A Preliminary Investigation towards the Development of an Emotion-Aware Partner Agent for Training Control

Federico Colecchia, Joseph Giacomin & Kate Hone
Brunel University London, UK

ABSTRACT
Simulator-based training platforms have become increasingly popular on the grounds of their potential to facilitate skill acquisition within safe and controlled environments. However, current technology is limited in its ability to adapt to individual trainees. Tailoring is in fact typically based on recorded simulation inputs and outputs, or relies on costly and time-consuming trainer-driven interventions, as opposed to direct monitoring of trainee state. This research explores whether automated detection of trainee emotional state can be used to drive real-time changes to the simulator control. The present paper reports on preliminary work to establish the technical viability of such an intervention using current emotion detection technology within a state-of-the-art fixed-base driving simulator environment. Data on the accuracy of the emotion detection software supports the feasibility of the approach, thereby suggesting the possibility of implementing emotion-driven training trajectories bespoke to the needs of individual trainees.

KEYWORDS
Human-computer interaction, emotion detection, emotion-augmented learning, training, driving simulators.

Introduction
The adoption of simulator-based platforms for training purposes has played a key role in enhancing skill acquisition within controlled environments capable of delivering reproducible training workflows (Goldberg *et al.*, 2012). However, the choice of simulator technology has typically been driven by the capabilities available on the market, sometimes resulting in training systems that are not in line with the needs of the trainees, because of a mismatch between platform capabilities and trainee characteristics (Farmer *et al.*, 2017).

The importance of optimising training workflows for individual trainees has been recognised as key to achieving training effectiveness (Goldberg *et al.*, 2015), where training effectiveness has been considered to be a function of the degree of skill proficiency and of the time and cost of reaching it. Existing protocols rely on trainer-driven, computer-tutored or adaptive simulation interventions. In particular, a significant body of research has focused on computer-based tutoring systems and adaptive learning environments (Sottilare & Goldberg, 2012). However, current automated adaptive training protocols are normally based on recorded simulator inputs and outputs as well as on post-activity estimates of the trainee's proficiency (Ministry of Defence, 1989), as opposed to direct real-time estimation of trainee state.

One technology that has the potential to facilitate automated training activities is that of automated emotion detection. Automated emotion detection has been a subject of extensive research over the past few decades (D’mello & Kory, 2015), including studies performed in conjunction with the development of computer-based learning platforms (Shen *et al*., 2009). However, there is still a significant knowledge gap relating to the integration of emotion detection technology within simulator-based training environments. In particular, the authors are not aware of any state-of-the-art simulator training system featuring closed-loop feedback capabilities for reshaping the training workflows based on real-time detection of trainee emotional state.

The potential synergies between simulator-based training and automated emotion detection are underpinned by studies supporting the existence of a link between emotion and performance in learning environments (e.g. Shen *et al*., 2009). Moreover, the hypothesis that feedback based on emotional valence, i.e. on an emotional dimension relating to a “pleasure-displeasure continuum” (Posner *et al*., 2005), can lead to improved training performance is supported by the observation that (i) positive and negative emotions have been shown to affect cognitive function (e.g. Pekrun, 2011), and that (ii) a link has been established between cognitive function and the acquisition of new skills (Fischer, 1980). This underlies the assumption that emotional valence can be used to drive simulator feedback control, thereby translating emotional states into relevant training points, where the term ‘training point’ is taken to refer to the dynamic selection of subsequent simulated scenarios based on real-time detection of trainee emotional state. Whereas real-time control is a well-studied topic in control engineering (Ng, 2016), the technical requirements for the implementation of real-time feedback within simulator-based environments are a subject of current research, and specifications are usually difficult to extrapolate from the original contexts (e.g. Sottilare *et al*., 2015).

The current investigation is part of a research effort that is evaluating the use of real-time emotion detection as a means of optimising training workflows. At its most elementary level, the concept is to adjust the difficulty or nature of training scenarios in real time, based on the emotional state of the trainee. For example, positive-valence emotions can be taken to indicate comfort and confidence about the requested skills and workload, while negative-valence emotions can be taken to indicate discomfort, lack of confidence or confusion. A number of steps are necessary in order to explore the utility of this approach in practice.

