Two approaches for the design of affective computing environments for education

Rafael A. Calvo, Dorian Peters, Lawrence Peters

School of Electrical and Information Engineering, The University of Sydney
Faculty of Education, The University of Sydney
Department of Theatre and Dance, California State University, Fullerton

This paper proposes two design approaches for developing affective tools. The first design is based on an information processing approach and describes a tool that utilizes classification algorithms for automated emotion recognition. The second design follows an interactionist philosophy and describes a visualization tool to support learner/instructor interpretation of affective feedback. These two approaches reflect two contrasting methodological arguments in HCI that have distinct impacts on pedagogy. We demonstrate these designs in the application area of actor training, due to the uniquely explicit link between emotion, emotion recognition and learning in this discipline. We describe a series of experiments, run with both professionals and non-actors, which demonstrate the feasibility of these tools for blended or intelligent learning environments.

Keywords: Physiological signal processing, data mining, affective computing, performing arts, acting.

1. Introduction

1.1 Designing Affective Learning Environments: Information processing and Interactionist approaches

Designing affective computing tools that support learning remains an open question. Herein, we propose, that drawing from the literature in Human Computer Interaction (HCI), we can develop design methodologies that are pedagogically driven and benefit from existing research. Specifically, we look at how two HCI-based design approaches [1] can be applied to the design of affect-aware learning technologies. The information processing approach treats emotion as an entity similar to information, that is communicated from one person to another, and for which designers can build models used to improve user interfaces by increasing the accuracy of adaptation. In contrast, Boehner [1] and others posit a second interactionist interpretation of emotions, in which they are co-created while subjects interpret visualizations of unspecified emotions.

The first approach has, by far, been the most widely employed up to this point. An example of researchers who follow this line of work are Gratch and Marsella [2] who

---

1 Corresponding Author: Rafael A. Calvo. School of Electrical and Information Engineering, Bldg J03, The University of Sydney, NSW 2006, Australia. E-mail: rafa@ee.usyd.edu.au
have developed a computational model of emotion based on appraisal theory, that they use to make more 'believable' agents or virtual humans. EMA (Emotion and Adaptation) uses the appraisal factors that trigger emotions and coping strategies, and has been used to adapt virtual humans that interact with users through natural language in high stress social settings (war scenarios). Work by Conati [3, 4] and Boehner [5] offer further examples of this affect-as-information approach. The underlying assumption of many of these systems is that automatically recognizing the learner's emotional state can enhance a system's capability to interact with the user effectively. For Intelligent Learning Environment development, there is the additional pedagogically driven assumption that an 'emotionally intelligent' tutoring system (or any learning technology) can support students to achieve better learning outcomes. Managing emotions was not the intent of any of these systems.

The interactionist approach was described and compared to the affect-as-information approach by Boehner [1]. The first system – Miró [5] was designed following the information processing approach and used sensor data to feed a model that was then used to create a visualization of the 'mood' of a group of people. The second system – Affector, followed the interactionist approach and the idea that systems can appear intelligent and exhibit complex behaviour without complex representation and processing of information [6]. Affector is an adaptation of a video conference tool connecting two offices, each with a screen and a webcam. Sensors in each office are used to distort each image in ways that can be ‘interpreted’ as representing different emotions. What is different in this approach is that the rules that map sensor signals to specific distortions can be defined by rules created by the users themselves.

1.2 Actor Training as an area of application.

As Jeanne Moreau once explained, “Acting deals with very delicate emotions. It is not putting up a mask. Each time an actor acts he does not hide; he exposes himself.” An essential element of actor education involves helping students learn to elicit and embody genuine emotions and not to rely solely on appearances. In fact, when an actor relies exclusively on the appearance of emotion he is said to be “indicating” [7] which is viewed as false or stylised. Affect influences all types of learning, but it is uniquely in the performing arts that the importance of affect is also at the core of learning outcomes. In no other discipline is the link between emotion, emotion recognition and learning so explicit. Thus it stands to reason that dramatic arts education stands to both a) benefit from emotion recognition and affective technology and b) provide a unique opportunity for experimentation and insight to researchers in the area. Just how these technologies could be feasibly integrated into the classroom and what part they might play within learning activities is an expansive question, but initial findings from a preliminary series of experiments with two distinct approaches to affective computing design, and ideas for practical educational applications are discussed herein.

