

1837  
2017  
YEARS



HELLENIC REPUBLIC

**National and Kapodistrian  
University of Athens**

**School of Economics and Political Sciences**

**Department of Communication and Media Studies**

# Design and Development of a Real-Time Bio- Sensing System Assessing Student Mental Workload and Engagement

By

Anthony Psaltis

**Supervisors:** Dr. Konstantinos Mourlas

Dr. Dimitrios Charitos

Dr. Bettina Davou

This thesis is submitted in partial fulfil(l)ment of the requirements for the  
degree of Doctor of Philosophy (Ph.D)

Athens, Greece

January 2018



## **Declaration of Authorship/Copyright Notice**







## **Abstract**

### **Abstract**

Facing the challenge of improving adaptive interaction in educational technologies scientists and educators have turned their focal point in diverse areas ranging from educational, teaching and behavioural psychology to cognitive, affective and perceptual neuroscience. The introduction of digital technologies and interactive media tools in education has shown improved learning efficiency, much higher memory activation and assimilation than verbal teaching, notably due to enhancing motivation achieved by employing approaches attracting student's attention. Excelling aspects of audio visual presentation proved highly valuable particularly in classes with multi ethnic groups of students, as for example consistency between definitions and objects which were verbally and visually defined, eliminating possible misconceptions caused by mishearing or misinterpretation by the learner.

Taking it all one step further as to how an educational system could be even more efficient, a new element would be needed revealing a credible judgment of learning scores and effectiveness of the learning process instantaneously as for example inner levels of activation and satisfaction. In fact, this could be made possible using existing technologies if subconscious neurophysiological responses of a learner could be ascertained and inferred to psycho-somatic conditions as they occur. A system including bio-sensing, data analysis and processing in real time able to provide quantified markers of psychosomatic states of a learner would help enormously in next generations of educational practice. Incorporating data of student engagement and active involvement could help to deduce the interest of a learner, which is known to improve sensitisation in implicit, incidental and also in classical learning. Experimental settings used in previous studies attempting to incorporate physiological responses and interpretations into responsive educational settings have faced major obstacles. Operational issues caused by the requirements of the devices used for the acquisition of physiological signals such as electrodes and movement restrictions have reduced the progress of such settings to laboratory environments. In such settings as described above, the effects of wiring harnesses and sensory components produced an additional psychological burden on the

## **Abstract**

participants. Consequently, the need to approach the physiological data acquisition from a new angle with seamless and unnoticeable operation is apparent. The challenge to design, develop and validate a system that being minimally obstructive and literally unnoticed by the user would uncover combined subconscious expressions of a learner was the primary objective of this research. Physiological data of Heart Rate and Skin Trans-Conductance (Electro-dermal Response) elected as vitally important and highly appropriate to produce the input of data required to evaluate a behavioural concept model. The behavioural assessment model entailed vector classifiers producing directional interpretations of measurements. Directional information (Gradient response) has been derived by comparison of measurements to previously measured values in real time. Assessing the effectiveness and accuracy of the adopted model to deduce attention and engagement of a learner in real time formed the second major objective. For this purpose, a series of relevant experimental methodologies have been employed. Data produced using formal personality assessments have also been investigated in conjunction with those derived from physiological responses in order to identify personality related particularities. The final part of this work has been supplemented by propositions and suggestions with regards to various applications of the system in accomplishment of the initial aims.

## **Acknowledgements or Dedications**

## **Acknowledgements or Dedications**

This paper presents a long time effort building on previous knowledge and studies in new disciplines. It would be disregarding to overlook the time spent in various research and employment throughout my career and particularly my tutors and professors that enabled me to gain the knowledge required to fulfil this undertaking.

Also I couldn't overlook that this work could not have been achieved without the patience shown from the part of my family and therefore I dedicate this writing to my beloved children Nigel, Marina, Perseas and Erato.

## **Acknowledgements**

### **Acknowledgements**

The author would like to thank his supervisors, Dr Konstantinos Mourlas, Dr Dimitrios Charitos and Dr Betina Davou, for their constructive guidance and effort put towards the fulfilment of this undertaking.

The researcher would also like to thank all those participated in the experiments and his colleague Dr Charalambos Rizopoulos, for his assistance in experimental sessions and discussions during the progress of this work.

## List of Abbreviations

### List of Abbreviations

**AAT:** Auditory Attention Test  
**ABL:** Activity Based Learning  
**ANN:** Artificial Neural Networks  
**AMi:** Ambient Intelligence  
**CBA:** Computer Based Assessment  
**CBL:** Computer Based Learning  
**DFA:** Discriminant Function Analysis  
**ECG:** Electro Cardio Graph  
**EEG:** Electro Encephalo Graph  
**EDA:** Electro Dermal Activity  
**EDR:** Electro Dermal Response  
**ERP:** Event related potential  
**fMRI:** functional Magnetic Resonance Imaging  
**FF:** Free Form  
**HR:** Heart Rate  
**HRV:** Heart Rate Variability  
**IAR:** Infrared Absorption Reflectometry  
**IHCI:** Interactive Human Computer Interface  
**NIR:** Near-Infra-Red  
**PR:** Phasic Response  
**PSD:** Power Spectral Distribution  
**RAT:** Receptive Attention Test  
**SC:** Skin Conductance  
**SR:** Skin Resistance  
**STC:** Skin Trans-Conductance  
**SVM:** Support Vector Machines  
**TL:** Tonic Level  
**VAT:** Visual Attention Test



## **Σχεδίαση και Ανάπτυξη Συστήματος Βιο-Αισθητηριακής Ανίχνευσης Νοητικής Προσπάθειας και Εμπλοκής Μαθητών σε Πραγματικό Χρόνο**

### **Περίληψη**

Ο εντοπισμός του επακριβούς επιπέδου προσήλωσης και εμπλοκής των μαθητών με το περιεχόμενο διδασκαλίας στην τάξη είναι ένας από τους πιο μεγαλεπήβολους στόχους των ερευνητών της εκπαιδευτικής και επιστημονικής κοινότητας. (Lang, 1995, Grossberg, 1987).

Σχετικές διεπιστημονικές ερευνητικές προσπάθειες προσαύξησης ενδιαφέροντος και εντοπισμού της αποτελεσματικότητας των διδακτικών πρακτικών βασίζονται σε μελέτες βελτιστοποίησης από τους χώρους της ψυχολογίας, της παιδαγωγικής, της παιδοψυχολογίας και της ψυχοφυσιολογίας. Πρόσφατες προσεγγίσεις, εισάγουν νέες τεχνολογίες χρησιμοποιώντας διαγνωστικές συσκευές δανεισμένες από τον χώρο της ιατρικής με σκοπό να εκμεταλλευτούν τις δυνατότητες μετρήσεων βιολογικών σημάτων. Βιολογικές μετρήσεις όπως αυτές των ρυθμών αναπνοής, καρδιάς και αποκρίσεις του παρασυμπαθητικού νευρικού συστήματος αποτελούν επιβεβαιωμένες εκφράσεις ψυχοφυσιολογικών καταστάσεων οι οποίες εκδηλώνουν καταστάσεις διέγερσης και διάθεσης, (Venables, Christie et al., 1980; Damasio 1994/2006).

Οι ιατρικές συσκευές απαιτούν εργαστηριακό περιβάλλον λόγω των αναγκών χρήσης ηλεκτροδίων, κινητικών περιορισμών, συγχρονισμού και ομοιομορφίας των στοιχείων που προκύπτουν και γι' αυτό τον λόγο δεν μπόρεσαν ποτέ να αποδώσουν μια προσιτή λύση εφαρμόσιμη ευρύτερα σε εκπαιδευτικό περιβάλλον. Στην παρούσα μελέτη, αναλύονται οι επιδόσεις μιας ειδικά κατασκευασμένης ηλεκτρονικής συσκευής, σχεδιασμένης ώστε να εξεταστούν οι δυνατότητες να εξαχθούν δείκτες ψυχοσωματικών εκφράσεων του χρήστη, με την δυνατότητα να χρησιμοποιείται εύχρηστα στην τάξη χωρίς ηλεκτρόδια και επηρεασμούς από προσαρτήσεις. Το ολοκληρωμένο σύστημα μέτρησης και αποτύπωσης συμπερασμάτων είναι βασισμένο σε μοντελοποίηση συμπεριφορών αλλαγής του καρδιακού παλμού και της

## Summary in Greek

ειδικής διηλεκτρικής αγωγιμότητας του δέρματος σε πραγματικό χρόνο. Η συσκευή χρησιμοποιεί οπτικούς και διηλεκτρικούς αισθητήρες επαφής και έχει μελετηθεί σε αντιπαραβολή με διαβαθμισμένα περιβάλλοντα προκλητών καταστάσεων νοητικής φόρτισης. Σειρές πειραματικών διαδικασιών εφαρμοσμένες σε διαβαθμισμένα σενάρια πρόκλησης ψυχοσωματικών διεγέρσεων έχουν ολοκληρωθεί για επικύρωση, μελέτη επιδόσεων και λειτουργία του συστήματος ακόμη και σε σύγκριση με εμπορικό προϊόν. Πειραματικά αποτελέσματα δείχνουν αξιόλογους συσχετισμούς του μοντέλου και των επιδόσεων του συστήματος με τις αναμενόμενες αποκρίσεις με ενθαρρυντικά ποσοστά ακρίβειας.

## Εισαγωγή

Πολυάριθμες μελέτες ανάλυσης των ανθρώπινων συναισθηματικών και συγκινησιακών λειτουργιών από τις επιστήμες της ψυχολογίας (Eysenck, 1982/2006; Frijda 1986; Strongman, 2003; Scherer, 2005), κλινικής φυσιολογίας (Venables, Christie et al., 1980; Damasio 1994/2006) και περιβάλλοντα διεπαφής ανθρώπου-υπολογιστή (Picard et al., 1997; Norman 2004; Brave & Nass 2008), έχουν αποδείξει ότι οι φυσιολογικές εκφράσεις του εγκεφάλου οι οποίες προκαλούνται από συναισθηματικές διεγέρσεις είναι συγκεκριμενοποιημένες, άμεσες, διαγνώσιμες και μετρήσιμες με συσκευές απεικόνισης λειτουργικής μαγνητικής συντονισμένης απήχησης (fMRI), ηλεκτρο-εγκεφαλογράφων, (EEG), τομογράφων εκπομπής ποζιτρονίων (PET) κ.α. Συσκευές μέτρησης δερμικής αγωγιμότητας, καρδιακής συχνότητας και ρυθμού αναπνοής επίσης σχετίζονται άμεσα με τις συγκινησιακές εκφράσεις του εγκεφάλου (Τσάτσου, 2006), με εξίσου υψηλή ακρίβεια όσον αφορά χρονισμό και ένταση, όπως αυτή που προκύπτει από ηλεκτρο-εγκεφαλογράφους και άλλες εργαστηριακές συσκευές υψηλής ακρίβειας.

Ο συνδυασμός δύο ιδιαίτερα άμεσων ως προς την έκφραση βιοσημάτων μελετάται σε αυτό το κείμενο ως προς την χρονική στιγμή εκδήλωσής τους και ως προς το σχετικό τους μέγεθος σε σύγκριση με την αμέσως προηγούμενη κατάσταση. Τα μεγέθη τα οποία εξάγονται από τις συγκρίσεις με τις προηγούμενες τιμές της ειδικής διηλεκτρικής αγωγιμότητας του δέρματος και του ρυθμού των καρδιακών παλμών εκδηλώνουν αυξητική ή μειούμενη τάση των δύο

## Summary in Greek

αυτών ποσοτήτων οι οποίες εξετάζονται σε όλους τους πιθανούς συνδυασμούς τους οποίους συνθέτουν.

Η μελέτη αυτή έχει βασιστεί πρωταρχικά στην σχεδίαση και ανάπτυξη ενός συστήματος με καινοτομίες όσον αφορά στην ευχρηστία και μηδενική παρενόχληση των χρηστών κατά την λήψη των βιοσημάτων που απαιτούνται. Με αυτό τον τρόπο, έχει επιτευχθεί η αμερόληπτη μέτρηση των δυο προαναφερθέντων υποσυνείδητων νευροφυσιολογικών μεγεθών καθώς οι χρήστες των Η/Υ ή κινητών συσκευών αλληλοεπιδρούν χωρίς επηρεασμό από την επίγνωση της διαδικασίας των μετρήσεων όπως συμβαίνει σε παρόμοιες μελέτες οι οποίες χρησιμοποιούν εργαστηριακές συσκευές. Οι μετρήσεις απεικονίζονται σε πραγματικό χρόνο κατά την διάρκεια της λειτουργίας του συστήματος σε μοντελοποιημένη μορφή δύο αξόνων που αναπαριστούν διάθεση και διέγερση. Οι τιμές των δύο αξόνων αναπαρίστανται σαν προβολή σε χώρο κατανομών τεσσάρων επιπέδων νοητικής εμπλοκής (περιλαμβάνοντας γνωστική, συγκινησιακή και συναισθηματική διέγερση) προσομοιάζοντας χωροταξικά το μοντέλο του Russell circumplex model of affect, (Russell 2003), καθώς και εκείνο χρησιμοποιεί παρόμοιο τρόπο κατανομών.

### Στόχοι της μελέτης

Απώτερος στόχος αυτής της ερευνητικής προσπάθειας είναι η βελτιστοποίηση και εξέλιξη εκπαιδευτικών συστημάτων με προσαρμογή του μαθησιακού περιβάλλοντος ανάλογα με την ψυχοφυσιολογική κατάσταση των μαθητών. Πρωτεύων στόχος της μελέτης είναι η εισαγωγή καινοτόμων νέων τεχνολογιών και επικοινωνίας στη μάθηση με σκοπό την μελλοντική ανάπτυξη ευέλικτων εκπαιδευτικών συστημάτων με δυνατότητες αυτο-διαμόρφωσης. Οι τεχνολογίες οι οποίες έχουν εφαρμοστεί βασίζονται σε μετρήσεις υποσυνείδητων βιοαισθητηριακών ποσοτήτων και στην μελέτη απόδοσης δεικτών ψυχοφυσιολογικών εκφράσεων εμπλοκής του μαθητή/χρήστη Η/Υ σε γνωστικό ή μαθησιακό περιεχόμενο.

Η απόδοση μιας έγκυρης απεικόνιση της εμπλοκής ενός χρήστη σε πραγματικό χρόνο δημιουργεί πρωτοποριακές δυνατότητες στην διαμόρφωση προσαρμοζόμενων εκπαιδευτικών περιβαλλόντων με στόχο την βελτιστοποίηση μαθησιακών επιδόσεων ανάλογα με το επίπεδο ενδιαφέροντος του χρήστη είτε σε περιβάλλον τάξης είτε εξ' αποστάσεως. Για τον σκοπό αυτό

## Summary in Greek

ολοκληρώθηκε η από μηδενική βάση σχεδίαση ανάπτυξη και λειτουργική επικύρωση ενός συστήματος το οποίο μπορεί να μετρήσει και να επεξεργαστεί βιο-αναδραστικές εκφράσεις καρδιακού παλμού και ειδικής διηλεκτρικής αγωγιμότητας του δέρματος με στόχο να μπορούν να μελετηθούν αντιδράσεις των μαθητών σε νοητικά ερεθίσματα δια ζώσης. Η υλοποίηση του συστήματος επέτρεψε την δυνατότητα μελέτης υποσυνείδητων εκφράσεων των συμμετεχόντων σε πειραματικές διαδικασίες σε σχέση με επικυρωμένα σενάρια βασισμένα σε εικόνες και βίντεο διαβαθμισμένων γνωστικών, συναισθηματικών και συγκινησιακών φορτίσεων. Η ανεμπόδιστη λειτουργία του συστήματος έχει καταστεί δυνατή για πρώτη φορά σε περιβάλλον εκτός εργαστηρίου. Το σύστημα παρέχει την δυνατότητα ομοιογενούς συλλογής και επεξεργασίας στοιχείων επιτρέποντας την πραγματοποίηση των επιμέρους στόχων οι οποίοι περιλαμβάνουν:

- την δυνατότητα ανάπτυξης ενός συστήματος το οποίο θα έχει την δυνατότητα αμερόληπτης λειτουργίας στην τάξη υπερνικώντας προβλήματα προσαρτήσεων, αρχικών βαθμονομήσεων, συνεχούς ή διακεκομμένης λειτουργίας και το οποίο θα αποδίδει δείκτες νοητικής εμπλοκής για χρήση σε αυτόματα προσαρμοζόμενα και διαμορφούμενα περιβάλλοντα μάθησης
- την διερεύνηση των χαρακτηριστικών των διανυσματικών κατευθύνσεων (Gradients) οι οποίες συμπεραίνονται από τις μετρήσεις των νευρο-φυσιολογικών μεγεθών σε σχέση με προκλητά δυναμικά διαβαθμισμένων νοητικών λειτουργιών,
- την μελέτη εγκυρότητας του αλγορίθμου ο οποίος βασίζεται σε ανίχνευση επιπέδων νοητικής εμπλοκής μέσω συνδυασμών διανυσματικών μετρήσεων φυσιολογικών ποσοτήτων αποδιδόμενοι ως δείκτες διάθεσης και διέγερσης,
- την δημιουργία δεικτών οι οποίοι μπορούν να παράγουν έγκυρες ενδείξεις ψυχοσωματικών καταστάσεων για αυτο-διαμορφούμενα περιβάλλοντα μάθησης,
- την δημιουργία ενός καινοτόμου συστήματος το οποίο θα παρέχει υποσυνείδητες πληροφορίες για το ενδιαφέρον και την εμπλοκή των μαθητών για χρήση στην τάξη, σε εκπαίδευση εξ' αποστάσεως και επίσης σαν εργαλείο έρευνας για περαιτέρω μελέτη ερευνητικών μοντέλων και εφαρμογών στην εκπαίδευση και όχι μόνον.

### Ιστορικό υπόβαθρο

Από τις αρχές του προηγούμενου αιώνα, (Jung, 1924), όταν έγιναν γνωστές οι ηλεκτρικές ιδιότητες του ανθρώπινου σώματος και κατέστη δυνατή η μέτρηση της ηλεκτρικής δερμικής δραστηριότητας δόθηκε το έναυσμα μελετών φυσιολογικών εκφράσεων όπως αυτές εκδηλώνονται από το σωματικό νευρικό σύστημα. Αντιδράσεις του συμπαθητικού νευρικού συστήματος και επαναφορά ομοιοστατικών ισορροπιών από το παρασυμπαθητικό σύστημα μελετήθηκαν εκτενώς σε σχέση με αποκρίσεις σε προκλητά δυναμικά (Tortora & Grabowski, 2006). Συνδυασμοί νευρο-φυσιολογικών ποσοτήτων όπως αυτοί των καρδιακών παλμών και δερμικής αγωγιμότητας διερευνήθηκαν καταλήγοντας σε σημαντικές ενδείξεις ειδικά σε καταστάσεις ζωτικών συναισθηματικών καταστάσεων όπως θυμού, φόβου, χαράς κλπ., (βλέπε Table 1 Σελ. 6 στο Αγγλικό κείμενο). Οι συμβολισμοί (+) και (-) στον πίνακα δείχνουν την κατεύθυνση (ανοδική ή καθοδική) των ποσοτήτων, (Gradients), σε σχέση με την προηγούμενή τους μέτρηση, σε διάταξη Καρδιακός ρυθμός / Ειδική διηλεκτρική αγωγιμότητα.

Ο αλγόριθμος ο οποίος έχει υλοποιηθεί στο λογισμικό της απεικόνισης των συναισθηματικών διεγέρσεων βασίζεται στην ίδια παραδοχή, δηλαδή στην διανυσματική έκφραση η οποία προκύπτει από τις μετρήσεις με αντιστοίχιση του καρδιακού ρυθμού στην οριζόντια διαβάθμιση (θετική ή αρνητική διάθεση), και της ειδικής διηλεκτρικής αγωγιμότητας του δέρματος στην κάθετη διαβάθμιση (δραστηριοποίηση ή χαμηλή ενεργοποίηση). Η απόδοση των προβαλλόμενων ποσοτήτων σε τέσσερα τεταρτημόρια τα οποία σχηματίζονται από τους δύο ως άνω άξονες αναπαριστούν συνδυαστικά απεικόνιση επιπέδων εμπλοκής του χρήστη με το περιεχόμενο ως ακολούθως:

- στο πρώτο τεταρτημόριο (επάνω δεξιά) υποδηλώνουν την κατάσταση ‘υψηλής προσήλωσης’, όπου εμφανίζεται θετική διάθεση και θετική διέγερση
- στο δεύτερο τεταρτημόριο (κάτω δεξιά) αποτυπώνουν την κατάσταση “ικανοποίησης” όπου εμφανίζεται θετική διάθεση αλλά αρνητική διέγερση

## Summary in Greek

- στο τρίτο τεταρτημόριο (κάτω αριστερά) την κατάσταση “δυσκολίας” όπου εμφανίζεται αρνητική διάθεση και αρνητική διέγερση
- στο τέταρτο τεταρτημόριο (επάνω αριστερά) την κατάσταση “αδιαφορίας” όπου εμφανίζεται αρνητική διάθεση αλλά θετική διέγερση.

## Περιγραφή του συστήματος

Για τον σκοπό της μελέτης αυτής σε φυσικό χώρο όπως αυτόν της τάξης στον οποίο η χρήση ηλεκτροδίων και προσαρτήσεων είναι απαγορευτική, σχεδίασα ένα ηλεκτρονικό κύκλωμα το οποίο μετρά τον καρδιακό παλμό με αισθητήρες ανάκλασης δέσμης στην περιοχή συχνότητας κοντινών υπέρυθρων (Near-Infra-Red-Sensors) και ένα κύκλωμα το οποίο μετρά την ειδική αγωγιμότητα του δέρματος με την τεχνική της βιο-αναδραστικής διηλεκτρικής αγωγιμότητας (Trans-Conductance Biofeedback). Τα κυκλώματα μέτρησης καρδιακού ρυθμού έχουν υλοποιηθεί εις διπλούν ώστε να αποφευχθούν είδωλα και ανακρίβειες λόγω της ευαισθησίας του σήματος το οποίο μετράει τις αλλαγές στην απόχρωση του δέρματος σε κάθε καρδιακό παλμό. Η διηλεκτρική αγωγιμότητα του δέρματος αναπαριστά ως ποσότητα την έκφραση του ψυχοφυσιολογικού στρες όπως αυτό εκδηλώνεται από την απόκριση των νευρονικών δικτύων του εγκεφάλου ως συνέπεια εξωγενών συμβάντων. Κατά την διάρκεια της δραστηριοποίησης ή και ενεργοποίησης αντιδράσεων του εγκεφάλου σε ζωτικά ερεθίσματα μέσω ορμονικών, νευρικών ή άλλων μηχανισμών, έχουν διαπιστωθεί μεταβολές στις διαμέτρους των υδατικών πόρων του δέρματος. Αυτό έχει σαν αποτέλεσμα την ποσοτική αλλαγή της εφίδρωσης του δέρματος σε σχέση ευθέως ανάλογη με τις αντιδράσεις σε νοητικά και συναισθηματικά γεγονότα. Η αλλαγή των χαρακτηριστικών περιεκτικότητας ηλεκτρολυτών στο δέρμα είναι ανιχνεύσιμη και μετρήσιμη μέσω της ποσότητας της διηλεκτρικής αγωγιμότητας του δέρματος. Το σύστημα περιλαμβάνει αισθητήρες οι οποίοι μπορούν να προσαρμοστούν στο ποντίκι ενός Η/Υ ή σε οποιαδήποτε θήκη κινητής συσκευής με μόνη απαίτηση από την πλευρά του χρήστη να διατηρεί την επαφή των δακτύλων του με τους δύο αισθητήρες. Τα δύο ηλεκτρονικά κυκλώματα, ένα αυτό που αποδίδει την κυματομορφή του καρδιακού παλμού και ένα το μέγεθος της διηλεκτρικής αγωγιμότητας ενσωματωμένα στο ποντίκι, έχουν σαν έξοδο δύο προ-διαμορφωμένα αναλογικά σήματα τα οποία επικοινωνούν με τον Η/Υ μέσω της θύρας

## Summary in Greek

USB, ανιχνευόμενα ως δύο περιφερειακές συσκευές, ένα ποντίκι και μία κάρτα ήχου. Λογισμικό το οποίο εκτελείται στον Η/Υ αναλαμβάνει την περαιτέρω επεξεργασία των δύο σημάτων ανιχνεύοντας συχνότητα καρδιακών παλμών από το πρώτο και συχνότητα απόδοσης επιπέδου διηλεκτρικής αγωγιμότητας του δέρματος από το δεύτερο κανάλι της ενσωματωμένης κάρτας ήχου. Το λογισμικό με συνεχή και αδιάλειπτη λειτουργία ανάγνωσης των σημάτων εκτελεί έναν αλγόριθμο έμμεσης απεικόνισης των ψυχοσωματικών καταστάσεων με ανάλυση σε δύο διαβαθμισμένους άξονες, τον άξονα της διάθεσης (οριζόντια διαβάθμιση) και τον άξονα διέγερσης (κάθετη διαβάθμιση). Η επεξεργασία των μετρήσεων του καρδιακού παλμού αποτυπώνονται στον άξονα της διάθεσης και της διηλεκτρικής αγωγιμότητας στον άξονα της διέγερσης αντίστοιχα. Η ειδικά κατασκευασμένη συσκευή παρέχει διαπιστωμένες αποκρίσεις των φυσιολογικών ποσοτήτων που προαναφέρθηκαν, με μέγιστα πλεονεκτήματα:

- την αμερόληπτη μέτρηση των δύο ποσοτήτων χωρίς ο χρήστης να επηρεάζεται από την αντίληψη πως γίνονται μετρήσεις όπως θα ήταν στην περίπτωση που χρησιμοποιούνται ηλεκτρόδια και προσαρτήσεις,
- την λειτουργία χωρίς την ανάγκη για αρχική περίοδο βαθμονόμησης από την πλευρά του χρήστη και ανοχή σε διακοπτόμενη λειτουργία χωρίς την ανάγκη αρχικοποίησης,
- την αδιάλειπτη λειτουργία από διπλά κυκλώματα ανάγνωσης του καρδιακού παλμού και αξιόπιστη λειτουργία μέτρησης της διηλεκτρικής αγωγιμότητας για μεγάλα χρονικά διαστήματα,
- την ομοιομορφία διαβάθμισης και τον απόλυτο συγχρονισμό των στοιχείων.

## Πειράματα

Τέσσερα στάδια πειραματικών διαδικασιών οργανώθηκαν για την επικύρωση της λειτουργίας και την μελέτη των πειραματικών επιδόσεων. Κατά σειρά εκτέλεσης περιλάμβαναν την εργαστηριακή επικύρωση της σωστής λειτουργίας του συστήματος, την πειραματική διερεύνηση της λειτουργικότητας και ευχρηστίας του συστήματος, την πειραματική μελέτη των επιδόσεων του συστήματος με χρήση επίσημα βαθμονομημένων σεναρίων και την

## Summary in Greek

πειραματική διαδικασία αντιπαραβολής του συστήματος με εμπορικά διαθέσιμο προϊόν διερευνήσεων παρόμοιων συμπερασμάτων.

Το πρώτο στάδιο περιλάμβανε έλεγχο λειτουργικότητας και ακρίβειας των μετρήσεων βασισμένο σε εργαστηριακούς ελέγχους των ηλεκτρονικών κυκλωμάτων. Επίσης μια ομάδα εθελοντών 8 ατόμων έλαβε μέρος σε μια σειρά ειδικά σχεδιασμένων σεναρίων για την επιβεβαίωση των μετρήσεων, την βελτιστοποίηση των χρονικών διαστημάτων κατά την διάρκεια απόκτησης των στοιχείων και άλλων εσωτερικών βαθμονομήσεων του συστήματος. Οι μελέτες λειτουργικής επικύρωσης απέδωσαν ικανοποιητικά αποτελέσματα όσον αφορά την ακρίβεια και την χρηστικότητα του συστήματος, ενώ βοήθησαν στην βελτιστοποίηση των χρονικών διαστημάτων μέτρησης και την επιβεβαίωση της σωστής λειτουργίας των γραφημάτων και καταγραφής των στοιχείων σε πραγματικό χρόνο.

Το δεύτερο στάδιο περιλάμβανε δύο πειραματικές διαδικασίες. Η πρώτη πειραματική διαδικασία περιλάμβανε πειράματα με σενάρια ενασχόλησης των συμμετεχόντων σε εργασίες με υψηλές γνωστικές απαιτήσεις μαθηματικού λογισμού. Στόχος ήταν να καταμετρηθούν μετρήσεις φυσιολογικών αποκρίσεων κατά διαστήματα νοητικής προσπάθειας σε τέσσερα διακριτά στάδια, i) νοητικής προσπάθειας, κατά την διάρκεια επεξεργασίας και εύρεσης των ενεργειών οι οποίες απαιτούντο από το περιβάλλον, ii) ικανοποίησης κατά την διάρκεια της επιτυχούς ολοκλήρωσης ενός επιπέδου, iii) αμηχανίας κατά την διερεύνηση για την εύρεση του επόμενου βήματος και iv) απογοήτευσης στην περίπτωση αποτυχίας. Η δεύτερη πειραματική διαδικασία περιλάμβανε την παρακολούθηση προβολών βίντεο τα οποία είχαν επιλεγεί από συγκεκριμένες βιβλιοθήκες ώστε να προκαλούν έντονες συναισθηματικές καταστάσεις οι οποίες αντιστοιχούν σε ερεθίσματα κατανεμημένα από μοντέλα απεικόνισης συναισθηματικών καταστάσεων όπως αυτό του Russell circumplex model of affect. Η επιλογή αυτή κρίθηκε κατάλληλη και χρησιμοποιήθηκε ως μοντέλο αντιστοίχισης καθώς το προαναφερθέν μοντέλο βασίζεται επίσης στην αποτύπωση της διάθεσης και της διέγερσης σε δύο άξονες ταυτόσημα με το εν λόγω σύστημα. Στόχος των πειραμάτων ήταν η επιβεβαίωση της λειτουργίας του αλγορίθμου βασισμένου στην υπόθεση ότι οι τάσεις που εκδηλώνονται από κοινές κατευθύνσεις των μετρήσεων εκφράζουν καταστάσεις διέγερσης (π.χ. ανοδική

## Summary in Greek

πορεία καρδιακού παλμού και ταυτόχρονα ανοδική πορεία φυσιολογικού στρες υποδηλώνουν υψηλά επίπεδα συγκέντρωσης και εμπλοκής του συμμετέχοντος). Τα πειράματα περιλάμβαναν 15 συμμετέχοντες, (6 άνδρες, 9 γυναίκες), με ΜΟΗ: 27 έτη και διάρκεια 20 λεπτά. Η εκτίμηση των αποτελεσμάτων έγινε με χρήση της μεθοδολογίας GOMS (Goals, Operators, Methods, Selection Rules) για την βαθμονόμηση συμβάντων/αποκρίσεων. Οι ενδείξεις από τα αποτελέσματα των πειραμάτων έδειξαν υψηλό βαθμό επιτυχών μετρήσεων με μέσο όρο 91,35% μετρήσεων χωρίς σφάλμα ανάγνωσης. Αναλυτικότερα τα αποτελέσματα έδωσαν κατά μέσο όρο 71% στην κατανομή “Ενεργής Εμπλοκής”, 9% στην κατανομή “Ικανοποίησης”, 13% στην κατανομή “Αντιλαμβανόμενης Δυσκολίας” και 7% στην κατάσταση “Αδιαφορίας” κατά την συνολική διάρκεια των πειραμάτων. Οι διαφορές μεταξύ γυναικών και ανδρών δεν βρέθηκε να είναι στατιστικά σημαντική όπως επίσης οι διαφορές μεταξύ πειραμάτων γνωστικής φόρτισης και βίντεο. Το αποτέλεσμα αυτό έδειξε πως ο αλγόριθμος κατανομής επιπέδων εμπλοκής στο περιεχόμενο έχει βαθμό ακρίβειας ικανό προς περαιτέρω διερεύνηση.

Το τρίτο στάδιο πειραματικής διακρίβωσης ακρίβειας του συστήματος περιλάμβανε πιο ισχυρά καθορισμένα προκλητά ερεθίσματα. Το σενάριο στο οποίο είναι βασισμένο το πείραμα διακρίβωσης της αποτελεσματικότητας του συστήματος χρησιμοποιεί ομάδες εικόνων διαβαθμισμένης συναισθηματικής διέγερσης με ποσοτικά χαρακτηριστικά διάθεσης και διέγερσης όπως έχουν αποτυπωθεί από μετρήσεις οι οποίες περιλαμβάνουν σχεδόν όλες τις δυνατές φυσιολογικές ποσότητες του ανθρώπινου σώματος. Οι διαβαθμίσεις έχουν γίνει με εργαστηριακές μετρήσεις functional Magnetic Resonance Imaging (fMRI), Electro Encephalo Gram (EEG), Electro Myo Gram (EMG), SC (Skin Conductance), ακουστικές αποκρίσεις και μια πληθώρα ακόμη νευρο-φυσιολογικών ποσοτήτων (Bradley & Lang, 2007), από το Πανεπιστήμιο της Φλόριντα, γνωστό ως International Affective Picture System (IAPS). Η σειρά πειραμάτων συνοδευόταν από μετρήσεις κατά την διάρκεια παρακολούθησης βίντεο τα οποία επίσης είχαν σκοπό να προκαλέσουν έντονες συναισθηματικές φορτίσεις αναδεικνύοντας τις διαφορές μεταξύ εκδήλωσης συναισθημάτων προκαλούμενες από στατικές και αυτές από συναισθήματα προκαλούμενα από αλληλουχίες οπτικών ερεθισμάτων (Gross & Levenson 1995). Ψυχομετρικές αναλύσεις ερωτηματολογίων συμπληρώθηκαν από τους συμμετέχοντες πριν από το πρακτικό μέρος των πειραμάτων. Τα ερωτηματολόγια

## Summary in Greek

συναισθηματικής διαχείρισης (NEO-FFI emotion regulation) περιλάμβαναν μεθοδευμένες πρακτικές διάγνωσης συμπεριφοριστικών (personality) και γνωστικών (cognitive) προφίλ όπως επίσης και ιδιοσυγκρασίας (trait anxiety), (Δημητριάδου & Σταλίκας, 2012).

Η επιλογή των ομάδων των εικόνων έγινε ώστε να υπάρχει ποσοτικά υψηλή διάθεση και διέγερση στον πίνακα αντιστοίχισης (IAPS), ώστε να αποφευχθούν πιθανές αμφιβολίες στις τοποθετήσεις των ενδείξεων στην απεικόνιση του μοντέλου. Επίσης κριτήρια ενδιαφέροντος συμπεριλήφθηκαν στην επιλογή των εικόνων ώστε να υπάρχει συνοχή συναισθημάτων σύμφωνα με την Ελληνική κουλτούρα, π.χ. μια εικόνα που παρουσιάζει ένα μπισκότο δεν προκαλεί χαρά σε κάποιον μαθητή Ελληνικής καταγωγής όσο θα προκαλούσε σε έναν Αμερικανικής καταγωγής. Το πρακτικό μέρος των πειραμάτων περιλάμβανε δύο μέρη.

Στο πρώτο μέρος των πειραμάτων του τρίτου σταδίου υπήρχε η παρουσίαση 64 εικόνων επιλεγμένες διαφορετικά για άνδρες και διαφορετικά για γυναίκες, προβαλλόμενες για χρονικό διάστημα 6 δευτερολέπτων, με ενδιάμεσες ουδέτερες εικόνες οι οποίες προβάλλονταν για 5 δευτερόλεπτα. Ο λόγος για την τοποθέτηση ενδιάμεσων εικόνων χωρίς περιεχόμενο ήταν η δημιουργία εκκενωτικών εντυπώσεων οι οποίες προκαλούν απελευθέρωση της μνήμης από την προηγούμενη εικόνα ώστε να υπάρχει πιο καθαρή αποτύπωση της μετάβασης των συναισθημάτων η οποία προκαλείται από το επόμενο ερέθισμα. Οι εικόνες ήταν ταξινομημένες σε τετράδες οι οποίες είχαν επιλεγεί ώστε να προκαλούν συναισθηματικές καταστάσεις οι οποίες ανήκουν στο ίδιο τεταρτημόριο του μοντέλου. Οι συνδυασμοί παρουσίασης των εικόνων ακολουθούσαν συγκεκριμένη σειρά δειγμάτων απεικόνισης εξαντλώντας όλους τους πιθανούς συνδυασμούς περιπλοκότητας, ενώ υπήρξε μέριμνα ώστε οι μισοί από τους συμμετέχοντες να παρακολουθήσουν τους συνδυασμούς με αντίστροφη σειρά.

Στο δεύτερο μέρος του ίδιου πειράματος περιλαμβανόταν η προβολή δίλεπτων βίντεο με ενδιάμεσες περιόδους παραγραφής προηγούμενων εντυπώσεων κατανεμημένων από άποψης περιεχομένου σε αντιστοιχία με τα τεταρτημόρια. Σενάρια που προκαλούσαν συναισθήματα χαράς αντιστοιχούσαν στο πρώτο τεταρτημόριο, ικανοποίησης στο δεύτερο και ούτω καθεξής. Η διαδικασία των πειραμάτων περιλάμβανε εγγραφή των προβολών στην οθόνη με ζωντανή

## Summary in Greek

απεικόνιση του μοντέλου και των μετρήσεων, καθώς επίσης και καταγραφή των στοιχείων με χρονική επισήμανση σε αρχείο για κάθε συμμετέχοντα. Είκοσι επτά συμμετέχοντες πήραν μέρος στα πειράματα, (11 άνδρες και 16 γυναίκες) ηλικίας από 18 έως 50 ετών με μέσο όρο ηλικίας 28.5 και τυπική απόκλιση 9.47. Δεν απαιτήθηκε προπαρασκευαστική διαδικασία εφόσον το σύστημα δεν χρειάζεται αρχική βαθμονόμηση αλλά είναι κατασκευασμένο έτσι ώστε να δημιουργεί αυτόματα την βαθμονόμηση που απαιτείται (auto-calibration), κυρίως στην μέτρηση της ειδικής αγωγιμότητας (Skin Trans-Conductance). Τα ποσοστά επιτυχών μετρήσεων όσον αφορά τις τοποθετήσεις στον σωστό χώρο συναισθηματικών καταστάσεων από το σύστημα στο μοντέλο απεικόνισης απέδωσαν κατά μέσο όρο για τις εικόνες: 75% στην κατανομή *”Ενεργής Εμπλοκής”*, 37% στην κατανομή *”Ικανοποίησης”*, 41% στην κατανομή *”Αντιλαμβανόμενης Δυσκολίας”* και 22% στην κατάσταση *”Αδιαφορίας”* και για το βίντεο: 66% στην κατανομή *”Ενεργής Εμπλοκής”*, 21% στην κατανομή *”Ικανοποίησης”*, 27% στην κατανομή *”Αντιλαμβανόμενης Δυσκολίας”* και 14% στην κατάσταση *”Αδιαφορίας”*.

Τα αποτελέσματα των πειραμάτων εκδηλώνουν υψηλότερα ποσοστά ακρίβειας του συστήματος στις περιοχές απεικόνισης με τις υψηλότερα διακριτές τιμές μετρήσεων όπως αυτές με ταυτόχρονα θετικές τάσεις (δηλαδή θετική διάθεση και θετική διέγερση), 66-75%, ή ταυτόχρονα αρνητικές τάσεις (δηλαδή αρνητική διάθεση και αρνητική διέγερση) 27-41%. Οι απεικονίσεις στις άλλες δύο περιοχές οι οποίες συντίθενται από διαφέρουσες τάσεις εμφανίζουν μικρότερη ακρίβεια. Εξετάζοντας τις συνολικές απεικονίσεις του συστήματος, εκτός από την περίπτωση της συνεδρίας των βίντεο και μόνο για το τρίτο τεταρτημόριο στο οποίο τα ποσοστά ακρίβειας ήταν πολύ πιο υψηλά για τους άνδρες από αυτά των γυναικών, (ανάλυση ανεξάρτητων δειγμάτων t-test:  $t=2.143$ ,  $p=0.049$ ), δεν βρέθηκαν διαφορές στην ακρίβεια μέτρησης με στατιστική σημαντικότητα μεταξύ των δύο φύλων.

Τα αποτελέσματα των ψυχομετρικών στοιχείων μετά από την δέουσα επεξεργασία τέθηκαν σε σύγκριση με αυτά του συστήματος. Τα αποτελέσματα της βαθμολόγησης σε σχέση με τους πέντε παράγοντες προσωπικότητας όπως ορίζονται από το NEO-FFI και η κατηγοριοποίηση αυτών των βαθμολογιών (*“very low”, “average”, “high”, and “very high”*) παρουσιάζονται στον Πίνακα 1 που ακολουθεί.

## Summary in Greek

*Πίνακας 1: Σύγκριση ακρίβειας μετρήσεων μεταξύ φύλων*

| Τεταρτημόριο    | Άνδρες    |                 | Γυναίκες  |                 |
|-----------------|-----------|-----------------|-----------|-----------------|
|                 | Μέση Τιμή | Τυπική απόκλιση | Μέση τιμή | Τυπική απόκλιση |
| Q <sub>1P</sub> | 38.18     | 4.285           | 34.81     | 4.956           |
| Q <sub>2P</sub> | 20.36     | 2.024           | 16.13     | 1.516           |
| Q <sub>3P</sub> | 22.09     | 4.969           | 18.06     | 5.836           |
| Q <sub>4P</sub> | 12.18     | 4.875           | 9.13      | 3.631           |
| Q <sub>1V</sub> | 67.55     | 10.093          | 66.25     | 15.597          |
| Q <sub>2V</sub> | 24.09     | 10.425          | 19.75     | 7.523           |
| Q <sub>3V</sub> | 34.82     | 18.236          | 21.69     | 10.806          |
| Q <sub>4V</sub> | 18.00     | 12.008          | 11.31     | 9.965           |

Σε σχέση με τους παράγοντες προσωπικότητας σαν αριθμητικές τιμές, δεν βρέθηκαν στατιστικά σημαντικές διαφορές μεταξύ φύλων, αλλά βρέθηκαν σημαντικές συσχετίσεις μεταξύ παραγόντων προσωπικότητας και φύλου στις περιπτώσεις “Εξωστρέφειας” (Extraversion) και “Ευσυνειδησίας” (Conscientiousness), (Fisher’s Exact Test,  $p=0.023$  και  $p=0.047$  αντίστοιχα) Πίνακας 2.

*Πίνακας 2: Μέσες τιμές και τυπικές αποκλίσεις παραγόντων προσωπικότητας μεταξύ φύλων*

| Παράγοντας προσωπικότητας | Άνδρες    |                 | Γυναίκες  |                 | Σύνολο    |                 |
|---------------------------|-----------|-----------------|-----------|-----------------|-----------|-----------------|
|                           | Μέση τιμή | Τυπική απόκλιση | Μέση Τιμή | Τυπική απόκλιση | Μέση τιμή | Τυπική απόκλιση |
| N                         | 20.00     | 8.44            | 25.06     | 8.23            | 23.00     | 8.54            |
| E                         | 35.27     | 8.27            | 30.50     | 4.37            | 32.44     | 6.56            |
| O                         | 27.73     | 7.24            | 32.13     | 6.73            | 30.33     | 7.15            |
| A                         | 29.00     | 7.25            | 29.69     | 3.77            | 29.41     | 5.34            |
| C                         | 31.55     | 10.55           | 33.63     | 4.90            | 32.78     | 7.59            |

Σχετικά με τα ψυχομετρικά δείγματα ιδιοσυγκρασίας (trait anxiety) οι γυναίκες (μέση τιμή = 44.19, τυπική απόκλιση = 10.73) βαθμολογήθηκαν ψηλότερα από τους άντρες (μέση τιμή = 35.18, τυπική απόκλιση = 9.622). Δοκιμή ανεξάρτητων δειγμάτων (independent t-test) έδειξε

## Summary in Greek

πως η διαφορά στην βαθμολογία του παράγοντα ιδιοσυγκρασίας μεταξύ φύλων είναι στατιστικά σημαντική ( $t=2.277$ ,  $p=0.032$ ). Η μέση τιμή για όλα τα δείγματα βρέθηκε 40.52 και η τυπική απόκλιση βρέθηκε 11.061. Όπως αναμενόταν, υψηλές επιδόσεις σε ιδιοσυγκρασία (trait anxiety) συσχετίζονται αρκετά υψηλά με αυτές του Νευρωτισμού (Neuroticism) ( $r = 0.647$ ,  $p = 0.001$ ), ενώ δεν βρέθηκαν συσχετισμοί μεταξύ ιδιοσυγκρασίας και των άλλων τεσσάρων παραγόντων προσωπικότητας.

