Computers in Human Behavior 48 (2015) 615–625

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system ☆

Jason M. Harley [a,b,*], François Bouchet [c,d], M. Sazzad Hussain [e], Roger Azevedo [f], Rafael Calvo [e]

[a] Université de Montréal, Department of Computer Science and Operations Research, 2920 Chemin de la Tour, Pavillon André-Aisenstadt 2194, Montréal, QC H3C 3J7, Canada
[b] McGill University, Department of Educational and Counselling Psychology, 3700 McTavish Street, 614, Montréal, QC H3A 1Y2, Canada
[c] Sorbonne Universités, UPMC Univ. Paris 06, UMR 7606, LIP6, F-75005 Paris, France
[d] CNRS, UMR 7606, LIP6, F-75005 Paris, France
[e] The University of Sydney, School of Electrical and Information Engineering, Sydney, NSW 2006, Australia
[f] North Carolina State University, Department of Psychology, 2310 Stinson Drive, Poe Hall 640, Raleigh, NC 2765-7650, USA

ARTICLE INFO

Article history:
Available online 4 March 2015

Keywords:
Emotions
Affect
Computer-based learning environments
Intelligent tutoring systems (ITS)

ABSTRACT

This paper presents the evaluation of the synchronization of three emotional measurement methods (automatic facial expression recognition, self-report, electrodermal activity) and their agreement regarding learners' emotions. Data were collected from 67 undergraduates enrolled at a North American University whom learned about a complex science topic while interacting with MetaTutor, a multi-agent computerized learning environment. Videos of learners' facial expressions captured with a webcam were analyzed using automatic facial recognition software (FaceReader 5.0). Learners' physiological arousal was recorded using Affectiva's Q-Sensor 2.0 electrodermal activity measurement bracelet. Learners' self-reported their experience of 19 different emotional states on five different occasions during the learning session, which were used as markers to synchronize data from FaceReader and Q-Sensor. We found a high agreement between the facial and self-report data (75.6%), but low levels of agreement between them and the Q-Sensor data, suggesting that a tightly coupled relationship does not always exist between emotional response components.

1. Introduction

Emotions are a critical component of effective learning and problem solving, especially when it comes to interacting with computer-based learning environments (CBLEs; multi-agent systems, intelligent tutoring systems, serious games; Azevedo & Aleven, 2013; Baker et al., 2012; Calvo & D'Mello, 2011; D'Mello and Graesser, 2012; Graesser, D'Mello, & Strain, 2014; Grafsgaard, Wiggins, Boyer, Wiebe, & Lester, 2014; Harley, Bouchet, & Azevedo, 2013; Pekrun, 2011; Sabourin & Lester, 2014). In recent years there has been a surge in interdisciplinary research leading to a plethora of new approaches (including tools/devices and analytical techniques) to measure emotions (e.g., physiological sensors, automatic facial expression analysis software, concurrent state self-report measures; Alzoubi, Hussain, D'Mello, & Calvo, 2011; Baker et al., 2012; Calvo & D'Mello, 2011; D'Mello and Graesser, 2012; Grafsgaard et al., 2014; Harley et al., 2013). The variety of tools and analytical techniques available to researchers enables studies to examine emotions from different modalities (e.g., physiological signals, audio, and video). *Multimodal* approaches (using more than one modality to measure emotions) are aligned with theories that define emotions as multi-componential; in other words, that emotions are expressed and experienced in different ways (e.g., an open mouth, elevated heart rate, *feeling* surprised; Gross, 2010, 2013; Pekrun, 2006, 2011). Multimodal approaches also afford researchers the opportunity to circumvent the constraints of individual channels; particularly those associated with self-report data (e.g., Hawthorne effect; physiological channels cannot be socially masked), and therefore achieve greater construct validity and reliability (Harley, in press; Pantic & Rothkrantz, 2003; Utthara, Suranjana, Sukhendu, & Pinaki, 2010).

