusion, and anger).

RGB camera and a depth camera. It provides skeleton tracking APIs and is capable of establishing positions of 20 skeleton joints. The skeleton points are derived from the processing of depth images collected by the depth camera using algorithms such as matrix transforms. We have developed an algorithm in C++ based on these standard APIs to especially identify the positions of seven joints (e.g., head, hand right, hand left, elbow left, elbow right, hip left, and hip right) in real-time interactions. OpenCV (Open Source Computer Vision Library) is also used to provide efficient real-time image processing. Especially, the Kinect sensor's skeleton tracking engine is also able to perform well on a partially occluded body (such as when a person is sitting near a table).

We employ a neural network algorithm, Adaptive Resonance Theory (ART-2), to perform emotional gesture recognition. Briefly, ART is a collection of models for unsupervised learning and mainly used to deal with object identification and recognition. It simulates the human learning process by linking new concepts with existing knowledge. A new structure is formed when failing to find the link with existing knowledge. ART-1 and ART-2 represent such human learning abilities. ART-1 has the ability to maintain previously learned knowledge (stability) while still being capable of learning new information (plasticity). It is capable of creating a new cluster when required with the assistance of a vigilance parameter. This parameter may help to determine when to cluster a feature vector to a "close" cluster or when a new cluster is needed to accommodate this vector. ART-2 extends the capabilities of ART-1 to support continuous inputs. ART algorithms generally identify the hidden structure in the data by finding how the data is clustered.

In our application, in order to obtain gesture vectors, Kinect is used to gain 30 frames per second and we use a 2-second interval as the length for each collected gesture. Thus, 60 observations are used as inputs to ART-2 to determine the final gesture:

The following attributes that best describe gestures in this application context are chosen.

- (i) the distance between the left hand joint and the left hip joint. It is used to indicate if the user is touching his/her hip, which gives clues to the emotional state of the user such as anger (both hands on the hips) or skepticism (one hand on the hip);
- (ii) the distance between the right hand and the right hip joint. The purpose is the same as the above,

- (iii) the distance between the left hand and the right elbow joint. This also gives clues to the emotional states, such as nervous or shy behavior (one hand holding one elbow) or aggressive behavior (arm-cross);
- (iv) the distance between the right hand and the left elbow joint. The same purpose is as the above;
- (v) the distance between the left hand and the head joint. This can indicate frustration (both hands holding forehead) or confusion (one hand holding forehead);
- (vi) the distance between the right hand and the head joint. It has the same purpose as the above;
- (vii) the distance between the left and right hands.

During recognition, each gesture feature vector is compared to each cluster, and the best match which satisfies the vigilance and similarity test is accepted into the cluster. If no suitable matches are found, a new cluster is created for the vector.

For the testing of the five selected gestures, we use 25 sets for each gesture with each set including 60 observations. Recognition results indicate that all the emotional gestures are well recognized with averaged 0.90+ high precision and recall scores. However, sometimes, arm-cross gestures are misregarded as left hand on right elbow gestures, while sometimes a hands-on-the-hips gesture shows high similarities to a stationary pose as well. In future work, other features will also be incorporated, such as the speed of the hand touching forehead (fast hitting indicating forgetting about something or very slow movement showing potential sadness), in order to extend the current system's recognition capabilities. Other gesture could also be incorporated such as checking time indicating boredom and applause showing excitement or agreement. As mentioned earlier, we intend to use a cluster of gestures in order to better reveal the emotions embedded in body language signals. The research presented here shows a non-intrusive channel to evaluate users' experience. We also have incorporated affect detection from verbal communication mentioned above with the emotional gesture recognition presented here in order to draw a more reliable conclusion on affect detection in social contexts. The overview system flow chart is provided in Figure 5.

We have used the following simple strategies to especially combine contradictory emotions detected from the above two (verbal and nonverbal) channels. For example, as we discussed earlier, when users show contradictory emotions from verbal dialogue contexts and body language in real-life

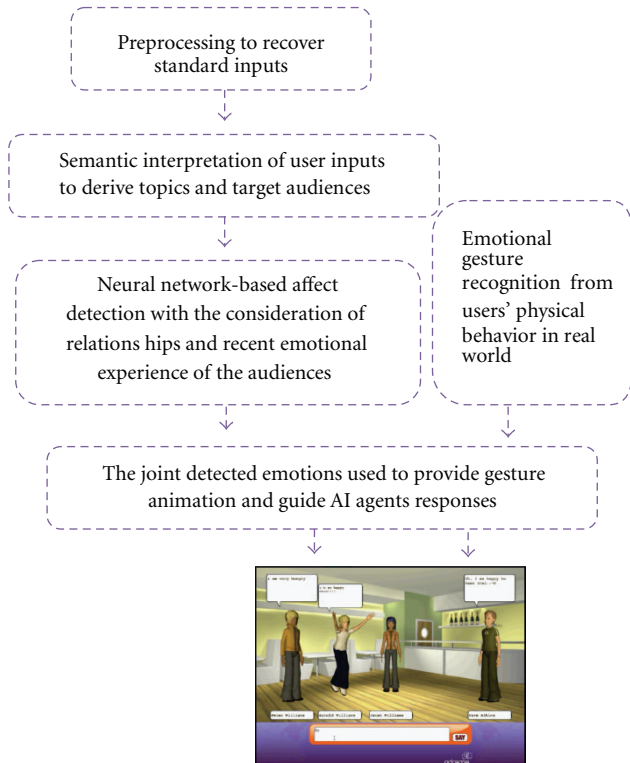


FIGURE 5: Multimodal affect detection processing and the control of characters.