### Ανάλυση αποτελεσμάτων

Η στατιστική επεξεργασία των αποτελεσμάτων έδειξε χαμηλούς έως μέσους συσχετισμούς μεταξύ δύο παραγόντων προσωπικότητας, Νευρωτισμού και Δεκτικότητας στην Εμπειρία, (Neuroticism and Openness to Experience), καθώς επίσης και ακρίβειας στις αποκρίσεις συμμετεχόντων σε κάποια από τα βίντεο. Ειδικότερα:

Δεκτικότητα στην εμπειρία βρέθηκε να σχετίζεται με την απόκριση στο τεταρτημόριο  $Q_{2V}$  ( $r=0.384$ ,  $p=0.048$ ) και  $Q_{4V}$  ( $r=0.493$ ,  $p=0.009$ ).

Νευρωτισμός, συσχετίζεται με αποκρίσεις στο  $Q_{4V}$  ( $r=0.426$ ,  $p=0.027$ ).

Κάποιες υποκλίμακες βρέθηκαν να συσχετίζονται με την ακρίβεια των μετρήσεων των συμμετεχόντων όπως:

- Τάση υποβιβασμού (Self-reproach), υποκλίμακα του Νευρωτισμού, σχετίζεται με τις μετρήσεις στο τεταρτημόριο εικόνων  $Q_{4P}$  ( $r=-0.383$ ,  $p=0.024$ ).
- Αισθητικά ενδιαφέροντα (Aesthetic Interests), υποκλίμακα της δεκτικότητας στην εμπειρία, σχετίζεται με  $Q_{1P}$  ( $r=-0.391$ ,  $p=0.022$ ) και  $Q_{4V}$  ( $r=-0.47$ ,  $p=0.006$ ).
- Διανοητικά Ενδιαφέροντα (Intellectual interests), υποκλίμακα της δεκτικότητας στην εμπειρία δείχνουν σημαντικότητα στα τεταρτημόρια  $Q_{1V}$  ( $r=-0.44$ ,  $p=0.011$ ),  $Q_{2V}$  ( $r=-0.338$ ,  $p=0.043$ ) και  $Q_{4V}$  ( $r=-0.349$ ,  $p=0.037$ ).
- Αξιοπιστία (Dependability), υποκλίμακα της ευσυνειδησίας (Conscientiousness) δείχνουν σημαντικότητα στο τεταρτημόριο  $Q_{2P}$  ( $r=-0.404$ ,  $p=0.018$ ).

## Summary in Greek

Όταν χρησιμοποιήθηκε δομή τριών επιπέδων παραγόντων προσωπικότητας (“κάτω του μέσου”, “μέσου” και “άνω του μέσου”), ο Νευρωτισμός βρέθηκε σαν παράγοντας επηρεασμού στο  $Q_{3P}$ :  $F(2.28) = 3.463$ ,  $p = 0.048$ . Περαιτέρω επεξεργασία πολλαπλών συγκρίσεων (Tukey HSD) αποκαλύπτουν πως η διαφορά μεταξύ επιπέδων Νευρωτισμού μέσου και κάτω του μέσου είναι σημαντική ( $p = 0.037$ ). Επιπλέον, η Δεκτικότητα στην εμπειρία εμφανίστηκε να επηρεάζει αποκρίσεις στο τεταρτημόριο  $Q_{4V}$  (Kruskal Wallis  $H(2) = 7.246$ ,  $p = 0.022$ , 95% CI[0.019, 0.025]). Συγκρίσεις ανα ζεύγος δεν ανέδειξαν σημαντικότητα, αλλά συμμετέχοντες με άνω του μέσου επιπέδου δεκτικότητας στην εμπειρία είχαν χαμηλότερη ακρίβεια στο τεταρτημόριο  $Q_{4V}$  (MdBelow = 0.24, MdAvg = 0.13, MdAbove = 0.06).

Η βαθμολογία Ιδιοσυγκρασίας των συμμετεχόντων (Trait Anxiety), συσχετίζεται σημαντικά μόνο με τις αποκρίσεις στο τεταρτημόριο  $Q_{4V}$  ( $r = -0.460$ ,  $p = 0.016$ ).

Μόνο η ακρίβεια των αποκρίσεων για την τέταρτη ομάδα εικόνων συσχετίζεται σημαντικά με τις αποκρίσεις στο τεταρτημόριο  $Q_{4V}$  ( $r = -0.460$ ,  $p = 0.016$ ).

Μόνο η ακρίβεια των αποκρίσεων για την τέταρτη ομάδα εικόνων συσχετίζεται σημαντικά με την Διαχείριση Συναισθημάτων, ( $r = 0.396$ ,  $p = 0.041$ ) και δύο υποκλίμακες, Συναισθηματικές εμπειρίες ( $r = 0.428$ ,  $p = 0.026$ ) και Συναισθηματικές εμπειρίες μαζί με Συναισθηματικές Εκφράσεις ( $r = 0.44$ ,  $p = 0.022$ ).

Το τέταρτο στάδιο πειραματικών μελετών περιλαμβάνει επιδόσεις του συστήματος σε συνδυασμό και σύγκριση με το σύστημα διάγνωσης ψυχοσωματικών καταστάσεων Noldus το οποίο βασίζεται σε συμπερασματική ανάλυση των εκφράσεων του προσώπου. Σε αυτό το στάδιο επιλέχθηκε η εφαρμογή μελέτης σε δύο χαρακτηριστικά πεδία στο χώρο εξέλιξης παιχνιδιών. Οι δύο βασικοί τύποι περιβαλλόντων που χαρακτηρίζουν τα παιχνίδια εν γένει διακρίνονται σε ανοικτής δομής (Free Structured) - όπου δεν υπάρχουν χρονικοί περιορισμοί και πιέσεις – και τυπικής δομής (formally structured) - όπου συγκεκριμένοι στόχοι επιτυγχάνονται μόνο μέσα σε συγκεκριμένα χρονικά περιθώρια και κάτω από πιέσεις καταδίωξης και αποφυγής εμποδίων.

## Summary in Greek

Ο αριθμός των συμμετεχόντων αποτελείτο από 25 μαθητές, (18 αγόρια και 7 κορίτσια), ηλικίας μεταξύ 9 και 12 ετών με μέσο-όρο ηλικίας 10.77 και τυπική απόκλιση 0.74. Συνολικά ως προς ολόκληρο το πείραμα, τα αποτελέσματα έδειξαν ότι τα δύο συστήματα μπόρεσαν να λειτουργήσουν απροβλημάτιστα συνυπάρχοντας στον ίδιο υπολογιστή και να αποδεικνύουν με τελείως διαφορετική αρχή λειτουργίας πανομοιότυπα αποτελέσματα σε υψηλό βαθμό σύγκλισης. Συγκεκριμένα οι αποδόσεις των ψυχοσωματικών καταστάσεων σε σχέση με το μοντέλο κατανομής του Russell, βρέθηκε να αντικατοπτρίζεται και από τα δύο συστήματα με υψηλή συμπτωματικότητα (63.48 – 100.00%, μέση τιμή 89.60% και τυπική απόκλιση 9.63%). Το σύστημα HR-STC απέδιδε την προβολή επιπέδων προσήλωσης σε πραγματικό χρόνο ενώ το Noldus απέδιδε συναισθηματική επίδοση η οποία βρέθηκε να αντιστοιχεί στις ίδιες περιοχές απεικόνισης. Επίσης και τα δύο συστήματα απέδιδαν την κατεύθυνση (Gradient) της διάστασης της διέγερσης η οποία φάνηκε να συμπίπτει σε υψηλά ποσοστά, (Εικόνα 7.4.1 στο Αγγλικό κείμενο)

Περαιτέρω στατιστική επεξεργασία των αποτελεσμάτων με την μέθοδο Spearman's απέδωσε υψηλές θετικές συσχετίσεις μεταξύ των μεταβλητών διάθεσης και διέγερσης στο παιχνίδι ανοικτής δομής ( $r=0.769$ ,  $p=0.05$ ) και ( $r=0.758$ ,  $p=0.05$ ) για το τυπικά δομημένο παιχνίδι. Αυτό επιβεβαιώνει πως αυξομειώσεις στα επίπεδα διάθεσης σχετίζονται με αυξομειώσεις σε επίπεδα διέγερσης και στα δύο περιβάλλοντα. Επιπλέον αναλύσεις των στοιχείων από ερωτηματολόγια που αφορούσαν φύλο, ηλικία, συχνότητα ενασχόλησης με παιχνίδια και προηγούμενη εμπειρία, έδειξαν πως υπήρχε σημαντική διαφορά μόνο για την μεταβλητή προηγούμενης εμπειρίας. Αυτό παρουσιάζει ένδειξη πως έμπειροι παίκτες εκδηλώνουν υψηλότερα επίπεδα ικανοποίησης σε παιχνίδια ελεύθερων σεναρίων και ανοικτής δομής από τους αρχάριους παίκτες. Ανάλυση Wilcoxon έδειξε πως τα παιχνίδια είχαν μεγάλη επιρροή στις νευροφυσιολογικές μετρούμενες ποσότητες και ιδιαίτερα στην μέτρηση του STC ( $Z = -2.127$ ,  $p=0.033$ ). Οι συμμετέχοντες, κατά την διάρκεια του παιχνιδιού κλειστής δομής (Subway Surfers) εκδήλωσαν σημαντικά μεγαλύτερα επίπεδα φυσιολογικού στρες από αυτά κατά την διάρκεια ενασχόλησής τους στο παιχνίδι ελεύθερης δομής (Minecraft).

## Summary in Greek

### Παρατηρήσεις και Συμπεράσματα

Το σύστημα που σχεδιάστηκε και αναπτύχθηκε για να καταστεί δυνατή η μελέτη ανίχνευσης δεικτών επιπέδου εμπλοκής μαθητών και γενικότερα χρηστών Η/Υ ή άλλων συσκευών βασισμένο σε μετρήσεις καθαρών νευροφυσιολογικών μεγεθών σε πραγματικό χρόνο έχει εκπληρώσει τις αρχικές απαιτήσεις λειτουργικότητας και ευχρηστίας. Αξίζει να σημειωθεί ότι δεν έχουν υλοποιηθεί αλγόριθμοι βελτιστοποίησης στο λογισμικό, αλλά τα αποτελέσματα αποδίδονται σε καθαρές μετρήσεις στιγμιαίων ψυχοφυσιολογικών εκφράσεων με σκοπό να μελετηθούν καθαρά οι δυνατότητες του συστήματος.

Συμπίπτουσες εξάρσεις καρδιακών ρυθμών και επιπέδων ψυχολογικού στρες παρατηρήθηκαν σε περιβάλλοντα νοητικών φορτίσεων και συναισθηματικών διεγέρσεων και αξιολογήθηκαν όσον αφορά τις δυνατότητες απόδοσης δεικτών εμπλοκής στο περιεχόμενο. Τα αποτελέσματα δείχνουν ότι το σύστημα αποδίδει με μεγαλύτερη ακρίβεια σωστές αποκρίσεις όσο οι χρήστες βρίσκονται σε ψηλότερα επίπεδα διέγερσης τα οποία προκαλούνται από μεγαλύτερη συγκέντρωση ή προσήλωση στο περιεχόμενο (Άνδρες:75%, Γυναίκες:66%). Η ακρίβεια των αποκρίσεων αποδεικνύεται μικρότερη σε καταστάσεις χαμηλότερης εμβάθυνσης (κάτω από 50%). Αυτό αποδίδεται στο ότι οι συνδυασμοί και ο μεγαλύτερος αριθμός συναισθημάτων τα οποία αναπτύσσονται σε αυτές τις περιοχές αποδίδουν λιγότερο έντονες και περισσότερο καθυστερημένες μεταβάσεις καθιστώντας την διαπίστωση των σωστών αποτυπώσεων πολύ δυσκολότερη χωρίς την χρήση αλγορίθμων βελτιστοποίησης.

Η ακρίβεια του συστήματος στην ανίχνευση εμβάθυνσης ουσιαστικά βασίζεται σε μέτρηση και άθροιση χαρακτηριστικών διάθεσης και διέγερσης στιγμιαίων συναισθηματικών καταστάσεων. Εφόσον ενέχεται η περίπτωση διαφορετικά συναισθήματα να εκφράζονται με παρόμοια χαρακτηριστικά παρεμφερούς έντασης, γίνεται εμφανής η ανάγκη χρησιμοποίησης αλγορίθμων οι οποίοι θα προβλέπουν μεταφορά από μια συναισθηματική κατάσταση στην επόμενη ανάλογα με χρόνο και τρόπο έκφρασης από άποψη φυσιολογίας. Συμπληρωματικά εργαλεία όπως η ενσωμάτωση συστήματος μετρήσεων αλλαγής του μεγέθους της ίριδας της κόρης του ματιού, ή της κατεύθυνσης και χρονικής διάρκειας παρακολούθησης της οθόνης (eye gaze) μπορούν να αυξήσουν σε πολύ υψηλά ποσοστά την ακρίβεια του συστήματος,

## Summary in Greek

καθώς οι δύο φυσιολογικές ποσότητες (καρδιακού ρυθμού και διηλεκτρικής αγωγιμότητας) θεωρούνται ελλιπείς για να επιτύχουν υψηλότερη ακρίβεια.

### Περιορισμοί και Επόμενα Βήματα

Λειτουργικοί περιορισμοί δεν εμφανίστηκαν κατά την διάρκεια των πειραμάτων και χρήσης του συστήματος. Ένας σημαντικός περιορισμός στις πειραματικές διαδικασίες αυτής της μελέτης υπήρξε το θέμα επιλογής διαβαθμισμένου υλικού για προβολές βίντεο. Παρότι οι εικόνες είχαν βαθμονόμηση από διεθνώς αποδεκτά πρότυπα, τα βίντεο έπρεπε να επιλεγούν με παρόμοια κριτήρια αλλά χωρίς να υπάρχει δυνατότητα επιλογής από διεθνώς εγκεκριμένες βιβλιοθήκες ερευνητικού υλικού. Για παράδειγμα, ένα βίντεο το οποίο παρουσίαζε την απόγνωση που ένοιωθαν δύο γονείς πιγκουίνοι προσπαθώντας να επαναφέρουν στην ζωή το νεκρό τους νεογνό δύσκολα θα μπορούσε να αμφισβητηθεί ότι προκαλεί συναισθήματα λύπης και απόγνωσης αλλά δεν είναι διεθνώς καταχωρημένο ώστε να κατοχυρώσει απολύτως την αξιοπιστία της επιλογής του για αυτόν τον σκοπό. Οι συσχετισμοί μεταξύ εικόνων και βίντεο δεν έδειξαν την αναμενόμενη σημαντικότητα για τα αντίστοιχα τεταρτημόρια. Επιπρόσθετα, δείγματα ζευγών (paired samples t-tests) έδειξαν ότι οι διαφορές σε ακρίβεια αποκρίσεων για κάθε ζεύγος εικόνων-βίντεο ήταν σημαντικές:  $t_1 = 2.623$ ,  $p = 0.014$ ;  $t_2 = 5.400$ ,  $p = 0.001$ ;  $t_3 = 4.422$ ,  $p = 0.001$ ;  $t_4 = 3.029$ ,  $p = 0.005$ . Αυτό δείχνει πως η επιλογή των βίντεο επέφερε αποτελέσματα σε λιγότερο υψηλά ποσοστά από αυτά των εικόνων. Επίσης αυτό μπορεί να αποδοθεί στο χαρακτηριστικό της ηπιότερης και αργότερης μετάβασης σε διαδοχική ψυχοσωματική κατάσταση η οποία προκαλείται από το βίντεο σε σχέση με αυτής των εικόνων.

Σημαντικές βελτιώσεις μπορούν να επέλθουν σε σχέση με την ακρίβεια ανίχνευσης εμπλοκής στο περιεχόμενο και κατά συνέπεια ενδιαφέροντος και προσοχής με τις ακόλουθες τροποποιήσεις:

- Αν η γραμμή αναφοράς (Baseline) η οποία χρησιμοποιείται για την αρχική αυτορρύθμιση του συστήματος χρησιμοποιηθεί στον υπολογισμό του αλγόριθμου προβολής σαν σημείο αναφοράς, αυτό θα δημιουργούσε ακριβέστερη απεικόνιση

## Summary in Greek

- Ο αλγόριθμος απεικόνισης δεν διαχειριζόταν τις τιμές οι οποίες ήταν ίδιες με την προηγούμενη μέτρηση ( $\text{Gradient} = 0$ ). Βελτίωση μπορεί να επέλθει αν σε τέτοιες περιπτώσεις λαμβάνονταν υπόψη οι πιο προηγούμενες τιμές (trait assessment).
- Αν συμπεριλαμβάνονταν στον αλγόριθμο απεικόνισης παράμετροι επηρεασμού ανάλογα με το προφίλ προσωπικότητας και ιδιοσυγκρασίας του χρήστη είναι πιθανόν να αυξανόταν η ακρίβεια του συστήματος (χρειάζεται περαιτέρω διερεύνηση).
- Σύστημα ανίχνευσης μεταβολών της ίριδας της κόρης του ματιού θα μπορούσε να βελτιώσει τις επιδόσεις του συστήματος σαν επιπρόσθετος δείκτης επιβεβαίωσης η απόρριψης των ενδείξεων όπως αυτές συμπεραίνονται από τις υποσυνείδητες φυσιολογικές μετρήσεις.
- Σύστημα διερεύνησης χρονικής επεξεργασίας σε διάφορες περιοχές της οθόνης επίσης θα μπορούσε να ενσωματωθεί στο σύστημα αποδίδοντας επιπλέον πληροφορίες σχετικά με την εμπλοκή του χρήστη στο περιεχόμενο. Ένα τέτοιο υποσύστημα όμως θα απαιτούσε εξάρτηση και συσχετισμό του περιεχομένου με τις αντίστοιχες μετρήσεις, δημιουργώντας πολυπλοκότητα στην δημιουργία διδακτικών σεναρίων.
- Περαιτέρω έρευνα στον χώρο της φυσιολογίας η οποία θα παρείχε περισσότερες πληροφορίες και προσδιορισμούς σχετικά με τον τρόπο, την ένταση και την χρονική διάρκεια της έκφρασης των συναισθημάτων θα βοηθούσε σημαντικά την βελτίωση αξιοπιστίας και ακρίβειας του συστήματος.
- Τέλος εξειδικευμένα πειράματα με σενάρια εμπλουτισμένων προκλητών ψυχοσωματικών καταστάσεων μπορούν να εμπλουτίσουν το σύστημα με επιλογές υποδειγμάτων (lookup tables) οι οποίες μπορούν να προάγουν το επίπεδο του συστήματος με δυνατότητες προβλέψιμων αντιδράσεων (predictive processing).

### Συνεισφορά στην Επιστημονική κοινότητα

Το σύστημα HR-STC όπως σχεδιάστηκε και αναλύθηκε σε αυτή την ερευνητική εργασία εκτός από τις καινοτόμες δυνατότητες που παρείχε για να πραγματοποιηθεί αυτή η μελέτη αποτελεί επίσης ένα κεφάλαιο σαν εργαλείο εφαρμόσιμο σε εκπαιδευτικά και ερευνητικά περιβάλλοντα. Η ευχρηστία και αποτελεσματικότητα του συστήματος σε συνδυασμό με το χαμηλό κόστος κατασκευής και τις δυνατότητες προσαρμογής σε πολλές συσκευές δίνουν σημαντικά πλεονεκτήματα για περαιτέρω αξιοποίηση.

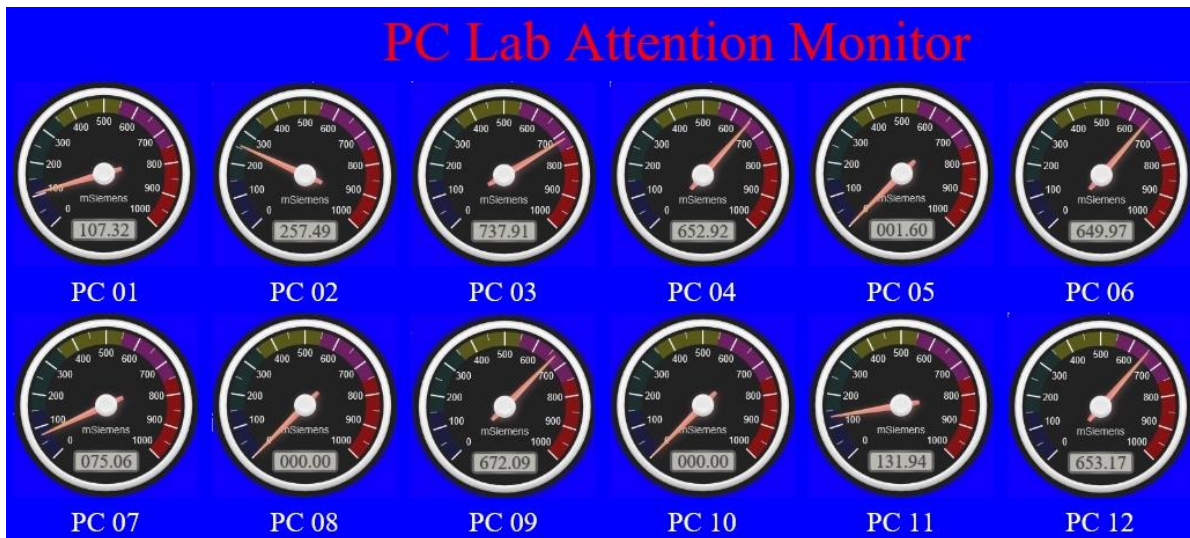
Τα χαρακτηριστικά της αδιάλειπτης και συνεχούς μέτρησης του συστήματος HR-STC με την ίδια ακρίβεια και αξιοπιστία είναι ένα επίτευγμα έναντι ήδη υπάρχοντος εργαστηριακού εξοπλισμού τα οποία δεν μπορούν να διατηρήσουν την μέτρηση διηλεκτρικής αγωγιμότητας σε αυτά τα επίπεδα. Αυτό έχει επιτευχθεί λόγω μιας καινοτόμου σχεδίασης βασισμένη σε κύκλωμα διηλεκτρικής διαπερατότητας (trans-conductance) αντί της συνήθους χρήσης κυκλωμάτων μέτρησης αντίστασης του ιστού του δέρματος. Το πλεονέκτημα του κυκλώματος της διηλεκτρικής διαπερατότητας έναντι άλλων τεχνολογιών είναι ότι λόγω χρησιμοποίησης εντάσεως αντί τάσεως έχει την δυνατότητα να μην επηρεάζεται από τις ηλεκτρολυτικές αντιδράσεις του δέρματος οι οποίες εξουδετερώνουν την τάση διέγερσης μετά από μερικά λεπτά όπως ακριβώς συμβαίνει με ηλεκτροστατικά φορτία.

Η ερευνητική μελέτη ανίχνευσης εμπλοκής μαθητών σε περιβάλλοντα μάθησης βασισμένη σε πραγματικές και ανεπηρέαστες μετρήσεις χρηστών αποτελεί ανάγκη στην σημερινή εποχή ερευνών για την θεμελίωση των δυναμικών συστημάτων εκπαίδευσης και αλληλεπίδρασης των επομένων γενεών. Οι δυνατότητες διερεύνησης της αποτελεσματικότητας εξειδικευμένων νευροφυσιολογικών μετρήσεων, η μελέτη νέων μεθόδων με τις οποίες μπορούν να είναι εφικτές και εφαρμόσιμες στην εκπαίδευση και η διερεύνηση ως προς το πόσο αποτελεσματικές και ακριβείς μπορεί να είναι, θεωρείται θεμελιώδης ανάγκη για την δημιουργία των επόμενων γενεών εκπαιδευτικών συστημάτων τα οποία θα συμπεριλαμβάνουν αυτοματοποιημένα αυτο-διαμορφούμενα περιβάλλοντα.

## Summary in Greek

Μία πολύ σημαντική καινοτομία του συστήματος είναι η δυνατότητα άμεσης λειτουργίας και από διαφορετικούς χρήστες χωρίς να απαιτείται αρχική βαθμονόμηση όπως συμβαίνει με εργαστηριακές συσκευές, ιδιαίτερα ιατρικές συσκευές οι οποίες μετρούν αγωγιμότητα του δέρματος. Το σύστημα HR-STC αποδίδει τις συμπερασματικές ενδείξεις ακόμη και αν εναλλάσσονται οι χρήστες, με την ίδια ακρίβεια, χωρίς να επηρεάζεται από τις αλλαγές των τιμών οι οποίες μπορεί να είναι σημαντικά υψηλότερες ή χαμηλότερες από το ένα άτομο στο άλλο. Αυτό έχει επιτευχθεί με την καινοτομία υλοποίησης του αλγορίθμου ώστε να επεξεργάζεται διανυσματικές τιμές των μετρήσεων και όχι απλά τις απόλυτες ποσοτικά τιμές των μετρήσεων.

Η πρόνοια σχεδίασης σαν αναλογικό κύκλωμα χωρίς την ανάγκη ενσωμάτωσης επεξεργαστή (CPU) επέφερε μια συμπαγή μορφή του συστήματος ως ανεξάρτητου λειτουργικά τμήματος το οποίο μπορεί να υλοποιηθεί σε μικρογραφικές διαστάσεις και να προσαρτηθεί σε διάφορες συσκευές κινητών ή επιτραπέζιων συστημάτων εκπαίδευσης. Επιπρόσθετα, σαν αναλογικό κύκλωμα προσφέρει άμεση απόκριση χωρίς αρχικοποιήσεις όπως αυτές που απαιτούνται από αντίστοιχα ψηφιακά κυκλώματα. Η επεξεργασία των σημάτων και οι επιλογές αλγορίθμων για την χρήση των μετρήσεων μπορεί να γίνει από ένα μεγάλο αριθμό περιβαλλόντων ανάπτυξης από ερευνητές ή εκπαιδευτικούς χρησιμοποιώντας εργαλεία ανάπτυξης εκπαιδευτικών, ψυχαγωγικών ή άλλων σχετικών εφαρμογών. Το σύστημα μπορεί με την σειρά του να χρησιμοποιηθεί σε διάφορες μορφές εκπαίδευσης, μάθησης, ψυχαγωγίας και περαιτέρω ενεργειών, αφήνοντας μεγάλα περιθώρια έρευνας είτε για καθοριστική ανίχνευση είτε για συμπληρωματική απόδοση στοιχείων υποσυνείδητων χαρακτηριστικών τα οποία δεν υπήρχαν σε αυτή την μορφή πριν από αυτή την μελέτη. Για παράδειγμα ένα σύστημα όπως αυτό που απεικονίζεται στο παράδειγμα εφαρμογής που ακολουθεί το οποίο μπορεί με μεγάλο ποσοστό εγκυρότητας να απεικονίσει τα επίπεδα εμπλοκής μαθητών στο περιεχόμενο σε μία τάξη σε πραγματικό χρόνο, είναι κάτι το οποίο δεν υπήρχε μέχρι σήμερα και θα μπορούσε μόνο να καταχωρηθεί σε σενάρια συστημάτων του μέλλοντος.



*Δείγμα εφαρμογής η οποία προβάλλει ενδείξεις εμπλοκής ανά χρήστη σε περιβάλλον τάξης*

Η κατά αποδεκτά υψηλό ποσοστό ακρίβειας απόδοση μιας εικόνας εμπλοκής και ενδιαφέροντος των μαθητών κατά την διάρκεια μιας διδακτικής διαδικασίας είτε με ΗΥ, είτε με ταμπλέτες, νέτμπουκ ή λάπτοπ είναι μία από τις αρχικές δυνατότητες που μπορούν να επιτευχθούν με την χρήση του συστήματος HR-STC.

Ακόμη πιο εντυπωσιακό είναι το αποτέλεσμα όταν το περιβάλλον μάθησης μπορεί να περιλαμβάνει απεικονίσεις μαθητών οι οποίοι ευρίσκονται σε διάφορα απομακρυσμένα μέρη της γης σε πραγματικά σενάρια διδασκαλίας μέσω Ίντερνετ και πλατφόρμες cloud.

Παρόμοιες εφαρμογές στον χώρο της ψυχαγωγίας και διαδικτυακής επικοινωνίας μπορούν να εκμεταλλευτούν τις ενδείξεις εμπλοκής των χρηστών και να αποδίδουν επιπρόσθετες δυνατότητες σαν μπόνους σε παίκτες οι οποίοι επιδίδονται με μεγαλύτερο βαθμό απορρόφησης και βοηθητικές λειτουργίες σε παίκτες που δείχνουν λιγότερο ενδιαφέρον. Περιβάλλοντα φτιαγμένα για παιδιά με ειδικές αισθητηριακές δυσκολίες μπορούν να αποκτήσουν διαδραστικότητα και να αποδίδουν μεταβαλλόμενα επίπεδα δυσκολίας ανάλογα με τις ενδείξεις του συστήματος. Εφαρμογές μπορούν επίσης να σχεδιαστούν για την βελτίωση πασχόντων από παθολογικού τύπου στρες με διαδραστικότητα η οποία μπορεί να δημιουργηθεί για πρώτη φορά, με την ανάλογη διαχείριση των μετρούμενων βιοσημάτων.



## Table of Contents

### TABLE OF CONTENTS

|   | Page |
|---|------|
| Abstract  | i    |
| Acknowledgements or Dedications                             | iii  |
| List of Abbreviations                                       | iv   |
| Summary in Greek  | vi   |
| 1. Preface  | 1    |
| 2. Problem Definition and Aims                              | 11   |
| 2.1 Research assumptions                                    | 13   |
| 3. Literature Review  | 16   |
| 3.1 Scientific Evolution and Relevant Technologies          | 18   |
| 3.2 Physiological Correlates of Psychosomatic Disposition   | 22   |
| 3.3 Biological Instrumentation                              | 23   |
| 3.4 Evolution of Educational Technologies                   | 27   |
| 3.5 Personality Profiles and their Role in Education        | 31   |
| 3.6 Psychosomatic Disposition During the Course of Learning | 35   |
| 3.7 Mental Workload and Attention                           | 36   |
| 3.8 Physiological Expression of Emotion                     | 38   |
| 3.8.1 Theories of Emotion                                   | 39   |
| 3.9 Psychological Measures of Attention                     | 40   |
| 3.10 Physiological Measures of Attention                    | 42   |
| 3.11 Formal Psychological Assessment                        | 44   |
| 3.12 Dynamically Adaptive Educational Environments          | 47   |
| 3.13 Analogue and Digital Signal Processing                 | 51   |
| 3.14 Proprietary Development vs. Commercial Products        | 52   |
| 3.14.1 Commercial Bio-Feedback Products                     | 54   |
| 3.15 Need for Bio-Sensing Technologies                      | 55   |
| 3.16 A Concluding Summary                                   | 62   |
| 4 Research hypotheses                                       | 65   |
| 4.1 Operational definitions                                 | 65   |

## **Table of Contents**

|   |           |
|---|-----------|
| <b>5. Systems Analysis</b>                            | <b>69</b> |
| 5.1 Systems Design Overview                           | 72        |
| 5.2 Description of Sensing Circuitry                  | 73        |
| 5.3 Skin Trans-Conductance Circuit                    | 73        |
| 5.4 Heart Rate Circuit                                | 76        |
| 5.5 Computer Software                                 | 77        |
| 5.6 Electronic Circuit                                | 85        |
| 5.7. Functional and Operational Description           | 87        |
| 5.7.1 Functional Description                          | 88        |
| 5.7.2 Operational Description                         | 89        |
| <b>6. Experimental evaluation</b>                     | <b>94</b> |
| 6.1 Phase One - Device & Systems evaluation           | 95        |
| 6.2 Phase Two - System operational evaluation         | 100       |
| 6.2.1 Assessment Process and Data Analysis            | 111       |
| 6.2.2 Experimental Procedure                          | 117       |
| 6.2.3 Data Assessment and Evaluation                  | 121       |
| 6.2.4 Findings  | 125       |
| 6.2.5 Use of Self-Report Instruments                  | 126       |
| 6.3 Phase Three - Application oriented evaluation     | 127       |
| 6.3.1 Data Evaluation and Selection Criteria          | 128       |
| 6.3.2 Experimental Framework                          | 130       |
| 6.3.3 Experiments                                     | 132       |
| 6.3.4 Analysis of the findings                        | 145       |
| 6.3.5 Discussion                                      | 146       |
| 6.4 Phase Four - Comparison with commercial products  | 147       |
| 6.4.1 Free-Form and Formally Structured Digital Games | 150       |
| 6.4.2 Psycho-Physiological Metrics                    | 152       |
| 6.4.3 Research objectives and hypotheses              | 155       |
| 6.4.4 Methodology                                     | 157       |
| 6.4.5 Participants                                    | 158       |

## **Table of Contents**

|  |     |
|--|-----|
| 6.4.6 Non-Linear Game Play vs. Linear Game Play                                | 158 |
| 6.4.7 Multiple Solutions vs. Unique Solution to Challenges                     | 159 |
| 6.4.8 Free vs. Fixed Sequence of Challenges                                    | 159 |
| 6.4.9 Selection vs. Pre-Defined Set of Challenges                              | 159 |
| 6.4.10 Criteria for Selection of Digital Games                                 | 160 |
| 6.4.11 Data Collection Methods   | 163 |
| 6.4.11.1 Self-Assessment Manikin   | 163 |
| 6.4.11.2 Bio-Feedback Measuring Device   | 164 |
| 6.4.11.3 Face-Reader Emotion Detection Software                                | 165 |
| 7. Discussion and Interpretation of Results                                    | 167 |
| 7.1 Results: Phase One – Device and Systems Evaluation                         | 168 |
| 7.2 Results: Phase Two - Systems operational evaluation                        | 169 |
| 7.2.1 Findings: Phase Two - Systems operational evaluation                     | 170 |
| 7.2.2 Discussion: Phase Two - Systems operational evaluation                   | 171 |
| 7.3 Results: Phase Three - Application oriented evaluation                     | 178 |
| 7.3.1 Findings: Phase Three - Application oriented evaluation                  | 178 |
| 7.3.2 Discussion: Phase Three - Application oriented evaluation                | 190 |
| 7.4 Results: Phase Four - Comparison with commercial products                  | 191 |
| 7.4.1 Findings: Phase Four - Comparison with commercial products               | 192 |
| 7.4.1.1 Impact of FF and FS Digital Games on<br>Satisfaction and Activation    | 192 |
| 7.4.1.2 Impact of FF and FS Digital Games on Stress Level                      | 193 |
| 7.4.1.3 Impact of FF and FS Digital Games on HR                                | 194 |
| 7.4.1.4 Impact of FF and FS Structured<br>Digital Games on Engagement          | 195 |
| 7.4.1.5 Impact of FF and FS Structured<br>Digital Games on Positive Experience | 196 |
| 7.5 Discussion: Phase Four - Comparison with commercial products               | 198 |
| 8. Conclusions   | 203 |
| 9. Applications  | 207 |

## **Table of Contents**

|   |     |
|---|-----|
| 9.1 School personalised learning            | 208 |
| 9.2 Cloud based educational personalisation | 209 |
| 9.3 Gaming interaction                      | 210 |
| 9.4 e-learning adaptive content             | 210 |
| 9.5 Communication and social media          | 211 |
| 10. Limitations and Future Developments     | 213 |
| 10.1 Improvements                           | 221 |
| 10.2 Future work                            | 223 |
| 10.2.1 Pupil size detection                 | 224 |
| 10.2.2 Virtual and augmented reality        | 225 |
| 10.2.3 Mobile devices and applications      | 226 |
| 10.2.4 Gaze analysis                        | 227 |
| 10.2.5 Cloud computing                      | 227 |
| 11. Contribution to Science and Technology  | 229 |

## **References**

## **Appendices**

**Appendix A: Functional topology of the human brain**

**Appendix B: Physiological process of emotion**

**Appendix C: Psychophysiology of Electro-Dermal Response**

**Appendix D: Literature on psychometric tests employed**

**Appendix E: Pictures of the electronic prototype**





**1. Preface**

The essential concept investigated in this research study entails the design and implementation of an integrated bio-sensing system assessing the effectiveness of simultaneous excitation of two physiological signals when used to deduce states of engagement and focusing involvement of a learner. An unobtrusive proprietary system has been developed as a tool for assessing validity and accuracy of a long standing modelling algorithm claiming that psychosomatic conditions can be deduced from behavioural patterns of certain physiological expressions. Innovative design concepts and flexibility allowed for effective use of the system in experiments proving its capabilities to be used in educational environments as an integral part of computer desktop, portable computers or mobile devices. Classifiers produced by the system in real time, presenting subconscious cumulative involvement of a user can be provided as inputs to dynamically adaptive learning environments. Although the risk of inconsistencies and precision concerns caused by instantaneous measurements were apparent, no additional correction or averaging algorithms were used. This was made in order to assess the accuracy of the algorithm used for the interpretation of user engagement, the actual operational capabilities of the system and the validity of the measurements with regards to revealing subconscious expressions in that respect.

Several scientific reviews in effective communication and learning, attempt to identify human emotional predisposition, affect or feelings and convey derived qualitative markers in social activities and interactive user environments. Methodologies employed in research areas such as adaptive learning, educational technologies (McCrickard & Chewar, 2006, Groth-Marnat, 2009), HCI, and affective computing (Picard, 2003; Fairclough, Moores, Ewing, Roberts, 2009), rely predominantly on formal tools for psychological assessment. Typically, data obtained from self-reporting questionnaires, interviews and personality assessment, form the initial collection (Costa & McRay, 1992). Data derived from initial collection of targeted samples are then used to develop and identify inferences which are subsequently correlated to form or prove a hypothesis. Influencing classifiers developed from inferences are then incorporated into a dynamic model for further event related experimentation and data produced

## **Chapter 1: Preface**

are statistically processed in an investigation of conclusive interpretations. Although carefully selected assessment methodologies are being used, one can argue that input data largely present an estimate of emotional state of individuals as they express their views subjectively. Since humans have no accurate perception of their emotional or affective state, its particular constituents and their specific intensity (Salovey & Mayer, 1990), their subjective evaluations can be partial and decline from the actual condition. Such cases are more obvious particularly when one tries to express an estimate of combinations of sentiments that produce similar sensation with other combinations (Ochsner, Gross, 2005). Additionally, emotional responses estimated by individuals are influenced heavily by their emotional regulation skills frequently inhibiting emotional and affective components.

Evolutionary biology and psychology have offered a number of theories of emotion, further analysis of which is deviating from the scope of this text, but it is worth mentioning that clearly emphasize their analogies either with propositional judgments or with perception. Different ontologies of emotion express various concerns disputing the rational and cognitive contribution towards their respective emotional and affective formation; however, in all their experimental conclusions were in agreement on that emotions are inter-related and have a serious impact on cognitive processing (Salovey & Mayer, 1990). The aforementioned suppositions are met frequently in several scientific fields where influencing factors in human communication and behaviour is the subject under investigation while they have been verified by numerous forms of experimentation. Consequently it has been widely accepted that emotion, cognitive involvement and affect can be considered as interrelated activities comprising the formation of a wider concept of involvement subsequently revealing intensified attention and engagement.

Various approaches identified in older theories regarding the relationships of emotion with cognitive processes and previous experiences (Lazarus, 1982; Oatley & Jenkins, 1996) have hindered a commonly accepted definition and subsequently forced scientists to seek established foundations for new classifiers by using instruments employing measurement and instrumentation methods. Appraisal of reactions to evoked emotional events constitutes the

## Chapter 1: Preface

most common method to deduce conclusive judgements between expressions of emotional or affective origins. Efforts have been made to assess all influencing indicators of emotional reactions ranging from analysis of visual activities such as pupillary motion, gazing intensity, focusing duration and direction in respect to their associated subconscious physiological responses to emotional and cognitive loading stimulations. What makes things even more complex in attempting to deduce human emotional status is the fact that there is no means to identify structures of emotional bindings that produce identical expressions in people with similar cultural and social profiles. Additionally, the two universally accepted emotional constituents of satisfaction and intensity provide insufficient detail to distinguish precisely emotional components emphasising the need of complementary descriptors (Fontaine, Scherer, Roesch, Ellsworth, 2007). Lack of precision when looking at emotion as a construct composed by insufficient psycho-physiological expressions becomes apparent as in many cases certain combinations of emotional constituents produce expressions identical to those caused by some other combination of emotional or affective stimulations. Therefore the mission to identify discreet sentimental constituents that cause individual emotional condition has been very intricate and therefore quite rarely attempted.

A credible alternative to self-reporting formal methodologies has also gained acceptance in the scientific communities using physiological instrumentation in effort to detect subconscious reactions to focus evoking stimulations. This approach can provide a more impartial method of ascertaining the users' psychological state than the above instruments of formal assessment as it cannot be distorted. The abovementioned technologies have elevated the field of detecting correlates of affect, emotion or cognitive workload to further our knowledge beyond the already known self-reporting or visual observation methods. Existing systems facilitating human physiological measurements for assessing psychosomatic correlates have been using medical devices obviously designed for use in laboratory experimentation and therefore refrained from use in open environments like a school class due to discrepancies caused by sensing instrumentation (e.g. electrodes), postural and in most cases kinetic restrictions. A system designed and developed specifically for the purposes of this study has overcome those major hurdles described above, by using a novel approach not compromising accuracy when

## Chapter 1: Preface

measuring user responses in any environment. Micro-electrical signal acquisition required to obtain physiological responses was achieved by incorporating a set of re-useable sensors in a minimally obstructive and unobtrusive fashion. Optimised bio-signals strongly related to psycho-physiological expressions can be captured with no physical or mental inconvenience whatsoever by holding a typical apparatus such as a computer mouse or a handheld device (e.g. mobile phone). The major advantage of this purpose built system over other systems used in studies of physiological assessment is that this system being inconspicuous circumvents additional errors induced by stress and psychological strain imposed on the user by them being aware of the fact that a measurement is taking place.

Numerous settings for scientific experimentation measuring physiological indicators have been performed successfully in laboratory setups; however, there is still a gap between obtaining efficiently data in those restricting environments and achieving to obtain similar results in non-clinical environments such as those in a school class or outdoors. Bridging this gap was one of the aspirations initially considered and finally achieved by the development of a purposely designed physiological acquisition system that could be used without constraints in any environment in every day's life, free of attachments, electrodes or kinetic restrictions whatsoever. The system was developed based on similar first principles of measurement as those used in medical devices holding a significant advantage by accomplishing unnoticed operation of bio-sensing. This was made possible by implementing an optimised sensing interface to acquire the user's heart rate (HR) and psychological stress as they were expressed by the human somatic system in real time. The sensing part of the system was embedded in a computer mouse. Optimisation of bio-signal types yielding a credible detection of vital psychosomatic expressions that inherently combine unobtrusive operation and dependable functionality led to the selection of a combination of HR and Skin Conductance (Venables, Christie, 1980).

