International Journal of Innovations in Engineering and Technology (IJIET)
http://dx.doi.org/10.21172/ijiet.173.03

Improving Data Convergence Across Multimodal Affect Assessment in Aviation Training

Leo Holton, Tianshu Li, and Susanne Lajoie
McGill University, Montreal, Canada

***Abstract*—Affect impacts learning by influencing various cognitive, psychomotor and motivational processes. Recent research on affect has explored multimodal means of assessment. However, multimodal data convergence and interpretation remains a relatively new field. This paper examines affect's role in aviation training using a multimodal affect assessment protocol. As participants performed ten simulated aviation tasks, various dimensions of affect were recorded. We examined convergence among subjective report of affective correlates, physiological response (EDA), behavioral affect changes (facial expression), and performance metrics (log file) with and without baseline removal, and the results were compared. We found that data convergence becomes more evident after baseline subtraction from facial expression and EDA measures, and becomes slightly less evident after baseline subtraction from performance metrics. These findings support the use of baseline removal when using facial expression or EDA data in a multimodal assessment protocol for aviation. We also found that self-report variables had a very similar level and pattern of convergence with data from the subsequent task as with the previous task. This conclusion holds with the Control-Value Theory of Emotions; self-report measures not only correlate with previous tasks, they also predict future affect change, physiological response, and performance.**

***Index Terms*—Affect, Aviation, Multimodal, Control, Value**

I. Introduction

Affect, which includes emotional states, moods, and stress responses [1], impacts learning through cognitive, psychomotor, and motivational processes [2]. In the field of aviation, affect has been shown to affect pilots' ability to perform under demanding circumstances [3]. However, affective data remains underutilized in flight simulation because the implications of affective changes on learning during flight training are not fully understood. Different components of affect can have varying impacts on performance [1]. This paper explores a novel multimodal analysis method by subtracting baseline from multi-componential affect data.

Multimodal assessment of affect is an emerging domain. Although techniques are improving rapidly, many multimodal methodologies remain relatively untested [4]. This paper focuses on multimodal affect research within the realm of aviation training. Previous research detailed a preliminary convergence analysis of collected facial, physiological, and self-report measurements in an aviation setting [5]. Several correlations were found between self-report of affective correlates and physiological responses; skin Conductance Response (SCR) peak count correlated positively with physical workload, effort, and fatigue. The work also showed significant positive correlations between anger and all six self-report variables, and significant negative correlations between surprise and fatigue, and between fear and perceived control. The current paper looks to improve on those results by exploring a new multimodal analysis method.

II. Literature Review

To inform our investigation of the relationship between affect, self-report variables, and performance outcomes in aviation training, we reviewed literature on theories and methodologies most pertinent to this research, namely control value theory of emotions [6, 2] and multimodal frameworks to assess emotional states [4, 7, 10, 12].

A. Control-Value Theory

Pekrun's control-value theory of achievement emotions outlines how appraisals of value and control predict emotions [6] which in turn influence performance in learning activities. Value appraisals relate to perceived importance, usefulness, and interest of an activity. Control appraisals refer to feeling causal agency over one's actions [1, 2]. Learners who value an activity highly with a high sense of control will experience the most positive emotions (e.g. happiness) during and after an activity. Learners with low value, control, or both will experience more negative emotions (e.g. anger). Both value and control appraisals act as affective correlates applicable to multimodal data convergence assessments.Both appraisals are defined as **proximal antecedents**; they are the only factors with direct impact on achievement emotions under Pekrun's theory. Distal antecedents (factors further in the past), such as childhood experiences or past achievements, only influence achievement emotions through their effect on proximal antecedents [6]. Therefore, distal antecedents do not need to be measured in our experiment as long as value and control appraisals are collected.

The emotions elicited by value and control appraisals include activity-related emotions (e.g. enjoyment, frustration) and outcome emotions (e.g. pride, disappointment). The difference between these two types of emotions is **object focus**; object focus describes where a learner's attention is directed, either on an activity in the past (retrospective), present (concurrent), or future (prospective) [2]. Activity-related emotions are concurrent. Outcome emotions are retrospective. Since FaceReader operates while participants fly X-Plane, a concurrent situation, we define the collected emotions as activity-related in the context of our experiment. In our data analysis, **object focus** was taken into account; value and control appraisals were analyzed for convergence with FaceReader and performance data from the previous task (retrospective) and the subsequent task (prospective). Through this method of analysis, we compare different **object focuses**. If Pekrun's theory applies aptly to our aviation-training setting, we should expect similarly high correlations regardless of whether survey appraisals are correlated with the previous or subsequent task.