☆ *Note.* An earlier version of the synchronization approach used and the agreement rate reported in this manuscript for FaceReader and the EV self-report measure was published in: Harley, J. M., Bouchet, F., & Azevedo, R. (2013). Aligning and comparing data on learners' emotions experienced with MetaTutor. In C. H. Lane, K. Yacef, J. Mostow, P. Pavik (Eds.), *Lecture Notes in Computer Science: Vol. 7926. Artificial Intelligence in Education* (pp. 61–70). Berlin, Heidelberg: Springer-Verlag. This manuscript extends our work synchronizing different methods to a physiological measurement device, provides more detailed results for agreement rates, and elaborates upon our discussion of them.

* Corresponding author at: Université de Montréal, Department of Computer Science and Operations Research, 2920 Chemin de la Tour, Pavillon André-Aisenstadt 2194, Montréal, QC H3C 3J7, Canada. Tel.: +1 (514) 561 3724.

E-mail addresses: jason.harley@umontreal.ca (J.M. Harley), francois.bouchet@lip6.fr (F. Bouchet), sazzad.hussain@sydney.edu.au (M.S. Hussain), razeved@ncsu.edu (R. Azevedo), rafael.calvo@sydney.edu.au (R. Calvo).

http://dx.doi.org/10.1016/j.chb.2015.02.013
0747-5632/

The use of multiple methods to measure emotions in the context of student-CBLE interactions has, however, led to several emerging conceptual, theoretical, methodological, and measurement issues that need to be resolved before empirically driven prescriptions pertaining to learners' emotions can reliably and validly be made (Harley, in press). Challenges include: (1) differences in the sampling rate of emotional data (e.g., frame rate for automatic facial recognition vs. pre-determined time intervals for self-report measures); (2) variation in the detail and kind of emotional information that different methods provide (e.g., dimensional [activation and valence information] for bracelets measuring electrodermal activity (EDA) vs. discrete emotional states from facial expressions); (3) disagreement among theories regarding whether data from different emotional responses should implicate the same emotional state (e.g., if a participant is biting his lip and reports that he is experiencing anxiety should there also be a spike in his physiological arousal data?; Gross, Sheppes, & Urry, 2011); and, (4) day variations in physiological measures due to factors such as environmental changes and sensor placements. The purpose of this paper is to address challenges one through three in the context of research with CBLEs using emotion data from learners' interactions with MetaTutor, a multi-agent-adaptive hypermedia learning environment (Azevedo, Behnagh, Duffy, Harley, & Trevors, 2012; Azevedo et al., 2013; Taub, Azevedo, Bouchet, & Khosravifar, 2014; Trevors, Duffy, & Azevedo, 2014; see Section 2.2).

1.1. Theoretical framework

We view emotions as goal-related and appraisal-driven multicomponential psychological processes that mediate effective learning (Gross, 2010, 2013; Pekrun, 2011). In line with other widely accepted qualities of emotions, we assert that discrete emotions can be categorized by the broad dimensions of arousal (i.e., activation) and valence (Pekrun, 2011; Russell, Weiss, & Mendelsohn, 1989). Valence refers to the intrinsic pleasantness (e.g., enjoyment) or unpleasantness of an emotional state (anger), while arousal refers to the physiologically activating (i.e., arousing; anxiety) or de-activating nature of an emotion (e.g., boredom).

We use Pekrun's (2006; Pekrun & Perry, 2014) control-value theory of achievement emotions, which highlights the role of learners' appraisals of value and subjective control in eliciting emotions that are related to and influential regarding the success of students' academic achievement activities such as, taking tests, studying, and attending class. Pekrun (2006, 2011) distinguishes these two types of appraisals as follows: learners' *appraisals of subjective control* include one's perception of the causal influence they exert over their actions and outcomes. In contrast, *appraisals of value* concern the merit of an activity and its outcome(s), or more broadly, the perception that an action or outcome is positive or negative in nature. For example, it is expected that students who make appraisals of both positive value and sufficient (e.g., high) control will have the most positive emotions while engaging in an academic activity (e.g., enjoyment). On the other hand, students who make appraisals of negative value and high control will experience negative emotional states such as anger. Learners who make appraisals of positive or negative value and low control will experience negative emotional states such as frustration. Students who don't appraise an academic situation as possessing any value are likely to feel bored irrespective of their appraisals of control (Pekrun, 2006; Pekrun & Perry, 2014).