situations, it probably indicates ironic social behaviors. Such a contradictory emotional display is usually used as a means to emphasize or disguise the intended negative emotion. Therefore, in this research, we select the negative emotion derived from either channel as the primary drive to control the intelligent agent's response and other human controlled characters' animation generation.

## 6. Evaluation and Conclusion

We conducted an intensive user test with 160 secondary school students, in order to try out and refine a testing methodology. The aim of the testing was primarily to measure the extent to which having the AI agent as opposed to a person playing a character affects users' level of enjoyment, sense of engagement, and so forth.

The experimental methodology used in the testing was as follows, in outline. Subjects were 14–16-year-old students at local Birmingham and Darlington schools. Forty students were chosen by each school for the testing. There was no control of gender. Four two-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 (SB) improvisation and the remaining five in a Crohn's disease (CD) improvisation; a Second Improvisation Phase in which this assignment is reversed; filling out of a questionnaire by

the students; finally a group discussion acting as a debrief phase. For each improvisation, characters were preassigned to specific students. Each Improvisation Phase involved some preliminaries (background familiarization, appearance choosing, etc.) followed by ten minutes of improvisation proper.

In half of the SB improvisations and half of the CD improvisations, a minor character called Dave in each case was played by one of the students, and by the AI agent in the remaining improvisations. When the AI agent 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.

Also, among the two improvisations within one test session, a minor character was played either by the AI agent or by a human player. This was either the first session or the second. This AI agent-involvement order and the order in which the student encounters SB and CD were independently counterbalanced across students.

Moreover, we concealed the fact that the AI-controlled agent 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 the AI agent 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. This also indicated that users indeed noticed Dave's improvisational inputs during the test sessions. Furthermore, it surprised us that few users appeared to realize 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 realize this.

We have taken previously collected transcripts recorded during our user testing to evaluate the efficiency of the updated affect detection component with contextual inference. In order to evaluate the performances of the topic theme detection and the neural network based affect detection in social contexts, three transcripts of the bullying scenario are used. Two human judges are employed to annotate the topic themes of the extracted 300 inputs from these test transcripts using these 13 topic categories. Cohen's Kappa was used to measure the interannotator agreement between human judges, and the result was 0.83. Then the 265 example inputs with agreed theme annotations are used as the gold standard to test the performance of the topic theme detection. A keyword pattern matching baseline system was used to compare the performance with that of the LSA. We have obtained an averaged precision, 0.736, and an averaged recall, 0.733, using the LSA while the baseline system achieved an averaged precision of 0.603

and an averaged recall of 0.583 for the 13 topic detections. The detailed results indicated that discussion themes of “bullying,” “disease,” and “food choices” were very well detected by our semantic-based analysis. The discussions on “family care” and “suggestion” topics posed most of the challenges. Generally the semantic-based interpretation achieves reasonable and promising results.

The human judges have also annotated these 265 inputs with the 10 frequently used emotions. The interannotator agreement between human judge A/B is 0.63. While the previous version achieves 0.46 in good cases, the new version achieves 0.56 and 0.58, respectively. Inspection of the annotated test transcripts by the new version of the AI agent indicates that many expressions regarded as “neutral” previously were annotated appropriately as emotional expressions. 50 articles from the Experience website were also used to evaluate the semantic-based topic detection. The processing achieved a 66% accuracy rate in comparatively unfamiliar contexts.

Moreover, in order to provide initial evaluation results for the neural network-based affect detection, the human judges’ previous annotations are also converted into positive, negative, and neutral. Then 230 inputs with agreed annotations are used as the gold standard with 37% negative, 33% positive, and 30% neutral expressions. The annotations achieved by the neural network are also converted into solely positive and negative. With the consideration of relationships and most recent emotions expressed by the target audiences, it achieved an average precision of 0.826 and an average recall of 0.813.

Inspection of the transcripts collected indicates that the AI agent also usefully pushed the improvisation forward on various occasions. Figure 6 shows an example about how the AI actor contributed to the drama improvisation in Crohn’s disease scenario. In this illustrated example transcript, Dave was played by the AI actor, which successfully led the improvisation on the desirable track. In another scenario (school bullying) used for the testing, example transcripts also showed that the AI actor has helped to push the improvisation forward.