B. Multimodal Methodologies

Multimodal methods of assessment involve "information input from at least two input sources" [8]. For complex tasks like flight simulation, successfully integrating multiple channels can yield a stronger predictive model of how affect influences performance [8]. Our research analyzes affect convergence along four dimensions; **self-report** (appraisal and workload), **behavioral cues** (facial expression), **performance metrics** (log files), and**physiological data** (electrodermal activity - EDA). EDA is further divided into skin conductance level (SCL), which represents baseline arousal, and skin conductance response (SCR),which measures fluctuations in arousal. Previous research found agreement across all these channels in a medical setting, but differentiated the strengths of each channel: self-report measures provide nuanced ratings for discrete emotions; facial expressions partition basic emotions; log data details context-relevant information; physiological patterns provide precise temporal data [7].

Control and value appraisals predict change in affect and performance [6]. Authors [5] examined this relationship solely using convergence metrics; we sought to improve this analysis by employing **baseline removal**, which isolates change in each parameter by subtracting the parameters observed in a preliminary task. This technique is widely applied to Electrocardiogram [9] and EDA, removing noise and tonic drift [10]. Baseline could also be subtracted from facial expressions to isolate affect change [11]. Therefore, our research isolates changes in EDA and facial expression during aviation performance over time by subtracting the measurements of the aviation task of minimal difficulty (the baseline task – task 0) from all tasks. We aim to remove the possible confounding effect of disparity in tonic facial and EDA output across participants.

III. Research Questions and Hypotheses

The following research questions will assess this potential improvement:

1. Will baseline removal increase the number of significant correlations between two EDA measures (SCR and SCL) and self-reported workload and appraisal?As appraisals predict changes in arousal [6, 12], and baseline-removal isolates arousal changes while removing EDA noise and tonic drift [10], we therefore expect that baseline removal will increase the number of significant correlations [2, 7].
2. Will baseline removal increase the number of significant correlations between affective distribution as measured by FaceReader and self-reported workload and appraisal? As appraisals predict change in emotion[6, 12], and baseline-removal controls for participant disparity in emotions [11], we expect that baseline removal will increase the number of significant correlations [2,4,13].
3. Will baseline removal increase the number of significant correlations between X-Plane log file performance metrics and self-reported workload and appraisal? As appraisals predict change in performance[6], and baseline-removal controls for participant disparity in preliminary skill, we expect that baseline removal will increase the number of significant correlations [2, 4,13].

4. Is a similarly strong correlation obtained as when answering the above three questions when EDA, facial expression distributions, and performance metrics are instead compared with self-reported data from subsequent tasks? As appraisal predicts affect and performance regardless of object focus [2], we expect that the number of significant correlations will be similar regardless of whether object focus is retrospective or prospective [6].

IV. Method

This study is part of a larger project conducted in conjunction with several universities and CAE Inc., an aviation company. Working with aviation experts, we designed an experiment that utilized X-Plane, a software especially useful for orienting beginner pilot trainees that allows aviation instructors to adjust flight tasks to match participant skill and experience. We built an acquisition protocol incorporating four measurements: facial expression recordings, EDA, performance metrics, and self-report questionnaires. These measures are described below. IRB approval was obtained for this study.

A. *Participants*

To test these protocols, 19 participants (11 females) were recruited from undergraduate and graduate student populations at a North American University and compensated $10/hour. The participants, aged 19 to 35 (X = 24.368, σ = 5.814),represent diverse ethnic backgrounds: 4 Asian, 5 White, and 5 from other ethnicities.