This paper addresses the first step, which is an assessment of emotion detection accuracy in open loop, i.e. in the absence of real-time emotion-driven simulator feedback control. Assessing emotion detection performance, although not sufficient for feedback control validation, is in fact a necessary requirement for the further development of this line of research. This study was carried out with a view to confirming the performance of the emotion detection technology employed, as well as to investigating different emotion signals in order to guide the future implementation of closed-loop simulator control. The assessment was performed within the context of a state-of-the-art fixed-base driving simulator. While it is recognised that driving simulators are not currently used for standard training purposes, they share many characteristics with simulators adopted for other training applications, and can therefore be used as a testbed. The approach discussed in this paper required the implementation of a dedicated control system architecture, a high-level representation of which is provided in Figure 1 with reference to the driving simulator used for this study.

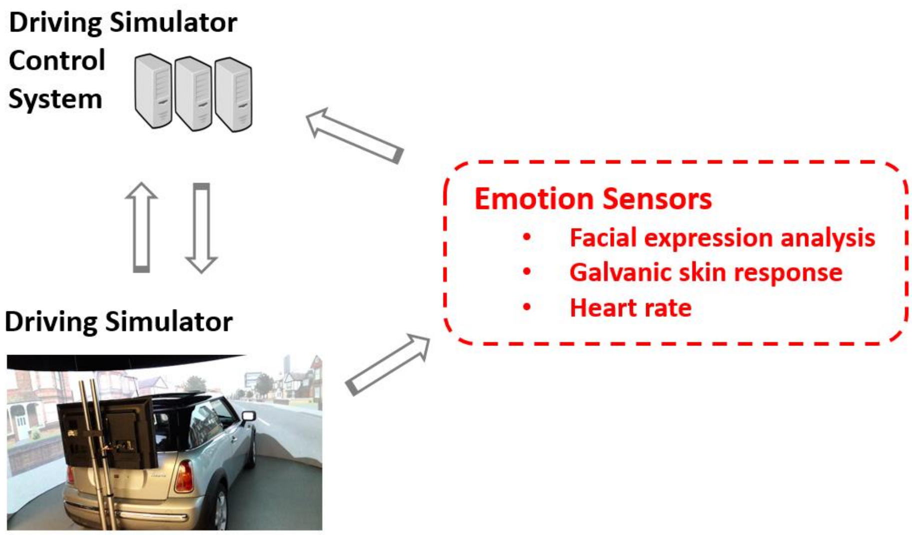

Figure 1 – Emotion-augmented driving simulator high-level architecture.

Specifically, the research objectives of the current investigation were the following:

- O1 To compare the trainee's facial expression signals to the outcome of an emotion self-assessment stage at the end of each driving simulator task;
- O2 To test whether data collected from a limited number of participants could be sufficient to estimate the accuracy of the emotion detection technology;
- O3 To assess the appropriateness of emotional valence to drive the simulator control feedback loop;
- O4 To compile a list of candidate real-time emotion signals of potential interest for further investigation.

Methods

An emotion-augmented simulator platform was set up whereby automated emotion detection technology was embedded within a state-of-the-art fixed-base driving simulator (XPI).

Emotion sensor technology

The choice of emotion detection equipment was driven by the need to (i) maximise the information content of the data in terms of emotion detection, and (ii) minimise interference with the execution of the simulator driving tasks. The underlying requirements relating to the high-definition webcam employed for facial expression detection were (i) low degree of intrusiveness, (ii) sensitivity at relatively-low ambient illumination with a view to exploring different ephemeral settings within the simulation, and (iii) availability of autofocus functionalities. The corresponding selection criteria for the data acquisition system and application software, which were based on a commercial facial expression analysis module (AFFDEX) and biometric research platform (iMotions), were (i) the availability of facial expression signals associated with 'valence', 'engagement' and the six basic emotions of 'anger', 'sadness', 'disgust', 'happiness', 'surprise' and 'fear' (Ekman, 1992), (ii) the inclusion of real-time emotion signal visualisation functionalities, and (iii) scalability with reference to future integration of multiple sensor data channels. Dedicated software was used in conjunction with the iMotions platform for purposes of data exchange, data visualisation, and simulation control. A summary of the emotion detection equipment employed is provided in Table 1.

Table 1 – Outline of the relevant emotion detection technology.

	Hardware	Sampling rates	Resolution	Software
Facial expression detection	Logitech C525 HD Webcam	5-30 frames/s	640x480	Affectiva AFFDEX SDK 3.4.0.1308
System integration	In-house platform	N/A	N/A	iMotions 6.3 API and in-house software

Driving simulator

The driving simulator was selected to meet a set of key requirements, namely (i) flexibility of the software with a view to implementing the emotion-driven feedback control at a later stage, and (ii) the possibility of implementing a range of driving scenarios and environmental conditions in order to elicit a sufficiently-broad spectrum of emotional responses in the driver.