This paper explores the feasibility of a new type of educational technology specially designed for supporting performing artists, particularly actors. We are applying the two different approaches described earlier, using physiological sensor data as input to an affective computing system.

Section 2 of this paper reviews the pedagogical issues commonly raised in actor training that relate to the support offered by the tools. Section 3 describes a visualization produced with features extracted from physiological signals. These features represent a form of summary that, instead of relying on a computer-generated mapping, allows for student/instructor interpretation as part of an educational activity. Section 4 describes
recording and recognition algorithms as they are commonly used in the affective computing literature, discussing an evaluation of the accuracy of different classification techniques and of data from different subjects (trained actors and non-actors). This section also discusses how they could also be used in supporting the training of actors. In section 5 we conclude, discussing future work.

2. Pedagogical issues in actor training.

Previous research in Affective AIED has focused on supporting student learning with content specific matters. Emotion has only been another variable to be accounted for in complex student models, with the goal of improving their accuracy. In contrast, within the discipline of acting, emotion is at the core, so ‘emotional intelligence’ issues that remain obscured in other training contexts are exposed more openly in actor training.

Other related research has been around how to reliably collect information about students emotions [8] and the so called ‘emotional cartography’ or the visualization of intimate biometric data and emotional experiences using technology [9].

Much of modern actor training is based on the approach created by the renowned Russian acting teacher, Konstantine Stanislavski. Key elements of Stanislavski’s system, include relaxation, emotional memory and concentration of attention. Stanislavski describes a series of elements, but the tools discussed in this paper focus on the enhancements of these three.

The first element, relaxation, is critical to an actor’s success. “Even the slightest tension of muscles can paralyze an actor’s creative state.” [10] Actors are famous for their process of relaxation before going on stage, and virtually every acting text will offer methods for finding this state of relaxed muscle tension, or neutrality.

The second element, referred to as emotional memory, makes use of an actor’s real life experiences in the elicitation of on stage emotions. “The actor must be capable of bringing out the imprint of a past experience and of making it respond to the conditioned stimulus on stage at the moment he needs it.” [10]

Finally, the third element is concentration of attention, which is also referred to in Stanislavski’s writings as “public solitude.” According to Stanislavski, actors must develop the skills to, through concentration of attention, represent actions and emotions realistically, while being watched by any number of audience members. Moore states that, “The more an actor exercises his concentration, the sooner it will become automatic; finally, it will become second nature to him.” (p34)

The practice and development of relaxation and concentration of attention, and the power of these abilities in finding and clarifying emotions through sense memory, is an essential component of Stanislavski’s system. A tool that validates an actor’s work in this area, providing physiologically based feedback could prove effective in the often vague and inconclusive nature of this work. A deeper level of relaxation may be found and focus of attention may be strengthened, facilitating the discovery of emotional memory, which is often elusive and fragmentary. A tool that allows an actor to find consistent personal emotion may help young actors to develop this technique, and make emotional memory more easily accessible. Initial experiments in supporting these elements with affective computing tools are described below.
3. Visualizations for an Interactionist approach

Boehner described [1] design principles that could be used to design this type of systems:

- **Affect as a social product.** The examples show how the idea that emotions can be constructed in the process of a person interacting (with a human or a machine).
- **Elicitation and Interpretative flexibility.** The information processing approach leads to applications where a model is trained with data from subjects who must assess their emotions as one or a combination of a finite set, or as a point in a multidimensional space (e.g., valence/arousal). These models are then used in other instances, possibly by other people. The interactionist model allows users to create their own meanings, and change them in other circumstances.
- **Supporting an extended range of communication acts.** By not constraining users to a finite number of emotional expressions, they can create their own communication acts.
- **Experience centered:** Instead of making systems more aware of emotions, make people more aware of emotions through system design.

Visualization tools that provide feedback have been developed to support teaching in a variety of scenarios, from group activities to writing support. Using the physiological sensor data as input data, we can produce 2 or 3 dimensional representations of physiological features that describe a ‘state’ that the actor and coach can use to discuss.