B. *Procedure*

After giving consent, participants completed a briefing and a demographic questionnaire. They were then trained to and perform basic maneuvers (adjusting speed, altitude, and heading) using a joystick on X-Plane. participants were able to arrive at and maintain the parameters at baseline, they began experimental tasks; EDA and facial expression recording started as the experimenter relayed separate verbal instructions for 10 aviation tasks for the participant to attempt (e.g. ‘turn left at a 15-degree banking angle to heading 30, raise speed to 275, maintain altitude’). The first task is to maintain baseline parameters. The subsequent tasks are paired maneuvers, such that the second task of each pair undoes the first task, returning the simulation to baseline. Each pair increases in intended difficulty. The final task is also to maintain baseline parameters. Participants fill out 6 identical questionnaires reporting on affect; one after the first task (task 0), one after each additional pair of tasks (tasks 2, 4, 6, 8), and one after the final task (task 9), with breaks after questionnaires if needed.

C. *Measurements*

Behavioural Cues: Facial Expressions through FaceReader 6.0. Footage from the Microsoft LifeCam hd5000 recorded throughout the experiment is processed at 15 Hz by FaceReader 6.0 [14]. FaceReader returns the proportion and intensity of 7 fundamental emotions– neutral, happy, angry, sad, surprised, scared, disgusted.

Physiological Arousal: Electro-dermal activity and BioPac. The BioNomadix EDA module records electrodermal activity [10]. Its sampling rate is 1000 Hz. Using Makowski’s EDA processing algorithm within a Python toolkit, Neurokit [15], SCL and mean SCR peaks per trial are extracted.

Performance Metrics: Log File Data. Performance data was extracted from X-Plane log files. Measures of altitude, heading, and speed accuracy were translated into whole number scores from 1-4, 4 being the highest performance, then averaged into aggregate performance scores [5]. Performance scores were analyzed alongside the three more direct measures of affect.

Experimental Appraisal for ‘Grounded Truth’: Subjective self-reports. Self-reports consisted of: (1) demographic questionnaire about age, gender, ethnicity, and relevant past experience; (2) an adapted perceived task value, control, and workload questionnaire designed in accordance with the expectancy-value theory of achievement emotion [16,1], the Perceived Control Scale [17, 1], and the aviation-focused NASA Task Load Index (TLX) questionnaire [12]. The TLX assesses workload, which is the overall cost to an individual executing a task [12].Workload influences affect regulation through physiological and neurological measures [18]. Our questionnaire contains items on four TLX workload parameters – mental demand, physical demand, effort, and fatigue. Physical and mental workload assess resources spent moving and thinking. Effort assesses an estimation of total physical and mental activity. Fatigue assesses the degree of tiredness that a participant experiences from a task.The questionnaire was administered between tasks and served as the “grounded truth” – a standard determinant of affective experience for comparison with other channels – during convergence analyses with EDA and facial affect data [4].

http://dx.doi.org/10.21172/ijiet.173.03

V. RESULTS

The results of this experiment are divided into four sections. The first three examine the effect of baseline removal on correlation analyses of learners' physiological arousal, facial expressions, and performance metrics with self-report data respectively. The fourth section examines the effect of changed object focus (previous vs subsequent task) on all correlation analyses.

To answer our first three research questions, bivariate correlations were performed between our grounded truth questionnaire variables (mental workload, physical workload, effort, fatigue, perceived value and perceived control) and the measures central to our first three research questions: 1) Will baseline removal increase the number of significant correlations between two EDA measures (mean SCR peak count per minute, mean tonic SCL)and self-reported workload and appraisal; 2) Will baseline removal increase the number of significant correlations between affective distribution as measured by FaceReader (neutral, happy, angry, sad, scared, surprised, disgusted)and self-reported workload and appraisal; 3) Will baseline removal increase the number of significant correlations between X-Plane log file performance scores and self-reported workload and appraisal. After these preliminary analyses, baselines (task 0 data) were subtracted from EDA, facial expression, and performance data before a second round of convergence analysis. For example, mean tonic SCL from task 0 was subtracted from participants' mean tonic SCLs from each task. All bivariate correlations were conducted for all participants and for all tasks that preceded the administration of a questionnaire ($N_{participants} = 19$).

A fourth round of bivariate correlations were conducted to answer our fourth research question:How will a changed object focus affect the correlations between all three areas of prior focus (EDA, facial expression, and performance measures) and self-reported affective correlates. To change object focus from retrospective to prospective, we conducted these bivariate correlations using baseline-subtracted data for EDA, facial expression, and performance from each task that *followed* a "grounded truth" questionnaire ($N_{participants} = 19$).