A fixed-base driving simulator was used. The experimental setup consisted of a BMW Mini bodywork, supplemented with 10 rack-mounted desktop computers dedicated to the execution of the driving simulation software and to management of the operator station computing infrastructure in a separate control room. Image rendering was achieved by means of a 270° wraparound screen of 2 m in height and five WUX4000 projectors (resolution 1920x1200 with a 60 Hz refresh rate). This setup was supplemented with a rear window liquid crystal display (LCD) screen and two LCD wing mirrors. Pedal resistance was supplied by tension springs and pick-ups, and gear selector pick-ups were interfaced with the in-car embedded computer while retaining the original gearbox. Working speedometer and revolution counter were also included, and steering was connected to a shock-resistant force feedback unit. Sound effect reproduction relied on in-car front speakers as well as rear radio speakers. Communication between the control room and the car was enabled via an intercom system combined with an in-car closed-circuit television (CCTV) camera.

A list of relevant simulated driving scenarios was compiled, incorporating recommendations from the academic literature and motor press, as well as input from senior researchers at the UK Transport Research Laboratory (TRL) (Kinnear, 2017) and from UK driver training managers (Born, 2017). The simulated driving scenarios were selected based on the desire to induce a broad range of emotional responses and realistic levels of perceptual workload in the driver, as well as in order to implement realistic training environments. Table 2 lists the scenarios employed. Preliminary pilot testing and experimentation with the simulator suggested the viability of these driving scenario specifications with reference to the objectives of the present study.

Table 2 – Summary of the simulated driving scenarios, along with the emotional states that were expected to be elicited in the driver.

DRIVING SCENARIOS	EMOTIONAL STATES

Town centre, rural	**Anger**	**Sadness**	**Disgust**	**Happiness**	**Surprise**	**Fear**
Pedestrian crossings	✓	✓				✓
Unanticipated events	✓				✓	✓
Other cars, larger vehicles	✓					✓
Road obstructions	✓	✓	✓			
Vulnerable road users	✓		✓			✓
Cornering				✓		✓

Table 3 – Summary of simulated scenario settings

Road geometry	Town centre, rural
Weather conditions	Snow, fog
Visibility	10 to 30 m in daylight
Average road segment duration	600 s
Average number of vehicles	10 within 300 m
Average car velocity	20 mph
Maximum car velocity	60 mph

The simulator settings relating to the activated environmental conditions are presented in Table 3. Weather conditions, visibility, and traffic density were selected with a view to increasing the perceptual workload on the trainee. The approach adopted was to achieve relatively-high perceptual workloads so as to increase the probability of triggering emotional responses to the context. Simulated events included (i) unexpected road obstructions, e.g. parked vehicles and barriers, (ii) pedestrians unexpectedly crossing the road, e.g. from behind an obstacle, and (iii) other vehicles not abiding by right-of-way rules, e.g. at junctions and roundabouts. Average and maximum car velocity as reported in the table were not enforced on the drivers as a requirement, but were instead the values observed during the simulator driving sessions as a result of the scenario configurations implemented.

At the end of each driving task, which ranged from 2 minutes to about 15 minutes in duration depending on the specific scenario, participants were asked to watch a recording of the events that took place within the simulation. A set of reference simulated events was selected by the researcher based on the emotional responses elicited in the driver, and the researcher asked the participant to rate his or her response to each reference event using a Self-Assessment Manikin (SAM) (Bradley & Lang, 1994). Participants were asked to assess their emotional state using all three SAM dimensions of 'valence', 'arousal' and 'dominance' (Mehrabian, 1980). Valence was defined as relating to a "pleasure-displeasure continuum" (Posner *et al*., 2005). Arousal was defined as "a mental activity describing the state of feeling along a single dimension ranging from sleep to frantic excitement" (Bakker *et al*., 2014), and dominance was defined as relating to "feelings of control and the extent to which an individual feels restricted in his behaviour" (Bakker *et al*., 2014).