The preliminary statistical analysis results of the user testing also indicated that the involvement of the improvisational AI actor made no statistically significant difference to the overall users’ engagement and enjoyment and it has usefully stimulated the improvisation under various circumstances. The preliminary results from statistical analysis also indicated that when the AI actor was involved in the improvisation, users’ abilities to concentrate on the improvisation were somewhat higher in Crohn’s disease scenario than school bullying scenario. When the AI actor was not involved in the improvisation, users’ abilities to concentrate on the improvisation were a lot higher in school bullying than Crohn’s disease. This seems very interesting, as it seems to be showing that the AI actor can make a real positive difference to an aspect of user engagement when the improvisation is comparatively uninteresting.

Moreover, in future work, we intend to extend the emotion modeling with the consideration of personality

#### Example transcripts - Crohn’s disease scenario

Dave - the AI character

Peter Williams: IM THE ILL ONE

Dave Atkins: Arnold, Peter needs ur support and u can’t just ignore it.

Janet Williams: nobody cares peter

Arnold Williams: we know peter now stop talking about it

Peter Williams: yes i need your support

Dave Atkins: Hey guys, lets all calm down a bit.

Arnold Williams: we will discuss it later

Janet Williams: well you have got mine

Peter Williams: help me daddy

Arnold Williams: not now son

FIGURE 6: Part of an example transcript produced in the user testing.

and culture. We are also interested in topic extraction to support affect interpretation, for example, the suggestion of a topic change indicating potential indifference to the current discussion theme. It will also ease the interaction and make human characters comfortable if our agent is equipped with culturally related small talk behavior. We believe these are crucial aspects for the development of effective personalized intelligent pedagogical agents. Emotional gesture recognition will also be extended to collect more users’ experience automatically so that such user experience will be used to contribute to a more reliable affect interpretation. In this work, we have initially integrated the affect detection results obtained, respectively, from the semantic-based interpretation of the improvisation and body language signals together in order to provide initial understanding of complex social interactions when unimodal affect sensing is not reliable. In the long term, we intend to build a “thinking” machine by equipping it with the capabilities of drawing affect conclusion from more multimodal channels to enable it to understand human emotions, possess human-like behaviors, and gain social bonding with human users.

## Acknowledgment

The authors thank the support of the Alumni Funding of Northumbria University.

## References

- [1] X. Sun, J. Lichtenauer, M. F. Valstar, A. Nijholt, and M. Pantic, “A multimodal database for mimicry analysis,” in *Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction (ACII ’11)*, pp. 367–376, 2011.
- [2] A. Kappas, “Smile when you read this, whether you like it or not: conceptual challenges to affect detection,” *IEEE Transactions on Affective Computing*, vol. 1, no. 1, pp. 38–41, 2010.
- [3] L. Zhang and J. A. Barnden, “Affect and metaphor sensing in virtual drama,” *International Journal of Computer Games Technology*, vol. 2010, Article ID 512563, 2010.
- [4] B. Endrass, M. Rehm, and E. André, “Planning small talk behavior with cultural influences for multiagent systems,”

- Computer Speech and Language*, vol. 25, no. 2, pp. 158–174, 2011.
- [5] M. Ptaszynski, P. Dybala, W. Shi, R. Rzepka, and K. Araki, “Towards context aware emotional intelligence in machines: computing contextual appropriateness of affective states,” in *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI '09)*, pp. 1469–1474, July 2009.
  - [6] K. R. Scherer, “Vocal communication of emotion: a review of research paradigms,” *Speech Communication*, vol. 40, no. 1-2, pp. 227–256, 2003.
  - [7] E. Mower, A. Metallinou, C. C. Lee et al., “Interpreting ambiguous emotional expressions,” in *Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops (ACII '09)*, Amsterdam, The Netherlands, September 2009.
  - [8] G. Castellano, L. Kessous, and G. Caridakis, “Emotion recognition through multiple modalities: face, body gesture, speech,” in *Affect and Emotion in Human-Computer Interaction*, C. Peter and R. Beale, Eds., vol. 4868 of *Lecture Notes in Computer Science*, Springer, 2007.
  - [9] R. Billon, A. Nédélec, and J. Tisseau, “Gesture recognition in flow based on PCA and using multiagent system,” in *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '08)*, pp. 239–240, October 2008.
  - [10] T. K. Landauer and S. Dumais, “Latent semantic analysis,” *Scholarpedia*, vol. 3, no. 11, article 4356, 2008.
  - [11] D. Widdows and T. Cohen, “The semantic vectors package: new algorithms and public tools for distributional semantics,” in *Proceedings of the 4th IEEE International Conference on Semantic Computing (ICSC '10)*, pp. 9–15, September 2010.
  - [12] Z. Wang, J. Lee, and S. Marsella, “Towards more comprehensive listening behavior: beyond the bobble head,” in *Proceedings of the International Conference on Intelligent Virtual Agents*, 2011.
  - [13] S. Hareli and A. Rafaeli, “Emotion cycles: on the social influence of emotion in organizations,” *Research in Organizational Behavior*, vol. 28, pp. 35–59, 2008.
  - [14] A. Pease, *Body Language—How to Read Others' Thoughts by Their Gestures*, Sheldon Press, London, UK, 1984.
  - [15] Kinect for Windows, April 2012, <http://www.microsoft.com/en-us/kinectforwindows/>.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