A. Physiological Arousal

We ran two convergence analyses, one with baseline and one without. Baseline subtraction increased the number of significant correlations between self-reported affective correlates and both EDA measures. Before baseline subtraction (Table 1), SCR mean showed significant negative correlation with mental workload (r = -.187, p< .05) and effort (r = -.195, p< .05), and SCL mean showed significant negative correlation with physical workload (r = -.152, p< .05). After baseline subtraction (Table 2) SCR mean showed significant negative correlation with mental workload (r = -.238, p< .05), physical workload (r = -.366, p< .01), effort (r = -.236, p< .05), and fatigue (r = -.296, p< .01), and significant positive correlation with perceived value (r = .325, p< .01) and perceived control (r = .349, p< .01), while SCL mean showed significant negative correlation with mental workload (r = -.322, p< .01), physical workload (r = -.289, p< .01), effort (r = -.285, p< .01), and fatigue (r = -.228, p< .05), and significant positive correlation with perceived control (r = .273, p< .01). The answer to our first research question – will baseline removal increase the number of significant correlations between EDA measures and self-reported affective correlates – is yes. Our first hypothesis is correct. The consistency of R-value directions for significant correlations strengthens our findings regarding convergence; all R-values are negative for mental and physical workload, effort, and fatigue, and are positive for value and control appraisals except for one outlier (SCR mean and perceived value, r = -.0134).

B. Facial Expressions

We ran two convergence analyses, one with baseline and one without. Baseline subtraction increased the number of significant correlations between self-reported affective correlates and the emotions neutral, happy, scared, and disgusted. Before baseline subtraction (Table 3), the only significant correlations were between angry and perceived value (r = -.271, p< .01), surprised and fatigue (r = -.231, p< .05), and scared and perceived control (r = -.189, p< .05). After baseline subtraction (Table 4), angry, surprised, and sad had the same number of significant correlations with self-report variables. Neutral is significantly correlated with fatigue. Happy is significantly correlated with mental and physical workload, effort, fatigue, and perceived control. Scared is significantly correlated with mental and physical workload, effort, and fatigue. and disgust gained three significant correlations (with physical workload, fatigue, and perceived control). The answer to our second research question – will baseline removal increase the number of significant correlations between facial expressions

TABLE I: BIVARIATE CORRELATIONS BETWEEN **EDA MEANS** AND SELF-REPORT VARIABLES

Measures	**Mental Workload**	**Physical Workload**	**Effort**	**Fatigue**	**Perceived Value**	**Perceived Control**
SCR Mean	-0.1865*	-0.1622	-0.1948*	-0.0697	-0.1343	0.05901

SCL Mean	-0.151	-0.1517*	-0.1758	-0.1333	0.1121	0.1337

*p<.05, **p<.01

TABLE II: BASELINE-REMOVED -BIVARIATE CORRELATIONS BETWEEN **EDA MEANS** AND SELF-REPORT VARIABLES

Measures	**Mental Workload**	**Physical Workload**	**Effort**	**Fatigue**	**Perceived Value**	**Perceived Control**
SCR Mean	-0.2381*	-0.3664**	-0.2355*	-0.2956**	0.3248**	0.3494**
SCL Mean	-0.3217**	-0.2885**	-0.2846**	-0.2278*	0.007507	0.2726**

*p<.05, **p<.01

TABLE III: BIVARIATE CORRELATIONS BETWEEN THE INTENSITIES OF EMOTIONS IN FACIAL **EXPRESSION** ANALYSIS AND SELF-REPORT VARIABLES

Measures	**Mental Workload**	**Physical Workload**	**Effort**	**Fatigue**	**Perceived Value**	**Perceived Control**
Neutral	-0.01071	0.02885	-0.1062	0.06535	0.08732	0.1039
Happy	0.1254	0.07364	0.09235	-0.05064	0.03215	0.05416
Angry	0.008553	-0.1215	-0.01896	0.1595	-0.2713**	-0.0324
Sad	0.07433	0.1725	0.1289	0.04811	0.07867	-0.1627
Surprised	-0.0524	-0.08282	-0.002789	-0.2307*	0.09865	0.07794
Scared	0.08016	-0.0204	0.08664	0.107	-0.01767	-0.1886*
Disgusted	0.1003	0.07784	0.1631	-0.08192	-0.01937	0.1276