As a proof of concept, sensors were used to record heart activity (electrocardiogram - ECG), face muscle activity (electromyogram - EMG), and skin conductance. Picard [11] developed a recognition system using these features and blood volume pressure (photoplethysmograph). We asked 3 subjects to enact 8 emotions (same used in [11]): (no emotion, anger, hate, grief, platonic love, romantic love, joy, and reverence). The recordings were made with all the benefits of a lab setting, where subjects were located in a quiet room, with relative privacy and were not required to do any simultaneous cognitive activity (i.e., remember the lines of a play). The subjects were aware that the project was about studying emotions. Naturally, as we approach more realistic scenarios these factors may all affect the performance of the technologies.

After subjects were prepared for the study, the eight emotions mentioned earlier were elicited in order. Each emotion was elicited for a three-minute period, separated by a period of rest. In this study subjects were not told exactly what was meant by each emotion (beyond its name) allowing individual, subjective, interpretations of each emotional label. After each emotion was elicited, subjects were asked to rate the emotion in terms of Arousal, Valence and Dominance on the Self Assessment Manikin pictorial scale [12]. Three sessions were recorded for each of three subjects (an experienced male actor and a female and male with no acting experience). Each session took place on a different day. The sessions with Subject 1 were recorded at 40Hz, while the sessions of Subjects 2 and 3 were recorded at 1000Hz, after the decision was made to see the effect of a higher sampling rate on classification performance.

The signal data for each emotion was organised into 30 overlapping 30-second windows. For each window 120 features were extracted using the Augsburg Biosignal Toolbox [13]. The features extracted were primarily the mean, median, standard deviation, maxima and minima of several characteristics in each signal.

For the sake of visualization, the 120 features were reduced to two and plotted using Fisher Projections [12]. Figure 1 shows the clusters of eight emotions produced by one of the subjects in the study discussed in Section 4. Other feature selection techniques such as
Principal Component Analysis (PCA) may be used. In this representation there is no actual model mapping features to emotions, the subjects would know which emotions they intended to feel and would see the points appearing on the screen. If the student is not focused the clusters might be more spread, since the spread might be interpreted as not being able to ‘sustain’ an emotional state. In collaboration with the coach the meanings of overlaps between clusters would be discussed. It is in the process of discussing with the coach, and reflecting on what they see that they learn.

Figure 1 Fisher projection of physiological representation of 8 emotional states.

The tools proposed would not be developed with the goal of making a na"ive attempt at classifying ‘good’ or ‘bad’ acting. This is neither realistic nor desirable. Instead, what we envision are tools that could provide visualizations of emotions that could be helpful in prompting student reflection on their own practice. For example, the tool might provide visualizations that reveal attempts at eliciting emotions to be distinct or ambiguous, embodied with greater strength or less conviction, etc. These could encourage student reflection and provide support for teacher observations and suggestions. For example, an instructor may observe that there is little notable difference between a student’s portrayal of anger, hatred and anxiety. She could discuss this with the student, but if she were able to demonstrate this lack of distinction by showing a visualization of this overlap, the problem is suddenly more clear. This could then prompt an exercise in which the student takes a closer look at their definition of these emotions. How does one differentiate hatred and anger? The teacher might suggest exercises to improve focus and concentration, or improvisations based on situations that elicit each discretely. It could also prompt a deeper look at the student’s techniques for eliciting these emotions. Is she relying on a personal memory or experience, for example? Perhaps another technique would be more effective in this situation or for this emotion. After making amendments to his approach, a new session may reveal altered results. Here again, teacher feedback could be sufficient to communicate the student’s progress, but a visualization that now shows three clearly differentiated emotions where once they all overlapped could provide uniquely effective reinforcement. In short, the tools could provide a concrete and visual representation of concepts usually confined to the abstract and subjective world.