Assessment of essential technologies in view of an optimisation which led to the selection of the above physiological quantities including a review of most relevant physiological measurements appropriate for the detection of psychosomatic expressions has been included in

## Chapter 1: Preface

following chapters of this text. Successful development of the above system made then possible for this research work to perform an investigation of a generic approach that attempted to distinguish quantities of emotional, cognitive, affective and conative congregations. Neurophysiological modalities effectively measured as reactions to pre-validated emotional stimulations were treated according to their manifestations of coinciding excitations. Essentially, the system developed for this research study has made possible to develop the optimal prerequisites for accurate and minimally obtrusive measurement of cardiac elevation or demotion in conjunction with Skin Trans-Conductance (STC) responses and study the behaviour of the excitation of those two biological expressions in relation to engagement and focused involvement of a person.

Essential requirements for instant acquisition and processing of purely unimpeded physiological signals could not be met if commercial products were used. The major aspects that could not be overcome by laboratory devices which forced the development of an integrated custom made device were in summary:

- requirements for non-obstructive operation without electrodes and attachments
- physiological expressions of a person should be measured in a natural environment
- the strict requirements for absolute synchronisation of signal acquisition
- requirements for zero latency during the reading of instantaneous input signals
- instantaneous real time processing and derivation of related classifiers for adaptive personalisation environments as well as projection of psychosomatic condition on the screen
- flexibility to incorporate more algorithms and additional components in the processing cycles
- elimination of initial calibration and familiarisation period
- ability to embed the electronics into a small device

In early studies involving physiological instrumentation has been claimed that intensified heart rate and increasing stress levels exposed at the same time can be reliable indicators of the

## Chapter 1: Preface

emotional state of an individual. Those claims have been derived from several scientific experiments, at different times since 1953 to date employing different technologies. In those citings, scientists posit that HR and SC can help to deduce states of basic emotion (Table 1).

**Table 1: A review of literature since 1953 (adapted from Psaltis, Mourlas, 2013)**

|                                      | Correlations between HR / SC gradients and basic emotion |       |         |           |         |           |          |
|--------------------------------------|--|-------|---------|-----------|---------|-----------|----------|
|                                      | Fear   | Anger | Sadness | Happiness | Disgust | Amusement | Surprise |
| <b>Ax (1953)</b>                     | + / +  | - / 0 | - / -   | - / -     | 0 / 0   |           |          |
| <b>Christie (2002)</b>               | + / +  | + / - | + / 0   | + / 0     | 0 / 0   |           |          |
| <b>Ekman et al. (1983)</b>           | + / 0  | + / 0 | - / 0   | + / 0     | 0 / 0   |           |          |
| <b>Fredrickson et al. (2000)</b>     | + / 0  | + / 0 | + / 0   | 0 / 0     |         |           |          |
| <b>Levenson et al. (1990)</b>        | + / +  | + / 0 | + / 0   | + / 0     | 0 / +   | - / +     | + / 0    |
| <b>Nasoz et al. (2003)</b>           | + / 0  | + / 0 | 0 / 0   | + / 0     | 0 / +   |           |          |
| <b>Palomba &amp; Stegnano (1993)</b> | + / 0  | + / 0 | + / 0   |           | + / 0   |           |          |
| <b>Palomba et al. (1999)</b>         | + / +  | + / 0 |         | + / 0     |         |           |          |
| <b>Picachin et al. (1999)</b>        | + / 0  | + / 0 |         |           | + / 0   |           |          |
| <b>Sinha et al. (1992)</b>           | + / 0  | + / 0 |         |           |         |           |          |
| <b>Sherer (2000)</b>                 | + / +  |       |         |           |         |           |          |

(HR / SC), +: Increase, -: decrease, 0: no significant

The above collection of published results reveals a connection between the two traits of HR, SC and basic emotion and the justification for selecting the above physiological modalities in order to deduce involvement from related emotional and affective activation. Further elementary analysis indicated that many of the concluding opinions coincide, thus strengthening the motivation and prospects to further investigate the findings advocated by using devices employing recent technologies. Further observation of the above table reveals imperfections in matching some results of the findings as different assessment methodologies, acquisition technologies and testing practices have been employed at different times during that fifty year period. For example it has been generally accepted that anger can be distinguished by increased heart rate, however, two cases in the above table yielded a different result.

## Chapter 1: Preface

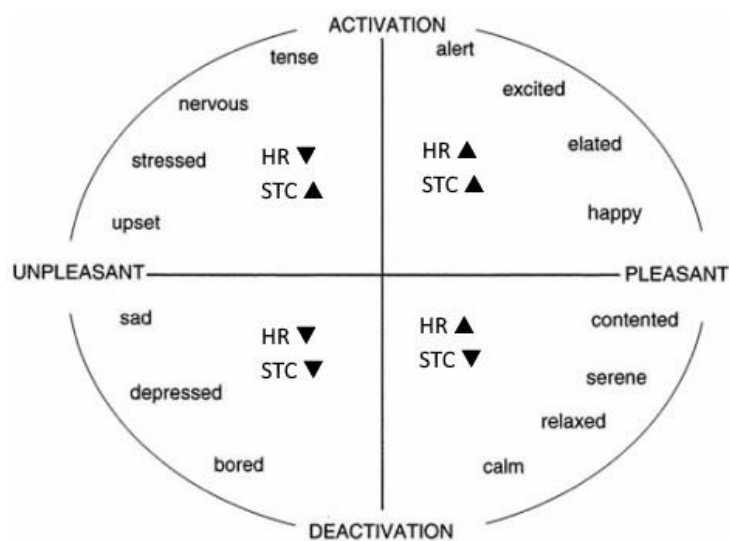
According to the above conclusions shown in Table 1, interpretation of actual combinations of the previously mentioned physiological quantities can produce valuable indicators that could reveal cumulative mental activation in a similar fashion that it could be detected by more complex and sophisticated settings. In that sense, measured quantities of bio-signals were treated as indicators of increasing or decreasing traits of values rather than as actual quantities. An increasing rate of cardiac frequency produced a positive gradient and a decreasing rate produced a negative gradient respectively. Similarly for stress levels, a positive gradient denoted by an increasing trend of measured values of STC was indicative to elevating awareness and a negative gradient shown a decreasing stress level or awareness, subsequently showing diminishing attentiveness.

The transition from measured physiological quantities to the notion of cumulative brain workload incorporating emotional, cognitive and affective congregations as states of engagement was attained through a classification model based on attributes expressed by all four possible combinations of the above gradients. Two variables such as HR gradient holding the positive (+) or negative (-) values and similarly for STC, could therefore produce four possible combinations (i.e. (+,+) : (+,-) : (-,-) : (-,+)). The above four combinations prompted for an assessment model based on a classification of corresponding levels of brain activation in four distinct areas of clustered characteristics resembling compliance with one of the most widely accepted models of emotional allocation (Russell, 1980). For the purposes of the assessment of states of engagement in a bi-axial representation of measured values as required by the above model, a relationship between the four combinations of gradients and the four quadrants of emotional classification of Russell's circumplex model was devised.

Real time workload analysis carried out according to the above algorithm provided a visualized assessment of the user's states of engagement displaying on screen the results derived from purely subconscious physiological expressions in an unrestricting and unobtrusive manner.

## Chapter 1: Preface

Effectively the user's HR provides a credible indication of the state of contentment while the STC provides an indication of the state of activation seen as a response to sympathetic and parasympathetic activity respectively. As constituent values, the above two physiological quantities composed the two major coordinates of emotional composition (i.e. satisfaction and activation), seen in Figure 1. Each quadrant characterised therefore by the corresponding gradients of HR and STC respectively (▼: descending, ▲: ascending).



*Figure 1: Nomenclature of emotional classification on a bi-axial model*

It has to be stressed that the term emotion throughout this assessment is looked as the instantaneous response expressed by the sympathetic nervous system as a first reaction to a stimulus as opposed to the longer duration component composed by new and existing sentimental constituents, (Oatley & Jenkins, 1996). The HR/STC system developed and used for the experiments and this study has not been designed to measure or identify emotion as it is perceived in its most widely accepted term that includes influences from longer term affective, impulsive or conative components such as mood, worries and such like. Cognitive, emotional or affective components are being treated as a cumulative quantity projected onto a two dimensional representation of estimated levels of engagement. Subliminal physiological responses of a learner derived from the HR/STC system are being represented as coordinate

## **Chapter 1: Preface**

values in a model of attention and engagement resembling the corresponding allocation of emotional components in Russell's circumplex of emotion; as it also uses the concept of bi-axial representation of valence and arousal. The quantity deduced by the system is processed to indicate the instantaneous estimate of the focusing involvement of the learner including emotional, cognitive and memory workload, bearing no reference to emotion or emotional constituents; However, since the allocation of engagement coincides with the allocation of distinct emotional elements according to Russell, one could claim that the learner may be experiencing emotional experiences as those designated on the corresponding quadrant of the circumplex.

Experimental methodologies for assessing both the efficiency of the system and the effectiveness of the common gradient excitation behavioural algorithm were studied thoroughly and a series of four distinct phases of experimental settings was formed. Tests included system validation, evaluation of system's performance using previously validated content causing emotional stimulations, mock-up applications assessing efficiency in simulated class environments and comparison of system results with those produced by a commercial product based on facial analysis.

Formal psychological assessment data have been obtained and compiled in order to deduce relationships and identify individual idiosyncrasies of personality groups through their respective responses. Statistical analysis has been included emphasising on most important findings. A discussion concerning the results and conclusions revealed by this study followed by system limitations and views on future developments and improvements concludes the essential part of this research work. Contribution to the scientific community is following and the merits of the system have been laid out in the very last chapter of this text. Application areas of the system of psychosomatic assessment and examples either in its existing form or with additional functionality brings this writing to a close.

The structure of this document has been formed according to the following plan. Chapter 2, includes an in depth analysis of the problems this study was set out to solve and the targets set

## **Chapter 1: Preface**

by this research work. Chapter 3 contains a review of relevant literature and supporting contents highlighting issues that have been ensued in previous studies involving the implementations of neurophysiological measurements with respect to detecting affective states, also exploring the most important latest technologies. In chapter 4 the principles of bio-sensing, bio-feedback techniques and system analysis have been explored, explaining the design and composition of the electronics through to the software and other contributing components that have been constructed and incorporated into the system. In chapter 5, a functional description of the system has been sited followed by an operational description comprising a form of operators manual. In chapter 6, the experimental sessions conducted for this study have been described in detail, including all series of experimentation methodologies and procedures, followed by comments on respective outcomes. Subsequently, in chapter 7, a discussion on the findings and interpretation of results has been written, extending statistical conclusions to practical applicability and effectiveness of the system. Chapter 8 presents the problems and limitations that hindered the aforementioned study, exploring also several suggestions to remedy the above predicaments. Conclusions drawn from this undertaking and future improvements have also been included. Chapter 9 has been serving the purpose to portray the usefulness of the system when used as a tool in real life environments with suggestions and ideas for future application developments, enhancements and implementations. Towards the end of this writing, have been included relevant appendices, containing supplementary content with further explanatory materials and software code. Finally, the reference links of supporting material are concluding the document.

## **Chapter 2: Problem Definition and Aims**

### **2. Problem Definition and Aims**

Scientific endeavours in psychology, attempting to identify human emotional predisposition, affect or feelings for use in systematic educational schemes, rely predominantly on formal assessment based on self-report evaluations that can be influenced by subjective information and therefore susceptible to biasing, especially in complex emotional or affective conditions. Another vulnerability of the above approach becomes obvious when typical process and interpretation of results needs certain steps of evaluation and statistical analysis in order to deduce profiling and behavioural aspects rendering formal assessment particularly arduous for real time applications. Conclusions derived from the above methods subsequently treated as weighing factors in judgment of performance are rendered ineffective and time-consuming for use in responsive computer environments where immediate evaluation of psychosomatic condition is needed. Recent technologies in education and computer based learning also necessitate immediate and at the same time highly accurate descriptors that would be impartial and unaffected by the learner.

Advancements in medical instrumentation have been employed for experimentation in assessing physiological expressions pertaining to focusing attention. When the above results were looked as an index of learner's effective motivation have since gained enormous recognition. Psychosomatic constituents of attention and engagement directly related to executive mental effort exerted by the learner have been identified by means of assessing momentous physiological modalities. Effective use of bio-sensing is strongly depended on special attachments and peculiarities that render their application outside laboratories unusable.

Detection of bio-somatic indicators necessitate novel approaches of un-obstructive acquisition interfaces in order to eliminate the above obstacles. Such interfaces would then allow for further study and optimisation of physiological correlates that can produce the psychosomatic projections of a person accurately and with high precision. Subsequently, inhibited physiological contribution can be invaluable in responsive applications that can become dynamically adaptable. Hence the need for development of a new instrument as a flexible tool for assessing physiological responses is portraying the first problem this study had to resolve.

## **Chapter 2: Problem Definition and Aims**

The aim of this first part was to design a system from first principles and verify its functionality and effective use in free space environments. Fulfilment of this first part was imperative as it would provide an essential tool for accomplishing the next stages of this study.

Next step was to identify and apply approved methodologies that would infer the physiological readings obtained to their corresponding constructs of focusing involvement as they happen in real-life. Reviewing existing methodologies and finding the optimal procedures for assessing physiological responses and subsequently derive correlating factors in conjunction with activation and engagement was another essential challenge constituting the next aim of this study with numerous related questions to be answered as found in detail in literature review. Adaptation of optimised assessment techniques to an appropriate interface design was the next challenge, incorporating numerous preconceptions aiming to avoid ambiguous conditions during measurement, making thus the interpretation of data questionable. Identifying those quantities derived from the measurements that would furnish reliable and undisputed statistical parameters for the experimentation analysis was another intermediate objective involving excessive analysis of experimentation methodologies and optimisation thereon.

Physiological modalities have been combined in several studies and evaluated for effectiveness in deducing inhibited functionality of the brain, nervous system and clinical pathology applications. In the context of measuring cumulative brain activity as a method to derive the level of involvement of a learner, two physiological quantities highly related to detecting related expressions considered the heart rate and psychological stress descriptor. The above quantities would only present a useful combination of physiological indicators if used according to a concrete and established method revealing cumulative psychosomatic activation. This presented the second problem in this research.

A modelling of the integrated signal acquisition and processing incorporating a credible algorithm was then required for further study of expressions and verification of the system capabilities, constituting the next problem. Another major challenge was the optimal selection of the appropriate tools for validation and presentation of data that would hold the most

## **Chapter 2: Problem Definition and Aims**

important part of the research and produce the qualitative assessment of this work. An explanatory and highly descriptive representation of interpreted results was essential to project the user's physiological responses onto the screen. The system would also provide comprehensive classifiers in real time needed for actively adaptive or responsive applications for personalisation. This particular problem was approached by regarding the above two physiological quantities as correlates to respective values of valence and arousal, resembling the biaxial emotional allocation of Russell's circumplex model of affect (Russell 2003). Certain steps needed to clarify the ability of the system to perform well according to its design specification and whether the initial assumptions regarding the graphical representation of the real-time projection correspond to the actual levels of valence and arousal.

Testing efficiency and the actual capabilities of the system, presented the problem to identify a convincing set of experimental procedures and perform experiments that would expose thoroughly the merits of the aforementioned development. Comprehensive experimentation had to be defined in such a fashion that the credibility of the physiological classifiers would be evaluated in both independent assessment environments and also in comparison with existing commercial products. Finally statistical analysis of data obtained from the experiments and also from formal psychological assessment would be needed aiming to assess additional correlations between personality groups and their relationships with the system responses.

### **2.1 Research assumptions**

Introduction of digital technologies in education and learning necessitate advancing technologies extending further than teacher's skills and evaluations of individual capabilities of the learner. It was assumed that instruments revealing affective physiological responses could answer the problem of adaptive personalisation in education more efficiently than existing methodologies since they could provide valuable inner information of the learner, more frequently, instantaneously and for each particular event. Exposure of instantaneous responses with psycho-physiological origins was expected to produce a valuable asset to

## Chapter 2: Problem Definition and Aims

evolving personalised technologies whether they employ event based or trend analysis methods. Also it was assumed that the advantage of real time processing of the HR/STC instrument could provide an instant estimate of the active involvement of a learner not possible with formal educational methods which. Attempting the development of a tool with flexibility to be used in the class was considered of high importance providing the missing link between detrimental restrictions of medical devices used in past settings and effective application of the aforementioned tool in close proximity or distant and unattended environments.

Earlier citings such as those summarised in Table 1, posit that simultaneous excitation of HR and STC can be considered as credible indicators of emotional constituents. It was thought that the same excitations as above would verify our views for credible detection of cumulative involvement as seen from the level of mental involvement expressed in the corresponding emotional constituents. Experimental analysis was expected to reveal the efficiency of the system, the effectiveness of the algorithm adopted and the accuracy of inferring constituents of user involvement. In summary the essential assumptions considered for this research were:

- It was assumed that behavioural assessment of HR and STC seen as quantitative changes during a learning session is correlated to mental involvement.
- Behavioural assessment of gradient variability was assumed that it could produce credible indications of focusing involvement and engagement.
- Enhancements in the class utilising bio-sensing was assumed that they could provide an impartial real time input to dynamically adaptive educational systems when used as an indicator of assimilation and as a predictor of the subliminal capacities of the learner.
- It was assumed that the HR/STC device could improve performance in education, learning and communication when used to provide a feedback in responsive and adaptive personalisation.
- It has been assumed that communicating physiological expressions would be a complementary tool and an improvement over formal appraisal and teaching methods.

## **Chapter 2: Problem Definition and Aims**

- It was anticipated that the HR/STC device will provide the missing link between existing technologies and flexibility for use in the class for adaptive learning and personalisation as such an instrument is not available to date.
- Finally it was assumed that detecting sudden and large alterations in HR and STC would be indicators of vivid psychosomatic conditions that could not be identified previously by any other tools, opening doors to new areas of application in forms of education designed for people with special needs.

## **Chapter 3: Literature Review**

## **Chapter 3: Literature Review**

### **3. Literature Review**

This interdisciplinary research study spans into several and somewhat diverse scientific fields, more specifically, human physiology, biological signal acquisition, analogue and digital electronics, analogue and digital signal processing, medical equipment technologies, educational, cognitive and behavioural psychology, Human Computer Interaction, (HCI) and software development for adaptive computing educational environments. As the stretch into the above scientific areas entails a substantial amount of background supporting material it was considered appropriate to regulate the extent of each discipline accordingly. For this purpose, background material has been narrowed down to the breadth and depth of contribution and influence of each area in the overall research work. Along these lines long-drawn-out and in-depth descriptions have been circumvented, replaced by a selectively co-ordinated sequence of topics that has been adopted so that the gap that this study has attempted to fill would be demonstrated thoroughly.

Staying on course with the predominant tenets of this study while theories and supporting notes come from several scientific fields as mentioned above, diverse views of areas under discussion may obscure exemplification and that could be a shortcoming if not approached systematically. Therefore, the layout of this review was structured around a stem of the notion looking at how digital technologies have affected educational technologies; assess the effectiveness of existing methodologies and look at novel perceptions applied to adaptation and personalisation in learning, education and communication. Pivoted around technological capabilities and scientific tools that emerged in the disciplines mentioned previously, an assessment was made including applied concepts and instrumentation that educators could use in order to improve their educational aims. The analysis next has been focusing on the justification of the improvement of educational efficiency that could be achieved by novel ideas such as the one introduced in this study when intertwined so that they could enhance teaching and learning practices. Experimental approaches and data processing methods appropriate for each scientific area have been explored next as they were necessary to optimise the selection of methodologies adopted and employed herein. Eventually, as the reader reaches the final sections of this chapter

### **Chapter 3: Literature Review**

has been expected to have acquired all necessary background information required for understanding the effort, composition and the accomplishment envisaged in this text.

The formation of this review, onsets with an exploration of specific generic concepts, which led to the first principles of operation of devices subsequently applied in physiological measurements. The merits of the aforementioned concepts eventually were acknowledged and developed into established methods for the detection of inhibited human expressions. A reference to major discoveries, ideas and devised techniques employed for the acquisition of modals that could help to interpret psycho-physiological conditions was found next. Background information required for basic understanding of acquisition and transformation of biological signals for use in intelligent responsive and Adaptive Computer Environments was also quoted. Relevant definitions, approaches and methodologies that could be applied to educational frameworks were then investigated. Available rules and regulations in adaptive environments, psycho-physiological assessment platforms and pitfalls were then looked into.

#### **3.1 Scientific Evolution and Relevant Technologies**

Since the very beginning of the documented historical findings of human evolution, there are non-scientific references to electricity in some form of physical phenomenon like lightning or in paradoxical phenomena encountered in the animal kingdom; for example incidents like electrification by contact with a sheatfish or an electric eel. At first, an electric surge was treated as an enigmatic concept getting acknowledgement as a form of energy created by nature. Generating electricity and electrical properties were also discovered progressively and in various forms, (Alexander Volta 1800: electric accumulator, Michael Faraday 1831: electromagnetic properties), creating fascinating new areas for research. Scientists explored the possibilities to use electricity in order to carry out a multitude of experiments in many areas such as conductivity of materials, chemistry, physics, and biology in order to investigate possible applications in everyday life. Eventually, production of electric flux was achieved in the form of static electricity (Van Der Graaf, 1901). Progressively scientists attempted to study the effects of electricity in conjunction with the electrical properties of the human body in particular. One interesting research was dedicated to assess the effects and reactions of the

### Chapter 3: Literature Review

human body in relation to various forms of electrical signals. The latter led to the discovery of a new scientific area that exposed permeability characteristics of the human body (Lykken & Venables, 1971). For the first time the human electrochemical properties, the conductive properties of tissue, the electrical characteristics of the nervous system and successively the congregating potentials causing muscular activation and motion were investigated. In 1939, Hodgkin and Huxley demonstrated their discovery on electrical conductivity of neurons and eventually the electrochemical brain activity was exposed. As the electrical properties of the human body became uncovered, they were classified as either brain mediated stimuli to muscular tissue or inductively produced electrical activity originating from the neuronal conductivity taking place in the brain tissue (K - SO<sub>2</sub> depolarisation).

Eventually, a new science called 'radio engineering' was propagated from the invention of the telegraph in the end of 19<sup>th</sup> century and eventually progressed to the science of '*Electronics*' in the middle of the twentieth century when the tube and radio communication evolved. This new discipline moved the use of electricity forward into a new dimension, supplemented by inventions like the radio and television. In 1948 a new fundamental invention took place named '*Transistor*'. This electronic component allowed the implementation of signal amplification and transformation into different forms like audio and thermal printing. Around 1950's the first concept of Electronics for Medicine emerged, giving birth to a new scientific area of Medical Electronics, dealing with the detection processing and presentation of biological quantities for medical use. In 1960 the development of integrated circuits (IC's), enabled the rapid development of smaller in size, faster and more accurate electronic circuits like the "*instruction fetch-execute*" (IFE) model processing units and shrewd designs that brought enough modulation speed for colour television and high speed communication.

Progress in applied sciences and physics revealed that minute natural quantities can be detected by using chemical, magnetic or electrical sensing components. Physical quantities like force or temperature could be then not only represented from a distance but also be made available in other forms like records or printouts. Studies employing collaborating scientific efforts coming from the areas of medical electronics combined with physics, physiology and psychology produced the first direct readings of certain properties of the human body in form of electrical

### Chapter 3: Literature Review

activity such as cardiac pulsation, brain activity, muscular activity, psychological stress loading etc. Electrical conductivity taking place in the human body as a result of chemical reactions inside cells at a molecular level and conductors in a systems level have been thoroughly documented in corresponding literature for clinical physiology, (Lykken & Venables, 1971) and medical electronics literature respectively (Horowitz & Hill, (1989). The science of electronics, in conjunction with progress made in material properties advanced the revolutionary forms and applications of the transistor, giving greater design flexibility and enhanced precision. Boolean algebra and theoretical mathematics combined to form the logic algebra that when implemented in electronic circuitry evolved into computer logic and general purpose circuits that could execute program instructions like addition, subtraction, multiplication and division. Those specific circuits could then be integrated into the IFE models and eventually progressed to Computer Processing Units (CPU's) like those found today in Personal Computers (PC's). A vast amount of human resources dedicated to developing programs for computer applications ranging from the industries and manufacturing to production, logistics, banking, communication and literally every area of our life has been simplified and improved dramatically after the proliferation of computing and computer systems.

In the early years of computer evolution dictions like '*Computers can recognize inner thoughts of humans, like mental activity, affective state or emotion*' seemed far from real and contemplated as rather fictional. People of that time would rather react offensively and be very reluctant to accept such a prospective to happen in the real world since computer technology those days could cater for arithmetic calculations and limited graphics capabilities. Sensing and instrumentation was then in its infancy and that made it difficult to justify a method that would devise or sense quantities that even humans themselves lack the means to perceive and stipulate. One would have thought that as a computer is a programmable machine that can only perform pre-programmed tasks and since attributes like emotion and psychological conditions could not be expressed in physical quantities that notion was considered for many years a technological exaggeration even by people with substantial amount of knowledge of the computer operational capabilities. This reservation was understandable since the electronic

### **Chapter 3: Literature Review**

circuits that consequently composed the computer processor developed from one simple operational property; that of a switch.

Looking in depth at first principles, one can understand how a simple concept if used intelligently could help to build more complex systems like today's computers. For instance, two switches connected in cascade could let the input get through to the output if both activated/deactivated simultaneously, producing thus a logic result of an AND circuit. Similarly, two switches connected in parallel can present the input signal in the output if either of the two switches were activated/deactivated, producing thus a logic result of an OR circuit. Combinations of the above circuits can produce adders and calculation circuits, transfer of values (registers) when triggered by a pulse (clock), retain or forward last input value (latch), last state retrieval (forwarders) and many components of today's computer processor circuitry and storage components.

The first components using those primitive forms of computing power to present simplified logic functionality were the Finite-State-Machines (FSMs), used for dosage or item selection in vending machines and such like. The introduction of triggered servicing to the processors (Interrupt Service Routines) enabled sequential reiterating processing as stands until to date. Interrupt driven processing enhanced the static logic concept by introducing a dynamic repetitive instruction execution model that could perform a number of activities such as data and memory transfer, scanning of peripheral devices like a keyboard or a mouse and serve many processes and calculations for many tasks. Conclusively, although since then unprecedented progress has been made in computing and software applications improving most professional, scientific, communication, industrial and other areas, the detection of quantities like cognitive load, emotion, feelings or affect has proved to be very intricate.

Progress in medical electronics and physics eventually helped to establish the capability to interconnect processing units with sensing elements. Readings of ambient thermal energy or chemical composition of organic materials was made possible by employing resistive temperature sensors and electro-chemical PH sensors. Further advancements in signal processing and instrumentation have given birth to a vast area of applications in sensing and

### **Chapter 3: Literature Review**

monitoring systems. Improved processing speed and accuracy made processors ideal for applications in most areas of science, the industries and commerce, offering flexible and convenient representation of data. Applications in sensing of physical measures such as temperature, weight, altitude, flow etc., was based on alterations of chemical (e.g. oxidation), or electromechanical (e.g. stress/load sensors) properties of certain materials occurring under certain conditions. Just as sensing of primary quantities like temperature or weight was made possible, similarly, secondary quantities could then be derived indirectly. Human ingenuity and inventiveness, combining the above capabilities in application areas such as those in medical technologies and patient monitoring has developed a number of relevant devices improving dramatically patient diagnostics, clinical assessment and monitoring. In order to gain a better understanding of the human assessment model that would demonstrate and clarify the possibilities to achieve detection of some forms of the undisclosed inner self, a review of relevant biological and physiological narrative has been set out.

#### **3.2 Physiological Correlates of Psychosomatic Disposition**

To begin with, one should imagine a human body resolved in a perfectly balanced somatic, physiological, pathological and emotional condition termed in physiology as the ideal state of '*homeostatic balance*'. The brain achieves homeostasis by maintaining equilibrium in the somatic system which induces or suppresses activating agents via its sympathetic and parasympathetic subsystems respectively (Tortora & Grabowski, 2006). By nature human beings when attempting to observe or sense their current mental activity they recognize that they have no means to substantiate expounded or rationalised quantification of cognitive or other mental processes as they take place in their brain. Expressions of psychological reactions transmitted as hormonal, neuron-mediated or combined responses associated to emotional stimuli have no imprint that can be identified and recognised by human cognitive capabilities as it happens for example with hunger, thirst, pain or similar feelings. Although the homeostatic balance in both, the Autonomic Nervous System (ANS) and the Somatic System (SS) is disturbed and altered during an affective condition, the human brain reacts to sentimental stimulations similarly to that of many other physical stresses in an attempt to counterbalance the disproportionate equilibrium states with a response via the Sympathetic

## **Chapter 3: Literature Review**

Nervous System (SNS). While the above condition has been counteracted and begins to fade out, the somatic system counterbalances the after effects initiating an equalizing response via the Parasympathetic Nervous System (PNS). Studies in human physiology have established strong biological correlations between chemical secretions accomplishing mental processes and psychosomatic responses which develop into observable forms of expressions. The above processes arbitrated by the brain and expressed as interventions via the somatic nervous system have been the focusing field of study of psycho-physiology. Subliminal assessment has been a major subject of study by employing formal psychological assessment as well as facial expression appraisal. Verification of the presence of the above brain activities has been achieved by studying models of affective stimulation-reaction measurements in vitro using laboratory instrumentation like Electro-Encephalo-Gram (EEG) and functional Magnetic Resonance Imaging (fMRI). The above techniques are being used extensively in medical diagnosis and due its sheer size, only relevant literature has been included in the appendices accompanying this document.

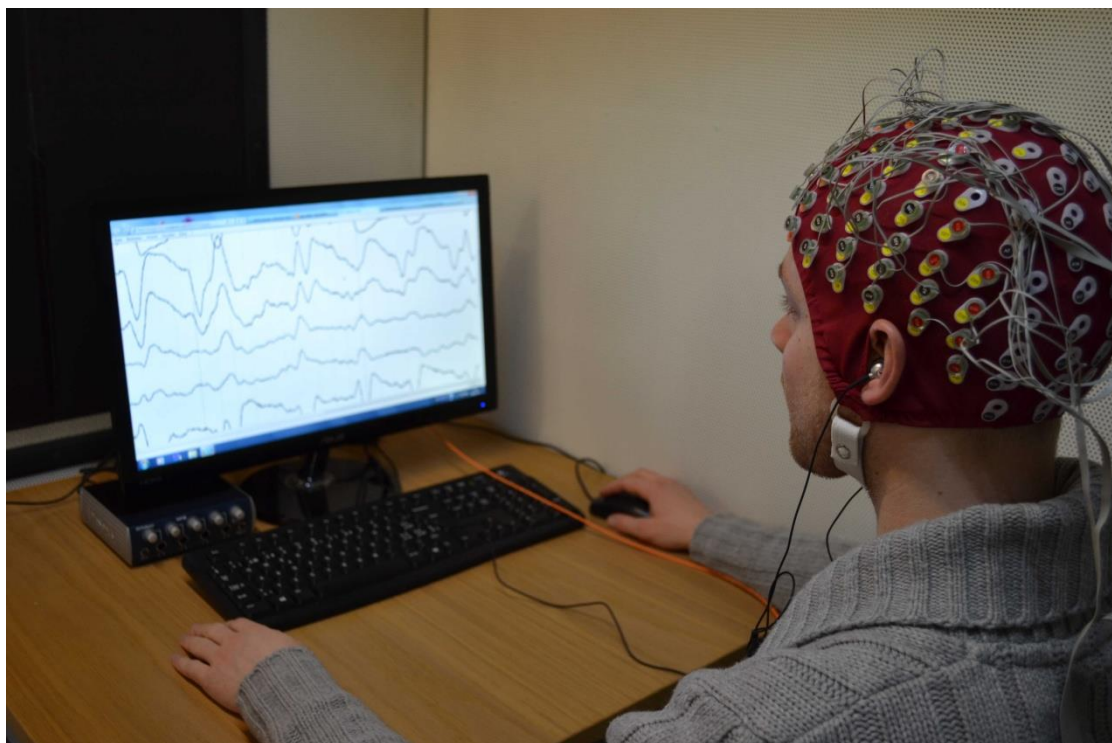
### **3.3 Biological Instrumentation**

Traditional applied methodologies for deriving state of attention and user engagement have been based upon interpreting external characteristics such as postural archetypes, facial expression, and neuromuscular activity in an effort to identify indirectly hedonic levels of satisfaction by using empirical protocols. Wedding conventional methodologies of learning, education and assessment with a systematic classification of physiological responses becomes therefore a promising area in improving educational effectiveness posing nevertheless a number of serious practical impediments. Studies to date (Phan et al., 2003) have produced corroborate proof that emotional reactivity can be identified and substantiated by using combined instrumentation for measuring Electroencephalography (EEG), Electromyography (EMG), functional Magnetic Resonance Imaging (FMRI), Electrocardiography ECG), Skin Conductance (SC) and Respiration Rate (RR). Technologies and principles of operation of the above equipment vary, however, accuracy and dependability of readings allowed to be established in use by medical doctors in clinical practice. Stringent rules for approval in medical sciences ascertain beyond any doubt the credibility and accuracy of related equipment.

### Chapter 3: Literature Review

An assessment of the capabilities of the above technologies against each other when used for the purpose of detection of psychosomatic condition is necessary, exploiting particular advantages and disadvantages of each technology. Also relations between combined technologies have to be analysed in order to reach an optimisation of nursing capabilities that can be achieved.

EEG, is based on a high sensitivity multi-node micro-electrical signal amplifier which is applying advanced signal processing in order to produce actual neuronal activity of the brain in a form of frequencies in the range 4-40 Hz. Figure 3.2 shows an EEG diagnostic test setup, demonstrating among others the significant limitations imposed on the subject as far as movement restrictions is concerned.



*Figure 3.3.1 EEG measurement session*

Frequencies deduced are often grouped in sequence and are known as bands. Theta band, for example, is the name given to frequencies ranging from 4 to 8 Hz. These bands reflect specific and different cognitive processing abilities in specific areas of the brain (Lubar et al. 1995).

### Chapter 3: Literature Review

Thus, the computation and analysis of frequency bands within power spectral density (PSD) combined with numerous researches on alertness and attention provides a powerful tool for monitoring and mapping mental engagement. As previously mentioned, (Pope et al. 1995) developed an engagement index using three EEG bands: Theta (4–8 Hz), Alpha (8–13 Hz) and Beta (13–22 Hz), at a ratio  $\text{Beta} / (\text{Alpha} + \text{Theta})$ . This ratio was found as the most effective when validated and compared with many other indices (Freeman et al. 1999). Engagement index is calculated by applying a Fast Fourier transformation to convert the EEG signal from each active site into a power spectrum. Bin powers (the estimated power over 1Hz) were summed together with respect to each band in order to compute total power and produce the EEG band ratio. Combined power is the sum of band power computed from each measured scalp site. The EEG engagement index at instant T is computed by averaging each engagement ratio within a sliding window of time preceding the instant T. Two methods have been used to deduce this index. 1) The slope method, i.e. the slope of successively derived engagement indexes (say every 2s) is calculated. More importantly the sign, negative (Low engagement tendency) or positive (high engagement tendency) is also considered. 2) The absolute method, where an engagement threshold is calculated by averaging engagement index values over a period of time prior to testing (baseline). During task performance, engagement index exceeding the threshold is considered positive and values below the threshold negative.

fMRI can be considered as an enhanced version of Magnetic Resonance Imaging (MRI) that measures metabolic function by detecting blood oxygenation. It employs a strong magnetic field of about 3 Teslas (around 50,000 times the size of the Earth's magnetic field) and relies upon the presumption that neuron activation and oxygen consumption in the brain are interrelated. Measurement is showing the hemodynamic response of brain tissue in activation maps and its value relies upon its own individual units known as 'voxels'. The 'activity' in a voxel is defined as how closely the time-course of the signal from that voxel matches the expected time-course. Voxels whose signal corresponds tightly (i.e. stimulus to hemodynamic response time is small) are given a high activation score, voxels showing no correlation have a low score and voxels showing the opposite (deactivation) are given a negative score. These can then be translated into colour coded activation maps.

### **Chapter 3: Literature Review**

Electro Myo-Graphy (EMG) is a technique based upon detecting electrical activity of kinetic muscles that is subsequently associated to specific actions. An EMG instrument is used for indirect derivation of emotional expression by measuring contractions of facial and eye muscle in particular. Similarly it is used in experimentation involving multimodal acquisition as a complementary recording for event synchronisation and verification of event timing. Other uses involve muscular efficiency and activation monitoring for clinical treatment in physiotherapy and rehabilitation.

Electro-Dermal Response (EDR) was considered by Venables and Christie (1980) and Fowles (2007) highly important because of the involvement of the sweat glands that essentially create this measurement. Since sweat gland activity, in turn, is controlled by sympathetic nerve activity, this measurement has been considered as an ideal way to monitor direct activity of the autonomic nervous system.

Indirectly EDR represents a proportionate manifestation to stimulations of the sympathetic nervous system. There is a distinction however whether the measurement is of the tonic level (L) type, or the time-varying (phasic) response (R) type. Tonic level is the trait of responses representing responses above or below a baseline while the phasic response indicates immediate measurements with amplitude representing vigour and pulse width the sharpness of the expression (Malmivuo & Plonsey, 1995). In continuous measurements when the skin resistance (SR) method is employed, the stability of measurement is decaying and also acquisition is affected by body temperature; therefore repeatability of measurements becomes unreliable. A major breakthrough has been achieved in the principle employed in our design of the psychological stress measurement system by implementing a novel approach for acquisition based on current measurement instead of voltage measurement as is the case in most devices. Measurement is obtained by a repeatedly fed back micro-current signal (in the order of  $0.06\ \mu\text{A}$ ) in combination with a trans-conductance feedback loop amplifier obtaining a self-regulated output with high linearity and stability. The aforementioned concept is proven from many high precision electronic circuits and it is particularly efficient in long term high sensitivity equipment like seismographic devices and laser alignment instruments used in medical equipment.

## Chapter 3: Literature Review

### 3.4 Evolution of Educational Technologies

In recent years the new era of digital and information technologies in general have transformed education significantly altering presentation of information and offering unprecedented capabilities and tools to the educators. Historically, the role of technology in education has always been closely linked with conveying information and teaching. Since the early days of scripting and writing people were seeking ways to leave indelibly their thoughts and messages to later generations. In the old days as stated in the Bible, Moses wrote the Ten Commandments carved into a stone around the 7th century BC, while later on the scripts were mainly written on papyrus. Regardless of all efforts made in ancient times to preserve written material for future generations the means of documentation was underdeveloped and cumbersome for the largest part of the society.

A method that seemed to be more easily accessible to everyone was oral education relying upon passing on learning to the next generations as found in ancient Greece where Homer's Iliad and the Odyssey were two indicative samples of the first audio learning methodologies. The only possibility that a context could continue to exist in the minds of the next generations was to be recited and memorised again. Scripts of the above work were found after the 5<sup>th</sup> century BC when the value of reminding the original story rather than distorting it progressively through generations was appreciated. In about 1200 AD, the term "*lecture*" evolved sourcing from the Latin "*to read*" as the scrolled manuscripts were first used written by monks. Around the same period of time, in India, had begun to use slate-boards, although the first use of blackboards has been reported to be used in schools only in the turn of the 18<sup>th</sup> century. During the 2<sup>nd</sup> world war the U.S. Army first used overhead projectors in training urged by the needs to train military personnel (Reiser, 2002) and around 1990 the electronic projectors and PowerPoint emerged as a revolutionary tool for presentation advancing educational technologies to the point of enhanced visual presentation.

The major leap forward brought by the aforementioned technologies was the matching of visual and oral expression that improved consistency in teaching considerably by eliminating possible misinterpretations of context since perception was verified by both oral and visual resources.

### **Chapter 3: Literature Review**

Although the use of technological tools for presentation was primarily invented for business and military briefings, the concept was soon established in educational fields as well.

In spite that other tools like the telephone which was discovered in the late 1870's, it wasn't used as an educational tool before around the 1970's when the concept of distance learning was born, (Harting & Erthal, 2005), helped by the drop in costs of telecommunication. Video-conferencing came out as the solution to bridge not only travelling costs but also importantly to cover educational or business needs in many places simultaneously. Progress in video compression technologies and cheaper computer storage has supplemented the outbreak of recorded lectures and entire distant learning courses as early as the beginning of the year 2000. After 2008 the above tools were offered in form of live-streaming in form of webinars and live lecture broadcasting allowing one to attend from literally any place on earth.

Recently introduced advanced interconnection facilities available in the developed world by the Internet can only be paralleled to the invention of the printing press in the 15<sup>th</sup> century which however, was the pioneer of formal education as it promoted literacy and the need to become more analytical and knowledgeable by spreading a surplus of written documents in all subjects of interest. Other concepts of education owing particularly to developments in transport and cost effective mail infrastructure formed by Universities, known as formal correspondence education. In 1969 the Open University (OU) was established in the United Kingdom incorporating a combination of video and written material from the British Broadcasting Corporation (BBC) that was involved in the first educational programs since the 1960's, offering integrated degree courses.

Latest developments of web-based educational platforms have displaced the until recently growing television based education diminishing the drawbacks of higher production costs, opposition from educators, local language stereotypes and cultural issues and even the lack of electricity in some places. Reduction of costs and the lecture capture technology has added to previous technologies the advantage to enable attendance and repeat viewing of lectures at a time and place of their choice, through an Internet connection. Massachusetts Institute of Technology (MIT) began a program offering its recorded lectures to the public free of charge via the Open Courseware project in 2002. This was the beginning of a new era since previous

### **Chapter 3: Literature Review**

dedicated Learning Management Systems (LMS's), required supervisory intervention by instructors and tutors for each effectively redesigned course environment, however, technologies for lecture streaming required no changes in course design whatsoever. YouTube was conceived in 2005 presenting an educational source in a multitude of topics from general, technical and advisory to scientific fields and has been used gradually until now for short educational clips that can be downloaded and integrated into complete courses.

Computer Based Learning (CBL) was aiming to produce a computerised version of learning by forming structured information integrating sessions for testing learners' knowledge, and it could therefore provide targeted feedback to learners, with no need of human intervention beyond the initial design of relevant hardware, software and content including teaching and assessment material. Fred Skinner started experimenting with teaching machines that made use of programmed learning in 1952, based on the theory of behaviourism. Skinner became famous for his theory of operand behaviour and operand conditioning and he demonstrated that humans behave according to consequences from previous familiar stimuli. Human behaviour has been shown to be prominently influenced by positive, negative reinforcement or both in the sense of strengthening their action by using experiences from others. Many more theories were developed explaining human perception, learning and thinking processes from famous psychologists (Freud, Grace et al.) providing constructive background material when it comes to designing graded CBL courses. One of the most widely accepted theory to date is that of Jean Piaget, who asserts that humans since their early age go through a number of fixed stages on their way to independent thinking, known as the "theory on cognitive development". Artificial intelligence (AI) environments started in the mid-1980's to replicate teaching processes on arithmetic and selected subjects with hardly any promising results so far. The reason that AI has proved difficult to produce respectable scores in teaching was that it was found difficult to cope with the extraordinary variety of ways and complexity in which students learn. Cognitive neurosciences have also attempted to designate the established rules of the science that govern and predict specific learning behaviours. (Koedinger et. al., 2004). For the purpose of this study however, educational technology developments were mostly studied as viewed in terms of assessing engaging conditions and vivid emotional reactions. 4

### **Chapter 3: Literature Review**

Learning Management Systems (LMS's) were essentially successors and outcome of the World Wide Web around the mid-1990's enabling users of the internet to create and link documents and digital media with flexible code independent tools incorporated into a web browser. Search engines fulfilled the task to simplify previously laborious tasks to cope with substantial amount of relevant information. LMS's apart from instant performance scoring introduced completely new types of information unknown before such as response times, study duration, optimal time before fatigue of the learner and more.