*p<.05, **p<.01

TABLE IV: BASELINE-REMOVED -BIVARIATE CORRELATIONS BETWEEN THE INTENSITIES OF EMOTIONS IN FACIAL **EXPRESSION** ANALYSIS AND SELF-REPORT VARIABLES

Measures	**Mental Workload**	**Physical Workload**	**Effort**	**Fatigue**	**Perceived Value**	**Perceived Control**
Neutral	0.05982	-0.0533	0.0486	0.3046**	-0.07092	0.04485
Happy	-0.203*	-0.3607**	-0.2606**	-0.353**	0.1114	0.2579**
Angry	-0.01969	0.1633	-0.03203	-0.2031*	0.02563	-0.06465
Sad	0.08499	0.0705	0.0696	-0.08897	0.1572	-0.02682
Surprised	-0.2568**	-0.1212	-0.1544	-0.09458	0.003457	-0.07331
Scared	-0.2428**	-0.2987**	-0.2441**	-0.2442**	0.1232	0.206*
Disgusted	-0.1451	-0.2214*	-0.09982	-0.3209**	0.03601	0.2748**

*p<.05, **p<.01

and self-reported affective correlates – is yes. Our second hypothesis is correct. R-value directions for significant correlations again show consistency.

C. Performance Metrics

Baseline subtraction decreased the number of significant correlations between performance and experiential affective correlates. Before baseline subtraction (Table 5), performance correlated significantly and negatively with mental workload ($r = -.357$, $p < .01$), physical workload ($r = -.399$, $p < .01$), effort ($r = -.389$, $p < .01$), and fatigue ($r = -.323$, $p < .01$), and significantly and positively with perceived control ($r = .226$, $p < .01$). After baseline subtraction (Table 6), performance correlated significantly and negatively with mental workload ($r = -.264$, $p < .01$), effort ($r = -.311$, $p < .01$), and fatigue ($r = -.263$, $p < .01$) The answer to our second research question – will baseline removal increase the number of significant correlations between facial expressions and self-reported affective correlates – is no. Our third hypothesis is incorrect. There are fewer significant correlations after baseline subtraction. R-value directions for significant correlations again show consistency.

D. Object Focus

Shifting object focus to prospective decreased the number of significant correlations between EDA and experiential affective correlates from 11 to 10 (Table 7, Table 8). The one correlation that became insignificant ($p > 0.05$) when analyzed with prospective focus was between SCR mean and effort. Correlations between SCR mean and fatigue and between SCL mean and effort became less significant, switching from $p < 0.01$ to $p < 0.05$ when analyzed with prospective focus.

Shifting object focus to prospective increased the number of significant correlations between facial expression and experiential affective correlates from 16 to 19 (Table 9, Table 10). Correlations between neutral and effort, sad and perceived value, disgusted and mental workload, and disgusted and effort became significant when analyzed with prospective focus. The correlation between happy and mental workload became insignificant ($p > 0.05$) when analyzed with prospective focus. Correlations between happy and effort, happy and perceived control, surprised and mental workload, and scared and fatigue became less significant, switching from $p < 0.01$ to $p < 0.05$ when

TABLE V: BIVARIATE CORRELATIONS BETWEENPERFORMANCEAND SELF-REPORT VARIABLES

Measures	Mental Workload	Physical Workload	Effort	Fatigue	Perceived Value	Perceived Control
Performance	-0.3574**	-0.3985**	-0.3889**	-0.323**	0.004948	0.2261*

*p<.05, **p<.01

TABLE VI: BASELINE-REMOVED - BIVARIATE CORRELATIONS BETWEENPERFORMANCEAND SELF-REPORT VARIABLES

Measures	Mental Workload	Physical Workload	Effort	Fatigue	Perceived Value	Perceived Control
Performance	-0.2641**	-0.07874	-0.3114**	-0.2627**	0.01213	0.1334

*p<.05, **p<.01

analyzed with prospective focus.