It is also worth mentioning that these tools could provide extensive opportunities, not only for supporting actor training, but more broadly for supporting research into acting as
a dramatic discipline, for example by providing a new method for understanding how acting works internally and how acting emotion manifests itself physiologically; for the comparison of different acting styles and techniques (e.g. Stanislavsky v. Alba Emoting); how the success of different techniques varies from actor to actor and among different dramatic styles (e.g. American Realism v. German Expressionism) or for different media (e.g. theater v. cinema), and how this correlates to perceptions of good or bad acting. Of course, a better understanding of acting as a discipline will also lead to improvement in actor training in the long run.

4. Affect-as-information and Recognition using physiology

Once the student has participated in an emotion elicitation session or sessions, this data can be used to train a model used for recognition and creating tutoring functionalities. If the system is able to accurately recognize emotions as he participates in a scene, a play or other activity a whole realm of opportunities open up for analysis of emotional dynamics.

A possible way of using an emotion recognition in a actor training scenario, assuming that trained actors ‘feel’ (physiologically) an emotion in similar ways would work as follows: 1) a group of trained actors elicit a number of emotions over different sessions. 2) data collected is used as a training set (‘gold rule’) for an automatic recognition system 3) The student enacts some of these emotions in a particular scene and the system points to ‘misses’, where the wrong emotion has been recalled, possibly due to wrong emotional memory, or has not been sustained for the whole scene due to lack of concentration.

For a system created with this design approach accuracy of the recognition is crucial. Picard [11] developed one such recognition system, training machine learning algorithms with data from 20 sessions where 1 subject elicited the same 8 emotions listed earlier, achieving 81% classification accuracy. They analyzed a total of 40 features, identifying the best 11 features for their subject. Although the set of features may differ for other subjects, the methodology described is quite general.

We investigated some of the effects on classification results of variations in factors such as: number of sessions, number of subjects, and algorithms used for classification. Eight classification algorithms were evaluated using 10-fold cross validation in WEKA [14]:

1. ZeroR: predicts the majority class in the training data; used as a baseline.
2. OneR: uses the minimum-error attribute for prediction [15].
3. Function Trees (FT): classification trees that could have logistic regression functions at the inner nodes and/or leaves.
4. NaiveBayes: A standard probabilistic classifier using estimator classes. Numeric estimator precision values are chosen based on analysis of the training data [14].
5. Bayesian Network: using a hill climbing algorithm restricted by sequential order on the variables, and using Bayes as optimisation criteria.
6. Multilayer Perceptron (MLP): using one hidden layer with 64 hidden units.
7. Linear Logistic Regression (LLR) using boosting.
8. Support Vector Machines (SVM): Finds the maximum margin hyperplane between 2 classes. Weka’s SMO with polynomial kernel was used [16].
Table 1 shows the accuracy of the best classifier (SVM). The columns 1,2,3 represent the results for each single session, while a dataset with the 3 together is in ‘combined session’. The rows show results for each subject with data standarized at 40Hz sampling.

<table>
<thead>
<tr>
<th>Subject</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Combined Sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – 40Hz</td>
<td>96.3%</td>
<td>92.1%</td>
<td>95.4%</td>
<td>80.4%</td>
</tr>
<tr>
<td>2 – 40Hz</td>
<td>94.2%</td>
<td>97.5%</td>
<td>95.8%</td>
<td>74.7%</td>
</tr>
<tr>
<td>3 – 40Hz</td>
<td>90.5%</td>
<td>95%</td>
<td>92.1%</td>
<td>68.1%</td>
</tr>
<tr>
<td>All Subjects (40Hz)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>42.2%</td>
</tr>
</tbody>
</table>

Table 1: Percentage of samples correctly classified for data sets using the SVM Algorithm.

An underlying hypothesis of this study was that the classifiers’ performance gives an indication of an internal ‘consistency’ of the data. If the performance is bad for all algorithms, the data is harder to model. A number of specific experimental issues arise, including:

1. **Intra-Subject, Single Session – concentration**

Classifiers trained and used on a single session for a single subject are the most accurate, and in that sense represent what we would like to achieve in a more general use recognition system. On the other hand we hypothesized when a classifier shows less accuracy for a particular subject/day it could be due not just to the classifier’s properties but also to the fact that the data is less consistent. In pedagogical terms this corresponds to Stanislavski ‘concentration’, described earlier.