Results

Data was collected from two participants, one male and one female, in the 25-30 age range, with 5-10 years of driving experience. The data corresponded to a total of 28 emotion-eliciting events collected over the four scenarios driven by each participant. This investigation was performed without using real-time detection of the driver's emotional state for the purpose of simulated scenario selection. The accuracy of the emotion detection technology was assessed (Objectives O1 and O2), and different emotion signals were investigated with a view to guiding future closed-loop simulator feedback control design (Objectives O3 and O4).

The AFFDEX facial expression analysis module provides output along different dimensions relevant to emotional response, including 'valence', 'engagement', and the six basic emotional states of 'anger', 'sadness', 'disgust', 'happiness', 'surprise' and 'fear' (Ekman, 1992), in addition to 'contempt'. With a view to estimating detection accuracy, the AFFDEX signals were compared to the outcome of self-assessment using the SAM with reference to the two participants from whom data was collected within this study (Objective O1). Each of the emotional states considered was defined as 'detected' if the corresponding signal either exceeded 30% of the maximum range allowed or was sustained over at least 5 s, whereby the trigger points were selected heuristically based on preliminary experimentation. 'Correct detection' related to the given emotional state being either observed or not observed based on both SAM and AFFDEX, whereas 'incorrect detection' corresponded to a mismatch between the two.

The results are summarised in Table 4 (Objectives O1 and O2), where the legend is as follows: 'engagement' (E); 'anger' (A); 'sadness' (Sa); 'disgust' (D); 'happiness' (H); 'surprise' (Su); 'fear' (F); 'contempt' (C), V_+ (positive valence); V_- (negative valence); V_0 (null valence). The numbers in the table are counts of simulated events. The performance of the eventual emotion-monitoring training system is expected to improve as a function of emotion detection accuracy, with a minimum target being some value greater than 50% probability of correct detection, i.e. better than random chance. This is particularly important regarding emotional valence, which the literature suggests as suitable for driving scenario selection at a later stage on account of observed links between valence and cognitive function (Pekrun, 2011), as well as between cognitive function and skill acquisition (Fischer, 1980). Table 4 reports a preliminary value of valence-related detection accuracy of 74% within the simulator-based platform employed, which suggests the feasibility of using the 'valence' signal to drive the simulator feedback control (Objective O3). The table also points to additional AFFDEX signals worth investigating in more detail (Objective O4), as discussed below.

Table 4 – Summary statistics from the preliminary data presented in this article. Additional information is provided in the text.

	E	A	Sa	D	H	Su	F	C	V_+	V_-	V_0
Correct detection (number of events)	24	26	26	20	27	21	25	26	2	10	8
Incorrect detection (number of events)	4	2	2	8	1	7	3	2	1	6	0

Discussion

This paper has reported on a preliminary investigation of emotion detection accuracy within the context of a state-of-the-art fixed-base driving simulator. The study is a first step towards developing training scenarios the difficulty and nature of which can be adapted based on the trainee's emotional state detected in real time. Data from two participants was collected in open loop, i.e. without real-time simulator feedback control, corresponding to a total of 28 emotion-eliciting simulated events.

The output of a commercial facial expression analysis module was compared to the outcome of emotion self-assessment at the end of the simulator driving tasks (Objective O1). Based on the rates of agreement between emotion detection output and self-assessment results, detection accuracy was estimated (Objective O2). The results are in line with performance figures reported in the literature relating to similar facial expression software (McDuff, 2016; McDuff & Soleymani, 2017; Senechal *et al.*, 2015, Tobin & Hedgcock), although the difference in context makes a comparison difficult. The data confirms known features, e.g. erroneous detection of 'disgust' signals in the absence of the corresponding emotion. Such instances of incorrectly-detected emotional states are documented by the vendor, and are generally attributed to the sensitivity of the software to individual muscle movement patterns, particularly relating to eyes and mouth. The observed rate of correct detection associated with the 'valence' signal also suggests the feasibility of using emotional valence to drive future closed-loop simulation feedback control, thereby justifying further research in this direction (Objective O3). Finally, regarding the identification of candidate real-time emotion signals of potential interest for further investigation, the data has pointed to 'engagement', 'anger', 'sadness', 'happiness', 'fear' and 'contempt' as worth pursuing, on account of the lower incorrect detection rates reported (Objective O4).

The results, albeit preliminary at this stage, suggest the viability of this line of research and encourage further development. The data showed that emotional expressions naturally occurring in the simulator environment could be detected with a degree of accuracy in line with the performance figures reported by the software vendors. Future work will focus on (i) analysing data from a larger number of participants and on (ii) assessing the impact of the emotion-driven simulation feedback control in terms of training effectiveness.