Shifting object focus to prospective decreased the number of significant correlations between facial expression and experiential affective correlates from 3 to 0 (Table 11, Table 12). Correlations between performance and mental workload, effort, and fatigue became insignificant ($p > 0.05$) when analyzed with prospective focus.

The answer to our fourth research question – is there a similarly strong when EDA, performance metrics, and affective distributions are instead compared with self-reported data from subsequent tasks – is yes. Our fourth hypothesis is correct. Overall, the number of significant correlations between self-reported affective correlates and EDA, facial expression, and performance data only decreased by one (from 30 to 29) when object focus was shifted to prospective. However, shifting object made all correlations between performance and self-report data insignificant, suggesting there are some situations where changing object focus will significantly alter correlation results.

TABLE VII: BASELINE-REMOVED -BIVARIATE CORRELATIONS BETWEEN EDA MEANSAND SELF-REPORT VARIABLES

Measures	Mental Workload	Physical Workload	Effort	Fatigue	Perceived Value	Perceived Control
SCR Mean	-0.2381*	-0.3664**	-0.2355*	-0.2956**	0.3248**	0.3494**
SCL Mean	-0.3217**	-0.2885**	-0.2846**	-0.2278*	0.007507	0.2726**

*p<.05, **p<.01

TABLE VIII: BASELINE-REMOVED, PROSPECTIVE FOCUS -BIVARIATE CORRELATIONS BETWEEN EDA MEANSAND SELF-REPORT VARIABLES

Measures	Mental Workload	Physical Workload	Effort	Fatigue	Perceived Value	Perceived Control
SCR Mean	-0.2119*	-0.2885**	-0.2119	-0.2885*	-0.2119**	-0.2885**
SCL Mean	-0.3535**	-0.3361**	-0.3535*	-0.3361*	-0.3535	-0.3361**

*p<.05, **p<.01

TABLE IX: BASELINE-REMOVED -BIVARIATE CORRELATIONS BETWEENTHE INTENSITIESOF EMOTIONSIN FACIAL EXPRESSION ANALYSISAND SELF-REPORT VARIABLES

Measures	Mental Workload	Physical Workload	Effort	Fatigue	Perceived Value	Perceived Control
Neutral	0.05982	-0.0533	0.0486	0.3046**	-0.07092	0.04485
Happy	-0.203*	-0.3607**	-0.2606**	-0.353**	0.1114	0.2579**
Angry	-0.01969	0.1633	-0.03203	-0.2031*	0.02563	-0.06465
Sad	0.08499	0.0705	0.0696	-0.08897	0.1572	-0.02682
Surprised	-0.2568**	-0.1212	-0.1544	-0.09458	0.003457	-0.07331
Scared	-0.2428**	-0.2987**	-0.2441**	-0.2442**	0.1232	0.206*
Disgusted	-0.1451	-0.2214*	-0.09982	-0.3209**	0.03601	0.2748**

*p<.05, **p<.01

TABLE X: BASELINE-REMOVED, PROSPECTIVE FOCUS -BIVARIATE CORRELATIONS BETWEENTHE INTENSITIESOF EMOTIONSIN FACIAL EXPRESSION ANALYSISAND SELF-REPORT VARIABLES

Measures	Mental Workload	Physical Workload	Effort	Fatigue	Perceived Value	Perceived Control
Neutral	0.1835	0.003492	0.2214*	0.3466**	-0.07374	-0.03938
Happy	-0.1787	-0.3739**	-0.2381*	-0.3227**	0.1328	0.2452*
Angry	-0.126	0.1479	-0.1788	-0.2386*	-0.02115	-0.02529
Sad	0.06476	0.0338	0.07646	-0.1143	0.2078*	0.09091
Surprised	-0.2262*	-0.1	-0.1264	-0.08329	0.02824	-0.08048
Scared	-0.3047**	-0.3**	-0.3057**	-0.2494*	0.1335	0.2506*
Disgusted	-0.2144*	-0.2575*	-0.2231*	-0.36**	0.0324	0.3065**