More recently we have seen the development of adaptive learning, which analyses learners' responses then re-directs them to the most appropriate content area, based on rated performance criteria. Learning analytics, effectively collect data about learner activities and relate them to other data, such as student performance, progress assessment and scoring, (Mourlas, Tsianos & Germanakos, 2008).

Proving that a specific educational methodology is effective has been an arduous task due to the fact that general scope teaching methods cannot cater for particular learner capabilities or make provision for particular learning idiosyncrasies. Individual student abilities and needs, distinct personality, learning capacities, personal aptitudes, psychological profiles, personal motivation, environmental and cultural factors or even nutrition and physical condition and fitness are all contributing factors affecting the maximum potential of the learner's abilities to perceive, comprehend and memorize a particular context, (Haapalainen, et al., 2010).

Efforts to stretch beyond the conventional teaching methods employing behaviourism, constructivism or even methods based upon activity theories applied by tutors, have led to extensive research and the development of numerous psychological tests. Methodical assessment employing the above tests constituting batteries of formal psychological assessment can reveal distinctive characteristics for each student individually providing a discreet classification into personal profiles. Student profiling has been adopted by educators and teachers as an invaluable input for the selection of the above teaching techniques and methodologies in the classroom.

## Chapter 3: Literature Review

### 3.5 Personality Profiles and their Role in Education

Searching for optimised methods for classification and assessment of personality characteristics and aptitudes, psychologists have adopted a number of theories beginning with Allport and Odbert as early as in year 1936, next by Cattell in 1957 and later in 1967 by Norman who finalised the most prominent method in use to date standing as the “*Big five theory*” (Digman, 1990). In that, personal or personality profiles designate the levels of favourite inclinations towards one of five distinct personality indicators identified in the group areas of inventiveness / openness, conscientiousness / carelessness, extraversion / reservation, agreeableness / uncompromising and neuroticism / confidence. The Big Five Theory is effectively a template used to identify five distinct personality factors as a midpoint for each individual; Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (OCEAN). Each of the five factors has six distinct facets as referenced below since they have complemented the experimental procedures used in this text.

Openness to experience focuses on the tendency towards art, beauty and general perception to novelty. The facets for openness to experience are:

- Aesthetics: strong appreciation of art and beauty; enthusiastic about poetry, art, and music
- Actions: willing to explore new challenges, try new foods, and experiment with new activities
- Fantasy: active mental life and strong imagination; rich and creative personal life
- Feelings: interested in someone’s own feelings and emotions; feel emotions very intensely
- Ideas: intellectual curiosity; enjoys philosophical idealistic arguments and brain-teasers
- Values: ready to explore and evaluate other people’s social, political, and religious values

Individuals who score high on openness to experience are expected to be very creative and willing to new ideas and activities. They typically possess a rich and satisfying internal life, often spending their time thinking about concepts and artwork seen recently. In contrast, those

### Chapter 3: Literature Review

who score low on openness to experience are more inclined toward conventional thought. Their interests are supposed to be narrow and they tend to be practical and realistic.

Conscientiousness exposes characteristics pertaining to organization and perseverance. The six facets of conscientiousness are:

- Order: enjoys tidiness and methodical approach to tasks; neatness in life
- Competence: capability to deal with challenging experiences in life and difficult circumstances
- Achievement-Striving: willingness to work hard; strongly goal-driven personality
- Self-Discipline: ability to follow through with tasks while controlling distraction
- Dutifulness: governed by conscience; very dogmatic in values
- Deliberation: tendency to carefully contemplate decisions before acting

High scorers on conscientiousness tend to be hard workers and very dependable. You are likely to see a high scorer creating many to-do lists and breaking down large goals into achievable steps. High scorers pay attention to proper organization and take a methodical approach to achieve their goals. They are prepared to dedicate an enormous amount of effort to succeed. Those who score low on Conscientiousness tend to be more impulsive and laid-back. They are usually characterized as spontaneous in their approach to academic and vocational situations. They go with the flow and refrain from schedules and concrete plans.

Extraversion focuses on sociability and locating where individuals derive their energy from. Low scores on this dimension tend to indicate a more internal source of energy, while high scores denote an external source of energy. The six facets of extraversion are:

- Warmth: easily developing close attachments to other people; friendly and affectionate
- Excitement-Seeking: longing for excitement and stimulation; preference for intense life
- Positive Emotions: tendency to experience positive emotions; partiality towards optimism
- Gregariousness: preferring the company of others; avoidance of being alone
- Assertiveness: tendency to lead and dominate social situations

### Chapter 3: Literature Review

- Activity: energetic disposition; fast-paced lifestyle and tendency towards a busy life

Individuals who score high on extraversion tend to prefer to be with and in the presence of other people. In most cases they present the “life source of the party”. They favour being underlined focus of attention and frequently engage in thrill-seeking behaviours. They leave social occasions feeling excited and full of energy. Low scorers on extraversion tend to be more introverted, talking very little and reserved in nature. They feel drained and exhausted when surrounded by people. They prefer more solo pursuits, like reading. Their lifestyles are more slow and deliberate, and they possess an inclination towards quietness.

Agreeableness indicates the closeness of a person to the idea of trust, honesty and compliance. Individuals who are agreeable tend to be more straightforward and to the point by nature. The six facets are:

- Trust: inclined to believe that others are honest and well-intentioned
- Straightforwardness: sincere and genuine in expression of opinions and thoughts
- Modesty: humbleness in speaking of own accomplishments
- Altruism: strongly moved by and dedicated to the promotion of the well-being of others; extremely generous
- Compliance: inhibition of aggression, deference to others in interpersonal conflict
- Tender-mindedness: highly sympathetic and concerned about others

High scorers on agreeableness are typically mild-mannered interpersonally. They tend to remain loyal and look for the best in everyone they meet. They can be thought on to be generous, honest, dependable, and very concerned about the well-being of others. Individuals who score low on agreeableness tend to be more suspicious of the motives of those they come across. They are frequently cynical and sceptical about the world around them. Additionally, they are more willing to use flattery or craftiness to gain favour with others.

Neuroticism focuses on tendencies of an individual to concentrate on the experience of negative emotions. Individuals who fall in the neurotic category tend to be more prone to mood swings and emotional reactivity. The six facets of neuroticism are:

### Chapter 3: Literature Review

- Anxiety: fearfulness, restlessness, tenseness
- Depression: propensity to experience depressive symptoms, such as loss of energy, difficulty concentrating, and issues with sleep
- Self-Consciousness: discomfort around others; frequent experiences of shame and embarrassment
- Angry Hostility: tendency to experience frustration and bitterness, as well as anger
- Impulsiveness: inability to control cravings or urges
- Vulnerability: difficulty contending with stress; dependence on others for support

Those who score high on neuroticism tend to experience negative emotions very intensely and have difficulty controlling these emotions when they arise. They are more vulnerable to psychological distress than individuals who score lower on this facet and tend to pay a significant amount of attention to their own behaviour in interpersonal situations. Low scorers on neuroticism are typically more stable in their experience of emotions. They are more calm and relaxed in times of stress and tend to be quite slow to anger. They usually trust their ability to handle stressful situations and do not internalize awkward social situations.

It should be mentioned for the model of big five that a high or low score on any particular factor is not necessarily distinguishing between good and bad. For example, there are situations where being more compliant and inclined to trust others is beneficial and there are other situations where a more sceptical approach would be the wisest choice.

Another dimension of personality is explored by trait anxiety which expresses neuroticism versus emotional stability. Trait anxiety tests reveal a person's stable perception and tendency to attend to, experience and report stressful emotions like worries or fear during a situation of negative valence and arousal. Trait anxiety is thought to reflect a cognitive-perceptual bias. In the perceptual level, expresses an excessive attentional bias to threatening stimuli. At the cognitive level has been observed a distorted negative interpretation of information promoting anxious responses. At the level of memory there is a disproportionately high recall of threatening information, (Benetti & Kambouropoulos, 2006). In this study the use of trait

## **Chapter 3: Literature Review**

anxiety manikin is aiming to deduce adaptation levels of the learner in correlation with the direct measurement of psychological stress level deduced by STC.

### **3.6 Psychosomatic Disposition During the Course of Learning**

Although the topic of cerebral workload has been explored extensively by bio physiologists, medical scientists and psychologists for the past 40 years, there is no unanimously defined nor a collectively established definition. Wickens, (2012), refers to the term "workload" as not even used widely before the 1970's and that the operational definitions of workload from various fields are in conflict regarding its mechanisms of composition, quantified methods to reproduce and concrete measurement." Constituents of workload in the most commonly accepted theories have been divided in three major categories:

- I. the magnitude of work and number of tasks,
- II. time duration and segments of time where one is highly concerned,
- III. and finally the psychological experiences that the human operator is subjected to (Lysaght, Hill et al. 1989).

As a conceptual construct, workload is contemplated as a latent or an intervening variable, representing immediate demands for mental activity imposed on humans during interaction between engaging tasks. (Gopher & Donchin, 1986). As capabilities and effort vary between humans in the context of specific situations the workload was thought to be depending on multiple parameters and was characterized as multidimensional and multifaceted, (Hancock & Caird, 1993). Multiple resources that constitute the brain activities involved during workload have been identified as memory, cognitive, perception, visual, emotional, motor and evaluation processes, (Wickens, 2008). Memory has been defined in terms of physiology as a very broad function engaging various areas of the brain with different characteristics of potency and intensity. In terms of psychology it has been effectively approached as a capacity of the brain to store and revoke information with different attributes and therefore different abilities have been associated with distinct systems ( Craik & Jennings, 1997). Memory functions have been recognized selectively through impaired clinical cases and individual testing; however, there is no commonly accepted view as to the number of memory systems, as they are viewed from

### **Chapter 3: Literature Review**

different psychological perspectives. As a general rule, a view of five distinct types of memory, has been adopted; “Semantic memory” and “episodic memory” (jointly called declarative or explicit memory), “procedural memory” and “priming or perceptual learning” (jointly called non-declarative or implicit memory) and “working memory” or simply memory (Sherry & Schacter, 1987). The first four memory systems constitute the long term memory systems and working memory is known otherwise as short term memory. Semantic memory is dealing with processes related to facts. Episodic memory is related to historical and biographical sentimentalities. Procedural memory is activity pertaining to performance of skills. Priming memory is facilitated by prior exposure to a stimulus. Finally working memory is a form of short term manipulation of information involving processing of recently realised stimuli.

A number of scientific assessment methods of the above forms of memory are indicatively: Wechsler Memory Scale, Rey Auditory Verbal Learning, California Verbal Learning Test, Cambridge Prospective Memory Test, Rivermead Behavioural Memory test, Test of Memory and Learning and other. The subject of memory function has been cited extensively by a vast amount of literature and scientific material, methodically expanding into areas such as cognitive functioning, visuo-spatial information processing and memory involvement in emotion. For the purposes of this study, memory function and its short term component in particular is seen merely as one constituent of the cerebral workload that is assessed quantitatively by neurophysiological measurements of HR/STC. Also cognitive, visual, motor activities and associated processes have not been looked into any further as they are not differentiated in this study but instead they are seen as components of the overall cumulative involvement of the participant.

#### **3.7 Mental Workload and Attention**

Workload has been widely recognised as a mental construct incorporating the accumulated mental effort exerted when performing one or more tasks under specific conditions (Cain, 2007). The capacity of the operator to respond to certain demands of a task can also be affected by external factors such as disruption, noise etc., while emotional components have been found to also affect decision making (Bechara, 2004). Frequently psychologists classify emotion

### **Chapter 3: Literature Review**

within the conceptual framework of rational cognitivism, (Lazarus, 1982; Oshner & Gross, 2005). Cognition refers to mental processes pertaining to memory, language, attention, decision making and problem solving; however, emotion has been confirmed also as a cognitive activity (Ossner & Gross, 2005). References also made in the above literature demonstrate cognitive involvement of emotion in numerous experiments of emotional regulation. Experiments in neuroscience have demonstrated the interdependence between emotion and activities that previously required only rational thought such as decision-making and problem-solving (Damasio, 1995).

Analogous studies expand the effects of emotion beyond mental workload into attention and decision making, through experiments on humans with emotional impairments, (Schupp et al., 2006). Event-related emotional stimulation to the above participants has shown impairment in concentration, decision making and inhibiting attention. In the context of experimental attention assessment by using the HR/STC system all the above processes that require cognitive involvement can therefore be considered interrelated mental tasks involving different functional channels.

The effectiveness of mental workload assessment has been established by psychophysiological experimental methodologies such as psychometric testing, self-assessment, behaviour and personality assessment. A highly flexible, accurate and dependable integrated instrumentation system that would measure combined physiological responses in vitro has been considered a good step forward in order to correlate mental workload with actual subconscious physiological responses. Measurements that could produce an indication of mental workload necessitate readings of direct and indirect reactions to stimuli caused by properly designed workload tasks; however, studies carried out in past experiments were mainly restricted to laboratory environments as they were constrained by the need for use of attachments, electrodes, headsets or other devices. The above restrictions cause a multitude of negative effects in the accuracy of measurements as postural and kinetic limitations preventing the freedom of expression and also more important restrictions such as the psychological load imposed on the subject by consciously knowing that measurements take place. An ideal system should be able to acquire all necessary physiological measurements with no interference,

### **Chapter 3: Literature Review**

intrusion or obstruction whatsoever and most importantly with no psychological intervention in a completely natural everyday environment.

Concluding from the above analysis, emotion causing events, cognitive involvement task with mathematical content or decision making scenarios would cause respective physiological reactions identifiable by alterations in the Somatic Nervous System. Sympathetic System Response is activated in order to counteract an urgently detected homeostatic imbalance by raising heart rate and exciting sweat glands (Tortora, 2006). It must be emphasised that the psychological stress produces physiological expressions of the biological reactions closely related to corresponding psychosomatic conditions. The STC measurement being a direct measurement of psychological stress has been selected to be an input channel of the HR/STC system, (Scherer, 2005) providing a measure of physiological stress. Immediate expressions of the Somatic Nervous System (SNS) can be detected by using STC method with timely accuracy comparable to that of an EEG signal, (Tsatsou, 2006).

Emotion has been shown to be an important constituent affecting cerebral workload influencing residual performance, attention and cognitive processing (Salovey & Mayer, 1990). The HR/STC system is a typical example of a tool attempting to provide spontaneous responses of emotional, affective and cognitive involvement by looking at physiological stress and heart rate excitations.

#### **3.8 Physiological Expression of Emotion**

Numerous studies in emotion and affect ranging from the areas of psychology (Eysenck, 1982; Frijda, 1986; Strongman, 2003; Scherer, 2005; Eysenck, 2006), clinical physiology (Lykken & Venables, 1971; Hassett, 1978; Venables & Christie, 1980; Damasio, 1994/2006; Tsatsou, 2006), affective computing (Picard, 2003), and HCI (Norman, 2004; Brave & Nass, 2008; Fairclough, 2009; see also the vast literature on User Experience [UX] design and game studies – e.g. Kuniavsky, 2010; Calleja, 2011; Law & Sun, 2012 etc.) have demonstrated that emotional stimulations effectively drive the human brain to elicit physiological responses that are direct, instant, measurable and quantifiable. These reactions appear consequently as rising or fluctuations of physiological stress, elevated heart rate and reactions promoting activity of

## **Chapter 3: Literature Review**

certain muscles. When the above reactions have been instigated during processes involving learning tasks and problem solving have been demonstrated as highly accurate indicators of a person's mental effort. Physiological measurements of heart rate excitation and psychological stress can be measured during the course of a teaching session and show an additional variable not possible with any previous formal methods.

Applied psychology used in evaluation and scoring of targeted capabilities of individuals, employ test batteries which combine a range of trials in order to provide a quantified sketch of behavioural emotional and cognitive skills. Related work has verified the value of physiological assessment in conjunction with proof obtained by the above scientific areas. The greatest obstacle when designing adaptively affective environments employing physiological data acquisition has been the challenge to acquire data reliably and in such a way that measurements do not affect the user due to restrictions that may be imposed from sensors and their connections. Also a major concern for a system deducing physiological expressions has been the elimination of the unknown psychological impact on the subject while measurements take place.

### **3.8.1 Theories of Emotion**

Many theories attempt to produce a conclusive definition of emotion but since a deeper investigation is beyond the scope of this study, will only focus on four of them as an exhaustive understanding is still required. As a means of communication, emotional event are known to have a strong impact on the exposure of bodily reactions as supported by theories of emotion since 1880's. James and Lange as referenced by Walter Cannon (1927) have come up with their theory portraying emotion as the bodily response to an event rather than a quantity caused by the event itself. Another more recent theory suggests that experiencing an emotion requires both bodily response and an interpretation of the bodily response which rely upon the particular state the person is in at the moment (Schachter & Singer, 1962). For example one can experience rising heart rate when chased by a dog because of fear, or can experience similarly high heart beat because of say excitement.

### **Chapter 3: Literature Review**

Walter Cannon, (1927), demonstrated through his experiments that in certain animals like cats, emotion occurs even if the brain was cut off from the information about bodily responses. He also supported that the same bodily responses accompany many different emotions claiming that although high heart rate may mean that a person is angry, it may also mean that he/she is positively excited. Consequently, our brain cannot just rely solely on our bodily responses to know which emotion we are experiencing but rather there must be something else that helps to distinguish whether they are angry or excited. Luiz Pessoa (2013) demonstrated that Cannon's theory complemented by Philip Bards (1958), concluded that both the experience of the emotion and the bodily response occur at the same time independently of each other and that the experience of an emotion does not depend on input from the body and how this is responding.

The Opponent Process theory of emotion, (Solomon & Corbit, 1974), explains our experience of emotions in relation to its opposites. In their view, the experience of an emotion disrupts the body's state of balance and also our basic emotions typically have their opposing counterparts. Expressed differently, they argue that when we experience some emotion then that suppresses the opposite emotion.

#### **3.9 Psychological Measures of Attention**

Attention has been widely recognised as a mental process by which a person selectively registers some stimuli and ignores others. In 1884, William James claimed that: "Attention is the taking possession of mind in clear and vivid form. It implies withdrawal from some things in order to deal effectively with others." Scientific efforts before today's technology which provide proof of data through direct measurements used a variety of assessment methods with the most important ones laid out next.

The Attention Checklist (ACL) and the individually administered tests of attention that emerged as a part of the Cognitive Assessment System (CAS: Naglieri & Das, 1997) was a frequently used method for estimation of levels of attention. The CAS battery is still used in formal assessment models providing a view of intelligence. Intelligence assessment in this battery is established from neuropsychological discoveries and on psychological findings

### **Chapter 3: Literature Review**

interpreting cognitive functions (Hunt & Lansman, 1975). Due to the distinction that selective attention tasks are divided into those in which selection occurs at the time of receiving and encoding the stimuli, in contrast to the tasks in which selection occurs at the time of response or expression, the above methods employ a number of tasks to distinctly address the problem. Stroop tasks, Visual Attention tasks, Auditory Attention tasks and Picture and Name Matching tasks are used in tandem in order to distinctly identify the above types of attention.

Expressive Attention Tests (EAT) include activities originating from the Stroop task (1935) and are used as a measure of interference. Effectively they consist of templates combining colours either as words or shapes and measure the response time of the participant for naming the colour used to print the word or shape, rather than read the actual word, as quickly as possible.

Receptive Attention Test (RAT) was also developed by Das and Naglieri, and consists of templates including four pages with pairs of letters. The first two pages contained a mix of visually alike letters among other dissimilar letter distracters. The candidate was required to circle all pairs of matching letters that were physically the same, for example UU (not UO or UV). The other two pages similarly include letters but the candidate should circle letters according to taxonomy, for example pairs like Dd are valid but not Db or Nh. Scores in both accuracy and latency measures are assessed.

Visual Attention Test (VAT), includes presentation with a number of stimuli and asks the individual to respond selectively based on specific instructions to identify numbers that appear in one form rather than another. For example characters or numbers may be asked to be found in regular or in bold typeface of print and vice versa. The task allows to participants a 90-second time limit during the two conditions. Scores were accounted for the number of correct responses and the time elapsed.

Auditory Attention Test (AAT), includes a wide selection of auditory tasks testing auditory memory employing sequential instruction templates like for example the Continuous Performance Test (CPT) developed by Rosvold (1956). Researchers have reported that the

## **Chapter 3: Literature Review**

auditory format of CPT discriminates better between children who have attention deficits and children who do not. It is anticipated that the dimension of attention assessed by auditory tasks is more relevant to the classroom situation than the presentation aspect assessed by the visual tasks, particularly in young children. The Auditory Attention Test subjects the participants to listening to a tape recording of words presented at a rate of one per second over a five-minute period and to respond by tapping on the table when they hear a certain signal word. Signal words are defined by combination of voices and categories, for example, signal words may be male voice dictating names of cars and female voice talking about dressing accessories. Everything except the signal words is ignored. Scoring includes the number of signal words ignored by the participant and also the false detections.

### **3.10 Physiological Measures of Attention**

EEG data entail countless information regarding functional, supervisory and operational aspects of the human body. Traditional physiological methods used to estimate attention using EEG and evoked potentials employ the P300 wave method. The P300 wave also known as P3 is the most essential and studied component of ERPs, which can be measured after the stimulus presentation in an EEG. The P300 is observed in an EEG as a significant positive peak 300 ms to 500 ms after an infrequent stimulus is presented to a subject. Typical peak latency of this positive wave occurs around 300 ms for most users suggesting to be related to the end of the cognitive processing, to memory updating after information evaluation or to information transfer to consciousness (Bernat, Bunce & Shevrin, 2001; Gonsalvez & Polich, 2002). In a typical P300-based experiment, three different types of paradigms are being used; 1) single-stimulus, 2) oddball, and 3) three-stimulus paradigm. The single-stimulus paradigm includes one type of stimuli called target. In a typical oddball paradigm, the subject is normally presented with target and standard (or irrelevant) stimuli. The three-stimulus paradigm consists of target, standard and distracter. Distracters are also known as probes or novels. Novel stimuli in a three-stimulus paradigm are presented infrequently and produce a P300 component that is large over the frontal/central area and is different from the typical parietal maximum P300 discrimination (Comerchero & Polich, 1999). This invention was based on the fact that P300 is elicited when the candidate is confronted with a particular stimulus for which has prior

### **Chapter 3: Literature Review**

knowledge. During cognitive measurements, factors such as memory, attention span, and level of consciousness have also to be considered.

Pupillary response is a physiological expression identified as a change in the size of the pupil of the eye, originating from the optic and oculomotor cranial nerve. A constriction of the pupil termed miosis indicates a response which may be caused in long term by degeneration of the sclera and in short term as a relaxed or inattentive condition characterised by a low level of consciousness and attention. A dilation of the pupil termed as mydriasis refers to the widening of the pupil and in long term may be caused by biochemical agents or drugs, whereas in short term or instantly by focusing action caused by arousing awakening or intense attention. Dilation of the pupil occurs when the smooth cells of the radial muscle controlled by the sympathetic nervous system (SNS), contract, while constriction of the pupil occurs when the circular muscle which is controlled by the parasympathetic nervous system (PSNS), contracts.

The Test Of Variable Attention (T.O.V.A.) was developed by Dr. Lawrence Greenberg (Greenberg et al., 1994), Head of Child and Adolescent Psychiatry at the University of Minnesota and has become since the citadel in many forms of assessment of attention employing a variety of instruments such as EMG, SC and other. Tests are based in response time assessment for Variability and Inattention tasks. As technology and computers advanced, many types of T.O.V.A. became available since 1991, providing test batteries applicable to unexplored physiological modals. In visual setup the T.O.V.A test typically consists of simple geometric figures. When a figure is displayed the subject is indicating that by clicking or otherwise through a physiological response under investigation. When the subject responds to a "non-target" it is noted as an error of commission, or impulsive. The inability for the subject to inhibit themselves is measured producing an error of omission. The auditory test is the same process. The test taker clicks when they hear the target, which is presented as a single tone. The test taker should inhibit their response when the non-target is presented by a different tone. It should be noted that this test since it is not using colours, it is not depending upon sequences and it is independent of cultural or habitual aspects it is applicable to many areas in both clinical and screening versions.

## Chapter 3: Literature Review

The T.O.V.A. batteries measure a set of different variables to determine whether or not response times and attention is at the normal range for the sex and age of the test taker providing information for:

- *Response Time Variability*: A time measurement of how consistently the micro-switch is pressed.
- *Response Time*: A time measurement of how fast or slow information is processed and responded to.
- *d' Signal Detection*: A time measurement of how fast performance drops.
- *Commission Errors*: A measure of impulsive responses; how many times the non-target is pressed.
- *Omission Errors*: A measure of inattention: how many times is the target not pressed.
- *Post-Commission Response Time*: A time measurement of how fast or slow a response is after a commission Error.
- *Multiple Responses*: A measure of how many times the button is pressed repeatedly. (Indicator of other problems)
- *Anticipatory Responses*: A time measurement how often a person is guessing rather than responding.

### 3.11 Formal Psychological Assessment

Psychometric tests are well-defined scientific methods to determine mental capabilities, behavioural style and personal idiosyncrasies. Focusing on discreet sets of characteristics met in groups with common individualities, test batteries have been formed building up on optimisation and improvements of formal questionnaires. Effectively, the abovementioned test batteries span into three main types: Ability testing, Aptitude tests and Personality questionnaires.

Ability tests measure a person's potential, for example a person's ability to learn the skills needed for a particular job. Ability tests are frequently complemented by Tests of Attainment, which assess specifically what people have learnt for example in a subject like maths or dancing

### Chapter 3: Literature Review

skills. Essentially what people have learned depends entirely on their ability in the first place so the scores on the two types of test are conceptually linked. The major difference between tests of ability and tests of attainment is mainly in the way the scores from both types of test are used. Many items look identical on both tests but attainment tests are different in one crucial respect - they are retrospective, i.e. they focus on what the person has learned and on what a person can do now that he knows. Ability tests are prospective; they focus on what the person is capable to achieve as well as their potential to learn thus a test of attainment cannot be used to directly infer ability, (Groth-Marnat, 2009). .

General ability is effectively divided up into more specific abilities, representing the hierarchical structure of intelligence that is commonly accepted in this field. For this purpose a general ability test may be composed of specific numerical, verbal and spatial ability scales brought together in a test battery. The score can then be evaluated and interpreted individually as a specific ability or aptitude measure, or together as part of a general ability measure. As there is no widely accepted definition to distinguish the difference between ability and aptitude tests, people accept to a certain extent that the two terms refer to the same thing. Aptitude is thought to be referring to specific ability and ability to be referring to general aptitude.

Psychologists have assigned to personality a particular technical meaning. The term personality resembles that it is not a single independent mechanism but closely related to other human cognitive and emotional systems. Samples of personality questionnaire in the Appendices illustrate this idea further. Personality is described as the distinct, stable and enduring aspects of an individual which differentiate them from other people, making them unique, but which at the same time permit a comparison between qualities of individuals.

In general, psychometric tests differ from one assessment to the other, however largely include two sections:

- a behavioural psychological section which usually includes a personality test or questionnaire, and a relevant exercise.

### Chapter 3: Literature Review

- a testing unit which includes an array of aptitude tests, verbal critical reasoning tests, numerical reasoning tests, abstract (diagrammatic) reasoning tests and in some cases special tests.

A personal assessment involving the identification of student capabilities and classification of students into groups with common learning characteristics presents an optimised approach when taken into account for teaching practices. Although personal preferences regarding the optimal communication with students exhibit a stronger potential for improving learning as such, they have not been facilitated by any contemporary teaching practices. Today's teaching practices are largely based on "*Keyword mind-storm assessment*" for appraisal of the accomplishment of student learning, while bodily and facial expressions are frequently used by the assessor empirically as an estimate of the correct perception and comprehension of the taught material.

A major concern to educators and equally essential to educational establishments and consequently to integral educational systems has always been the quality, consistency and efficiency of their educational faculties. Equally important to designing an effective educational system have also been the norms set out to assess effectiveness of such a system. Extended models of educational practices employ assessments as the process of determining "*what core*" would be optimally included at each stage of educational process, (Palomba, Banta. 1999), "*how valid*" is the content and its exposure to each educational level and what is the reliability of the teaching methodology in producing consistent results. The complete set of information derived from an assessment, in fact represented the qualitative aspects of an applied educational methodology. Evaluation was then performed on the outcomes of assessments in order to determine degrees of credibility and causal interpretation of the conclusions.

Validity of an assessment describes a condition where a specific method, such as a handwritten test, assesses what it claims to assess and therefore produces results that can lead to valid inferences usable in decision making. For example, a classroom test claiming to measure higher-order thinking skill but actually assessing only memorized knowledge lacks validity.

## **Chapter 3: Literature Review**

Inferences and decisions concerning thinking skills of a student cannot be justified from evidence produced by such an assessment.

Reliability is the capacity of an assessment method to perform in a consistent and dependable fashion every time it is being used. Apparently, reliability is a prerequisite for validity. An unreliable indicator cannot produce trustworthy results. In the above processes of assessment, evaluation, validation and reliability stress tests, consideration has been concentrating on efficiency of certain educational structures sacrificing thus consideration of the individualities of a student.

### **3.12 Dynamically Adaptive Educational Environments**

Educational environments have changed dramatically over the last decades inevitably affected by new technologies gradually introduced. New technological tools including interactive screens and audio visual environments enhanced the quality of presentation of educational content improving at the same time the actual quantity of information that a student can manipulate and retain. Progressively through studying more aspects of execution of educational processes, the need for provision for adaptability of the educational scenarios according to student capabilities was revealed. Adaptive Automated Systems adjust task demands dynamically (i.e., continuously) in accordance with the current level of operator workload (Scerbo, Freeman & Mikulka, 2000). For example systems employing combinations of sensors and aids for automated decision may accelerate the sensor-to-shooter loop and enhance the decision process (Adams 2001; Rovira et al., 2007). Facial expression has occasionally been used as an educational appraisal assessment method but its effectiveness is proven weak. This has been mainly due to the presumption that psycho-somatic condition, feelings and affect of a person can be expressed perfectly, always and in an identical way in all students. Apparently this cannot stand always true since not all people express their frame of mind in exactly the same way. In fact, an assessment based on facial expression is lacking reliability and universality as it can vary enormously between students with different background and culture. Also differences may build while the level of attention varies amongst students too.

### **Chapter 3: Literature Review**

Established from the aforementioned studies regarding the use of instrumentation for accurate caption of psycho-physiological expressions and assuming that today's technology could overcome particular application hurdles, a dependable approach to identify student's deliberation and concentration could be achieved by detection of physiological signals in vitro. Responses obtained are entirely involuntary, subconscious and cannot be altered intentionally by the person providing thus a generic indication of states of active engagement. Adaptive computer environments in education have been focusing at the interpretation of psychosomatic conditions in dynamically changing environments initiating psycho-physiological dynamics that affect learning in directions that seemed impossible to be explored by conventional methods. Existing educational approaches could in a best case scenario take into account an optimized presentation to learners, using techniques of student cooperation, tailor made audio-visual context and targeted guidance according to individual student personal profile classification. Following a successful implementation of the system described herein, dynamically adapted educational environments have been seen as systems able to assess physiological and somatic responses and used in order to enhance the potential of the instantaneous student ability to absorb information and respond to the learning process. Subconscious indicators derived from physiological responses measured in real time can produce identifiable reactions that can be used to interpret the cognitive load and state of attention. Attention and focusing involvement could then be used as a real time input to applications that can modify the content according to predisposed scenarios related to concluding levels of satisfaction and engagement of the user.

The integration of physio-cognitive data has been one of the most important and promising challenges for developing and significantly improving human-technology interaction by enhancing skill acquisition, performance and productivity in educational, military and industrial fields (Parasuraman 2005). Indeed, several physiological sensors were incorporated in various systems for detecting changes in monitored emotional and cognitive states (Arroyo et al., 2009; McQuiggan & Lester 2009; Murugappan et al., 2007). Although the most reliable and accurate physiological signal for monitoring cognitive state changes remains the electroencephalogram (EEG), mapping patterns of neuronal activity to a distinct action

### Chapter 3: Literature Review

consistently and similarly for all humans remains to be seen. The EEG data has a high level of time resolution and precision hampered only by differences in activation areas from one individual to another. In fact, EEG information and features cannot be used for detection of specific events but rather data extracted from Power Spectral Distribution (PSD) bands and/or Event Related Potential (ERP). ERP components have been used as inputs for linear and non-linear models to identify and classify cognitive changes such as alertness, attention, workload, executive function, verbal or spatial memory and engagement (Berka et al., 2007; Russell et al., 2005). Quantitative interpretation of mental workload from EEG data receives a great deal of argumentative discussion since brain activity may be occasionally higher at rest rather than in cases of cognitive load; thus assessment models use statistical and artificial intelligence techniques such as Discriminant Function Analysis (DFA), Artificial Neural Networks (ANN), and Support Vector Machines (SVM), (Serman et al. 1993). In a prominent study by Wilson (2005), 38 measures derived from EEG and heart rate were used to classify with high accuracy workload level and verbal/spatial working memory. An engagement index based on EEG was proposed by Prinzl, Pope and Freeman (2001) and used on a closed-loop method to adjust modes of automation according to levels of engagement reporting performance improvement. As regards education and problem solving, Stevens, Galloway & Berka (2007) have produced an engagement index using EEG that attempted to identify and relate mental changes during problem solving. The aforementioned system however was lacking an affective analysis hence had no consideration of the emotional state of the learner.

Although EEG can help to deduce brain activity and fMRI even density and location of activity in the brain, it has not been possible to deduce precisely the levels of engagement and workload as the exact operations taking place during the above conditions remain unclear. (Brouwer et al, 2012). In many cases higher neuronal activity is detected in more or larger areas of the brain which does not necessarily reveal cognitive overwork but may involve a number of supervisory or physical actions instead. Also a major hurdle that cannot be easily overcome by the above assessment methods is that they mostly employ cumulative spectral analysis and comparison between responses of large samples, hence they cannot address specifics of an individual. In an assessment of physical workload for example the HR reveals the state of effort but the power

### Chapter 3: Literature Review

expended can be estimated by using the HRV and also the duration of time that HR is high. In order to achieve something similar when looking for mental workload by using EEG, an additional source is required that would produce a somewhat cumulative index of the trait of activity. But not all active areas in the brain mean mental overwork as motor cortex and other areas of the brain may show higher physical activity, so an empirical selective process has to be incorporated making even more difficult the task of real-time assessment. Since the operational and positional patterns of activity differ between people the above task becomes even more difficult. It would have been possible to deduce precisely each individual type of workload only if the brain activity was operationally common to all humans, was identifiable and well known. Deciding to use EEG and fMRI instrumentation, correlated with emotional paths and measurements of responses that would express the result of affective stimulus has been therefore more efficient in group studies rather than precision assessment of an individual.

Adaptive Automated Systems present the latest generation of active computer based educational or automation environments incorporating the functionality to adjust task demands dynamically in accordance with the current level of operator workload (Scerbo, Freeman, & Mikulka, 2000). Parasuraman (2000) also proposed an information-processing function model for the design of adaptive automation systems.

Due to technical problems encountered in the implementation of adaptive automation systems applicable outside laboratory settings and also because of the needs for a universal approach to the problem, a model for effective human automation interaction design was adopted which identifies four distinct stages: Information acquisition, Information analysis, Decision and selection of action and Action implementation. The first two are referred to as the information automation stage and the last two as decision automation respectively.

Psycho-physiological measures are brought into play in adaptive automation based on the presumption that an optimal state of engagement exists (Prinzel et al., 2000). Many psycho-physiological indicators have been shown to reveal differences in workload (Kramer, 1991; Wilson & O' Donnell, 1988). While electroencephalography (EEG) has proven to be a valid and reliable psycho-physiological measure for use as a decision criterion (Prinzel, Pope, &

## **Chapter 3: Literature Review**

Freeman, 2001), it is affected by recording artefacts caused by electromagnetic fields and body movements (Bucks & Boucsein, 2000). Despite the above drawbacks, to date EEG is one of the most investigated and validated psycho-physiological measures of workload, with Heart rate (HR) and heart rate variability (HRV) being of great importance showing to be sensitive to levels of difficulty of a task. HR and HRV indicate both central and peripheral activity of the sympathetic and parasympathetic nervous systems. Gaillard and Wientjes (1994) and Hockey (1986) found that an increase in workload which reflects an increase in arousal and mental effort, resulted in an increased HR observed as a decrease of inter-beat interval (IBI). Scerbo, Freeman, Mikulka, Parasuraman, et al., (2000) indicated that cardiovascular measures, such as HR and HRV, provide the strongest correlates of workload. They noted that increased workload is associated with decreased HRV and increased HR (typically measured as decreased IBI). Tecce (1992) reported that individuals paying close attention to visual events showed a decrease in eye-blink frequency. Tecce supported that this was possibly due to increased information processing. In other words, the more information participants were processing, the less they blinked. In agreement with Tecce's argument, Boucsein (1992) also backed the argument that in a dynamic flight environment eye-blink rate tended to decrease with increased visual demands. Thus, eye-blink increases are associated with concentrated thought, and eye blink decreases are associated with the processing of increased visual information. An important point to be taken in consideration for future developments was that the pupil of the human eye reacts in two tenths of a second, constricting to as little as 1.5 mm or dilating to over 9.0 mm in diameter (Guyton, 1977; Lowenstein & Loewenfeld, 1962). Hess and Polt (1964) found that a 4-30% increase in pupil dilation was associated with an increase in cognitive workload. Additionally, Beatty (1982) and Bucks and Walrath (1992) demonstrated that pupillary responses are sensitive to information-processing demands (i.e., cognitive workload). Later, Marshall (2000) developed an algorithm that showed a high correlation between dilatory pupil activity and cognitive workload.

### **3.13 Analogue and Digital Signal Processing**

Raw signals of physiological data are notoriously noisy and can be particularly difficult to extract from their original analogue counterpart. Indicatively EEG signals are of the range of

## **Chapter 3: Literature Review**

0.2 – 50  $\mu\text{V}$  peak-to-peak for a healthy individual with noise levels from muscular and other interferences reaching 1000mV (Teplan, 2002). Extracting the actual waveforms is therefore a huge challenge since the actual signal value will be 0.5mV in an environment with noise offset of 300mV. It has only been made possible resolved with advanced electro-biological filtering methods. HR signals float in the range of 0.2-0.5 mV hampered by electro-muscular noise of much higher amplitude, however, triangulation methods employing electrode pairs are used for error cancellation. In typical skin electrical activity measurements like EDR, sources of noise originate from movement of contact area creating a variation in contact surface area. This is caused because skin resistance changes if the device is measuring EDR by means of Voltage flow. Also voltage related responses tend to fade out after some time of continuous contact with the sensing element because the skin electrical potential compensates for the voltage difference causing disproportional transfer of ion charges of the electrolytes of the skin cells; as happens with electrostatic charges developing on clothing. Essentially this is the reason that most commercial products have tarnished inefficient and suffer from poor repeatability issues after continuous use.

It should be noted that in the implementation of STC measurement developed for this study, a different approach has been used employing micro-current sensing which has major advantages over the voltage amplitude method. This is because micro-current sensing provides a pure signal unaffected by muscular electrical noise while at the same time is not depending upon surface area and remains dependable when used for longer time intervals. Artefacts caused by movement of the palm of the fingers essentially simply disappear since the current flows in the same amount through a small or a larger skin surface area. A proof of the efficiency of the micro-current implementation can be found in industrial and military communication systems where 4-20mA current is used for transmission instead of signalling methods based on voltage pulse-train.

### **3.14 Proprietary Development vs. Commercial Products**

Typically as has been mentioned in other parts of this text, in comparable studies of bio-sensing and physiological data acquisition, commercially available laboratory equipment have been

### Chapter 3: Literature Review

used for the experiments. Such devices offer confined flexibility of the system and at the same time introduce a number of errors as explained below. Serious drawbacks in such cases are overlooked, reducing this way the credibility of the results for a multitude of reasons such as:

- the presence of electrodes has a negative effect on individuals
- imposed restrictions induce a psychological impact on individuals introducing hard to estimate factors of error and also with varying magnitude
- the cost can be extremely high when high accuracy is required
- synchronisation between equipment is difficult particularly since devices produce data at different rates and in most cases in different timing, format and range
- event triggering is difficult to be recorded in all devices simultaneously
- data produced in different format needs transformation so they become uniform
- different devices produce different amounts of data per unit of time
- data represent quantities often difficult to comprehend and therefore it is difficult to identify errors
- processing of data is performed by different proprietary software making it difficult to derive information by combining data from different devices in real time
- dependability of the system degrades to the device with the lower specification
- software development is difficult and error prone, so much rather developing actively adaptive platforms
- calibration of independent components as well as a system can be a concern

In contrast to the above solution where separate commercial devices were used, a system designed and developed specifically for a dedicated assessment just as the one used in this study has distinct advantages as an integrated system. The aforementioned disadvantages of the commercial products have motivated the author to design and develop the HR-STC system. The major advantages of the HR-STC system versus combined medical equipment have been laid out below:

- homogeneity of data and synchronised timing of acquisition

## Chapter 3: Literature Review

- design flexibility allows un-tethered operation with minimal obstruction that would otherwise affect the users physically or psychologically
- flexibility and expandability of the HR-STC system allows combinations of other sub-systems to be incorporated or used in parallel
- fully comprehensive data by design and therefore calamities can be easily detected
- cost has been minimised and better overall system optimisation has been achieved
- the system has no slack or delays during acquisition of both STC and HR
- data produced are in comprehensive format enabling the implementation of real time algorithms
- data sampling rate can be adjusted for higher precision as required
- processing algorithms can be enriched by incorporating provision for special cases
- real time visualisation of allocation data, trends or diagnostic graphs
- additional modules can be incorporated without affecting real time performance
- the system holds a great potential to be used as a tool assisting research in many fields

Deduced from the above, commercially available equipment have serious shortcomings demonstrating that simply could produce results in certain laboratory settings but definitely could not achieve the flexibility required so as to be used for applications like educational platforms and mobile devices in place of the HR-STC system.

### 3.14.1 Commercial Bio-Feedback Products

Seeking alternative implementation tools in order to fulfil the requirements for the experimental settings employed in this study, one has to address the major discrepancies laid out above before making a selection of available equipment. Biofeedback and physiological data acquisition devices have to be sought among products coming from the medical instrumentation areas and the clinical diagnostic devices in particular. For instance, in highly specific measurements such as skin conductance, available instruments include instrumentation amplifiers that need to be calibrated for ranges of measured values and then use electrodes; something that was excluded from the initial requirements specification. Data produced from

## **Chapter 3: Literature Review**

the above component could only be related to those from another device deducing heart rate (that also required gel electrodes) strictly on time stamps making the whole process of implementation of the common gradient algorithm almost impossible for real time processing. Another alternative option to compose a system for experimentation in this research was to choose devices for measurement and data logging of ECG as well as Electro Dermal Response (EDR) from the market area of medical research equipment. Some devices have been primarily constructed for precision measurements not simply providing accurate instant HR but additionally produce the detailed shape of the cardiac pulse that is used in diagnosis to determine operational and functional condition of the heart. Also the above devices are usually accompanied by their specific suit of software development, introducing an added component of familiarisation. Such devices, having been designed for clinical monitoring are particularly cumbersome for use in real time data acquisition as required in our experimental settings. Requirements for heart rate data in the experiments performed in this study did not really focus on the heart beat shape but rather on the frequency, so a precision device was literally overkill. Additionally the need for using electrodes could not be circumvented violating again the prerequisites for this study.

As far as the EDR goes, commercial products provide either audio output - that had subsequently to be converted somehow into a numerical form that would provide data logging capability. Data scaling would then be required with another major drawback the fact that such devices require calibration for each candidate before each session of the experiment.