Another factor that informs Pekrun's theory is the object focus, in other words, where a learner's attention is being focused regarding an academic situation that will take place (prospective), has already taken place (retrospective), or is presently taking place (concurrent or activity). The object focus has implications for the appraisals a student will make and the emotions they will subsequently experience. Object foci delineate whether an outcome is being reflected upon or whether an action (that may lead to an outcome) is the focal point. In this study we measured activity emotions: emotions that students report feeling while interacting with a CBLE. We do, however, draw on other emotional states (beyond those Pekrun lists as academic achievement activity emotions) because of the relevance of examining emotions that pertain to appraisals other than achievement standards, including epistemic emotions that relate to the cognitive and learning components of an academic task (e.g., information processing) and include curiosity and confusion (Pekrun, 2011). Examining a more comprehensive set of emotional states also allowed us to compare our findings (1) between modalities (which measure different types of emotions and emotional characteristics such as arousal) and (2) with the results of other researchers whom have identified a large number of emotional states in interactions with computer-based learning environments (Harley & Azevedo, 2014).

Although theories of emotion have different labels and numbers of emotional components, there is indication of agreement in behavioral (e.g., facial expressions), experiential (e.g., how an emotion makes one feel), and physiological (e.g., electrodermal activation) expressions of emotional states (Gross, 2010; Pekrun, 2006). Accordingly, there is growing recognition among researchers of emotions that there is a need to move beyond experiential, self-report measures to inform our theoretical and empirical understanding of emotions in the context of learning (Calvo & D'Mello, 2011; Graesser, D'Mello, & Strain, 2014; Harley, in press; Pekrun, 2006; Pekrun & Linnenbrink-Garcia, 2014). One of the caveats in this area of research is disagreement between theories of emotion regarding whether different components of emotions should provide similar or dissimilar emotional information; in other words, whether different components of emotions should provide a coherent (i.e., coordinated) response (Ekman, 1992; Pekrun, 2011). Pekrun's description of a student's anxiety before an exam illustrates a coherent response among emotional expression components comprised of: "nervous, uneasy feelings (affective); worries about failing the exam (cognitive); increased heart rate or sweating (physiological); impulses to escape the situation (motivation); and an anxious facial expression (expressive)" (Pekrun, 2011, p. 24). Other researchers, on the other hand, argue that a tight coupling (i.e., high level of coherence) between all components may not necessarily exist (D'Mello, Dale, & Graesser, 2012; Gross et al., 2011). This empirical study therefore contributes to the body of research being conducted using multiple modalities to examine emotions with educational, technology-rich environments as well as more experimental contexts. Both are briefly reviewed below.

1.2. Brief review of multi-modal emotion research

A number of empirical studies have used multiple modalities to examine different emotional components during learners' interactions with computer-based learning environments (AlZoubi, D'Mello, & Calvo, 2012; Arroyo et al., 2009; D'Mello, Dale, & Graesser, 2012; D'Mello and Graesser, 2010; Grafsgaard et al., 2014; Kapoor & Picard, 2005). Most, however, have focused on using them to predict learners' emotions compared to a single, separate modality that is used as the grounded truth measure (i.e., standard). Grounded truth measures are typically self-report measures or classifications of facial expressions from video data that are compared with other modalities post-experiment. These studies have focused on optimizing the cumulative accuracy ratings of multimodal measurement approaches and weeding out extraneous individual methods that bring little or no additive gain to the combined ratings. Taken together, results from these studies and a meta-analysis conducted by D'Mello and Kory (2012; that included studies from the