*p<.05, **p<.01

TABLE XI: BASELINE-REMOVED - BIVARIATE CORRELATIONS BETWEEN PERFORMANCEAND SELF-REPORT VARIABLES

Measures	Mental Workload	Physical Workload	Effort	Fatigue	Perceived Value	Perceived Control
Performance	-0.2641**	-0.07874	-0.3114**	-0.2627**	0.01213	0.1334

*p<.05, **p<.01

TABLE XII: BASELINE-REMOVED, **PROSPECTIVE FOCUS** - BIVARIATE CORRELATIONS BETWEEN PERFORMANCEAND SELF-REPORT VARIABLES

Measures	Mental Workload	Physical Workload	Effort	Fatigue	Perceived Value	Perceived Control
Performance	-0.1858	0.0475	-0.2003	-0.1792	0.03703	0.02495

*p<.05, **p<.01

VI. DISCUSSION AND CONCLUSION

This paper examines data analysis techniques applicable to the relatively new realm of multimodal affective analysis, specifically in an aviation-training setting. Our contention was that convergence would improve when we isolate the net change in EDA, facial expression, and performance. Our analysis confirmed our first two hypotheses; removing baseline from both EDA and facial expression data increases the number of significant correlations with self-report data. These results are a significant improvement to our multimodal analysis framework [5].

However, our analysis disproved our third hypothesis, as baseline data from performance measures led to a decrease in significant correlations. There are a few possible explanations for this that could be evaluated further: 1) It is possible that participants are still not familiar enough with the X-Plane program during Task 0 to perform predictably, and as a result task 1 or task 2 would function better as a baseline. 2) Performance might simply not respond well to any baseline removal as we have formulated it; expanding our 1 to 4 scale for performance means to a larger range of magnitudes (as our EDA and facial expression data had) could improve the effectiveness of baseline removal.

Changing object focus from retrospective to prospective had little effect on the results of convergence analyses, confirming our fourth hypothesis. The exception was a decrease in significant correlations between self-report and performance data. Further investigation into task 0 as a suitable performance baseline could also address this disparity.

This paper provides empirical evidence for applying control value theory to an aviation context. After baseline removal to eliminate base participant disparities, **perceivedcontrol** correlated significantly and positively with SCR, SCL, happy, scared, and disgusted. This upholds control as a predictor for change in affect and arousal [6]. **Perceived value** correlated significantly with only SCR and no facial expressions after baseline removal, calling into question its viability as a predictor for affect change assessed through physiological and behavioural cues.

The results of our convergence analyses support the inclusion of both facial expressions and EDA data in an eventual affect assessment framework for aviation training. Our research demonstrates that SCR, SCL, happy, and scared after baseline removal negatively predict all four workload parameters that we used. Pilot instructors should therefore focus on monitoring these EDA measures and two FaceReader emotions in pilot trainees to control workload during training and avoid burnout or boredom. The results also support inclusion of performance in such an assessment framework, but without employing baseline subtraction. Before baseline subtraction, we found that all four workload measures negatively predicted performance, while perceived control positively predicted performance. In order to improve performance, instructors should therefore look to lower trainee's perceived workload.

Our results hold several implications for future multimodal methodology research. First, we found that baseline subtraction substantially increased the total number of significant correlations between objective (EDA and facial expression) and subjective measures by factors of 3.67 (EDA) and 5.33 (facial expression). Baseline subtraction is used in several physiological areas [9,10], but remains underutilized, along with other analytical practices, in the emerging field of multimodal methodologies [4]. As such, replication of baseline removal in other contexts and with larger participant sample sizes is necessary to ensure the technique's validity. If successfully substantiated, baseline removal could ensure that important affective indicators are not missed in future multimodal experiments. Additionally, while confirming past findings detailing significant convergence between grounded truth and affect measures [8], our analysis raised several questions about facial expressions as affective correlates. After baseline removal, every significant correlation between FaceReader emotions and TLX workload parameters was negative. Every significant correlation between FaceReader emotions and perceived value and control parameters was positive. Emotional valence (positive vs negative emotion) seemingly had no effect on the directionality of correlations with workload and appraisal. This contradicts Pekrun's assertion [6] that "emotional intensity increases with increasing controllability (in positive emotions) or uncontrollability (in negative emotions)". To explain this disagreement, further investigation is needed into common action units that FaceReader uses to identify emotions like happy (positive valence) and scared (negative valence).