### **3.15 Need for Bio-Sensing Technologies**

A thorough investigation of educational technologies as to how they have evolved during the past couple of decades reveals a convincing trend towards broadening communication in the context of teaching and learning practices from teacher-centric into a learner-centric teaching model. Effectively the need for this change was made apparent as advanced methods of educational psychology were used in an effort to identify and exploit educational characteristics that have a heavy impact on improving learning efficiency. Focusing on the

### **Chapter 3: Literature Review**

student-centric approach it was made possible by utilising new systematic approaches employing recently acquired capabilities introduced by Computer Based Learning (CBL).

Communication enhancements have enriched teaching and learning by introducing audio-visual and interactive media tools. Teaching methods were promoted to educational technologies as an umbrella term that also incorporated utilisation and management of technological processes and resources. In further advanced environments like e-learning and web based unattended study platforms, for the first time, learners could have an option to select characteristics of the content, amount and speed of presentation for their studying material through the integrated Learning Management Systems (LMS's). Educators subsequent to this fundamental improvement however, not only have to be concentrating on exhaustive content exploration and evaluation according to their educational criteria set by the educational establishments but also have to analyse and pile on the study of learning in look for optimised paths of teaching practices; letting alone the concern for student's individual learning characteristics.

This multifaceted area eventually combined formal teaching methodologies with empirical performance evaluation and also shared advanced concepts derived from scientific approaches such as those coming from educational and learning psychology, cooperative communication, sociology, computer science and more (Natalie Descryver, 2006). Instruments required to construct a teaching environment included audio-visual and interactive tools in the classroom, such as interactive screens, clickers and responders for auditoriums and projector based interactive configurations. The above systems encouraged participation and interest of the students in game like scenarios and consequently have introduced a new approach that has been shown to improve efficiency in student learning. The benefits when using such systems come not only from notions like gamification of the teaching process that increased the interest of the students but also because the context was analysed further and illustrated more than used to be, focusing on a few targeted answers rather than many insinuated or possibly implied as it was in formal and verbal teaching methodologies. This latter step was necessary for adaptation

### **Chapter 3: Literature Review**

of teaching practices with the more recently introduced technologies of Computer Based Assessment (CBA).

Ameliorating towards the idea to identify student's individualities and use them to approach their learning needs accordingly, digital technologies could be employed focusing at the flexibility of certain educational scenarios so they could be adapted to each person and compose a customized and optimised learning environment. Smart Wearable devices like smart Glasses and Virtual-Reality (VR) devices have introduced a promising new way of enhancing learner's experiences through immersive exploration learning scenarios, but further research in this area is still pending. Information could be enriched like never before providing hierarchical information display and when combined with motion and pressure sensors could enter new areas of Activity-Based Learning (ABL). The flexibility presented to the students to attend a learning session at their own time has also made an invaluable contribution to novel learning methodologies whether the courses involved direct educational content (e.g. academic material) or indirect educational targets (e.g. improving response time, memory retention and practical skills). Another advantage of the CBL approach has been the fact that a student or a learner in general could attain excerpts or complete courses in a multitude of devices and platforms, ranging from phones and tablets to net-books, laptops and desktop computers. Cloud computing and collaboration enables more schools to use cloud-based tools like Google Classroom, making it easier for students and teachers to have access to information and on whatever device they use.

Ample use of social media resources for research and information sharing eventually is expected to make its impact within but also across schools and universities. In a simulation, versus remembering test, (Brame & Biel, 2015), learners remember 90% of what they do, just 10% of what they read or 20% of what they hear. Numerous case studies suggest that one of the substantial contributing factors in effective learning is increased interest caused by teacher personality and individual competency but most importantly increased interest instigated by the attraction that learners show to the content. Subsequently, notions pertaining to motivating learners through gaming scenarios, has led to a leap towards a new era of education

### Chapter 3: Literature Review

technologies introducing the concept of “*gamification*”. These particular methodologies facilitated transformation of content into a more entertaining and amusing presentation incorporating assessment questions in an environment that by default increased student attention and involvement and thus achieved more effective learning scores.

Digital technologies could make possible to produce adaptive educational material that would cater not only for simplification of teaching content but also for developing appropriate transformations according to student’s idiosyncrasies such as their individual learning and personality profiles. Specific preparative assessments, including personality, behavioural and aptitude tests of each individual student could also play a role as an input, providing a relevant classification of students into groups with common learning characteristics. Although the task of teaching can get even more complex because then multiple teaching scenarios had to be produced for each lesson catering for the individualities of those groups, significantly improved learning results could be achieved (Mourlas, 2009).

A well put together and optimised educational scheme employing a set of advanced components in order to take advantage of the above technologies such as audio-visual instruments, educational tools, context flexibility, integral management functionality and assessment methodologies can be expected to be offering to a student the optimal educational content in the best possible way to date, either in the class or from long distance as is the case in an e-learning environment. Despite the fact that the above scheme appears to be the perfect solution that can be applied to most educational needs, it misses a segment that has too heavy a bearing to be left out concerning the instantaneous ability of a learner to reach their optimum learning capacity. More specifically, an educational setting with the above constituents is lacking a valuable component that would include a response indicative to some form of involvement and comprehension from the part of the learner. This latter view is strengthened by many studies over the last decades revealing that learning capabilities are heavily influenced by the circumstantial emotional state of the learner (Snow, Corno, Jackson, 1996). Recent scientific research studies compare different physiological variables with respect to their association with mental workload, but is still remaining unclear which variables can best

### **Chapter 3: Literature Review**

produce information about focusing involvement of an individual and how to use combinations in order to deduce specific attributes of the workload (Maarten et al., 2014).

Conventional teaching methodologies have no means to obtain proof on learner's perception or degree of comprehension. Similarly, it is difficult to ascertain one's feelings, affect, sentimental condition or other interrelated factors influencing the student abilities to concentrate while learning. Physiological psychology was founded as a subdivision of neurosciences such as psychobiology and applied psychophysiology that has as its main task to analyse the biological interactions during the evolution of emotional and behavioural traits. The above scientific areas targeted specific mental processes and sequences in order to identify endocrine activities producing or being produced during discreet affective states, however, their scope was strictly applied for laboratory experimentation for medical research and the pharmaceutical industries (Appendix B).

Formal teaching practices have reached the maximum effectiveness possible that could be achieved by means of conventional teaching and technological tools available today. Effectiveness of teaching has reached its ultimate peak exercising all possible performance methodologies in the class and other types of education, slightly overlooking unattended or other forms of distance learning. The next level in education and learning has to present more radical approaches with concrete substantiation and confirmation of the learner's perception. Computer and applied technologies from biomedical sciences engaging in education present the only alternative to take more decisive steps forward. Models implementing adaptive content based on physiological biofeedback seem to be the optimal method for classroom settings as well as unattended education.

Apparently as warranted from the above analysis, the technology to detect the human subconscious reactive responses to psychological conditions characterised by specific feelings or affect and transform them into some substantiated form is available. A selection of bio-signals entailing vital information being qualitatively adequate to deduce identification correlates to basic or clustered emotion has been the next challenge emerging. Based on the

### **Chapter 3: Literature Review**

above physiological origin, requirement for the development of a tool that would produce physiological classifiers as emotional markers for use in the next generation of educational technologies has become apparent.

In order to introduce a possible solution in the above direction, a purpose built system has been developed attempting to deduce physiological responses and investigate the applicability, effectiveness and credibility of inner somatic expressions when used for learning or other applicable areas. It has been expected that the flexibility and minimally obstructive operation of the system assessing emotional or affective expression driven by the inner brain interaction could promote further research for evaluating models related to understanding and learning performance classification algorithms.

The bio-signal acquisition system offers the flexibility to obtain physiological signals by contact rather than attaching specific electrodes and that can produce a revolutionary input to many systems concerned with the task to observe or perceive affective reactions of a person. The value that would be added to existing educational and learning processes if a system could detect and utilise the information on mental predisposition could help to produce a significant improvement over existing conventional approaches. The route however to reach such an optimal state where some device could identify feelings, cognitive states, affect or emotion with high precision is some way ahead into the future.

A decisive factor for that lack of progress in this direction described above is today's limited research mainly because the existing technologies have not yet produced practical tools for use in every day's life. Largely, by using laboratory devices for research with major drawback the need for initial calibration and familiarisation period have discouraged researchers to investigate further possible application development. As a result of the important restrictions imposed by hardware presenting the only means that could produce data from physiological activities for further study, researchers have been focusing towards different approaches employing observation methods. Such methods including eye gazing and visual area observation attempt to deduce brain activity indirectly, by studying patterns of readings in time

### **Chapter 3: Literature Review**

and frequency domains. Development of a system that can operate in such a manner so that it would make possible the implementation of seamless reading of physiological quantities revealing human disposition was the missing link that could promote research in this area enormously. Apparently the benefits obtained if an appropriate bio-sensing device was included in an educational core system are extremely important as they could add information about the inner psychosomatic condition of the learner that essentially cannot be sensed, perceived or modified even by the person himself. Subsequently, such an input would allow for expert systems development including specific reactivity and adaptation of content in personalised environments. Methods of teaching can also be enhanced by incorporating information of student engagement during teaching, opening doors to possible new educational technologies not known to date.

The reason that a device that could measure physiological quantities in an unrestricted educational environment was so important and difficult to implement was not only because the use of electrodes or attachments was prohibitively impractical but also because the system would have to operate unobtrusively with no compromise in data integrity whatsoever. In an effort to identify the optimum biological indicators able to provide dependable information on the human state of alertness, vivacity and affective disposition one has to take into account the restrictions imposed by conditional factors influencing the method of acquisition. Exemplifying further the above predicament, it is clear that a physiological signal that necessitates strict mobility constraints as for example EEG, could not be considered as a candidate for such an implementation. A valuable quantity however like the ECG, although it is derived from minute electrical activity that requires high sensitivity instrumentation, can be considered as an appropriate modal quantity since it can be obtained by using alternative methods of acquisition like optical absorption reflectometry. A creative application method is then required to make the sensing of the above signal viable. Another equally valuable physiological quantity appropriate for determining psycho-physiological condition is psychological stress measurement that also holds the advantage that it can be easily acquired with no significant operational constraints.

### **Chapter 3: Literature Review**

The effort described herein promotes the conception to assess instantaneous physiological measurements and associate their significance with potentials portraying mental workload, state of engagement or perceptual involvement. Fulfilment of the above objective entailing the construction of an apparatus capable to acquire HR and STC was set out to introduce novel technologies into new forms of practical application in wearable or mobile apparatuses. Existing wearable devices have been developed for HR detection and local display of deduced data in forms related to training and fitness. The HR-STC system provides information on behavioural responses of the two physiological markers possessing the ability to communicate this data to a platform in such a way that has not yet been attempted in this scale. Also the HR-STC system has provision for expandability including visual assessment and additional functionality as it has been specifically designed for assessment of engagement and focusing involvement. It is worth noting that the technological novelties introduced by the system developed in the department of New Technologies have been the excelling practicality regarding minimal obstructiveness and the synchronisation of measurements that is of highest importance in this type of assessment. These advantages can be seen as independent entities regardless of the validity of the assessment methodology and the verification of the initial hypothesis.

#### **3.16 A Concluding Summary**

Effective interpretation of neural activity for contexts like cognition, perception, visualisation, and imagination are reaching diagnostic regions prohibitively difficult to identify by using available technologies and cannot be approached individually. They have therefore been excluded from the objectives of this study as they are expanding beyond the scope of this research. The essential quantity to be assessed in this research has been a quantified estimate of accumulated neuronal, affective, emotional and cognitive volume as it is expressed by the bio-physiological processes of an individual. Although facial expression has been included in sets of experiments, bodily expressions, movements, visual or eye responses, verbal and nonverbal activity or sound have not been included in the experiments. The reason for excluding the above as a complementary verification of emotional, affective or other expression was deliberate due to the need to isolate the capabilities of the chosen physiological

### Chapter 3: Literature Review

quantities to identify discreetly states of involvement as they are being detected purely from our proprietary system. Existing results and models based on simulating affective conditions of humans and successively studying expressions, psychological and postural reactions have difficulties to coincide or consistently produce conclusive interpretations due to the increased complexity of the subject, for example it is difficult for a visual analysis computer based system to distinguish reliably between a smile and a grin or even between male and female unless in the second case information have been supplied manually.

Primarily, computer systems introduced in many professional applications as a tool to enhance speed, accuracy, eliminate errors; improve overall performance, quality and consistency of data. In addition to the above improvements achieved by the computers, peripherals like cameras, wearable apparatuses and headsets introduced new capabilities for human psychophysiological sensing. Advancing technologies in medical electronics applied in scientific experimentation resulted in verified measurements of the reactions of the human body to emotional stimuli. The above observations permitted to substantiate improved high precision measurements of physiological quantities which were proved to be immediate and directly analogous to affective reactions. Further optimisation including miniaturisation and incorporation of electronics into an appropriate design allowing for the development of the system described herein, however, it should be noted that the processing power of a general purpose computer five years prior to this study wouldn't have been sufficient for the implementation of the software for the aforementioned system. To demonstrate the above issue further one has to allocate real time processing capabilities per measurement for the following processes:

- Acquisition of data from the two audio channels entailing two data packets consisting of 96000 bytes of 16 bit length and copying into data buffers.
- Digital noise filtering (2 X 96000 bytes \* algorithmic calculation)
- Digital smoothening function (2 X 96000 bytes \* algorithmic calculation)
- Fast Fourier Transform on the chunks of 96000 bytes for the STC data
- Savitzky-Golay filtering on each packet of HR packet of 96000 bytes

### **Chapter 3: Literature Review**

- Wavelet feature extraction on HR channel
- Pulse extraction algorithm of the HR packet with missing pulse and ectopic bit prediction algorithm
- Factor analysis and weighing threshold conversion algorithm for calculation of valence and arousal
- Display allocation onto Russel circumplex model of affect
- Recording function of raw data into a file

In view of the existing growth in neurophysiological and neuropsychological research further advancements may soon be achieved, bringing this effort to inconceivably higher precision. Advancements entailing details in manifestation and expression of emotion such as time, vigour, duration and superimposition attributes would allow to classify physiological responses with sufficient precision to recognise even emotional constituents.

A collaborating effort was made in this study entailing scientific resources from biomedical engineering, human physiology, cognitive, behavioural and educational psychology, affective computing, adaptive environments and software interface design. The portion allocated to each scientific area has been unfolded in this review providing sufficient background information for the reader to capture all areas involved.

## Chapter 4: Research Hypotheses

### 4. Research hypotheses

The HR/STC instrument is producing the two corresponding physiological quantities of HR and STC as they are being instigated by the learner during sessions of various tasks. HR and STC constitute the dependent variables. It is hypothesised that:

- HR and STC quantities compose a variable that represents factual effects caused by reactions of the Somatic Nervous System (SNS),
- increasing or decreasing quantities of measured responses of HR and STC with regard to their previous measured value constitute an index variable,
- the variable produced by the combination of HR and STC constitutes a fundamental two dimensional unit representing Cartesian coordinates in a bi-axial space,
- instruments used for the experiments have proven validity with regards to sentimental, cognitive, emotional or affective stimulations that can cause to a person,
- pre-validated test batteries can ensure the validity of tasks used for the experiments,
- external factors such as noise or other ambient disturbances did not affect the validity of measurements,
- corresponding the quantities of HR to valence and STC to arousal respectively was a valid approach to represent psychosomatic condition in a bi-axial space,
- comparing the HR/STC system performance with a commercial product assessing psychosomatic condition was a valid method to prove system's consistency,
- the independent variables including additional data for personality profiles and trait anxiety was a valid method to identify possible correlations between corresponding groups.

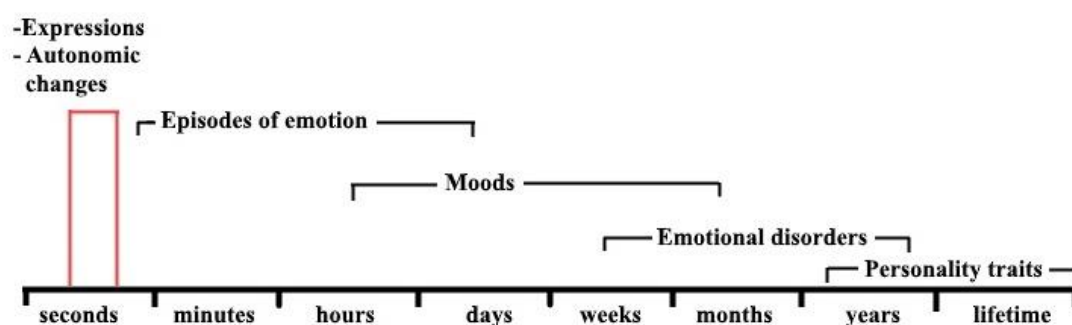
#### 4.1 Operational definitions

Qualitative attributes of critical neurophysiological processes administered in the human brain during conditions of affective and emotional activation, memory recollection and cognitive formulations have been blended for the purpose of simplification in this text with the umbrella term “*workload*”. This term has no relation to stress workload or executive workload as meant

## Chapter 4: Research hypotheses

in psychology but instead includes the instantaneous accumulation of cognitive, affective and emotional physiological expressions.

Measurements performed every 1.8 seconds represented the instantaneous perceptive reaction of the participant. In fact, derived values constitute primarily a good part of the spontaneous emotional reaction and cognitive response of the person as laid out in the spectrum of emotional events in Figure 4.1.1.



*Figure 4.1.1 Time spectrum of different emotional phenomena, (Oatley & Jenkins, 1996).*

The area designated in red, in Figure 4.1.1 shows the time that the measurements take place in relation to the full spectrum of emotional phenomena.

Effects of longer term emotional, affective sentimental, temperamental or other constituents had not been individually assessed but rather are taken cumulatively inclusive in the measurements.

Initial familiarisation period or user specific calibration is not necessary before an operational session of the HR-STC system as is the case with other devices. The need of the above tasks was effectively circumvented by looking at the measured quantities as descending or ascending values (gradients) rather than quantitative values. In essence the fact that some participants could show higher average values of heart rate or psychological stress than others, did not affect the measurements since only gradient values were used. Gradient values produced the

## **Chapter 4: Research Hypotheses**

psychosomatic index at a particular instance regardless if the participant was characterised by higher long term stress and heart rate values than another.

Concerns on the efficiency of the converging gradient algorithm in the first stages of development did not permit to set the initial aims too high regarding the precise distinction of specific affective or emotional constituents. Essential target was therefore set to attain a cumulative engagement index derived by detection of subconscious physiological quantities serving as classifiers of manifested states of engagement. Detection of precise affective or cognitive constituents or distinction of combinations of the aforementioned components cannot be attempted or claimed to be derived by using the HR-STC system.

Finally in statistical analysis, in places where parametric tests are shown, data distribution is normal. Similarly wherever non parametric tests are shown, data distribution is not normal.

## **Chapter 4: Research hypotheses**

## Chapter 5: Systems Analysis

### 5. Systems Analysis

The principles of measurement of Electro-Dermal-Activity (EDA) by measuring skin electrical permeability has been established as a dependable method of identifying responses to sympathetic and parasympathetic stimuli of the autonomic nervous system and verified in numerous scientific books (Venables & Christie, 1980; Cacioppo & Tassinari, 1990). Systems employing those principles have gained approval for use in clinical practice. Clinical studies have also correlated specific responses to stimuli with those of an emotional state of humans and the credibility and accuracy of measurement using various combinations of bio-medical instruments has been validated thoroughly recognizing the use of stress monitoring instruments for therapeutic procedures (Eysenck & Keane, 2005; Picard, 2003). Biomedical research technologies employing SC measurements and combinations of SC and HR have classified those instruments as approved devices for valid detection of verified subconscious physiological responses of humans. Just as in use for clinical assessment and diagnosis, these very instruments adapted appropriately have been employed in Interactive Human Computer Interface (IHCI) environments, whereby the responsiveness, functionality and visual context of the computer environment is modified actively according to predetermined threshold levels of psychosomatic responses. The adaptation of a specifically designed bio-feedback sensing environment, can respond to certain bio-electrical reactions of the user. Such an input can be considerably useful in dynamically adaptive environments allowing to perform specific modifications according to the actual user behavioural, cognitive or habitual profile achieving a personalisation of the interface that can improve dramatically the performance of an individual (Tsianos et al., 2009; Kort & Reilly, 2002; Kim, 2006).

The Pros and Cons amongst the principles and methodologies available for measuring psychosomatic responses through the variations of skin electrochemical composition, has also been the subject of numerous discussions with two main contenders dominating the principle of measurement; the Skin Resistance method (SR) and the Skin Conductance (SC) method. A form of improved SC method, Skin Trans-Conductance (STC) has been adopted as the method of acquisition of the generic quantity of measurement for our system. Implementing the system incorporating trans-conductance is another innovative characteristic not found in similar

## Chapter 5: Systems Analysis

systems so far providing a more stable reference signal and reliable long term data acquisition. Existing electronic products used for measuring electrical skin permeability employ one of two methods; that of voltage applying Skin Resistance and another applying a pulse modulation method. The first method become less efficient as the time goes by because the human skin appears to be getting used to the electrical characteristics of the sensing circuit. This is happening because the SC method applies voltage to the skin via the conducting elements and electron charges react with the corresponding electrical charges of the electrolytic properties of the skin making this method less efficient for long term measurements. The second method is more stable and used extensively in medical diagnosis; however, it requires expensive instrumentation and at the same time requires extensive signal processing power. An STC circuit is not using voltage charges but instead is supplying to the sensing elements a resonating current of 0.006 Amps that is not affecting the electrical characteristics of chemical charges as in the case of SC or Skin Resistance for that matter. The trans-conductance concept is originating from an ingenious technique of an electronic current feedback that is used in precision electronics and has been attempted in this type of measurement to the best of our knowledge for the first time. Additionally, a continuous micro-current measurement method has been preferably used as opposed to pulse-period measurement or other computationally intensive technique, providing a more robust solution for non-laboratory applications as it is faster, more efficient and allows better flexibility in measurements with slight movement of the contacting surface with the sensing elements. The strong parallelism of the responses of the SC in direct comparison with that of the EEG (Lewis, Critchley, Rotshstein, Dolan, 2007; Tsatsou, 2006) bolsters that the SC measurement provides comparably accurate results with those of EEG and fMRI coinciding in time and magnitude. Considering the advantage of the SC measurement to produce identifiable responses similar to higher quality medical instrumentation and at the same time to be obtainable with no restrictions in mobility or other, makes it an ideal candidate for use as an input to our attention awareness detection system. HR is the second quantity considered important when assessing subconscious mental, emotional or affective workload as it is directly dependent upon the brain auto-regulation process inducing higher or lower pacing in order to compensate for oxygen demand. HR was adopted as a

## Chapter 5: Systems Analysis

simultaneous input to our system. The sensing circuit for the HR is based on the principle of Infrared Absorption Reflectometry (IAR).

During the optimisation stage of the electronic design, it was considered appropriate to select between the available options that could accomplish the above functionality and provide at the same time the greatest advantages over alternative solutions. The first part of the electronic circuit was an optimised specifically designed analogue circuit that could not be substituted by anything else since it was a uniquely designed dedicated component. The next stage required to provide communication with the computer included two main options for implementation. The first option envisaged a high processing power microcontroller or Digital Signal Processor based component capable to process locally the signals fed from the two channels of the analogue front-end circuit and successively communicate with the computer via proprietary software. The second option was to feed the analogue signals produced by the front-end circuit directly to the computer via the two channel audio port.

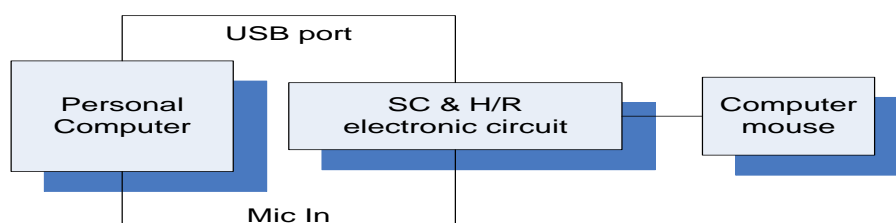
In order to produce a system that is cost effective, computer platform independent, involving minimal adaptations and relatively simpler to make, the use of a particular DSP or Microprocessor has been circumvented and the need for a special or proprietary communication protocol has been avoided. Instead, the preconditioned signal is fed to the computer via its universal sound card and further software processing components was done by using the processing power of the Personal Computer (PC). The software after processing the signals also completed the task to transform the indications derived from the user responses into parameters expressing the results of the measurements. Those parameters were then available to any adaptive platform for the modifications of the adaptive user environment accordingly. The system in its final form consists of a sensing part that is accommodated onto a typical computer mouse, an analogue electronic circuit that feeds the processed signal to a typical home computer and finally a software component that translates the measurements into a predetermined format appropriate for our adaptive application platform. STC is detected by direct skin contact of the thumb and ring fingers with two pellet shaped sensors located on the left and right vertical sides of the computer mouse respectively. The areas of the epidermis

## Chapter 5: Systems Analysis

(stratum corneum) of the fingertips as the densest parts of the human body in sweat glands and nerve endings (Gray, 1977) conveniently provide a highly reliable source for our instrument with optimal response to stress induced stimuli. The pulses of the heart are detected by infrared sensors measuring the reflective absorption of a Near Infra-Red beam that takes place during the changes of the coloration of the skin caused by the pulsation of the blood in the vesicles of the tissue. The sensors for the HR circuit are located in the centre of the pellet rings on the sides of the mouse. A dual sensing circuit of the HR provides a congregated signal, minimises interference or false readings and eliminates movement artefacts through correlation of the two measuring channels, while further analogue filtering and scaling ensures reliable pulse detection. Software algorithms for missing pulse detection, ectopic and offset pulse elimination, as well as inconsistencies between measurements improve further the accuracy of the system.

### 5.1 Systems Design Overview

The system connection diagram of the psycho-physiological assessment system is shown below in Figure 5.1.1. The computer mouse is connected to the electronic circuit box via a DB9 connector. The electronic box is then connected to the PC via the USB port that enables the communication and use of the mouse and feeds its output to the microphone input via a 3.5 mm microphone jack connector. No external power is required as the low power circuit is powered by the USB port that is used for the signature and operation of the mouse. Finally the psycho-physiological index assessment software is required to be installed on the PC to conclude the system operational requirements.



*Figure 5.1.1: System Connectivity Block Diagram.*

## **Chapter 5: Systems Analysis**

The HR front end uses two sensing elements providing two measurements at the same time, improving the reliability of measurement. The two raw signals for the HR and a current signal corresponding to the STC from the sensing elements are fed through to the STC & HR electronic circuit where they are processed, filtered, de-noised and conditioned appropriately. The two HR signals are hardware correlated producing a filtered ECG analogue calibrated output in the range of 0 – 1 Volt. The STC current is hardware converted into a linear frequency output signal with amplitude of 0 – 0.9 Volts. Outputs of HR and STC from the electronic circuit are then fed to the computer microphone input via the two independent audio input channels. Dedicated software identifies and configures the hardware sampler chipset found in every computer sound card subsystem, adjusting the type and range of input signal to unipolar with maximum amplitude of 1 Volt as well as other parameters such as sampling rate, buffer memory size pre-allocation and time interval for each measurement.

### **5.2 Description of Sensing Circuitry**

For the sensing of STC, Ag-Ag/Cl pellet type sensors are used for the skin surface contacts considered most appropriate due to their low oxidation characteristics, reasonably high gel-free electrolytic conductance and minimal susceptibility to ambient interferences, however, the circuit works sufficiently well by using low cost stainless ring washers. The sensing circuit used for the detection of the pulses of the heart is based on two packaged infrared reflective sensors with compensation for ambient light rejection and maximum reading distance requirements of 0.5 mm. The photo-reflective sensors provide two independent continuous current output signals, making the HR detection feasible even in case the user is in partial contact with only one sensor of the mouse.

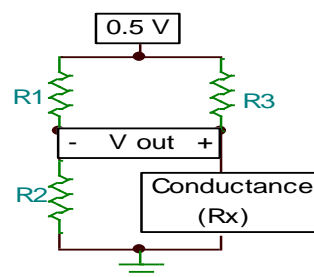
### **5.3 Skin Trans-Conductance Circuit.**

Prominently for the design of our sensing circuit we emphasised in linearity over a full scale of conductivity measurements ranging from 100 mho's to 4 mmho's. Linearity has been achieved by designing the current sensing circuit with increased sensitivity over a small range of output (typically 200 mV). Achieving linear response over the full scale of measurement essentially allows for a simplified system response in the later stages of the measurement,

## Chapter 5: Systems Analysis

where we have to deal with varying characteristics between different users. Near absolute linearity allowed us to eliminate the needs for pre-calibration of the system for each individual, by using an instant measurement baseline estimation technique that follows a repetitive re-alignment during the session of the measurement. A four quadrant current measurement bridge has been used for the SC measurement of the subject as shown in Figure 5.3.1, obeying the following equation for the sensing circuit response:

$$V_{out} = \left( \frac{R_x}{R_3 + R_4} - \frac{R_2}{R_1 + R_2} \right) * V_{in}$$

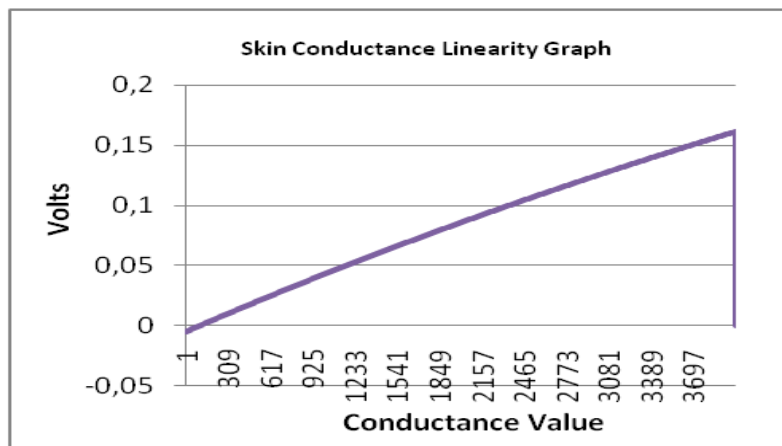


**Figure 5.3.1: Current Measurement Bridge.**

Values of R1, R2 and R3 were chosen so that the output  $V_{out}$  would be linear for the typical range of the human skin conductance, at the expense of a smaller range of output of the circuit and this way demanding from the electronic circuit a higher input sensitivity. The multiplication factor  $V_{in}$ , is the voltage supplied to the circuit, and it has been selected at 0.5V, provided by a constant voltage reference component. As the voltage passes through the resistors is converted to a constant current available to close the circuit when the user touches the sensors with the fingers (Rx). The constant low-current technique has been adopted instead of alternating current as the application requirements are more appropriately adapted to this technique eliminating the three major problems of sensing circuit de-polarisation, minimal Time-To-Settle (TTS) requirements as well as sensitivity problems in interference prone environments. The linearity graph over the full scale of measurement is given in Figure 5.3.2,

## Chapter 5: Systems Analysis

showing values from open circuit to 4 mmho's, (i.e. from non-contact up to the typically highest possible human skin conductance value).

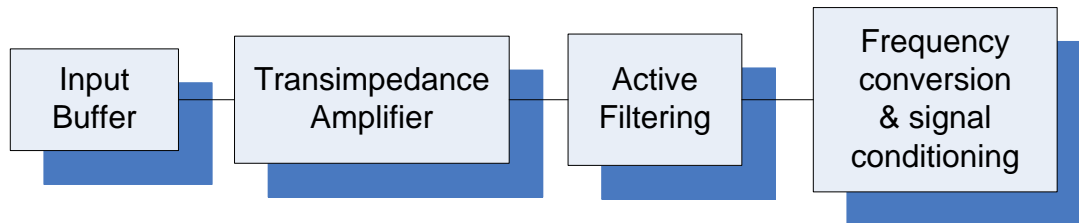


*Figure 5.3.2 Skin Conductance Input Response.*

It can be deduced from the above graph that for 500 mho's the output is 0.02V, for 1000 mho's is 0.04V, for 1500 mho's 0.06V and so on. Linearity in the initial measurement circuit ensures an appropriate scaling of the input for all human subjects (0.6 to 1.4 mmho's) and provides the flexibility for greater accuracy while it also simplifies the calibration of the electronic circuits in the intermediate stages as well as for the final signal conditioning. Unlike similar circuits that supply 3.5V at the point of contact, the current measurement bridge configuration is supplied by a stabilised 0.5V that ensures a very low current of 5  $\mu$ A for a typical STC value of 1mmho. This value is far below the requirements for compliance with the standards for medical equipment safety (IEC 60601-2-3, 2-10 and EN 60601-2-3, 2-10) qualifying this device as eligible for long term use. The system is not designed according to stringent medical standards for use in clinical applications and therefore there is no provision for optical or magnetic isolation of the point of contact with the user that would make it appropriate for use while a human subject is connected simultaneously with other medical therapy, monitoring or radiating devices, however, the extremely low current allows for safe use in long term and for long time intervals. An electronic system level block diagram for the STC electronic circuit is shown in Figure 5.3.3. The sensing circuit is fed to a buffering block that converts the current

## Chapter 5: Systems Analysis

to its corresponding voltage value and isolates back currents. The signal is then forwarded to the trans-impedance amplifier circuit, next to the signal conditioning and filtering circuit and finally via the signal to frequency conversion and signal conditioning circuit to the soundcard.



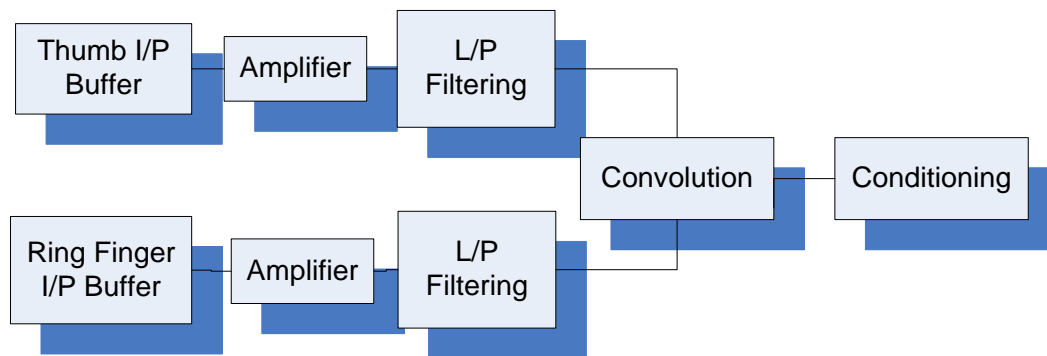
*Figure 5.3.3: STC Electronic System Block Diagram.*

All electronic components of the circuit operate at a single supply of 0–5V DC and the electronic components used are inexpensive popular IC's and off the shelf passive components. Provision has been taken for low power consumption, AC decoupling and the circuit can therefore be powered from the USB port of the computer mouse.

### 5.4 Heart Rate Circuit.

The HR sensing circuit is using a pair of Near-Infra-Red (NIR) reflective sensors TCRT1000 operating in the range of wavelengths between 600 to 1000 nm (peak enter. With centre frequency at around 840 nm and fixed angle of deflection of the photodiodes is an optimised good option providing a sound response at reflective variations of each pulse of the heart of a person. Other sensors with similar characteristics may be used with equally good results as long as they have good repeatability characteristics and they can produce dependable transistor output. The NIR sensors incorporate transmitter and receiver in the same package and provide a reflective transistor output. The integrated circuitry of the sensors incorporates a filter for ambient light compensation. Two of those sensors and circuits are operating in parallel and the signal is fed through an op-amp adder circuit before the instrumentation amplifier, filter and conditioning stage. The system block diagram of the electronic circuit, shown in Figure 5.4.1, denotes an initial signal buffering stage, an amplification stage, a Low Pass filtering and noise rejection stage, a convolution stage and a final stage for signal conditioning.

## Chapter 5: Systems Analysis



*Figure 5.4.1: HR Electronic Circuit Block Diagram.*

The output is an actual analogue heart pulse shaped almost identical in precision to that drawn from clinical ECG devices. It has to be noted that the precision of the heart pulse shape is not important for the process taking place in the system as the diagnostic capabilities that can be obtained from the details of the portrayed heart pulse are not utilised in this particular application. A very precise peak-to-peak interval detection is sufficient for the purpose of our study.

### 5.5 Computer Software

The software component of the system has been developed using MATLAB and utilises the computer sound card as an Analogue to Digital Converter (ADC). Source code has been included in the accompanying DVD. Firstly, the program detects all available sound hardware components of the computer and allows for selection of the hardware available and sampling rate that is to be used. Next deactivates the High-Pass filter that most cards apply to frequencies below 10-20Hz, it then configures the limits for amplitude and polarities of the signal and optimises the size of memory to be used for the buffers when operating in real time processing mode. For the STC measurement channel, a frequency signal is projecting in a linear scale the quantified value of measured STC. A baseline is calculated from the initial measurements with local maxima and minima of  $\pm 20\%$  of that initial value and maintained thereafter updated by the sequence of measured values. If the maximum or minimum value is exceeded, the range is re-adjusted producing an auto-calibrated range of measurements independent of the individual skin trans-conductance characteristics that may as well vary between different users. With this method the STC is allowed to have a positive or negative value with respect to the baseline as

## Chapter 5: Systems Analysis

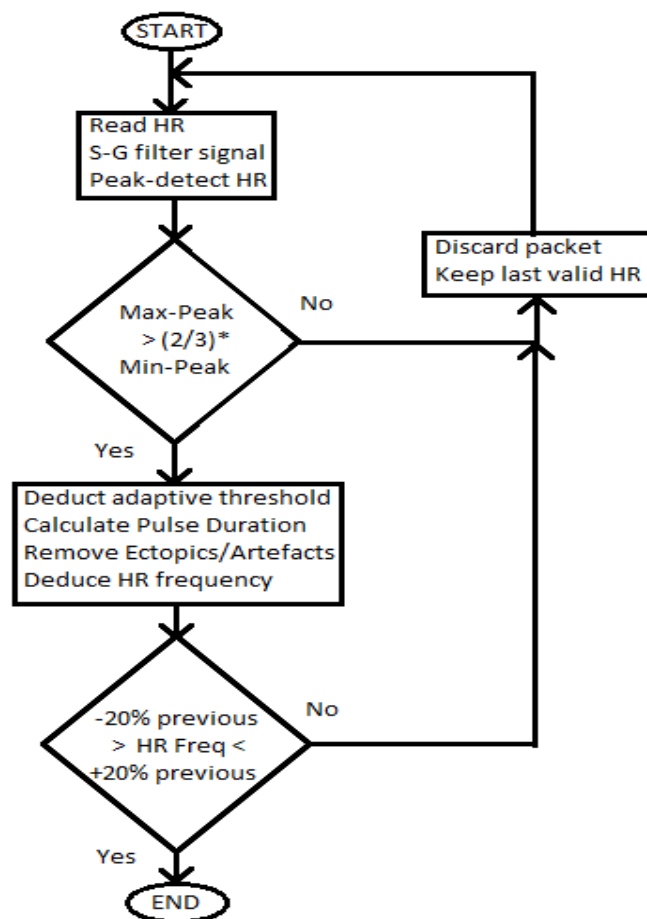
an independent variable, eliminating the requirement to subject the user through an initial calibration procedure. Tonic Level (TL), Phasic Response (PR), baseline (BL) and the tendency to improve or deteriorate the stress level (gradient of the STC), are the main parameters derived from the STC measurement. TL is constructed by accumulation of the mean frequency value i.e. adding the last value of frequency read to all previous and dividing the result by the number of valid measurements. TL provides a trail of measured frequency values and represents the variations effectively interpreted as the psychological level of arousal or active involvement. TL is rising when the subject demonstrates mental stimulation and is increasing further with raised mental activity, tentative engagement, mental effort and especially stress. PR is effectively the instantaneous frequency measurement representing the last measured value of trans-conductance. PR's are signed positive or negative intensifications in skin conductance that begin 1–2 seconds after stimulus onset and peak within about 5 seconds. The amplitude of the STC response varies according to the subjective impact of the eliciting stimulus and that in turn varies according to the intensity of the stimulus, its abruptness and unexpectedness of the subject, and its meaning or importance, however, an absolute one to one relationship between STC amplitude and intensity of affective response could not be assumed. Aroused subjects exert spontaneous skin conductance responses and therefore their corresponding frequency, (tonic level), increases proportionally to their level of arousal. The software algorithms for real time stress analysis include:

- audio input signal amplitude scaling,
- software low-pass-filtering with cut-off frequency  $f = \frac{F(\text{sampling})}{4}$
- signal period correction,
- frequency domain transformation
- and choices for real time plotting of the input signals, or the filtered input signal and its transformation into the frequency domain.

The algorithms for the HR pulse detection and correction as shown in Fig 5.5.1 below include: audio input signal amplitude scaling, signal normalisation, DC component filtering, software low-pass-filtering with cut-off frequency at 5 Hz, adaptive threshold calculation, edge detection, relational pulse interval correction in time domain and validation of detected pulses

## Chapter 5: Systems Analysis

as shown next. Also the trends for the STC and HR values as well as the physiological responses and psychosomatic states deduced by the combinations of measurements can be plotted either in real time or as recorded data. Figure 4.5.1 presents the flowchart of the HR software process.



*Figure 5.5.1: HR software process flowchart.*

The signal conditioned HR input channel is filtered using a fifth order Savitzky-Golay moving average filtering algorithm before the edge detection of the rising and falling edges of the signal. Signal noise is deducted and a valid signal is detected if peaks detected are at least two thirds higher than the highest noise level after filtering. Illustrating further, after each acquisition of HR, peak values are identified in the buffer holding the samples of the signal (typically 96000 samples per measurement) and then the highest numerical value is identified. The two thirds of that value is appointed as the threshold meaning that any sample with

## Chapter 5: Systems Analysis

numerical value above that threshold is considered as a valid pulse. Any value below Pulse detection is then determined by an adaptive threshold iterative process and validated by an ectopic beat or arrhythmia pulse correction algorithm. Pulse intervals are then compared to those of previous measurements and if they fall within acceptable time margins and the range of heart beat between 30 and 210 beats per minute the heart rate is validated positively. The heart beat rate is deduced by the formulae:

$$HR = \left( \frac{Rl * \left( \frac{t}{Tm} \right)}{As} \right)$$

where:  $Rl$  = Record Length

$t$  = 60 (secs per minute)

$Tm$  = duration of measurement (user selected,  
typically 1.84 sec's)

$As$  = Average Samples Between Peaks.

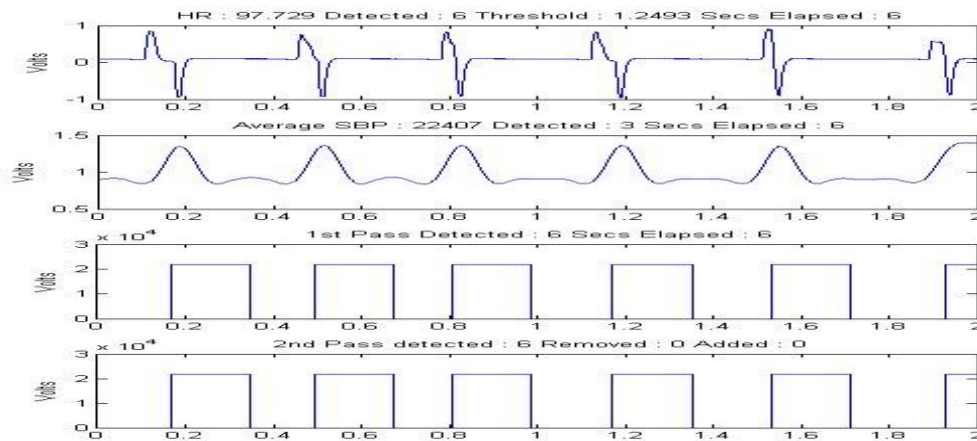
The accuracy and validation of HR responses were laboratory bench tested and verified by using a function generator with triangular waves and pulse waves with variable duty cycle and at predetermined frequency values. The following table (Table 5.5.1) shows some random simulated input values and the corresponding output values of the heart beat rate.