References

AFFDEX, www.affectiva.com, last accessed on 12th Dec 2017

Bakker, I., van der Voordt, T., Vink, P., & de Boon, J. (2014). 'Pleasure, Arousal, Dominance: Mehrabian and Russell revisited', *Curr Psychol*. 33:405-421.

Born, M., Automobile Association (2017), private communication.

Bradley, M. M., & Lang, P. J. (1994). 'Measuring Emotion: the Self-Assessment Manikin and the Semantic Differential', *J Behav Ther Exp Psychiatry* 25(1):49-59.

D'mello, S. K. & Kory, J. (2015). 'A review and Meta-Analysis of Multimodal affect detection systems', *ACM Computing Surveys (CSUR)*. 47(3):43.

Ekman, P. (1992). 'An argument for basic emotions', Cognition and Emotion 6(3-4):169-200.

Farmer, E., Van Rooij, J., Riemersma, J., & Jorna, P. (2017) 'Handbook of simulator-based training', Routledge.

Fischer, K. W. (1980). 'A theory of cognitive development: The control and construction of hierarchies of skills', *Psychological Review*. 87(6):477-531.

Goldberg, B. *et al*. (2012) 'Use of Evidence-based Strategies to Enhance the Extensibility of Adaptive Tutoring Technologies', *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, Orlando, Florida, DOI: 10.13140/2.1.1531.1367

Goldberg, B. *et al*. (2015) ‘Instructional Management for Adaptive Training and Education in Support of the US Army Learning Model – Research Outline’. ARL-SR-0345, tinyurl.com/yajox7et, last accessed on 12th Dec 2017

iMotions, www.imotions.com, last accessed on 12th Dec 2017

Kinnear, N., Transport Research Laboratory (2017), private communication.

McDuff, D. (2016), ‘Discovering Facial Expressions for States of Amused, Persuaded, Informed, Sentimental and Inspired’, *18th ACM International Conference on Multimodal Interaction*, pp. 71-75, Tokyo, Japan.

McDuff, D, & Soleymani (2017), ‘M. Large-scale Affective Content Analysis: Combining Media Content Features and Facial Reactions’, *IEEE 12th International Conference on Automatic Face & Gesture Recognition*, Washington DC, USA.

Mehrabian, A. (1980) ‘Basic dimensions for a general psychological theory’, pp. 39–53. ISBN 0-89946-004-6.

Ministry of Defence (1989), Defence Standard, 00-25(Part 12)/Issue 1, Human Factors for Designers of Equipment – Part 12: Systems.

Ng, T. S. (2016) 'Real Time Control Engineering: Systems And Automation (Studies in Systems, Decision and Control)', Springer

Pekrun, R. (2011) ‘Emotions as Drivers of Learning and Cognitive Development. New Perspectives on Affect and Learning Technologies’ – in ‘Explorations in the Learning Sciences, Instructional Systems and Performance Technologies’ book series (LSIS, volume 3), DOI: 10.1007/978-1-4419-9625-1_3.

Posner, J., Russell, J. A., & Peterson, B. S. (2005) ‘The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology’, *Dev Psychopathol*. 17(3):715–734.

Senechal, T. *et al*. (2015), ‘Facial Action Unit Detection using Active Learning and an Efficient Non-Linear Kernel Approximation’, *2015 IEEE International Conference on Computer Vision Workshop*, Santiago, Chile.

Shen, L., Wang, M., & Shen, R. (2009) ‘Affective e-Learning: Using 'Emotional' Data to Improve Learning in Pervasive Learning Environment’, *Journal of Educational Technology & Society* 12(2):176-189

Sottilare, R. & Goldberg, B. (2012) ‘Designing Adaptive Computer-Based Tutoring Systems to Accelerate Learning and Facilitate Retention’, *Cognitive Technology*, Vol 17, No 1, pp 19-34.

Sottilare, R. *et al*. (2015) 'Domain Modeling for Adaptive Training and Education in Support of the US Army Learning Model – Research Outline', ARL-SR-0325, tinyurl.com/ycpf49w6

Tobin, B. & Hedgcock, W. (no date). ‘Use of Facial Expression Recognition in Market Research’, tinyurl.com/y8ecjnx3, last accessed on 16th Sep 2017.

XPI (no date) ‘XPI DS2 Full Car Simulator’, tinyurl.com/ydbcg3d3, last accessed on 12th Dec 2017