Moving forward with our larger project, we hope to involve convergence analyses between identified affect measures and X-Plane performance to identify the relationship between affect and accuracy of specific maneuvers. Our research findings confirm the viability of EDA and facial expression as affect measures and support the importance of baseline removal, providing a strong base for further analyses. We are one step closer to a multimodal affective model that will improve pilot training.

CONFLICT OF INTEREST

Susanne Lajoie declares that she has previously worked with Professor Claude Frasson on a grant. ImeneJraidi, who contributed input to the research direction for this project, declares that Professor Claude Frassonserved in the past as her PhD supervisor.

AUTHOR CONTRIBUTIONS

Tianhsu Li led research for this project, with help from Leo Holton. Leo Holton conducted data analysis for this paper. Leo Holton also wrote this paper, with directional and editorial help from Tianshu Li. Susanne Lajoie provided oversight at all stages of the project – research, data analysis – and helped with the editing process.

ACKNOWLEDGEMENTS

We would like to thank Alejandra Ruiz Segura and ImeneJraidi for their time and input during the research process. This study is jointly funded by Natural Sciences and Engineering Research Council of Canada (514052-17), Consortium for Aerospace Research and Innovation in Canada (CARIC) and Quebec (CRIAQ) (OPR-1618).We also thank CAE Inc. and researchers at Concordia and Université de Montréal for their continued work on this larger project. Finally, we appreciate all participants who gave their time and mental energy towards flight-simulations needed for our research.

REFERENCES

[1] Gross, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology*, *39*(3), 281–291.

[2] Pekrun, R., Linnenbrink-Garcia, L. (2014). Introduction to emotions in education. *International handbook of emotions in education*, 11-20.

[3] Kaempf, G. L., Klein, G., & Klein, G. (2017). *Aeronautical Decision Making: The next generation*. Aviation Psychology in Practice; Routledge.

[4] Harley, J. (2015). *Measuring Emotions: A Survey of Cutting-Edge Methodologies Used in Computer-Based Learning Environment Research* (pp. 89–114).

[5] Authors. (2020). Details omitted for blind reviewing.

[6] Pekrun, R. (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review*, *18*, 315–341.

[7] Harley, J., Bouchet, F., Hussain, S., Azevedo, R., & Calvo, R. (2015). A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior*, *48*, 615–625.

[8] Omi, H. (2009). *Multimodal input method* (United States Patent No. US7630901B2).

[9] Chouhan, V., & Mehta, S. (2007). *Total Removal of Baseline Drift from ECG Signal*. 512–515.

[10] Braithwaite, J., Watson, D. G., Jones, R., & Rowe, M. (2015). *A guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments.* University of Birmingham.

[11] Li, X., Pfister, T., Huang, X., Zhao, G., &Pietikainen, M. (2013). *A Spontaneous Micro-expression Database: Inducement, collection and baseline*. 1–6.

[12] Hart, S. G., &Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology* (Vol. 52, pp. 139–183). North-Holland.

[13] Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements: *Psychological Science in the Public Interest*.

[14] Den Uyl, M., &Kuilenburg, H. (2005). The FaceReader: Online Facial Expression Recognition. *Proceedings of Measuring Behavior, Wageningen*.

[15] Makowski, D. (2016). *neurokit: A Python Toolbox for Statistics and Signal Processing (EEG, EDA, ECG, EMG...).* (0.2.0) [Python; MacOS, Microsoft :: Windows, Unix].

[16] Wigfield, A., & Eccles, J. S. (2000). Expectancy–Value Theory of Achievement Motivation. *Contemporary Educational Psychology*, *25*(1), 68–81.

[17] Perry, R., Hladkyj, S., Pekrun, R., Clifton, R., & Chipperfield, J. (2005). Perceived academic control and failure in college students: A three-year study of scholastic attainment. *Research in Higher Education*, *46*, 535–569.

[18] Regula, M., Socha, V., Kutilek, P., Socha, L., Hána, K., Hanakova, L., & Szabo, S. (2014, December 3). *Study of heart rate as the main stress indicator in aircraft pilots*. Proceedings of the 16th International Conference on Mechatronics, Mechatronika 2014.