*Table 5.5.1: HR Results on simulated input signals.*

| Freq(Hz) | Waveform   | Duty cycle(%) | Beats Per Minute |
|----------|------------|---------------|------------------|
| 0.6      | Pulse      | 10            | 36               |
| 0.75     | Triangular | 100           | 45               |
| 0.82     | Pulse      | 10            | 49.2             |
| 0.95     | Pulse      | 5             | 57               |
| 1.22     | Triangular | 100           | 73.2             |
| 1.29     | Triangular | 100           | 77.4             |
| 1.68     | Pulse      | 5             | 100.8            |
| 1.8      | Triangular | 100           | 108              |
| 1.9      | Pulse      | 5             | 114              |
| 2.12     | Pulse      | 10            | 127.2            |
| 2.28     | Pulse      | 5             | 136.8            |
| 2.3      | Pulse      | 5             | 138              |
| 2.75     | Pulse      | 10            | 165              |

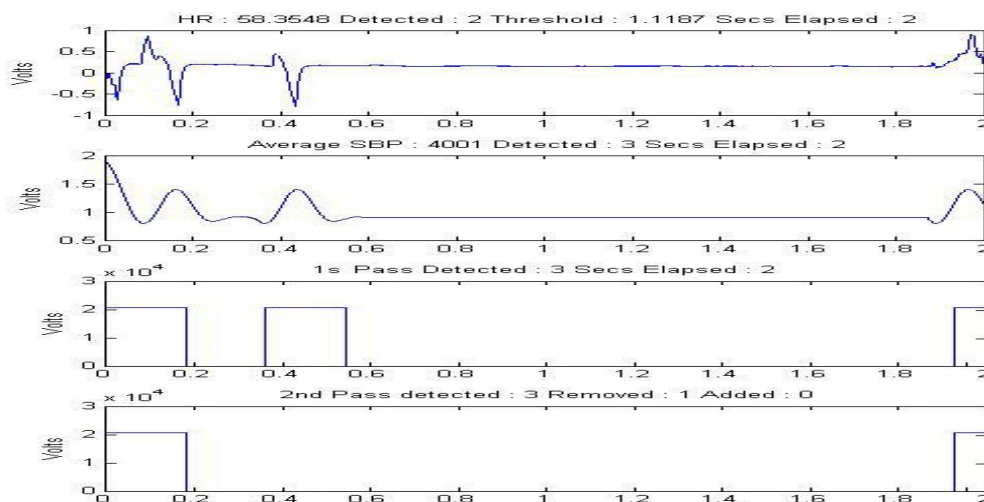
Table 5.5.1 exhibits the tolerance of the system on various simulated signals of different shape with amplitude of 1 Volt. Figure 5.5.2 presents graphically the stages of HR real time detection process as it can be configured from the configuration console to be displayed in real time if so desired.

## Chapter 5: Systems Analysis



**Figure 5.5.2: HR Pulse Detection Process.**

At the top window is the actual row pulse acquired from a human subject by the sensors in a frame time interval. The second window horizontally shows the pulses detected after the calculation of the adaptive threshold and fifth order Savitzky-Golay filtering, showing vertically the coinciding inverted peak value. The third window depicts the pulses after the rising/falling edge peak detection algorithm and the fourth window shows the pulses evaluated after the final check for irregularities of the intervals between pulses. Figure 5.5.3 portrays the process of irregular pulse removal for a frame time interval of 1.04 seconds.

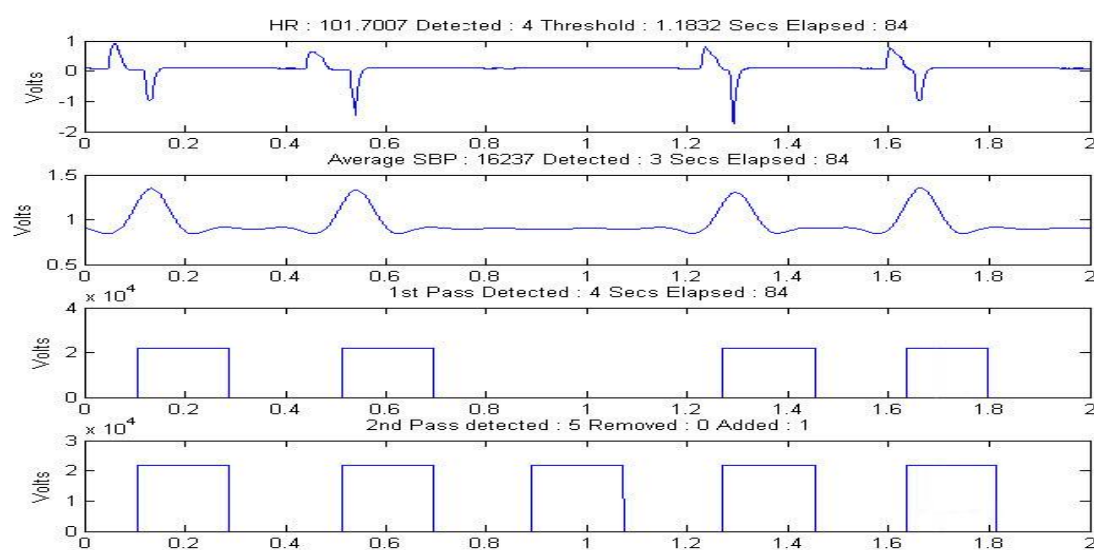


**Figure 5.5.3: HR Ectopic / Irregular Pulse Removal.**

## Chapter 5: Systems Analysis

When a pulse interval was too small (i.e. HR greater than 210 beats per minute) it was judged as an ectopic beat or an artefact originated pulse that was then being removed as rejected by the algorithm. For this judgement the algorithm compared the intervals between the first and third pulse as well as the second and third pulses respectively, in relation to immediately previous recording and determined whether the first and third or the second and third pulse intervals were closer to acceptable values according to previous measurements.

Figure 5.5.4 shows a real time case of the missing pulse detection and correction algorithm for a frame time interval. The first window shows the raw sensor signal, the second one down the adaptively detected pulses, while the third and fourth the graphical representation of the correction process as it was plotted in real time.

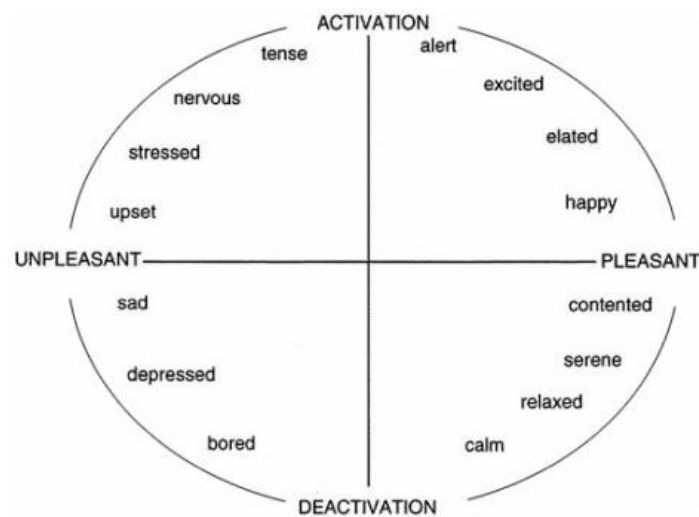


**Figure 5.5.4: HR Missing Pulse Detection.**

Data produced in real time and/or logged into a file are: HR, Mean Value of HR, Heart Rate Variability (HRV), HR Gradient, STC Tonic Level, STC Phase Response, STC Gradient, and also derived quantities as Valence, Arousal and Psychosomatic Index produced by processing the combination of the STC and HR simultaneous measurements. Positive valence was derived according to the loci of the values of the measurements if they fell in the area of positive STC

## Chapter 5: Systems Analysis

amplitude, positive HR gradient positive STC gradient and positive Tonic Level in relation to the Baseline. Arousal was derived from the above parameters but instead of Tonic Level was depending upon the Phasic Response of the STC. The above combinations of measurements were used for the real time validation of the circumplex model of affect (**Russell, 1980**), seen in Figure 5.5.5 below. During the initial system testing investigating the accuracy as well as the precipitation of the model, the system indicated high accuracy in the sense that; better than 80% of the cases when a subject was concentrating on the computer screen the system deduced levels of positive arousal while turning negative when the subject looked away from the screen.



*Figure 5.5.5: The Affective Computing Circumplex.*

The above observation is referenced only indicatively since the verification of the validity of the above model was a partial subject of investigation for this study. The main objective of this research has been the design and development of an integrated diagnostic system of the psychosomatic state of a learner rather than the validation of affective allocation models such as Russell's model adopted herein or any other model that could be selected for investigation and analysis in the future.

## Chapter 5: Systems Analysis

### 5.6 Electronic Circuit

The schematic diagram of the electronic circuit is shown in Figure 5.6.1 and the Printed Circuit Board (PCB) artwork of a double layer connectivity diagram in Figure 5.6.2. Inputs to the circuit arrive from the sensors (not shown). The front-end is decoupled and fed through a unity buffer stage for signal stability, before it enters the trans-impedance and amplification stage respectively. 5<sup>th</sup> order Butterworth filtering (Low-Pass at 10Hz for the HR circuit and High-Pass at 120 Hz for the STC) and signal conditioning is following ensuring a power grid induced 50 Hz interference rejection, cancellation of noise produced by artefacts and external sources and finally the output voltage is scaled by the final stage to between 0-1 Volt. Software is then undertaking the task to optimise the packets of data influenced by mishandling and frailty of contact of the fingers of the user with the sensing elements on the mouse. Contact of the fingers is detected by the STC input being detected above a threshold frequency value of 150 Hz. Below that threshold both the STC and HR values are reset to the value of -1. Only valid data are actively being used for determining the psychosomatic state with the non-contact state data being rejected. The configuration console allows for a selection of an option to display either a blank screen or last read values when the user does not make contact with the mouse.

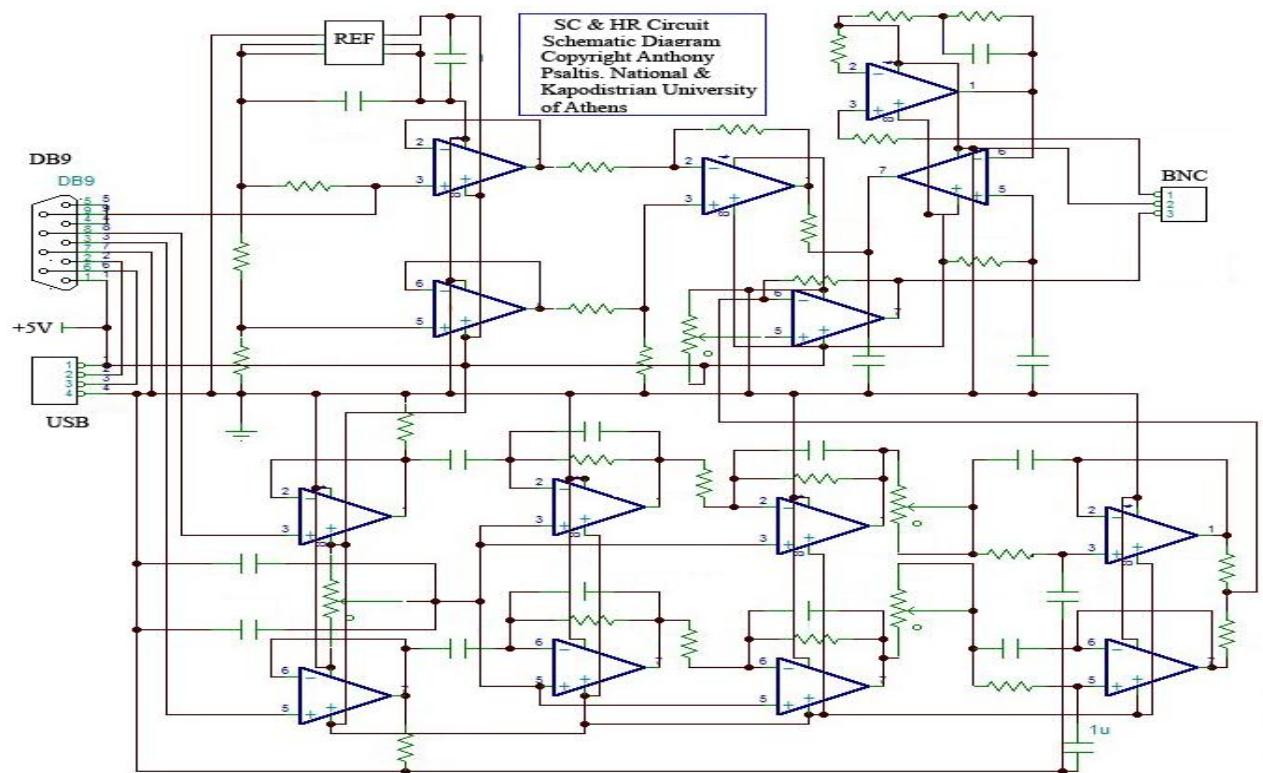


Figure 5.6.1: Electronic schematic diagram of the HR and STC detection circuit.

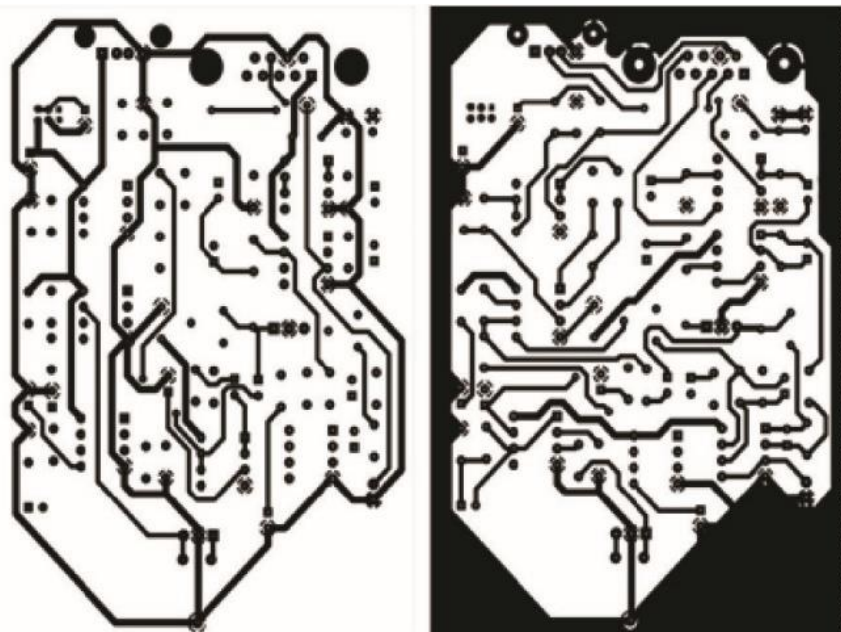


Figure 5.6.2: Electronic Printed Circuit Board artwork (dual layer).

## **Chapter 5: Systems Analysis**

The electronic circuit was optimized by using Tina Version 8.0 software suite from Texas Instruments. The schematic diagram was drawn by using ORCAD software suite and the PCB design was done by using the Auto-Router PCB software suite from the same bundle. The artwork provides a production ready material while the parts list and component placement lists (omitted) provide all information necessary for assembling the complete analogue electronic circuit.

### **5.7 Functional and operational description**

The integral system functionality involved certain operational requirements categorised in the distinct areas of software and HCI design, instrumentation and medical safety precautions and that for interpretation of physiological data. It was therefore judged appropriate to dedicate a complete chapter with main objective to illustrate the process flow employed for the identification of affect and mental state of engagement and clarify the operational methodology exercised. Initially anticipated problems or those occurring during the development stages are also analysed in this section.

As physiological metrics have been used for a while in the domain of HCI as a means to obtain a realistic input of mental effort or other affective response (Cacioppo & Tassinary, 1990) a major limitation has been suppressed by the advantages of the value of acquired information and taken as an inevitable concession. The deterrent of psychological interference during acquisition was the major obstacle as in all experiments so far, users were subjected to wires and attachments of sensors and instrumentation. This discounted variable entailed a conscious psychological discomfort that caused an inestimable amount of interference due to some part of the thoughts of the user being occupied by the apprehension of being part of an unknown instrument rather than taking part in the experiments in a totally free environment. This major limitation was turned into the major advantage of our system as explained in systems design section and was retained throughout the design of functional processes achieving a seamless operation eliminating methodological issues as addressed by Rosalind Picard, (2003).

## Chapter 5: Systems Analysis

### 5.7.1 Functional description

Participants taking part in our setup were seated on a typical computer spot, with a normal monitor, keyboard and mouse available to use, just as in everyday life. When participants used the computer mouse, physiological signals of their heart beat and trans-impedance of their body were fed to the system with no additional effort or concern from the part of the user. Biological signals were amplified and conditioned by electronics and then fed through to the computer in two separate channels and in different form as analogue pulses in one channel and audio frequency in the other. This configuration eliminated cross-channel interference and stray capacitance or current induced signal noise with signal to noise ratio better than 102 db between channels.

The software components for communication were storing the blocks of audio data from each channel into allocated buffers, available for processing while the acquisition of the next block was taking place. HR channel was stored into a buffer queue retaining two blocks improving the accuracy of peak detection algorithm applied on data using a wavelet technology function. STC channel had no need to store more data as the requirement for frequency detection was that for an instant measurement. Heart rate was deduced and compared to the previous measured value deriving thus the HR-gradient. Similarly, STC frequency was compared to the previous STC value deducing the STC-gradient. The two gradients above were employed in the algorithm of allocation of the measured value in the corresponding quadrant of the circumplex model, while the absolute value of the two coordinates was deduced by the algorithm described in the system description. Trend values of HR and STC were stored for every measurement and used to update the Heart Rate Variability (HRV) and STC baseline respectively. The HRV was crucial in the determination of a weighing factor affecting the absolute value of HR enhancing or diminishing the HR coordinate with respect to the minimum and maximum value of measured HR, effectively increasing or decreasing the distance of the displayed value from the mid-point (imaginary diagonal line on the graph). The baseline was effectively the mean value maintained by the SC and used on the graph to determine the distance of the second coordinate from the mid-point (imaginary diagonal line on the graph). The system produced a raw data per-measurement signature on the four quadrant circumplex-

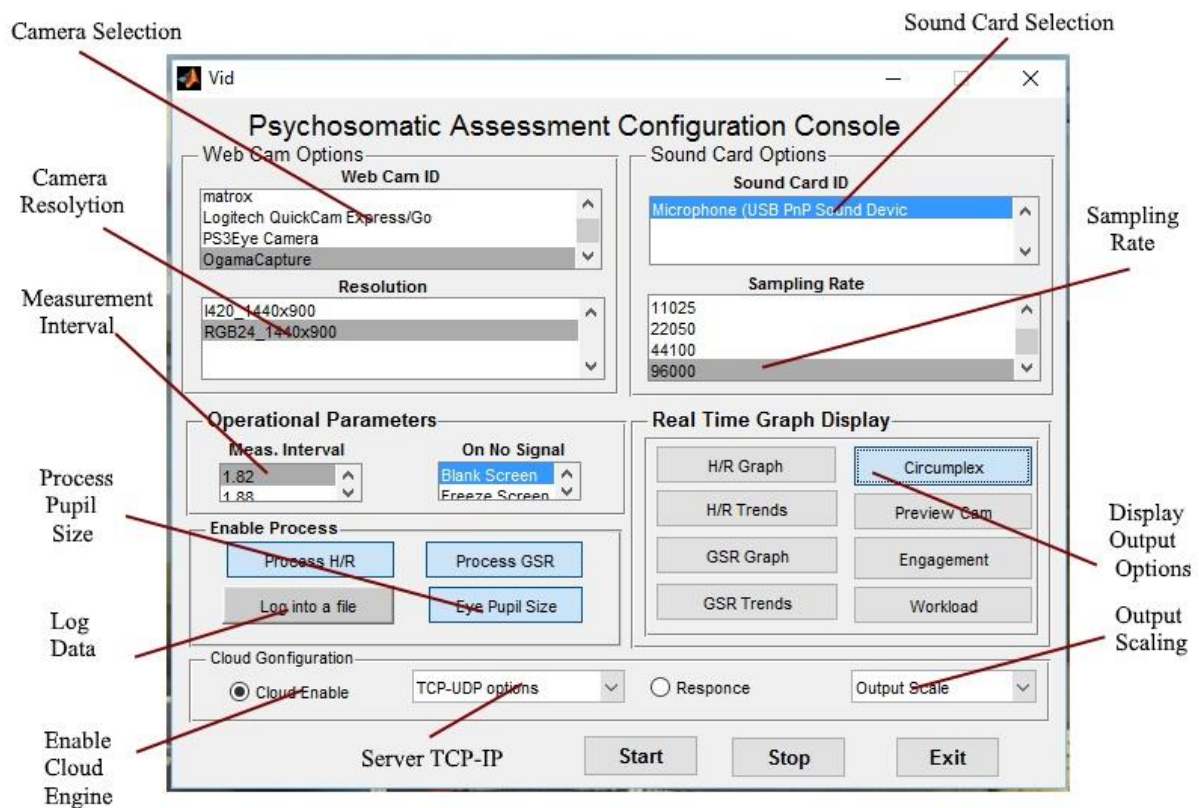
## **Chapter 5: Systems Analysis**

like representation. No prediction algorithm like Hidden Markov Models, smoothening or curve fitting was employed. Data were stored in detail including raw values, intermediate calculation data, weighing factors, quadrant allocation, coordinate values and parametric conversion values onto the graph.

### **5.7.2 Operational description**

When the system was connected to a PC according to the connection diagram illustrated in the systems description, the software component was executed either selected from within the MATLAB environment or executing the psycho-physiological configuration system console PPCSC.exe program in Microsoft windows environment. The program after detecting the peripheral devices connected to the computer such as cameras, network and audio cards as well as their individual capabilities, displayed the identification descriptors in a list of the system configuration console. Individual capabilities of each peripheral device depicted directly from the driver configuration of the Operating System, were made available to populate the lists provided to the user for selection. More specifically, for the camera, the system could detect all available cameras connected to the computer by system ID and all available resolution options for each one individually. For the sound card, information included the ID of all available cards (e.g. Realtek Sound Card) and all available sampling rates that they could operate (e.g. 9600 bps). Data acquired by the program during initialisation were used to populate the lists of options of the Psychosomatic Assessment Configuration Console as shown in Figure 5.7.2.1.

## Chapter 5: Systems Analysis

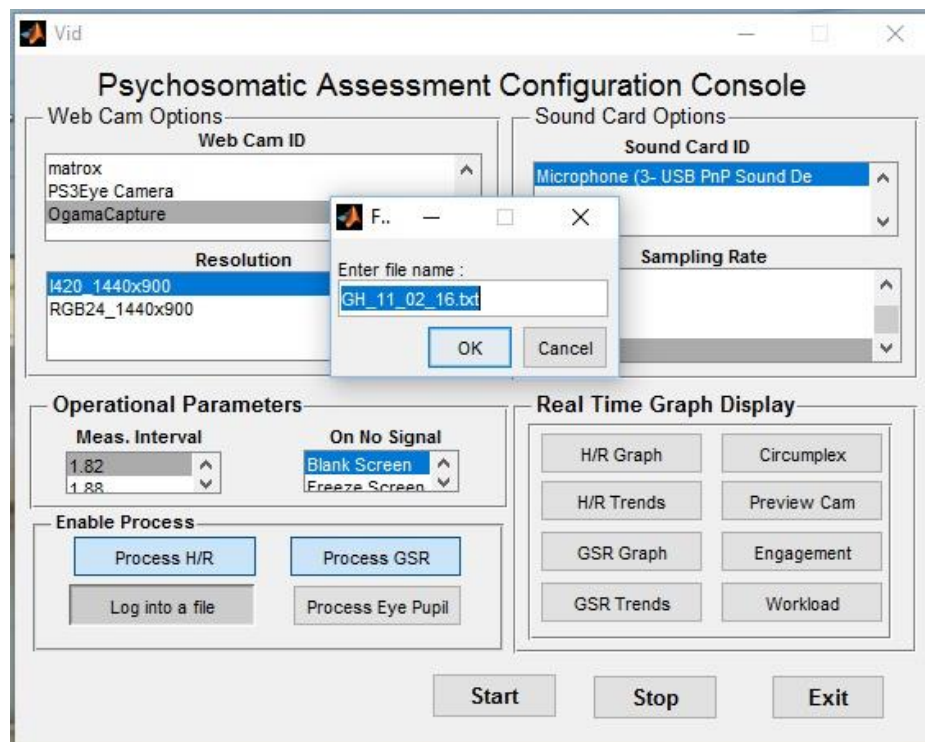


**Figure 5.7.2.1: Systems Configuration Console.**

During the initialisation of program execution, information about the network card and configuration parameters were also identified and used to populate the fields required for the configuration of “*Cloud Enabled*” applications, provided for future development. Preset values selected as default, allowed the operation of the system by selecting only the “*Start*” button for basic operation, however, a number of options were also available for fine tuning or debugging information. Options for the “*Web Cam ID*” and “*Resolution*” of the image capture on the top-left as well as the “*Eye Pupil Size*” provided for future development and their settings did not affect the present operation. The Operational Parameters tab allowed for modification of the time interval of the measurements and the optimised value was pre-selected at 1.82 seconds. The “*On No Signal*” selection list presents the options available for messages when the contact with the device has been interrupted. Default message is “\*\*\* *No contact*,

## Chapter 5: Systems Analysis

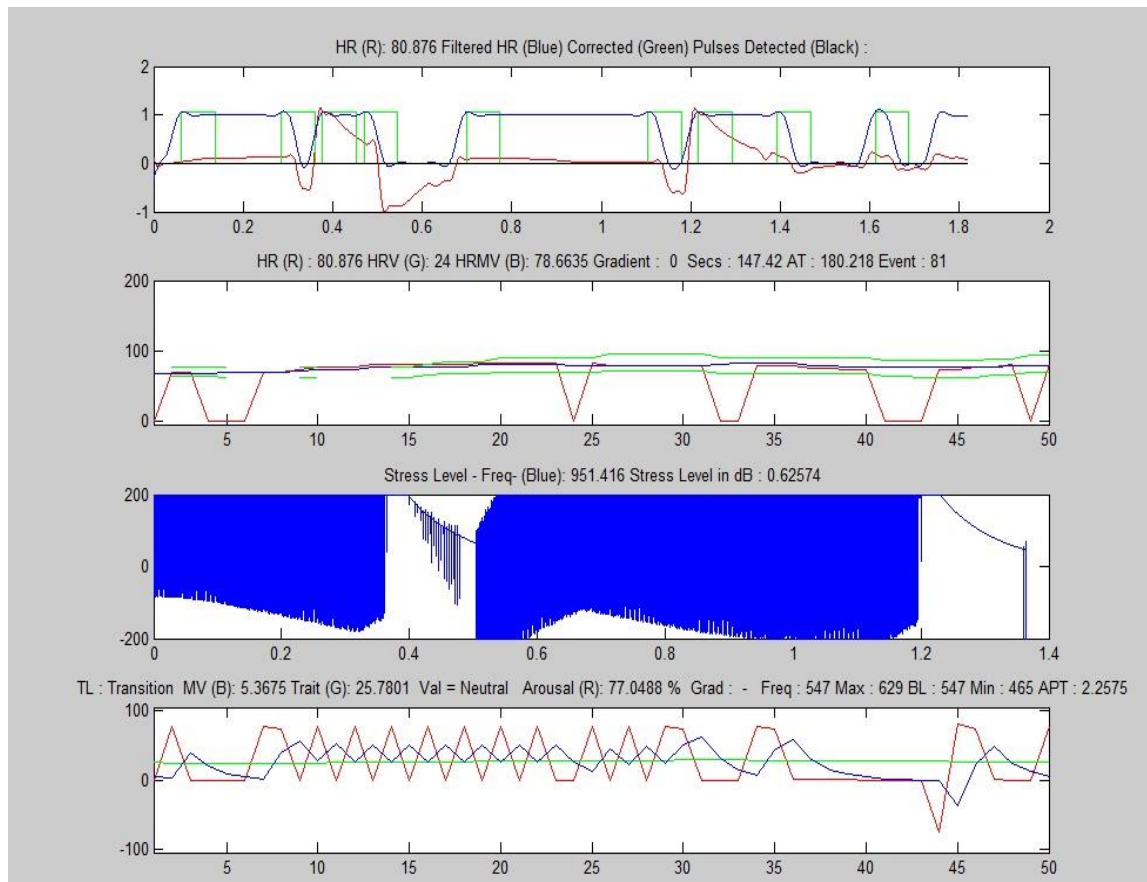
*Please reposition your fingers \*\*\** ". The two buttons on the tab "Enable Process", allow to select processes individually according to the demands of an experiment, i.e. HR, SC or both. The button directly below designated as "Log Into a File" allows for the composition of a filename where data for a particular session will be kept as required (Figure 5.7.2.2).



**Figure 5.7.2.2: File Save Option.**

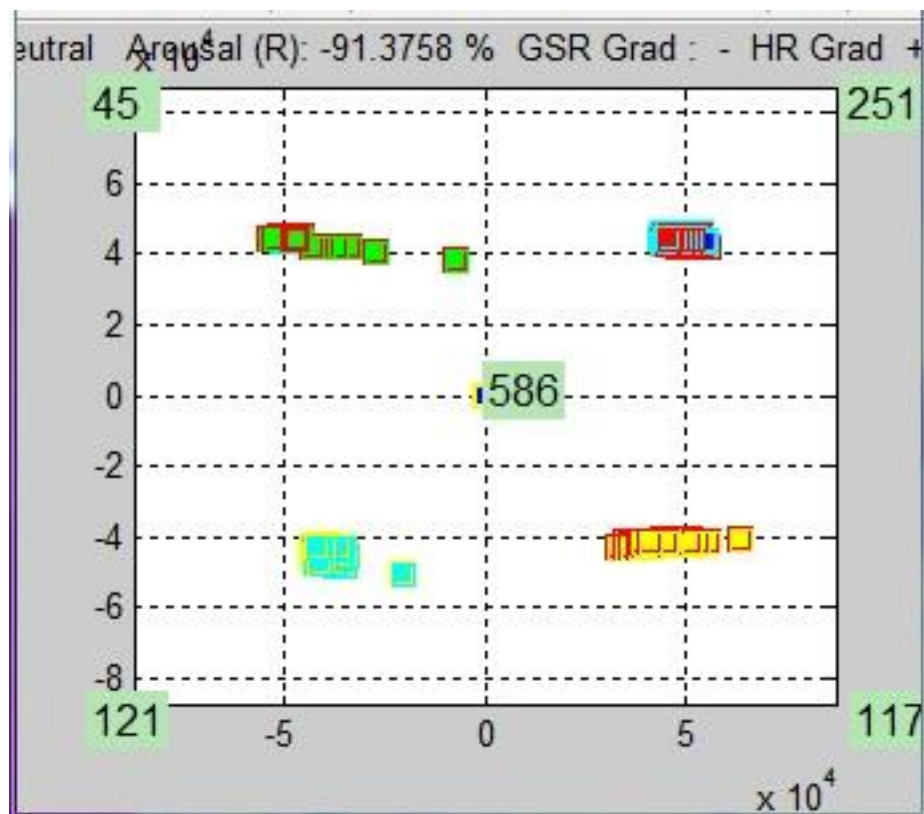
On the top-right corner the "Sound Card ID" and "Sampling Rate" tabs allowed for selection of the particular sound card and baud rate of the specific sound card where the system was connected. Default values were for the master sound card of the computer and the highest baud rate the device could support. The tab "Real Time Display Graph" allowed for the selection of individual windows to be displayed during measurements. A sample display used mainly for debugging purposes, displaying HR signal with colour coded graphs of the processing stages of pulse detection and correction, the STC frequency graph and their trends respectively can be seen in Figure 5.7.2.3 below.

## Chapter 5: Systems Analysis



**Figure 5.7.2.3: Real-Time Debug Data Caption**

Data presented in their final form were available either for visual comparison between captured involvement and their corresponding value represented in the circumplex allocation, or as raw data sequence recorded in the log file. A sample of data representation during an experiment is shown in Figure 5.7.2.4.



*Figure 5.7.2.4: Real-Time Data Allocation Model*

Each measurement was represented on its corresponding quadrant as it was derived by the calculated gradients of the physiological measurements. The latest displayed value was highlighted by a frame that was fading out just before the next value was allocated. Exact position on the graph was determined by the distance from the vertical and horizontal axis respectively. The above distance was an expression of the magnitude of the physiological measurements. The number on the centre of the picture in Figure 5.1.4, represented the count of total measurements, while the four peripheral values indicated the number of allocations for this particular quadrant, i.e. the top-left for the top-left quadrant, the top-right for the corresponding top-right quadrant etc.

Other information on the graph included values of valence and Arousal and gradients of HR and STC respectively.

## **Chapter 7: Discussion and Interpretation of Results**

### **6. Experimental Evaluation**

A strategic partitioning of the experimentation procedure of the aforesaid trial was crucial in order to substantiate and demonstrate the advantages of the innovative formation of the system conception and produce a convincing and at the same time truthful and accurate reflection of the real capabilities of the system. The entire set of experiments required for this study had to be carefully selected and optimised in a multi-faceted manner in order to

- corroborate the usability and validity of the system,
- fulfil a comprehensive evaluation, to compare results against widely accepted previously validated data and others produced by approved commercial systems and finally
- evaluate the overall efficacy of the system.

The research hypotheses reasoned and analysed in chapter 3, viewed in relation to established theoretical frameworks and previous scientific efforts explained in chapter three, have channelled this study through a course of individually optimised experiments. A series of tests were optimised according to their effectiveness, complexity and feasibility characteristics in producing clearly the parameters for testing our presumed hypothesis. For this reason an exhaustive program of experimental processes with distinct and progressive segments was stipulated. Each segment included an autonomous experimental plan committed to testing a particular system capability as well as assessing the essential operational parameters derived. Typical experimental procedures entailing design of test environments, data collection and statistical analysis of results investigating the research hypotheses are sufficient for most integral research studies; However, in this present work, which is including the design and development of the tools required for experimentation, provision for additional experiments was necessary. An essential mission was to optimise the system output so that it was shaped in a flexible format that would be capable to feed dynamically adaptive platforms for affective computing applications. Four distinct phases of experimental evaluation were adopted serving the purpose to prove operational, usability, applicability, accuracy and effectiveness of system

## **Chapter 6: Experimental Evaluation**

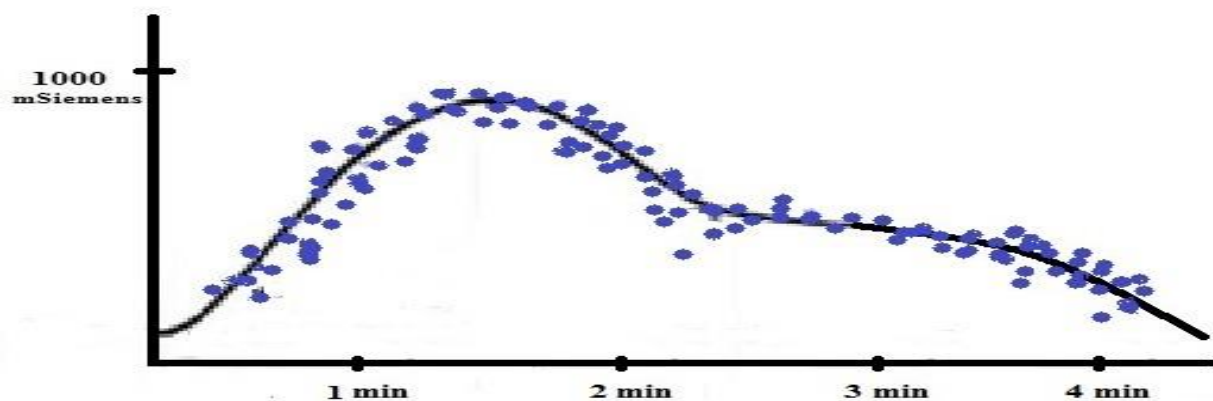
use. The first phase included laboratory validation and verification tests. The second phase involved real time experiments aiming to demonstrate the validity and repeatability of measurements in a non-laboratory environment and assess dependability of the results. The third phase incorporated experiments for a direct real time comparative assessment between pre-evaluated pictorial and audio-visual context data and their deduced affective and emotional system responses and finally in the fourth phase the system was used in an actual application, in combination and in direct comparison with an approved commercial product for assessing relationships in derived data on emotional and cognitive involvement.

### **6.1 Phase One - Device & Systems evaluation**

Since the specific combination of physiological signals (HR and STC) and the concept of gradient convergence (as analysed in chapter “System design”) has not been attempted in this fashion, using an apparatus designed specifically for this purpose and applied in a natural non laboratory environment (see also “literature review”), a system validation and verification stage was necessary. Internal validity tests were necessary to prove that the quantity measured a factual effect and also that the quantity taken as the dependent variable (DV) was a fundamental unit that could provide important interpretations. External validity tests would have to prove the repeatability of the system under similar ambient conditions, demonstrate applicability in all candidates regardless of race, skin colour or other social attribute and identify dependences that could possibly affect reliability in reproduction of the experiments. Initial stages of system validation were carried out by testing each individual component of the system for accurate and timely response. The STC subsystem was tested individually during the piloting phase in real time and responses were assessed during sessions of vigorous stress induced by force. The measurement of the STC was based on a continuous micro-electrical input signal directly proportional to skin conductive permeability (trans-conductance) that was then transformed by electronic circuitry into a regulated audio frequency range. As the signal could not be simulated by electronic components such as resistors or shunt capacitors but could only be produced when human fingers were in contact with the sensing elements, a number of people were used for the control group try-outs, providing data for further revisions of the electronic filtering and calibration in order to improve consistency and linearity. Full range measurements were

## Chapter 7: Discussion and Interpretation of Results

obtained from individuals performing the following clinical exercise: an individual holding the device in relaxed state was gradually forcing their breathing rate to irregularities, (E.g. holding their breath, longer duration or higher breathing rate). The system response in those instances was a corresponding climb of measured values, as an indication that the stress level was raising; thus verifying the correct operation of the system producing results exactly as they occurred naturally when humans anticipate a stress elevated condition. Real time response and speed of detection was observed and found to be instant and directly proportional to increase in HR that was measured independently. STC response data were displayed in real time and also logged into a file during the session, till the subject returned to its previous relaxed state. Data produced were plotted in a scatter plot view and best curve fit process was applied (Figure 6.1.).



*Figure 6.1.1: Data congregation during the physiological stress test of the control group*

Plotted data from the group of participants, displayed clearly the overshoot response that was typical in case of exaggerated stress condition followed by the slow recovery stage. The time required for each individual to return to a state of physiological relaxation after the induced excitement of the exercise varied enormously, as expected, due to differences in cardiovascular efficiency of each individual depending essentially on personal fitness and effectiveness of cardiac output.

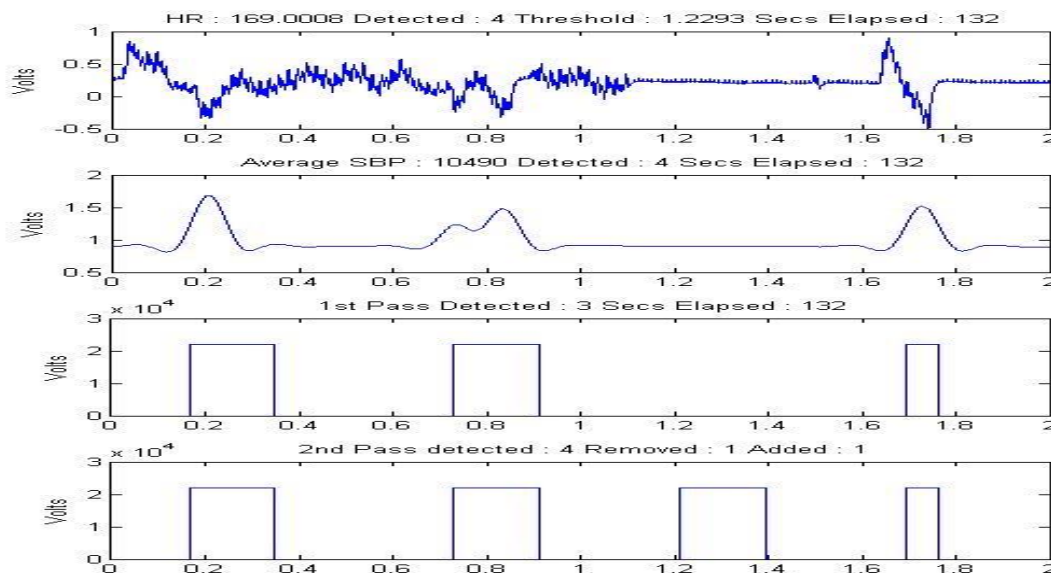
The above procedure also revealed the valuable capability of the system to perform continuous and uninterrupted measurement as well as its simplicity of uninterrupted operation. This was clearly demonstrated by the lack of discontinuity problems, even with excessive movement of

## Chapter 6: Experimental Evaluation

the fingers about the sensing elements as long as the contact remained continuous. Another worth noticing capability of the system revealed during workbench testing was that a sustainable and accurate measurement could be obtained for as long as the fingers were in contact with the instrument regardless of typical motion of the fingers. A correct STC signal could be obtained while the sensing part was impervious to possible variations in pressure, contact surface area or even level of induced perspiration or fingers (e.g. when deliberately were soaked into water).

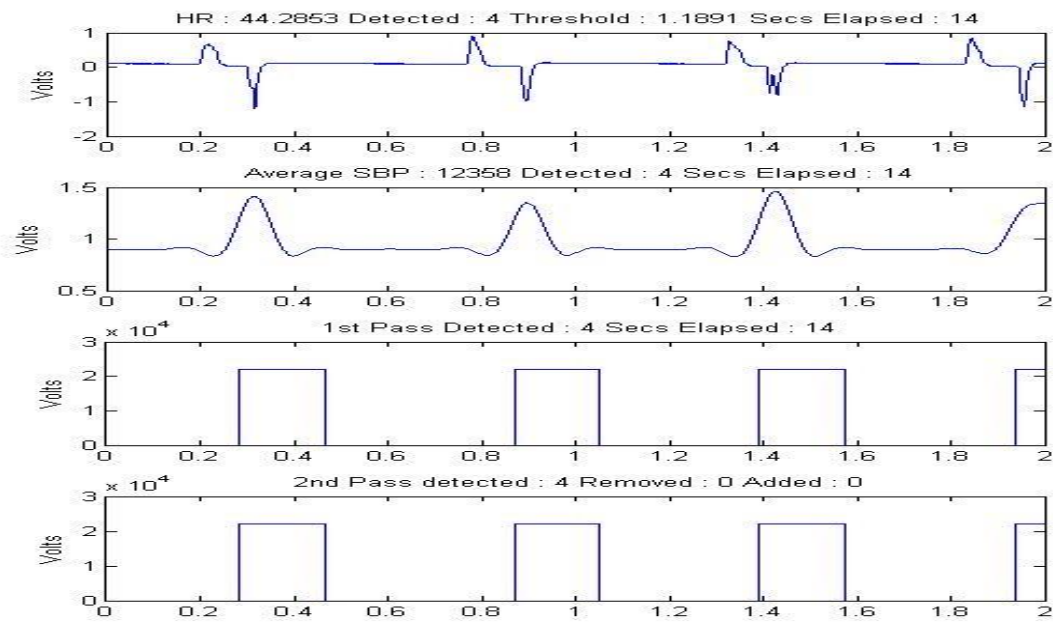
Pilot tests for the HR subsystem were performed by employing the following procedures:

1. by testing the electronic section for a count of a given pulse train,
2. by testing the Near-Infra-Red (NIR) elements fitted onto the sensing device producing the pulses caused by the imperceptible change of skin colour due to the subcutaneous variation of the blood pulsation and the oxygen content,
3. by testing the software that was translating the pulses into a count of heart beats, employing also missing and ectopic beat detection and rejection algorithms as explored in chapter “system description” and also illustrated graphically in Figures 6.2 –6.5 below.