The results (Table 1) showed that the physiological signals can be used to produce classifiers with over 90% accuracy. They did not show significant differences for the single session datasets of the professional actor than the single-session datasets of the novices. Although we cannot generalize from these results, they are interesting and require further evaluation.

2. **Intra-Subject, All Sessions – emotional memory**

The consistency with which a subject ‘feels’ an emotion between different sessions is similar to Stanivslasky’s emotional memory. In general, each subject will elicit emotions in different ways on different days. To build a classifier and to test it on data from a single session means excluding the factors of intra-session variation. A subject specific classifier can be trained and tested by combining data from a number of sessions. By combining the data from the three sessions for each subject into a single data set, the classifiers’ accuracy indicates how inter-session variation in emotional elicitation reduces the accuracy. This is probably caused by differences in the appraisal of emotion, intensity and quality of the elicitation.

The results in Table 1 show that the same classifier trained/tested on the multi-session data set is less accurate than the ones trained on a single session, but still achieves accuracy above 68%.

We can see in Table 1 that the accuracy on the expert actor (subject 1) data, at 40Hz, is much higher than the equivalent for other subjects. Comparing the results of individual sessions, some emotions were consistently poorly classified, others consistently well classified, and others varied from session to session. Platonic love and romantic love stand out as emotions that are often misclassified, while anger and the no-emotion baseline were
consistently well classified. Since the data collected for the 3 subjects is limited we do not make general inferences, but indicate that these aspects of the confusion matrix of our very accurate classifiers, are worth further study.

The consistency with which an emotion is elicited appears in the data set combining all of a subject’s sessions. Subject 3 (a novice), for example, shows very high classification success in each session individually, but displays the lowest classification success in the combined data set. This suggests that for each individual session, the consistency of emotion elicited for each 3-minute block was very good (the ‘concentration’), but that the character of emotion elicited from session to session was not as consistent (the emotional memory). Subject 1 (the expert actor), in contrast, shows greater variation within individual sessions, but better consistency across the three sessions.

5. Conclusions

The potential for creating education support tools for actor training that would incorporate emotion recognition technologies and emotion visualisation was discussed. Two design strategies, the affect-as-information most commonly used in the literature, and the affect-as-interaction recently proposed in the HCI literature where used to design two types of tools. Although the ideas described here have not yet being used in actual teaching scenarios, the data collected from professional actors and teachers (co-author in this paper) and novice non-experienced actors provides the first glance at the type of acting training systems that could emerge from the affective computing research. The two systems might complement each other and the efficacy of each is something to be determined. The similarities and distinctions between two approaches would be applicable to other teaching domains.

For the interactionist approach a visualization tool that presents a 2 dimensional representation of the physiological signals produced while eliciting different emotions showed data clustered for each emotion. Actors would be expected to interpret this data, and it is in the reflection process where learning would be expected to occur.

For the second type of systems, automatic recognition is required. The possibility of using real time recognition was evaluated. The strong consistency of classifier success (> 90%) across the nine primary data sets reported (Table 1) adds to the evidence that accurate classifiers are possible. The high accuracy shows that distinct emotions produce distinct and therefore classifiable physiological signals, and that, perhaps unsurprisingly, self-elicited emotions are particularly distinct and classifiable for the expert actor for which significantly better accuracy was achieved.

A noteworthy result was the consistency of misclassification within a subject’s data sets. Subject 3’s romantic love samples were often misclassified as joy, and all subjects showed some misclassification among negative emotions; anger, hatred and grief. The accuracy of the classifiers is reduced when data for all 3 sessions is combined which highlights the effect of intra-day variation noted by Picard [2].

This study provided a first step, but we plan to have a more accurate evaluation with more subjects, of how key factors such as the acting experience, number of sessions, number of subjects, and algorithms used for classification, play a role in the accuracy of classification. While this study lays down an important foundation for recognising the importance of these factors, a complete understanding of the ways they affect the results can only be properly obtained through more detailed studies. Future studies will look
deeper at the difference in results for experts and novices and how visualisation tools could most effectively be used to support student actors.

Acknowledgements

The authors are grateful to Iain Brown for his help in the recording and processing of the data.

References