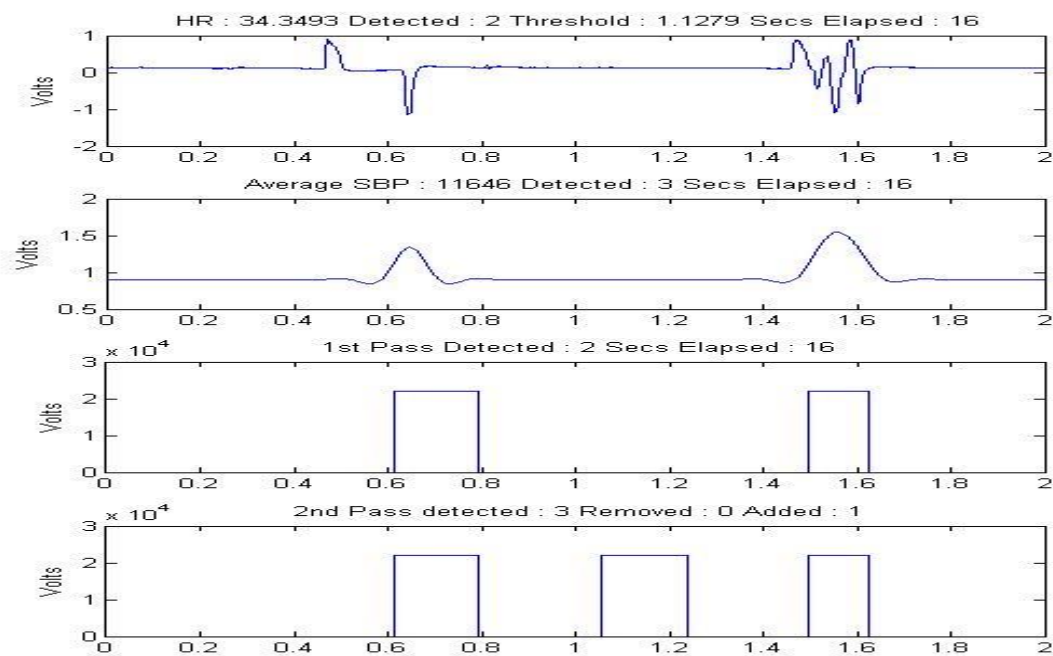


*Figure 6.1.2: Pulse detection and correction algorithm*

## Chapter 7: Discussion and Interpretation of Results

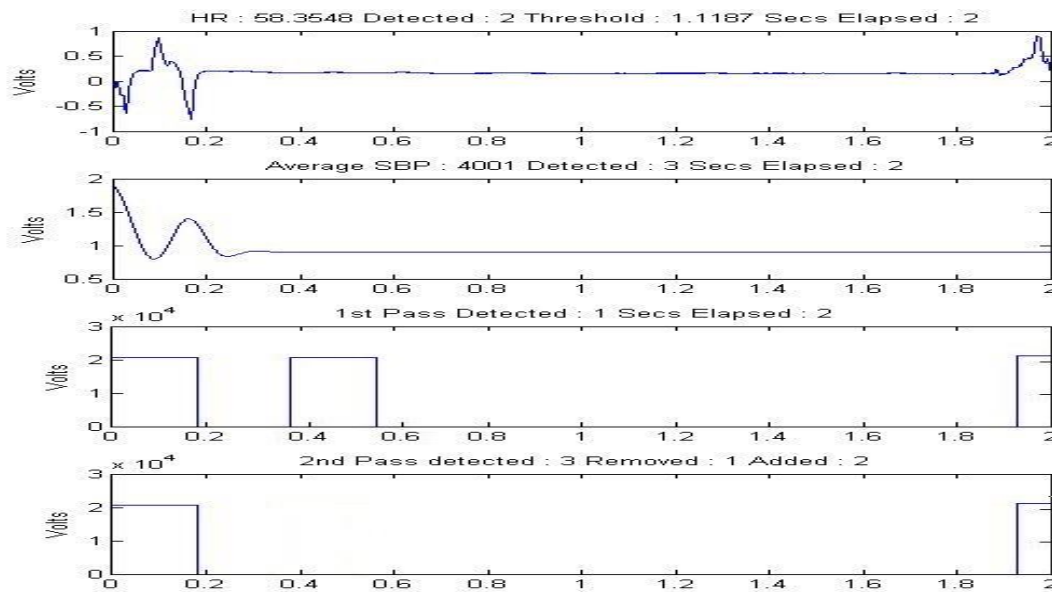


*Figure 6.1.3: Correction of imperfections in received signal*



*Figure 6.1.4: Missing pulse detection and correction algorithm*

## Chapter 6: Experimental Evaluation



*Figure 6.1.5: Ectopic beat detection and correction algorithm*

Plots above show the algorithms through the stages of processing with progress shown from the top to the bottom and were depicted while measurements were made in real-time. Display of the above plots can be activated as an option that could be selected in the configuration console. The system maintained also the capabilities to display raw input data, data through important process stages and result data in a variety of display forms or traits by enabling the appropriate selection in the console. (See also systems description).

Testing of the spectroscopic pulse sensors was made by simulated periodic masking of the sensors with materials impermeable to ambient light. Periodic covering was made at known frequency and the frequency and shape of output was assessed. The electronic signal processing part that filtered and provided the pulse train to the audio card of the computer was verified for precision and accuracy by simulating a known pulse frequency fed through a high precision function generator. The software component of the system was designed in object oriented approach using fully enumerated data types in MATLAB environment, implemented on a Windows 7, 32 bit operating system platform. While in the final software development stages, software profiling, memory usage, speed of execution, performance optimisation and integrity

## **Chapter 7: Discussion and Interpretation of Results**

tests were also performed, ensuring that the software components would not generate possible hardware conflicts or operational discrepancies on the host computer and therefore the software could be executed for infinite time without hindering by any means all other processes executed. Additional tests performed for worst case operational scenario, including execution of the software while many other processor and memory hungry applications were executed at the same time like video capturing, additional audio programs and internet casting applications.

Finally the overall system including the STC, HR sensing component, the electronic signal conditioning component and the software component took a series of verification key point tests for integration. More specifically consistency tests included testing the same person at different times in identical environment and in the same test-set, produced STC values with better than 98.92% consistency for each person, while HR variability boundaries were also nearly identical.

### **6.2 Phase Two - System operational evaluation**

Emotional condition has been considered to play a major role in human affect, cognitive involvement, attention and engagement influencing thus behavioural responses (Quartz 2009). It has been widely accepted in the scientific community of psycho-physiologists and cognitive psychologists that emotion not only plays an important role in cognitive processes and decision making but is greatly intertwined with it (Damasio 1994). Also decision making and learning is heavily depended upon spontaneous emotional state (Snow, Corno, Jackson 1996). Purely subconscious components play also a substantial role in learning (Davou, 2002) and maintain a prohibitively difficult component to be detected or assessed. As a consequence of the above being the most prevailing opinions in the scientific community, an attrition bias develops regardless of the precision of a given system to detect a particular emotional or affective component. By default, either there would be doubt about validity as it is affected greatly by multiple factors falsifying emotional condition or it would be less specific as it would be incorporated into a wider affective construct, thus making the detection less valuable. In an attempt to bridge that particular gap and at the same time averting the obstacle to reduce the validity of the precision of the system in detecting a specific psycho-physiological condition,

## Chapter 6: Experimental Evaluation

a reification of the combined construct obtained from the measurements was deemed necessary, termed “mental workload”. Taking also into consideration that following the detection, a precise distinction of an emotional component was not possible with the existing technologies and also due to multiple factors affecting an authentic emotion or sentiment, a congregation of distinct major emotions into factions with common major characteristics was used as it was frequently done in the past very often (Russell, 1987; Ekman, 1999). Although mental workload has been explored extensively by scientists and psychologists for the past 50 years, there is no unanimously defined nor a mutually accepted definition. Wickens (2012) refer to the term "workload" as a term not met even before the 1970's and also that the operational definitions of workload were in conflict in various fields as regarding their composition, quantified methods to reproduce and an established method of measurement.

Effectively the “mental workload” was defined as a construct including cognitive, psychological and emotional involvement entailing all components of each constellation such as motivation, predisposition, affect, sensitivity, responsiveness etc. All the above interrelated quantities were represented by two coordinate spatial dimensions signifying size and intensity. The “mental workload” by inheritance had to be expressed by two figures representing the amount of involvement quantitatively and the magnitude of dominance of the overall physiological and psychological workload. As a paradigm, mental workload was envisaged as a reactive quantity in response to immediate demands for mental activity forced on humans during brain tasks of sensory interaction (Gopher and Donchin, 1986:41-4). As the individual capabilities and endeavour differs between humans as far as their reaction to specific situations goes, the mental workload was thought to be incorporating multiple parameters and was characterized as an all-inclusive index of known and unknown constituent variables thus a highly inclusive DV for the purpose of this assessment.

In the most commonly accepted theories the known constituents of workload have been classified into three categories:

1. the number of tasks and accumulating magnitude of work,
2. the time duration and span of workload distribution, and

## Chapter 7: Discussion and Interpretation of Results

3. the psychological experiences that the human operator is subjected to (Lysaght et al., 1989).

Scientific studies about detection and quantification of mental engagement, have been conducted mainly in the areas of psychology (Eysenck 1982/2006, Hasset 1978, Slater 1998, Tsatsou 2006), clinical physiology (Ekman 1999, Insko 2003, Newman & Lieu 1998, Whitton 2003), affective computing (Peter 2006), and HCI (Brave & Nass 2008, Frijda 1986, Picard 2003) (see also the vast literature on User Experience [UX] design and game studies – e.g. (Damasio 1994/2006; Law & Sun 2012; Lessiter, Freeman, Keogh, & Davidoff 2001). As proven by conclusive results arisen from the literature cited above, although emotional stimuli cannot be as yet identified as individual emotions, physiological stimulations of the human brain elicited by those emotional events are direct, instantaneous, can be measured and quantified by using active bio-sensing such as functional Magnetic Resonance Imaging (fMRI) and electro-encephalography (EEG). Also the same stimuli can be detected with equivalent responsive accuracy (Venables & Christie 1980), by measuring electro-dermal activity (EDA), respiration rate, electromyography (EMG), electrocardiography (ECG) or combinations of the above. An additional endorsement of the established validity of bio-sensing and respective instrumentation becomes apparent when looking at the extensive use of such instruments in clinical practice. In clinical environment, biological signal measurements have been strongly established as the dominant method for non-invasive medical diagnosis (e.g. ECG, EEG etc.) and also in rehabilitation and therapy (e.g. pain alleviation, stress reduction, muscular therapy and more).

Visual expression and kinetic or postural assessment methods have been employed in affective computing research in order to reveal subliminal feelings and emotional states. These methods were not particularly efficient in detecting instantaneous responses but have been very successful in determining longer term expressions and emotional after-effects. The reason for this downside was effectively because of the time lag intervening between the moment when an emotional stimulus has taken place and has been perceptively identified and the instance following where a cognizance reaction has been physiologically established and subsequently

## Chapter 6: Experimental Evaluation

after a corresponding emotional expression occurred. An essential operational requirement in most Ambient Intelligence (AmI) and Ubiquitous Computing (UC) environments remains that the detection of contributing inputs of the system need not only be accurate but also timely and identified as early as possible. An early detection of emotional or psychosomatic state can make a difference in some cases such as for example in car driver monitoring systems or in a commanding reaction of a fighter pilot. Previous research addressing the issue of determining emotional characteristics by using physiological measurements (e.g. Ark, Dryer & Lu, 1999) have used the concept of physiological assessment from the early days of affective computing. The acquisition of physiological signals necessitated wires and electrodes to be worn overlooking or accepting the limitations of the impact that such an environment could have on the actual stress loading exerted on the subject. These restrictions not only caused discomfort and constrained the user's posture and mobility, but also had important psychological consequences, thus causing an unpredictable effect consequently affecting the precision of the measurements. In some laboratory fMRI and EEG assessment environments the subjects were even required to remain still for the duration of the experiments. Because of the above limitations one could never be absolutely certain that the responses acquired in the laboratory would be identical to those transpired in a natural environment. In the implementation described in this effort, users were completely free from binds of any kind, since they only had to use an ordinary computer mouse in the usual manner; therefore, they were unaffected by such limitations.

The concurrent excitation of two physiological quantities was assessed, namely skin transconductance (STC) and heart rate (HR), via a specifically designed interface testing strong emotional stimulation in an attempt to answer the following questions:

- Can the state of a human subject at instances where both HR and STC were increasing or decreasing simultaneously reveal a psychosomatic state?
- What is the explanation of the psychosomatic state of the subject when both the above conditions were met simultaneously?
- Is there a correlation between events that occur during the interaction with the system and the aforementioned quantities?

## Chapter 7: Discussion and Interpretation of Results

- Would the combinations provided by the two states of each of the measured quantities show a predictable pattern in response to certain emotional events?
- Can this system produce an effective mechanism for inferring a psychosomatic condition onto a descriptive model that would optimally represent clusters of emotional constructs?
- Can the emotional representation be correlated or in accordance with one of the established models of affect, namely Russell's circumplex model? (Russell & Pratt, 1980; Russell & Snodgrass, 1987).
- Finally, does this system constitute a flexible tool that could be used efficiently for detecting psychosomatic responses in AmI and intelligent adaptive or personalization interfaces?

Generating additional indicators as a derivative of generic physiological quantities such as HR and STC, has been used over time as a means of extracting augmented information from dimensionless quantities as in this case opted to be used for detecting specific psychophysiological states and / or mental workload. An added dimension providing more meaningful information about the actual value was the element of direction and it was conceived by comparing two successive values quantitatively. A lower previous value resulted in an increasing trend (+), while larger previous value showed a decreasing trend (-). In this study the above parameter was referred to as 'gradient'. On each measurement the values available for processing were therefore the values of HR, STC, HR gradient, STC gradient, Heart Rate Variability (HRV) and the trait of STC values representing the Tonic Level (TL). All the above were represented in real time revealing a more exemplified condition of the participant.

Basic emotion is termed in psychology as the distinct emotional composition that can be expressed to a similar degree of valence and arousal by all humans independent of individual cognitive, psychological behavioural or cultural characteristics (Eysenck M., 1982). As evident from a large scale assessment derived from various scientific reports through experiments performed on verbal and nonverbal expression of emotion by human subjects from various nationalities and cultural backgrounds, increasing or decreasing trends of physiological

## Chapter 6: Experimental Evaluation

responses of HR and STC can correlate to basic emotion. Relevant findings derived from extensive research representing experimentally verified observations on HR and STC gradients acquired when testing basic emotions are presented conclusively on Table 6.1 (adapted from Psaltis & Mourlas, 2011).

**Table 6.2.1**

*A review of relevant literature in tabular format (adapted from Psaltis & Mourlas 2011)*

|                               | Change in HR / STC measurement |       |         |                            |         |           |          |
|-------------------------------|--------------------------------|-------|---------|----------------------------|---------|-----------|----------|
|                               | Fear                           | Anger | Sadness | Contentment<br>/ happiness | Disgust | Amusement | Surprise |
| <b>Ax</b>                     | + / +                          | - / 0 | - / -   | - / -                      | 0 / 0   |           |          |
| <b>Christie</b>               | + / +                          | + / - | + / 0   | + / 0                      | 0 / 0   |           |          |
| <b>Ekman et al.</b>           | + / 0                          | + / 0 | - / 0   | + / 0                      | 0 / 0   |           |          |
| <b>Fredrickson et al.</b>     | + / 0                          | + / 0 | + / 0   | 0 / 0                      |         |           |          |
| <b>Levenson et al.</b>        | + / +                          | + / 0 | + / 0   | + / 0                      | 0 / +   | - / +     | + / 0    |
| <b>Nasoz et al.</b>           | + / 0                          | + / 0 | 0 / 0   | + / 0                      | 0 / +   |           |          |
| <b>Palomba &amp; Stegnano</b> | + / 0                          | + / 0 | + / 0   |                            | + / 0   |           |          |
| <b>Palomba et al.</b>         | + / +                          | + / 0 |         | + / 0                      |         |           |          |
| <b>Picachin et al.</b>        | + / 0                          | + / 0 |         |                            | + / 0   |           |          |
| <b>Sinha et al.</b>           | + / 0                          | + / 0 |         |                            |         |           |          |
| <b>Sherer</b>                 | + / +                          |       |         |                            |         |           |          |

(HR / STC), +: Increase, -: decrease, 0: no significant difference

Studying the above collection of published scientific work, it is evident that a connection between the two traits STC, HR and basic emotion is present and studied long ago but never granted a methodical in-depth exploration. This in turn imposed the justification for the use of the above physiological quantities in order to deduce emotional characteristics to an acceptable level. Further elementary analysis indicated that many of the concluding opinions coincide, thus strengthening the motivation and prospects to further investigate the findings advocated by using devices of recent technology. It is imperative however observing the above table to not have an absolute matching on all results of the findings as different assessment methodologies and testing practices have been employed. For example it has been generally

## Chapter 7: Discussion and Interpretation of Results

accepted that anger was distinguished by increased heart rate, however, two cases in the above table yielded the opposite result.

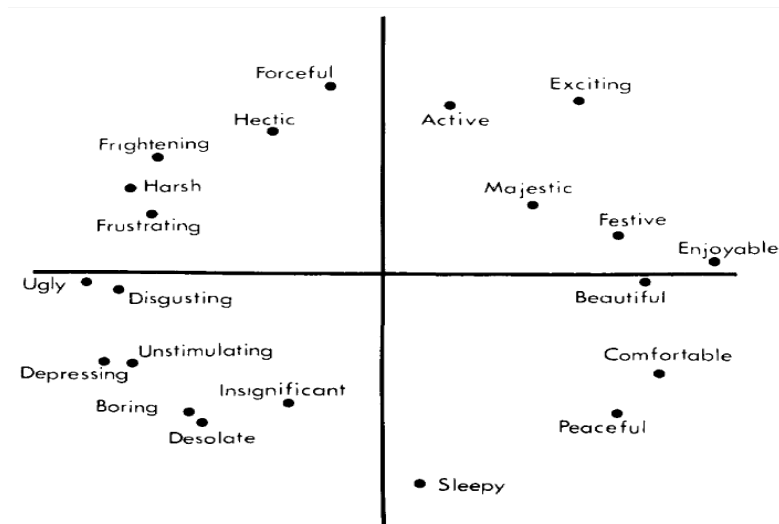
Regarding the representation and classification of affective or emotional responses, two predominant approaches may be identified.

The first is the ‘basic emotions’ approach, according to which there is a number of distinct, universally identical categories of verbal and nonverbal responses to emotion-inducing stimuli. According to Ekman (1999), undeniably the main representative of this approach, the basic emotions are anger, fear, happiness, sadness, disgust, and surprise (Adler & Rodman 2006; Scherer 2005). Although Ekman did expand the list to encompass fifteen basic emotions in total, whether all of these can be considered basic, the issue is open to debate.

The second approach for classification of emotion is a Cartesian dimensional representation, according to which affective or emotional behaviour is analysed into coordinate values. Typically the axis in such arrangement signifies horizontally the levels of satisfaction (valence) and vertically the levels of activation (arousal). Emotional constructs are thus characterised by their corresponding value pair within a domain of emotional clusters typically represented by divisions of a circular area. Such an approach can be said to lend itself better to biofeedback acquisition applications on account of allowing for an easier and possibly more intuitive quantification of affective reactions. The most prominent of these approaches is Russell’s circumplex model of affect (Russell & Pratt, 1980; Russell & Snodgrass, 1987), which posits that affective reactions can be analysed in the dimensions of valence, (evaluation of the emotion-inducing stimulus as beneficial or harmful) and arousal (activation level).

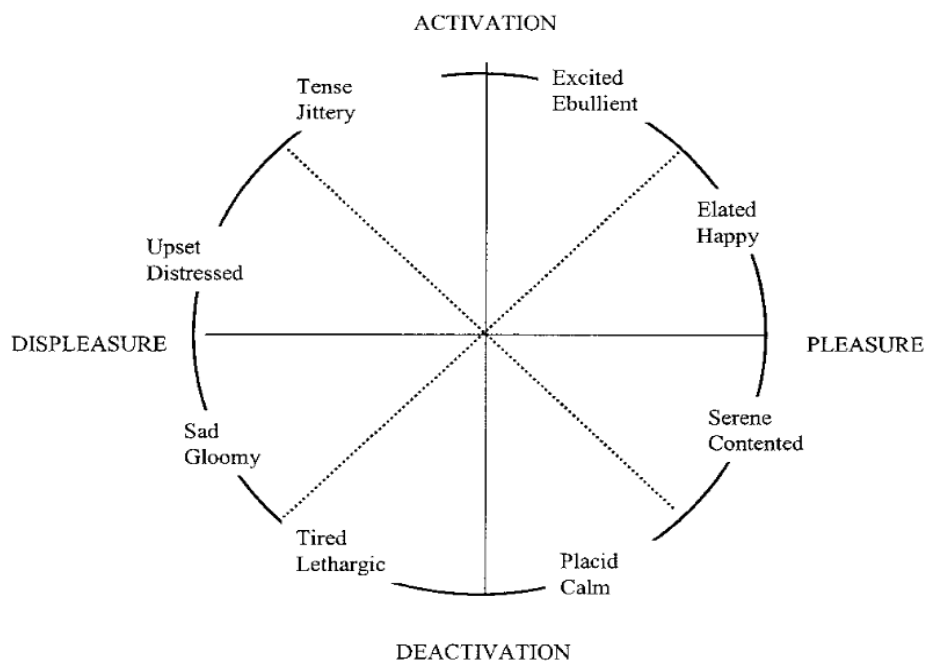
The aforementioned approaches are not necessarily at odds, as distinct emotion may be expressed in a biaxial coordinate system, as exemplified by Russell himself displaying the mappings of various terms that describe stimulus properties as combinations of arousal and valence, depicted in Figure. 6.2.1.

## Chapter 6: Experimental Evaluation



**Figure 6.2.1: Russell's circumplex model of affect (Russell & Pratt, 1980:312).**

Russell (Russell & Snodgrass, 1987) has also divided the coordinate system into a number of distinct zones each of which constitutes half a quadrant, as shown in Figure 6.2.2<sup>1</sup>.



**Figure 6.2.2: Russell's circumplex model of affect displaying eight distinct zones (Russell & Snodgrass 1987:148).**

<sup>1</sup> The oblique axes in figure 6.2.2<sup>1</sup> represent a complementary system for classifying emotions. Russell and Pratt (2006) have provided equations that allow for conversion of coordinates between the two sets of axes.

## **Chapter 7: Discussion and Interpretation of Results**

Both operational parameters influencing the composition of the system (STC and HR) were derived according to the basic principles of biofeedback measurements (Whitton, 2003). Psychological stress is manifested in the human body as a sympathetic physiological response of the autonomic nervous system which amongst other reactions induces vasodilatation of the skin's sweat glands (Tortora & Grabowski, 2006). This alteration causes measurable changes in electrical skin conductance and skin temperature that produce quantifiable indicators of stress levels.

During intensified conditions of stress, humans also experience cardiac elevation stimulated by brain hormonal secretion, abnormal respiration arrhythmias and muscular contractions such as in the area of the intestinal tract or other involuntary movement. An inclusive term used in psychology for all the above symptoms is reputedly known as anxiety, analysed further into state (chronic) and trait (personality) anxiety (Strongman, 2003). The above quantities have been chosen for the psychosomatic assessment in this study because they have been considered the most influential indicators of psychosomatic state. The gradient of HR was considered as an indicator of the direction of the sympathetic response of the subject, (fight or flight effect) attempting to compensate for immediate corrective measures taken by the brain via the autonomic nervous system and the gradient of the STC was considered as an indicator of the potency of the actions taken (rest and digest phase) as a parasympathetic response.

A specifically designed electronic device was used for the acquisition of HR and STC has been presented in previous publications of the authors (Psaltis & Mourlas, 2011). HR signal acquisition was based on the principle of Infrared Spectroscopy, using as sensing elements two near-infra-red sensors based on deflective absorption, acquiring HR pulses from the spectroscopic variation of the skin colour that occurs as a result of each cardiac pulsation. For optimized signal acquisition, minimization of artefacts and error cancellation two of those sensing circuits have been implemented employing hardware signal aggregation, complementing each other in cases of discontinuity e.g. when the contact by the user was lost in one of the two sensing elements. For stress measurement the STC method has been adopted as opposed to SR, providing a far more reliable indication of the measured quantity independent

## Chapter 6: Experimental Evaluation

of surface area of contact, temperature variations, finger pressure variations or perspiration on the part of the subject.

Stress level was identified by two silver Ag - silver chloride contact ring sensors measuring skin electrical permeability, placed on the sides of a computer mouse and at the points where the thumb and the ring fingers are normally resting during typical use of the mouse by a user. The two HR sensors were situated in the centre of the STC rings. Hardware redundancy was embedded in the system with high precision circuitry and provision for handling external noise or other sources of error such as USB noise, current leakages, optical mouse interferences etc.

In case of complete loss of contact, the system preserved the trait values measured up to the last valid measurement until the next valid measurement was detected. Since each particular measurement was evaluating the present event performing all necessary calculations for the new samples of data, discontinuities did not essentially affect the quality of assessment as each measurement was conclusive for each event. Minute losses of contact with the sensors or erratic movement of the fingers onto the mouse could produce an error in HR reading interpreted effectively in the scale of milliseconds between pulse readings, while the STC measurement was almost unaffected. Such errors were typically corrected by software algorithms implemented for curve fitting, signal smoothening, periodic pulse error detection and correction, as well as typical missing pulse and ectopic beat detection and correction. Essentially the software algorithm was processing a window of pulses and assessed consistency rules in every single measurement. In practice however, it has been justified that the contact to the sensing elements revealed a more than sufficiently robust operation as any user set to focus on visual aspects avoided distraction of their concentration or moving their fingers.

Considering that the bio-sensing system did not take into account detailed pulse shape and cardiac arrhythmias, but instead of that it was designed for accurate detection of the interval between the pulses of the heart and effectively derive the instantaneous HR and heart rate variability (HRV), reading error was within acceptable margins of tolerance of the system, as it produced a negligible signal attenuation of  $\pm 0.018$  of a pulse per reading in the worst case

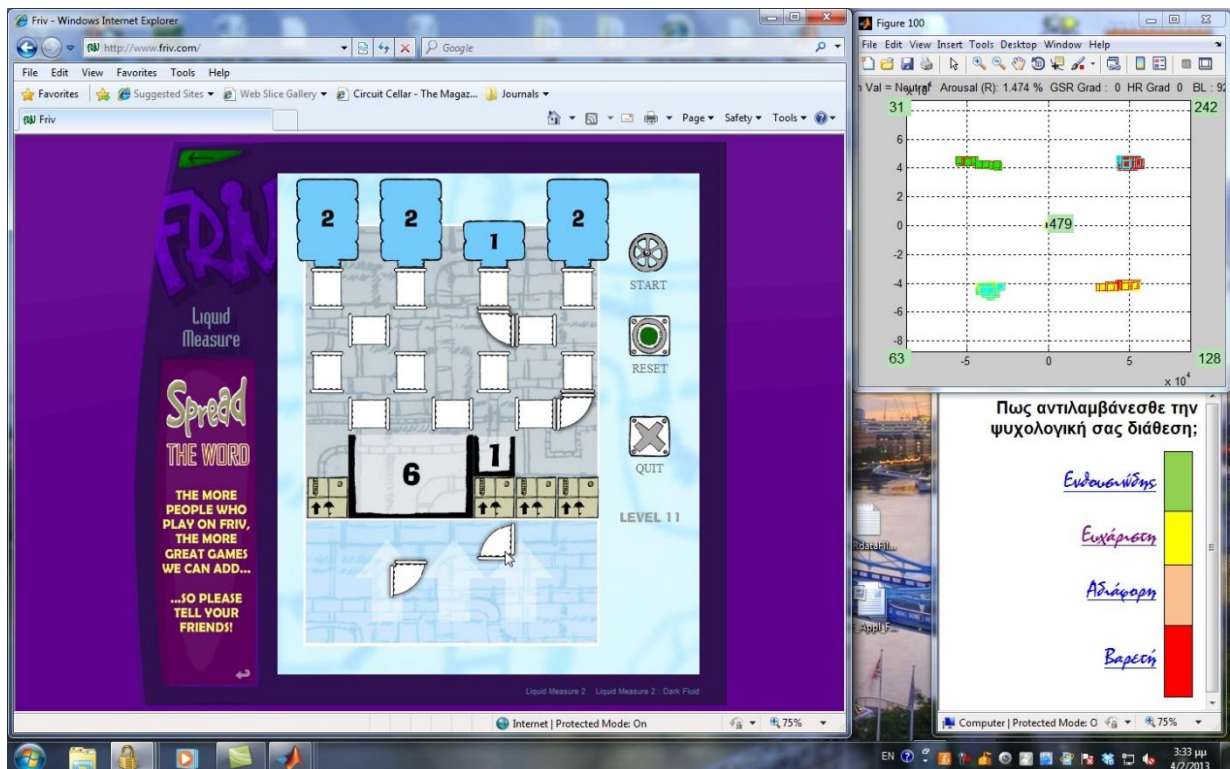
## **Chapter 7: Discussion and Interpretation of Results**

scenario. The mouse, including a signal preconditioning circuit, was connected to a computer via a specifically designed precision electronic interfacing circuit that filtered and conditioned the above primary signals and converted them into a form that could be read by the computer via the two channel audio input. A purpose-built software suite comprising the appropriate components required for a system configuration console including automatic peripheral identification and connection, as well as selection of optional recording and speed setting attributes was also developed. Finally, the signal acquisition, processing, recording and visualization algorithms were incorporated into the software suit. The use of this console was essential for the initial settings required, such as audio card selection and settings, acquisition time interval, selection of physiological signals as well as data display selection and format of storage configuration. The console provided also capabilities for the selection of row data display in a variety of formats and options for use in combination with cameras for the incorporation of gaze and eye pupil size detection software components for future development.

An STC auto-calibration algorithm was adopted, effectively providing a relative baseline for each subject independent of the actual stress levels of individuals. Since the main interest was not exactly how much stress was the user experiencing, but instead, the state of stress in relation to the previous level of stress, a method for auto calibration was devised, envisaging a reference point (“baseline”) at the mid distance between highest and lowest measured stress value weighed by the trait values. The baseline was continuously updated based on the measurements and the deviation from the mean value of STC (“tonic level”). This was important primarily because it eliminated problems like the need for initial calibration and similarly deducted continuation inconsistencies during measurements once the user released the mouse momentarily. Moreover, as an additional threshold, the baseline provided a method of distinguishing additional details in the attributes of our measurements, effectively indicating the zero point of transition during instantaneous reactions of the users (“phasic response”). Also the baseline approached quantitatively in a scale between one and ten was used in conjunction with the HRV in order to deduce emotional loading. Quantitatively the baseline could also provide a more detailed representation of emotional reactions if used to represent

## Chapter 6: Experimental Evaluation

the mid-quadrant-value (hence a diagonal axis) in each quadrant of the circumplex, subdividing the arrangement of the model into an eight area composition. Details on the use of the above parameters for projecting affect as well as user workload can be found in the section of process description and data analysis. A caption of the screen during an experiment was found below in Figure 6.2.3.



*Figure 6.2.3: Typical caption of a display during the experiment.*

### 6.2.1 Assessment process and data analysis

The primary objective of this research was to assess the conditions in which a certain physiological response was indicative of a strong emotional stimulus. Considering HR as a good indicator of inner body alertness and STC as an outer bodily expression, and given that both provide more immediate reactions compared to facial expressions, they have been treated as two independent quantities rather than values. In the aforementioned system, since two almost equally significant quantities (HR and STC) were available, a reasonable assumption was to examine conditions where both physiological quantities measured showed a common

## Chapter 7: Discussion and Interpretation of Results

tendency compared to their previous measurement and also whether they followed the same directional pattern simultaneously, i.e. both either increasing or decreasing at the same time. From the two measured quantities (HR and STC), five more secondary parameters were derived:

- Heart rate variability (HRV) derived from the recording of the difference between highest and lowest heart rate during the experiment (used in the formation of the loading factor applied to the values in the visualization process).
- Skin conductance tonic level (TL) effectively constructed as a trait of the values of measured STC raw values
- Skin conductance phasic response, comprised of the instantaneous response value of STC,
- STC gradient, refers to the vector form of STC measurements in relation to the last measurement expressed as '+' for increasing and '-' for decreasing
- HR gradient, represents the relationship between two consecutive HR measurements expressed as '+' for increasing and '-' for decreasing.

The above process was described by Eq. (1):

$$R = f[(G_{HR}(HR) \ G_{SC}(STC)] * Q_{(HRV/HR)} \quad (1)$$

where R denoted output coordinates of HR and STC values,  $G_{HR}$  was the Gradient of HR,  $G_{SC}$  was the Gradient of STC.  $Q_{(HRV/HR)}$  was described by Eq. (2):

$$Q_{(HRV/HR)} = (A_{HRV} / HR) * S_{SC} \quad (2)$$

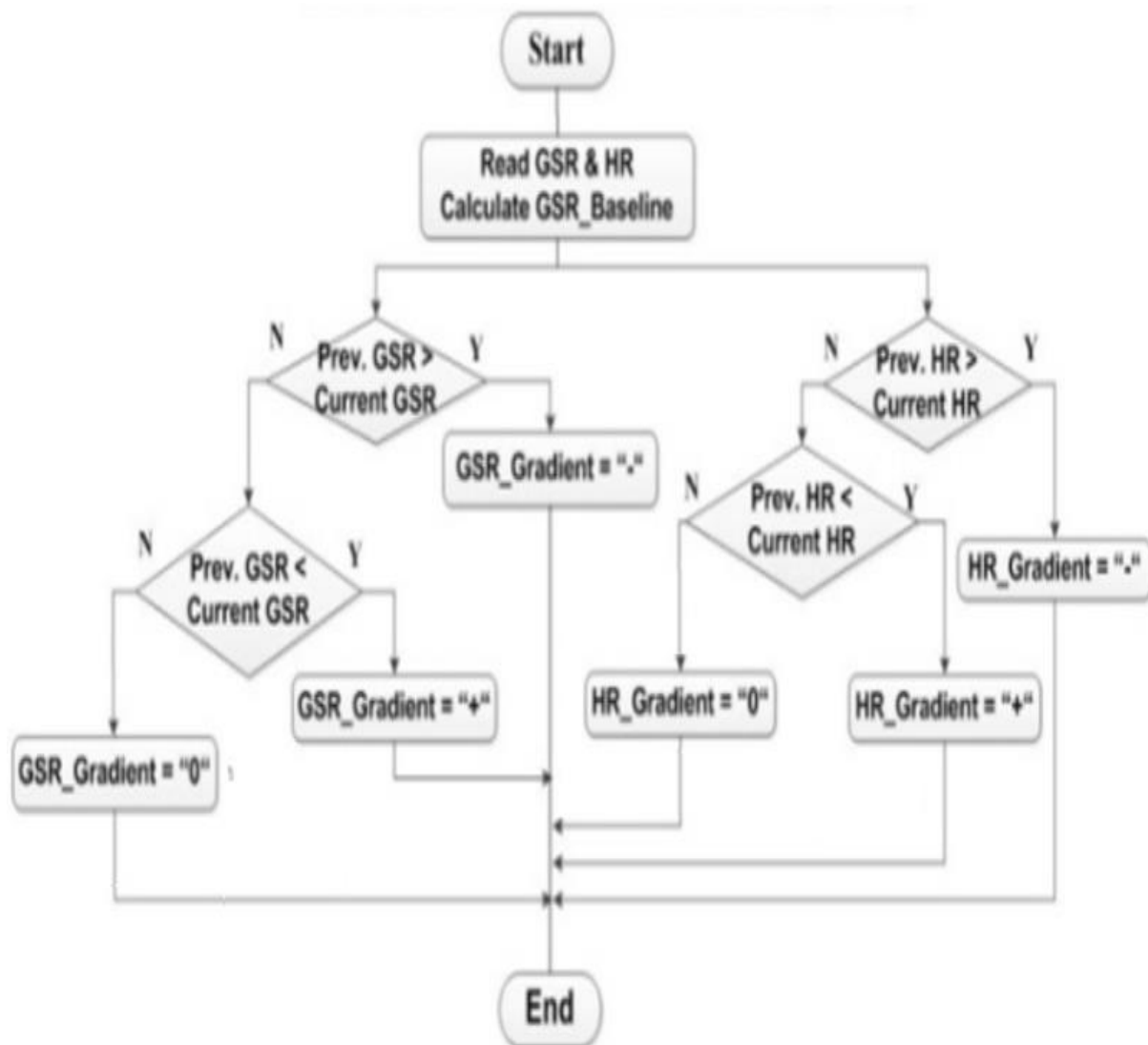
where  $Q_{(HRV/HR)}$  was the loading factor derived from the ratio of the average HRV value over the last two minutes of the session ( $A_{HRV}$ ) divided by the latest HR value, multiplied by the STC value scaled quantitatively between one to ten ( $S_{SC}$ ). Scale one was the lowest measurable stress level of the system and scale ten was the highest value respectively (Psaltis & Mourlas,

## Chapter 6: Experimental Evaluation

2011). The above factor was expressing the physiologically induced level of intensity as experienced by each individual according to their personal fluctuations of cardiac level and the quantified level of stress. The output in Eq. (1) was effectively a function of a logical AND relationship between the gradient values thus classifying each possible combination of gradients into the corresponding quadrant of the emotional context classification as illustrated further in this chapter. The actual values of HR and STC were the scaled coordinates transposed onto the appropriate quadrant. A diagonal axis that subdivided each quadrant into two octets corresponded to the baseline measurement for each user while the loading factor Eq. (2) determined the distance transposition of the coordinates from the diagonal axis (i.e. baseline value). The above distance was treated as the instantaneous emotional loading.

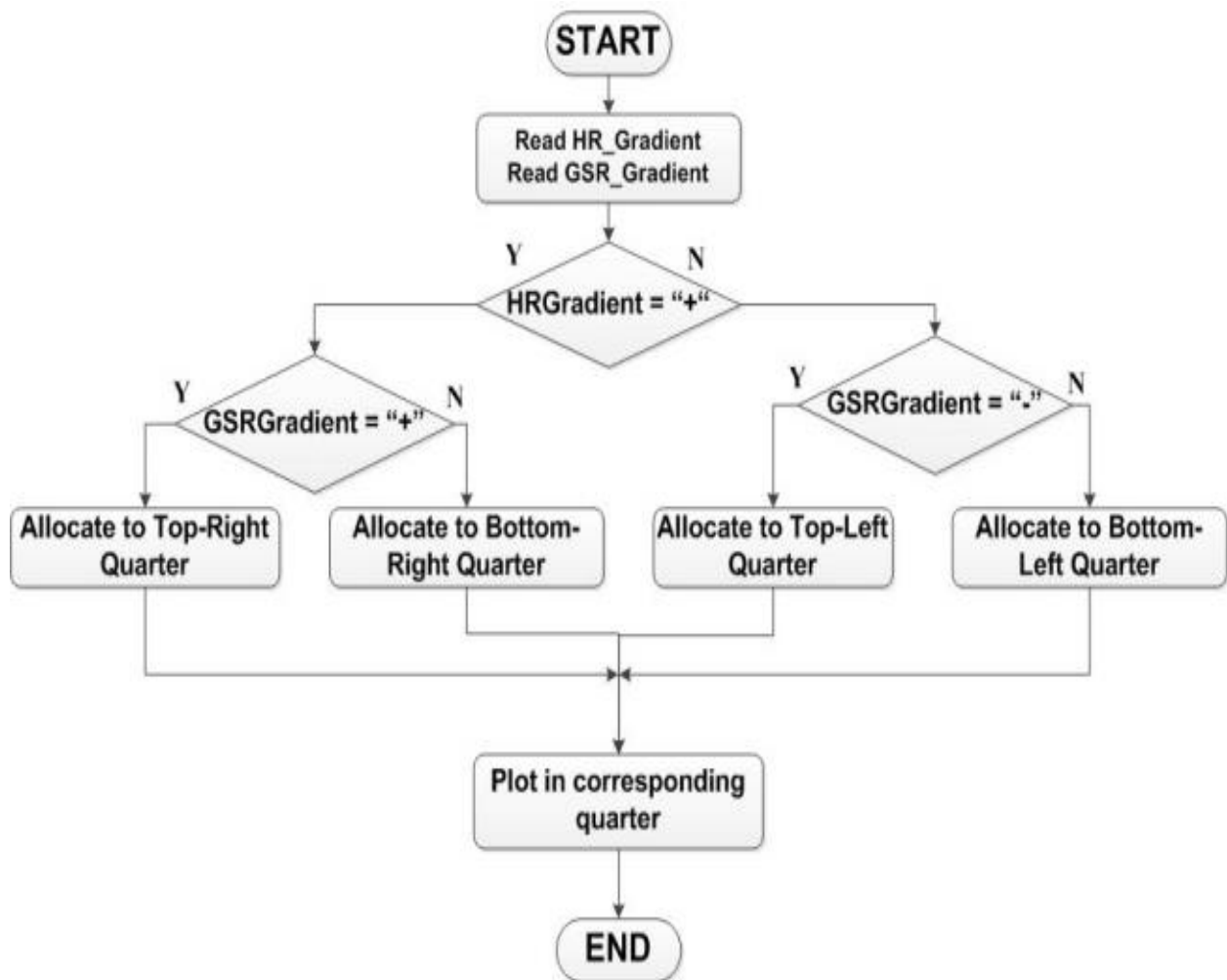
As such, the gradient value (i.e. direction indicator), indicated the tangent of the curve that depicted the user's psychosomatic state at any given moment in time. A gradient could be positive (i.e. raising pattern with regard to previous value), negative (i.e. falling pattern when the new value was lower than the previous) or zero (representing a flat pattern essentially indicating no change between the previous and current measured quantities).

The information flow diagram of the algorithm used for the classification of tendencies on each response acquired by the physiological measurement device was illustrated in Figure 6.2.1.1.



*Figure 6.2.1.1: Affective tendency classification algorithm*

The algorithm used for the visual representation of the active response to strong emotional stimuli may be seen in Figure 6.2.1.2.



*Figure 6.2.1.2: Attention cluster circumplex allocation algorithm*

Behavioural patterns of the user measured via the experimental platform were classified into four areas corresponding to four principal states of mental workload viewed as quantitative engagement, which in turn were hypothesized to correspond to the quadrants of Russell's circumplex model of affect representing also correlating clusters of emotion. It should be noted that no correlation existed between HR or STC value to either dimension of the circumplex model (i.e. arousal, valence); rather, data were mapped to the four quadrants (as shown in algorithm described in Figure 9) in relation to the states defined below.

## Chapter 7: Discussion and Interpretation of Results

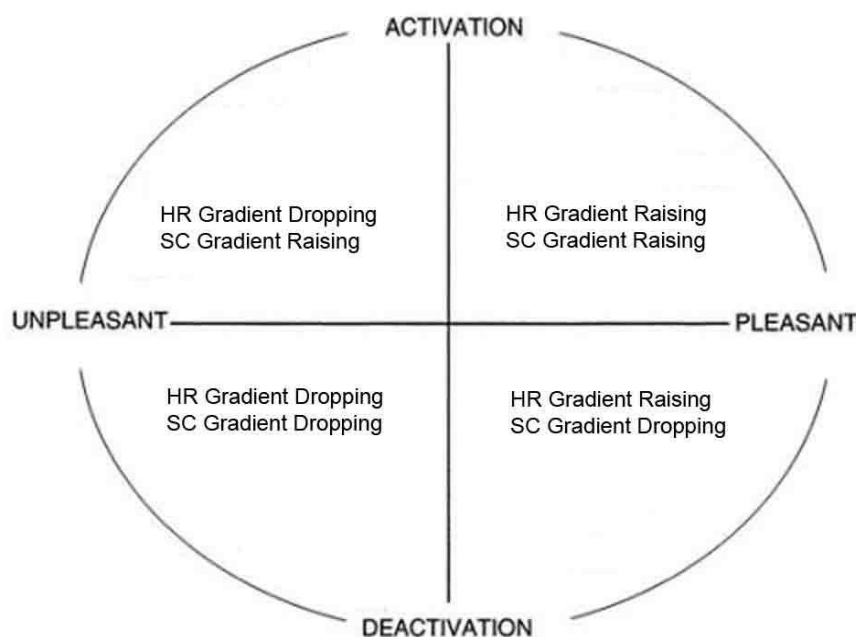
- State of *Focused Involvement / Engagement* (positive arousal and valence), where the user was expected to be in a pleasant condition and gratified as for example during or when fulfilling a task successfully.
- State of *Contentment* (negative arousal, positive valence), where the user interest was decreasing as for example when was unable to fulfil a task while still maintained a high level of amiable affect.
- State of *Perceived Difficulty* (positive arousal, negative valence), where the subject's focus on a task did not change significantly, while satisfaction was gradually diminishing.
- State of *Non-involvement / Apathy* (negative arousal and valence), where an uninterested and inattentive person performed a task in negative disposition and interest.

Evidently, a particularly precise and detailed detection of the emotional or affective states of a learner was hardly the aim of this work. Furthermore, although the two superficially acquired quantities measured in this appraisal have been considered in the existing scientific findings to date as very substantial, a detailed emotional identification was considered beyond the capabilities of the system. The system was developed in an attempt to design a novel system as a tool that can derive reliable indices of alertness and focusing engagement as they are being experienced by the user. Several patterns of interpretation were studied and adopted those that the stimuli produced by the experimental scenarios constituted a more realistic representation of the users' psychosomatic state compared to attempting to infer emotion in detail. The decision to test the system in real life environment for detecting user engagement in vivid real world applications and not in laboratory or mixed environments was made so that a realistic output of system responses was produced, and typical reactions to specific events could be studied, however, a more detailed study of carefully selected basic emotion stimuli was considered for testing at a future experiment.

During the experiments, processed data were displayed on-screen in real time, representing user responses mapped as loci onto a screen area subdivided into four quadrants. Each of those

## Chapter 6: Experimental Evaluation

quadrants dynamically represented the states described above as interpreted by the indices produced by measurements of HR and STC. One locus point was created every two seconds. The actual stress level as a raw quantity determined the position on the projected area and the magnitude showed the distance from the baseline value (imaginary diagonal axis), that was weighed by the loading factor Eq. (2) producing an orderly indicator of emotional loading. This can be explained in physiological terms as the percentage of excitation in relation to the maximum and minimum cardiac activity of the user. The relationship between the measurements and the pattern of the projected mapping was shown in Figure 6.2.1.3.



**Figure 6.2.1.3: Mapping of corresponding physiological derivatives**

### 6.2.2 Experimental procedure

Encountering the interesting challenge to design a front-end that would take full advantage of the capabilities of the system to perform assessment of emotional clusters in a thorough and convincing manner and at the same time explore to an optimal degree the latest concepts of emotion-inducing techniques in user interaction, a visual environment was designed so that all stimulating events and reactions of the user, as well as processed data from our system, could be displayed in real-time. A recording of data, as well as factors and parameters used for generic

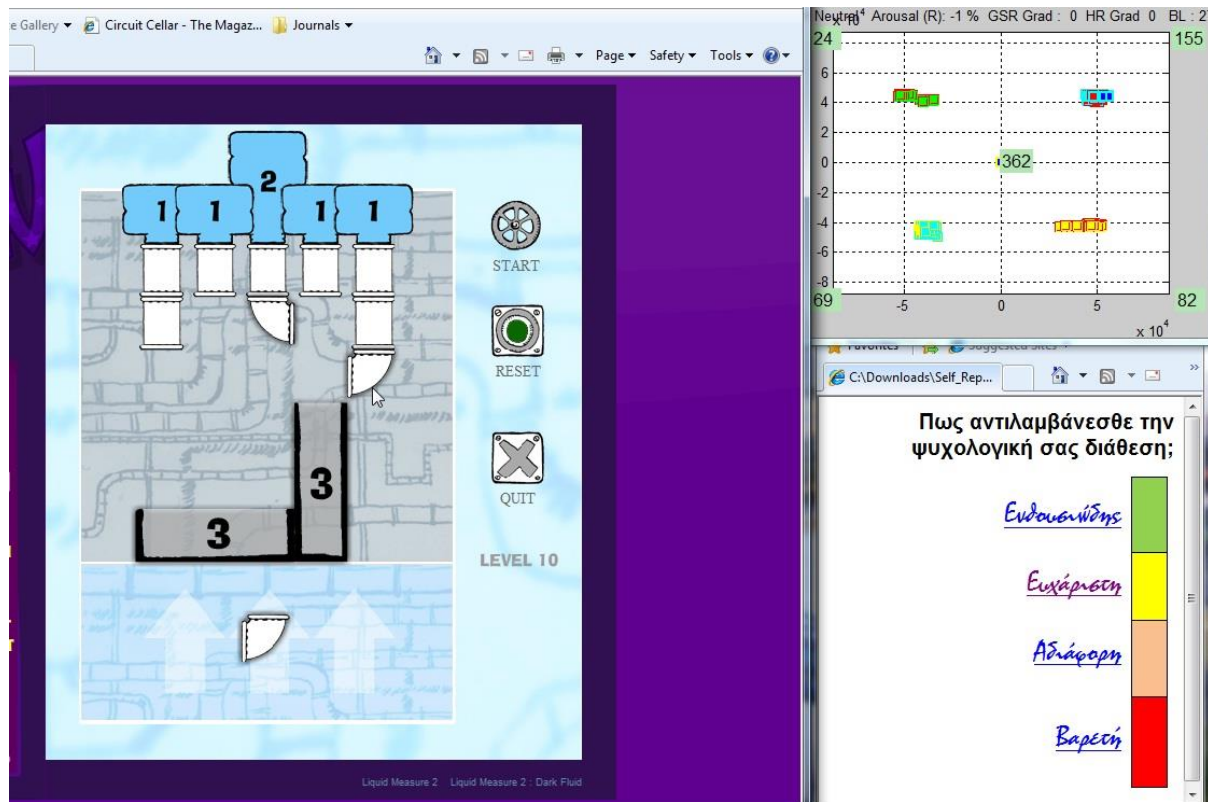
## **Chapter 7: Discussion and Interpretation of Results**

data transformation, were also produced for event tracing and post-processing. Thus the assessment and verification of the effectiveness of the system described herein was facilitated by the ability to reproduce at a later stage as well as observe in detail and verify specific user responses.

For the experiment, a typical desktop PC with average technical specification (Intel Dual Core processor, 2 GB memory) was used. The experimental sessions were prepared and tested in a computer desk setup in the laboratory of new technologies of the University of Athens, however, a typical classroom or computer room would suffice since no special laboratory equipment or restrictions were employed. Environmental conditions were approximately 25 degrees Celsius, similar during the experimental procedures and noise levels or external disturbances minimal and the same for all participants.

Each session of the experiments included two parts with an average total duration of twenty minutes (ten minutes per half session). The first part consisted of a puzzle game (“Liquid Measure”, available at <http://www.friv.com/>). The game was intended to elicit mental effort while the participant was trying to find their way in the unknown environment of the game, positive affect as a result of successful task completion, and negative affect in cases where the participant was facing difficulties hence was unable to progress. The setting was selected so that it provided to the user an unknown environment with tasks of progressively increasing difficulty. A screen caption during an experiment can be seen in Figure 6.2.2.1 below.

## Chapter 6: Experimental Evaluation



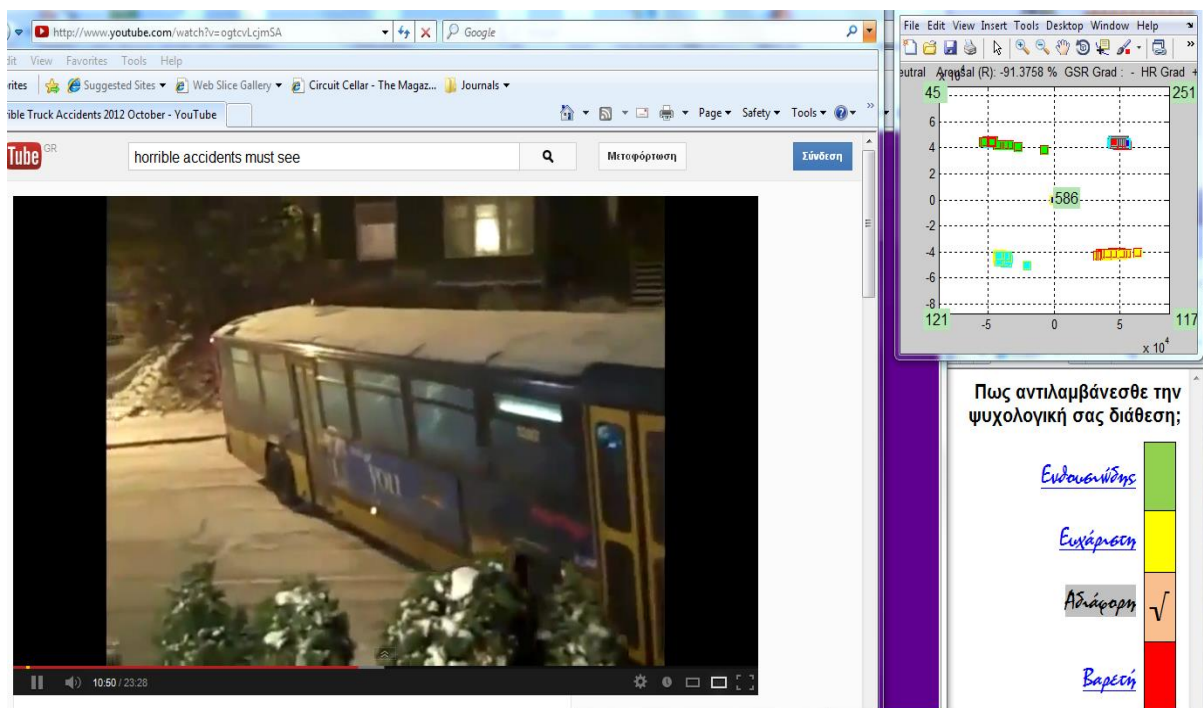
*Figure 6.2.2.1: Experiment assessing mental effort during a puzzle game*

No prior instructions were given; thus, subjects needed to concentrate and improvise in order to complete each of the twelve levels of the game. Not all levels were completed by the participants, as completion time varied between participants and the time limit was not sufficient in most cases.

The second part consisted of a video of motor fail and road traffic accidents found on “youtube” (available at <https://www.youtube.com/watch?v=26gTIQ1FDW4>). It was intended to elicit instinctive physiological responses to stimuli with negative valence (essentially indicative of distress and apprehension for the unknown). Some of the accidents were predictable, whereas others occurred suddenly. Two separate categories of emotion-inducing stimuli were included in this video. The first category consisted of accidents occurring at a distance, while the second category consisted of accidents (or near misses) involving the car

## Chapter 7: Discussion and Interpretation of Results

on which the camera was located. These two categories of stimuli were assumed to be evaluated differently. More specifically, accidents that occurred further away were assumed to be evaluated as less threatening (hence producing less vibrant measured values), compared to accidents occurring nearby (that was expected to produce higher values of HR and STC). A picture of the setting during the video experiment can be seen in Figure 6.2.2.2 below.



**Figure 6.2.2.2: Assessment during a video presentation**

In addition to the physiological measurements obtained, participants were able to express their estimate of psychosomatic condition at any time by selecting one of the four states displayed on the screen throughout the duration of the test session. It should be noted that additional information regarding user preferences on pictorial or textual assessment, personality or psychological profiles was neither requested nor taken into account in this part of the study. Apart from demographics (age and gender), the only type of personal information requested from the subjects concerned their driving experience.

## **Chapter 6: Experimental Evaluation**

The initial sample consisted of 17 university students ranging from 1<sup>st</sup> year to postgraduate students (10 women and 7 men). The age ranged from 18 to 47 years of age with a mean value of 26.2 years. The participants had no prior knowledge of our experiment setup, were given the option to refuse to participate, and were able to withdraw or refuse to complete a part at any point during the experiments. Two subjects were not included in the final assessment: one female user was excessively inconsistent in her effort to complete the experiment so that exceeded the time limits and one male participant was very talkative during the experiment. Although the consistency of the results obtained from the latter user was better than 83%, he exhibited respiration arrhythmias induced by speaking effort that could have affected stress measurements considerably; therefore, this participant was excluded from analysis. As such, the actual sample size for the experiments was reduced to 9 women and 6 men, age mean value 28.2.

### **6.2.3 Data assessment and evaluation**

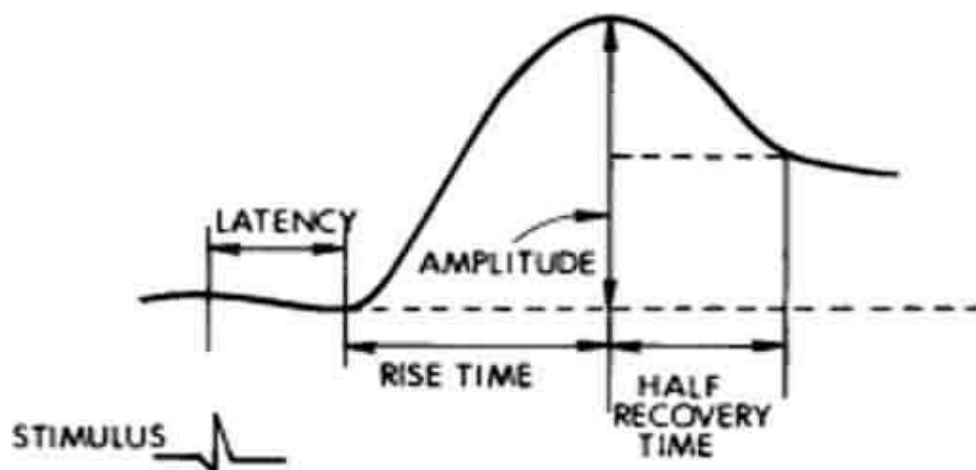
Data analysis was performed for each individual participant in three steps. Firstly, the visual content (both game and video) was weighted according to the perceived emotional impact of each particular event, providing a table of expected state of engagement, event number, and timestamp. This was essential as the timestamp of each measurement could be different for each event as users completed the tasks at different pace. Secondly, the participants' response to the emotional stimuli was assessed and the expected state of engagement for each specific event was compared to the state of engagement measured by the system. Lastly the accuracy of measurement for each participant was determined by the percentage of agreeing values between the predicted and observed states of engagement during the course of the experiment.

In fact, the correlation of the corresponding values of emotional intensity was assessed in comparison with values acquired as responses from the bio-sensing system. For each subject, data were analysed in order to formulate a distribution table that provided a more extensive indication of the convergence/divergence of emotional patterns derived from our system, thus deducing its accuracy.

## Chapter 7: Discussion and Interpretation of Results

During the development of the interface, provision was taken to integrate data post-processing algorithms and produce results in their final form ready for analysis in real time. For example, instead of post processing a vast amount of raw data after each experiment that then had to be transformed into a meaningful form, a display of processed data was integrated proactively. As such, the outcome of post-processed data was presented on the screen in real time. At the same time, detailed data recordings allowed for further exploration and tracing as required.

State of '*Contentment*' was recorded higher than expected. This was explained as in physiological terms STC decays slowly after an event stimulus, known as response overshoot as shown in Figure 6.2.3.1.

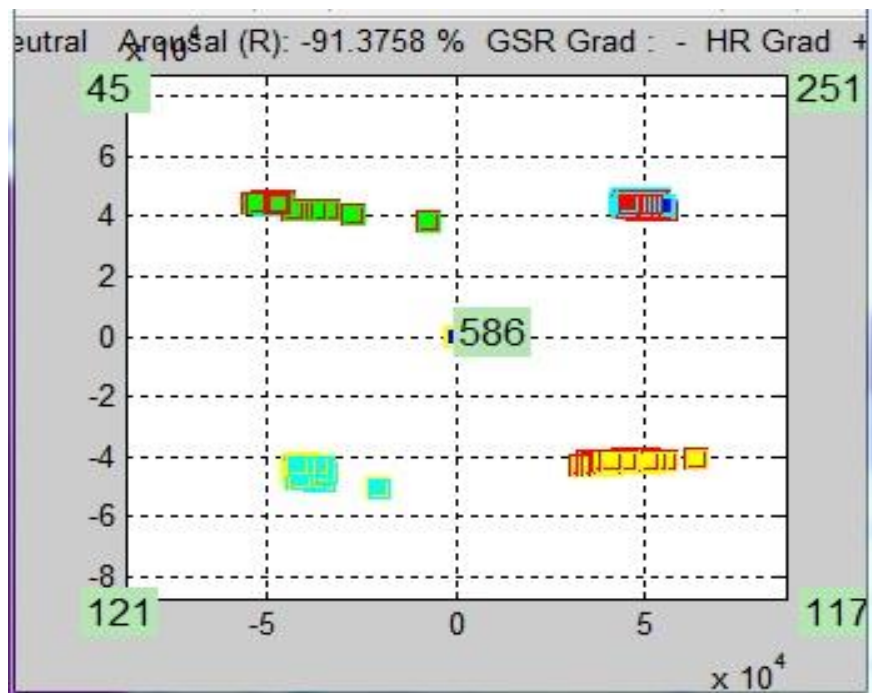


*Figure 6.2.3.1: Skin Conductance response overshoot*

The effect of slow decay of STC forced an error in situations that although the heart rate was still rising due to an event eliciting positive affect, the gradient of STC although still high was appearing to be negative, thus allocating the emotional event incorrectly to the lower right quadrant. A corrective algorithm was implemented sustaining the state of '*Focused Involvement / Engagement*' until a threshold value of the STC was reached and for as long as the HR gradient was positive.

## Chapter 6: Experimental Evaluation

A caption during a typical real time test representing the mapping of the physiological data as they were acquired was presented in Figure 6.2.3.2.



*Figure 6.2.3.2: Real-time data visualization*

The four numbers in the periphery of the graph (in the above figure the numbers 45, 251, 117 and 121 respectively) represented the actual number of occurrences of the physiological data acquired from the participant and classified by our system in this particular quadrant. The number in the centre of the graph was the index of measurement helping to identify the exact time and event during post data assessment.

At the top of the graph Tonic Level, Phasic Response as well as Arousal, Stress Level, Baseline and Gradients were displayed refreshing the newly calculated values after each cycle of measurements. The latest value was indicated with a characteristic border shown in a bright colour that was fading out as soon as the next value was mapped onto the graph. The relationship of each measurement to the baseline in each quadrant could be thought as the imaginary diagonal axis drawn from the centre of the graph towards the corners of the

## Chapter 7: Discussion and Interpretation of Results

rectangle. This additional information was considered to be remarkably valuable for future development since data presentation could be scaled so as to provide a more detailed analysis of the psychosomatic state of the user with precision of eight subdivisions of states of engagement or emotional clusters rather than the existing four. Data obtained from all 15 subjects were presented in Table 6.2.3.1.

*Table 6.2.3.1: Detailed data presentation*

| Subject ID  | Age         | No of Events |              | Response (Average) | Female        | Male          |
|-------------|-------------|--------------|--------------|--------------------|---------------|---------------|
|             |             | Game         | Video        |                    |               |               |
| F18a        | 18          | 42           | 51           | 91.38%             | 91.38%        |               |
| F18b        | 18          | 40           | 62           | 98.13%             | 98.13%        |               |
| F18c        | 18          | 48           | 69           | 89.08%             | 89.08%        |               |
| F23a        | 23          | 45           | 55           | 94.23%             | 94.23%        |               |
| F23b        | 23          | 38           | 60           | 87.76%             | 87.76%        |               |
| F24         | 24          | 44           | 60           | 88.91%             | 88.91%        |               |
| F25         | 25          | 43           | 57           | 89.40%             | 89.40%        |               |
| F27a        | 27          | 43           | 70           | 92.22%             | 92.22%        |               |
| F27b        | 27          | 40           | 67           | 94.68%             | 94.68%        |               |
| M32a        | 32          | 44           | 53           | 95.41%             |               | 95.41%        |
| M32b        | 32          | 44           | 59           | 90.26%             |               | 90.26%        |
| M35         | 35          | 48           | 55           | 86.74%             |               | 86.74%        |
| M37a        | 37          | 46           | 60           | 91.33%             |               | 91.33%        |
| M37b        | 37          | 43           | 51           | 90.61%             |               | 90.61%        |
| M47         | 47          | 41           | 49           | 94.71%             |               | 94.71%        |
| <b>Mean</b> | <b>28.2</b> | <b>43.27</b> | <b>58.53</b> | <b>91.66%</b>      | <b>91.75%</b> | <b>91.51%</b> |

In an additional empirical test conducted with a subset of the above participants whereby the users were requested to look away and divert their attention away from the computer, it was observed that values mapped the states of Perceived Difficulty and Non-involvement. As soon as the users redirected their line of sight back to the computer, their state of engagement returned to the positive valence state of Active Involvement and subsequently other quadrants according to their respective levels of involvement. This result occurred in more than 96% of the instances, making obvious that the system could detect with very high accuracy the fact

## Chapter 6: Experimental Evaluation

that a participant was taking their sight away from the screen, resembling states similar to those of non-involvement or uninterested.

### 6.2.4 Findings

Studies of data derived from the experiments indicated a high degree of proximity (91.66%) between optimally assigned values and those produced by the system with negligible instances of No-Reading errors. Participants described their psychosomatic state by selecting manually one of the states displayed on the bottom right corner of the screen as follows: ‘*Focused Involvement / Engagement*’ 71%, ‘*Contentment*’, 9% ‘*Perceived Difficulty*’ 13%, ‘*Non-involvement*’ 7% of the time of the experiment. The deviation between the values obtained through self-report and those obtained through measurements was approximately 17%, although it was worth noting that users were not inclined to use the manual selection feature very often. Positive user involvement during the experiments was represented on our model by the states of Focused Involvement and Contentment at the top right quadrant. The opposite states were actually mapping either the state where users were changing level of difficulty between scenarios or while they spent time waiting for the next event with diminished or diminishing involvement. Overall results classified all users in the above four cases at, Focused Involvement: 44.2%, Contentment: 12.6%, Perceived Difficulty: 31.9%, and Non Involvement / Apathy: 11.3% of the time of the experiment. Detailed data regarding allocation per quadrant and gender is shown in Table 6.2.4.1.

**Table 6.2.4.1: Mapping distributions per quadrant and gender**

|               | Emotional Mapping Distribution (%) |      |      |      |
|---------------|------------------------------------|------|------|------|
|               | AI                                 | C    | PD   | NI   |
| <b>Male</b>   | 44.6                               | 11.6 | 32.2 | 11.1 |
| <b>Female</b> | 43.9                               | 11.0 | 31.6 | 14.2 |
| <b>Mean</b>   | 44.2                               | 11.2 | 31.8 | 13.0 |

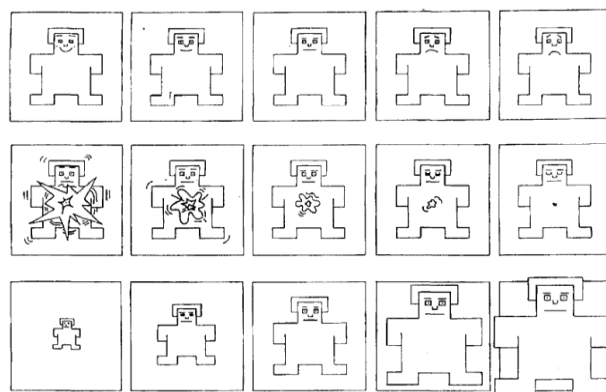
(*AI = Active Involvement, C = Contentment, PD = Perceived Difficulty, NI = Non-involvement*)

## Chapter 7: Discussion and Interpretation of Results

The difference in responses between male and female participants was trivial ( $< 1\%$ ), although event-to-response evaluation has shown coincidence between 87.76% for female and 98.13% for male users. Coincidences between predicted and measured values in the game task were similar for both male and female participants, however, differences were observed in the video session. In our view, this small difference has its origins in the profiles of male participants being affected more by aversive driving experiences (age of 32-47 with driving experience) than female participants (age 18-27 with little or no driving experience whatsoever). Differences in both male and female participants between gaming and video sessions indicated similarly accurate event / response matching during the gaming session data and that during visual observation, which was an indication that the system could perform equally well in a variety of types of tasks involving greater or lesser use of mouse movement.

### 6.2.5 Use of self-report instruments

Notwithstanding the problems associated with self-reported affective reactions as described in other sections, a cross validation of the system's output was planned with the aid of established self-report instruments, the most prominent of which being the Self-Assessment Manikin (SAM) (Bradley & Lang, 1996), and the International Affective Picture System (IAPS) (Bradley & Lang, 2007). The former is a pictorial method of assessing the arousal, valence, and dominance of emotional stimuli. The latter consists of a series of images. Each of these images corresponds to specific values in these dimensions in a manner that the images can be used as a baseline for validating emotional responses (Figure 6.2.5.1).



*Figure 6.2.5.1: The Self-Assessment Manikin (Bradley & Lang, 1996:51).*

## **Chapter 6: Experimental Evaluation**

### **6.3 Phase Three - Application oriented evaluation**

Fundamentally, the intensity of emotion varies enormously amid humans due to the fact that the subconscious judgment of an incident relies highly on a subjective evaluation by the person who is experiencing that emotion. Whilst many more factors play a key role for universally characterising emotion like different social, cultural, habitual and personal experiences it is difficult for a distinct emotion to be deduced by the proprietor and coincide each and every time with the commonly accepted definitions. People while they attempt to assess their emotional status, subconsciously, sweep scan their long term memory, in an attempt to identify and quantify their state of hedonic tone and the vivacity of this present experience (biphasic motivation– Bradley & Lang 2007). If the outcome was vague enough so that it could not produce an identifiable emotion, then the brain initiated a cognitive process comparing that very state to short term experiences of comparable emotional status in order to evaluate differences with their current psychosomatic state.

Apparently the two levels of analysis of affective reactions were purely subjective and insufficient in most cases to specify the correct emotion, and even when a person was certain that they had identified an emotional element, the case of combined emotional constituents superimposing a distinct emotion and thus producing misleading emotional identification was always a strong possibility. For example pinching, holds a great similarity in subtlety and first impression to the feeling of a contact with ice or a very hot surface and for a very short time interval could have been identified mistakenly at a first impression until it had been correctly identified in a second instance. Self-assessment of emotion also required that an individual had the ability and sensitivity to differentiate among discreet, overlapping, and often contradictory emotional states, however, this may not always be the case.

An alternative to self-report of emotion was the use of human physiology as a tool for identification of the expressions of the autonomic nervous system as a result of emotional experiences (Lykken & Venables, 1971). Studies in this area rely on identification of some form of neuro-signature of an emotion, using biometric (fMRI, CAT and IR thermo-graphic imaging), biofeedback, (EMG, ECG, EEG and STC), and extend to classification of gestures,

## **Chapter 7: Discussion and Interpretation of Results**

postural and facial recognition analysis. The present work outlines an attempt to develop a diagnostic device that can reliably and effectively identify emotional states, - thus presenting a viable solution for wide use and not restricted to laboratory use only as the majority of all known developments so far. The system has been designed so it can detect two physiological quantities considered as optimal components in process of identification of emotional state, namely HR and STC, completely unobtrusively by simply maintaining contact with a casing or a device such as a computer mouse or a mobile phone. Data were obtained through the system while users were exposed to emotion causing pictorial content selected from the pre-validated and scientifically approved IAPS material.

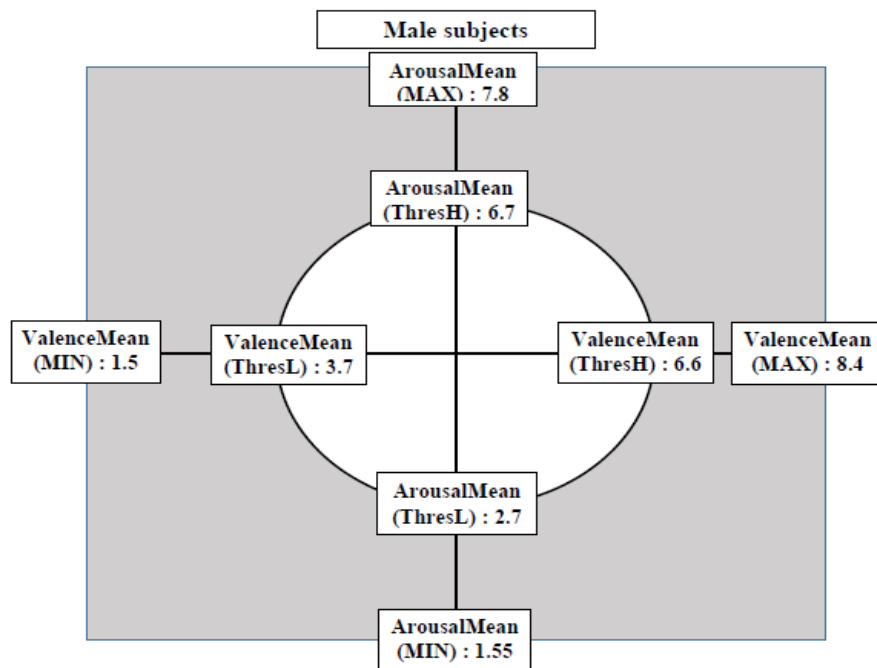
Measurements obtained were mapped onto Russell's Circumplex model of affect (Russell 2003) in real time as raw data, classified in four emotional groups determined by the precipitation of the tangents of HR and STC. Experiments have shown dissimilar patterns but an encouraging degree of absolute hit rate to the correct emotional classification area. In addition to the physiological measurements, the participants' personality was assessed by means of the NEO-FFI personality questionnaire (Costa & McCrae, 1992); Trait Anxiety was ascertained by means of Spielberger's (1983) State and Trait Anxiety Inventory (using only the Trait Anxiety form); additionally, Emotion Regulation was assessed. Detailed analysis of the data evaluation and selection could be found in the following section, followed by description of the experimental procedure, analysis of the findings, discussion, conclusion, and future enhancements.

### **6.3.1 Data evaluation and selection criteria**

Looking towards one of the major hurdles in the process of building experimental environments dealing with assessing emotion, the three parameter model of the attuned pictorial stimulus of emotion dataset provided by IAPS was adopted. Only two values derived from those parameters, mean valence and mean arousal, (the third being dominance) were used in our assessment, and those were corresponded to HR and STC precipitation, respectively. The images in the IAPS inventory were sorted in ascending or descending order according to the values of arousal and valence, in all four possible permutations (i.e. high arousal and high

## Chapter 6: Experimental Evaluation

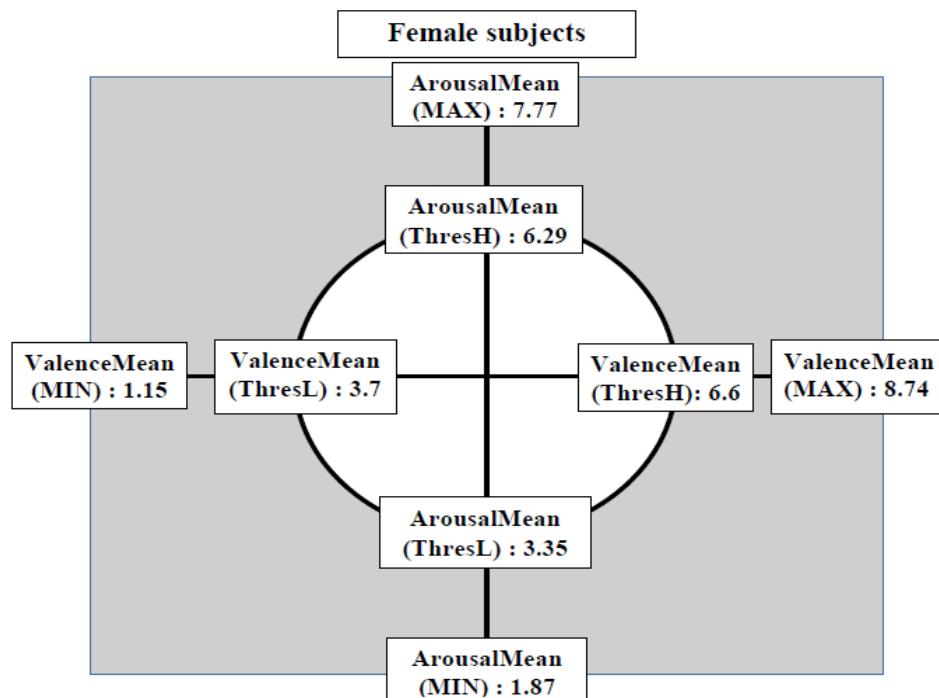
valence, low arousal and high valence, low arousal and low valence, and high arousal and low valence). In order to obtain pictures that would clearly fall into each classification area, we used a threshold value determined by the distribution of images per section area as illustrated in the following pictures (Figure 6.3.1.1 and Figure 6.3.1.2).



*Figure 6.3.1.1: Selection area of IAPS images for male participants*

Essentially, the cluster of selected IAPS content was picked from the areas of corresponding values of valence mean and arousal mean that fulfilled the criteria represented in the grey area. Figure 6.11.1 shows the areas of accepted values for the pictures found in the IAPS inventory for male subjects while Figure 6.11.2 the ones for female candidates respectively. These values were selected from the complete cluster of data in order to avoid pictorial content that would produce ambiguities during assessment as values falling into the inner circle were more likely to be detected in the wrong quadrant compared to images that fall in the grey area.

## Chapter 7: Discussion and Interpretation of Results



*Figure 6.3.1.2: Selection area of IAPS female content*

Similarly, Figure 6.3.1.2 shows the areas of accepted values for the pictures selected from the IAPS tools for female candidates.

### 6.3.2 Experimental framework

Prior to the main experiment, the participants completed the NEO-FFI, Trait Anxiety, and Emotion Regulation questionnaires. The experiment consisted of two parts. During the first part, the participants were exposed to a sequence of pictures depicted from the pre-evaluated inventory of IAPS. Two sets of sixty four pictures were used, one set for male and one set for female participants respectively. The groups of pictures were different for male than those for female subjects as the criteria of IAPS dictate. Each picture had values of valence mean and arousal mean that classified the picture within one of the four quadrants. The sixty four pictures were divided into groups of four; these four images were selected within values that would allocate them in the same quadrant.

## Chapter 6: Experimental Evaluation

Images were presented sequentially, one every 6 seconds, separated by a monochromatic blank page with blunt neutral content that was displayed for 5 seconds. The latter was necessary so as to even the transition between contents and minimize the effect of interference between emotional stimulations. Consideration had to be taken for dissimilarity effects on the individuals during the experiment such as order effects and consequential transition effects. For this purpose the pictures classified for each of the four main areas of emotion according to Russell's circumplex model of affect were grouped into four successive views for each quadrant and were spaced in an exhaustive order of combinations as shown in Table 6.3.2.1.

*Table 6.3.2.1: Order pattern of viewing of IAPS pictures.*

|                          |                         |                         |                       |
|--------------------------|-------------------------|-------------------------|-----------------------|
| Pos – Act                | Neg – Act $\Rightarrow$ | Pos – Pass              | Neg – Pass            |
| Pos $\downarrow$ Pass    | Pos $\uparrow$ Act      | Neg $\downarrow$ Pass   | Neg $\uparrow$ Act    |
| Neg $\downarrow$ Act     | Neg $\downarrow$ Pass   | Pos $\downarrow$ Act    | Pos $\downarrow$ Pass |
| Neg – Pass $\Rightarrow$ | Pos – Pass              | Neg – Act $\Rightarrow$ | Pos – Act             |

Valence was denoted as either positive (Pos) or negative (Neg) and arousal as either active (Act) or passive (Pass) respectively. During the process of the experiment half of the groups viewed the order shown in Table 6.3.2.1 while the other half viewed the pictures in reverse order.

The second part of the experiment involved the use of video clips – given that video was considered as most captivating when used as a medium for eliciting affective reactions (Gross & Levenson 1995). The video clip set consisted of four video clips with duration of approximately two minutes each. A duration of two-minutes was deemed sufficient for the participants to comprehend the content and express their emotional state, while at the same time being concise enough to avoid causing tediousness. A neutral grey background was displayed for five seconds after each clip.

The selection of images and video clips was made employing cultural factors in order to avoid biases that might be encountered across the Greek population and were carefully selected to represent a typical emotion causing stimuli based on Russell's Circumplex emotional

## **Chapter 7: Discussion and Interpretation of Results**

aggregation. The criteria for selecting the video clips were focusing in scenarios that would cause (in order of appearance):

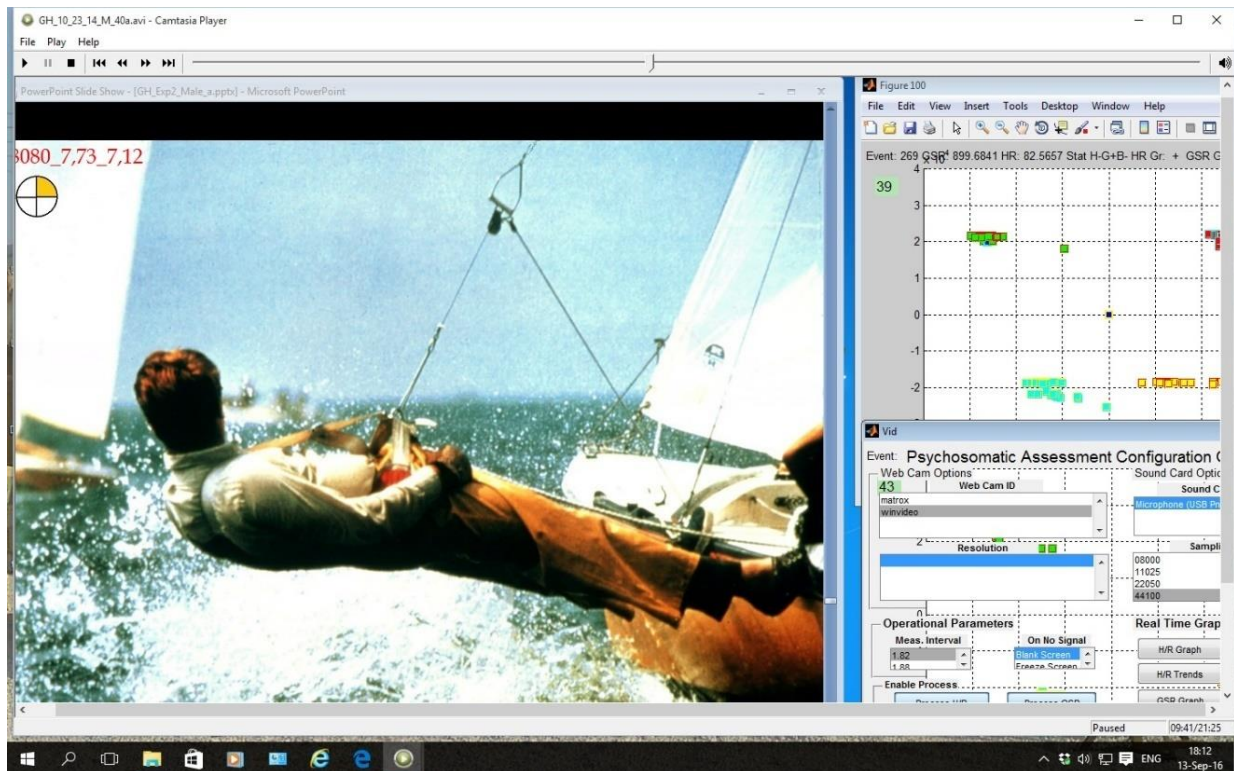
1. positive arousal and valence, presenting material causing satisfaction derived from widely accepted pleasant and exciting tasks, incorporating the emotional cluster of excitement, happiness and astonishment.
2. negative arousal, positive valence, where the scenario maintained a pleasant state while being in a low state of activation including emotions namely as contented, serene, calm and relaxed.
3. negative arousal and valence, with content biased towards a negative mood and activation incorporating the emotional cluster consisted of tension, anger, distress and frustration.
4. positive arousal, negative valence, presenting tasks with less satisfaction and clustering sadness, depression and fatigue.

### **6.3.3 Experiments**

Data obtained from the bio-feedback session incorporated a recording of the session of the experiment including the visual part and real time responses as the subject was viewing the content. A raw measurement data file was also recorded including generic measurement quantities as well as processed values for reference if required. While the participants were viewing the scenarios, the system was allocating their psychosomatic response in one of four quadrant areas representing clusters of emotion. Assessment was made for each individual by observation of the absolute hit ratio in the valid emotional cluster only. No other combinations of allocated responses were accounted for as for example adjacent quadrant weighing factors or subtracting weighing averages.

A screenshot of data capture during the experiment is shown in Figure 6.3.3.1. Pictures selected from the IAPS batteries including evaluated valence and arousal displayed on the largest area of the screen, while the real time responses of the bio-sensing system can be seen on the top right corner.

## Chapter 6: Experimental Evaluation



*Figure 6.3.3.1: Screenshot of experiment using IAPS evaluated pictures*

Twenty seven individuals took part in the experiments, (11 Male, 16 Female) ranging from 18 to 50 years of age (mean age = 28.04, s.d. = 10.282). A pre-session familiarization interval was barely necessary as the system did not require initial calibration for each participant. One measurement was obtained every second.

Data obtained from the psychosomatic response acquisition as absolute hit rates (i.e. exactly as expected) were shown in Table 6.3.3.1 for IAPS pictures and Table 6.3.3.2 for video clips respectively.

Wherever parametric tests are used it is implied that the distribution of the variables in question is normal (as ascertained through Shapiro-Wilks test of normality). Otherwise non-parametric tests are used.

## Chapter 7: Discussion and Interpretation of Results

**Table 6.3.3.1: Correct detections of emotional clusters using IAPS pictures.**

| Gender   | Age   | Absolute hits per quadrant |                     |                     |                     |
|----------|-------|----------------------------|---------------------|---------------------|---------------------|
|          |       | Q <sub>1P</sub> (%)        | Q <sub>2P</sub> (%) | Q <sub>3P</sub> (%) | Q <sub>4P</sub> (%) |
| Male     | 18    | 90                         | 46                  | 44                  | 29                  |
| Male     | 20    | 88                         | 42                  | 46                  | 31                  |
| Male     | 21    | 83                         | 38                  | 52                  | 17                  |
| Male     | 21    | 88                         | 42                  | 42                  | 19                  |
| Female   | 23    | 77                         | 27                  | 56                  | 19                  |
| Female   | 28    | 83                         | 44                  | 19                  | 21                  |
| Female   | 46    | 85                         | 38                  | 27                  | 19                  |
| Female   | 50    | 73                         | 35                  | 42                  | 17                  |
| Female   | 50    | 75                         | 29                  | 38                  | 23                  |
| Male     | 39    | 77                         | 25                  | 63                  | 21                  |
| Male     | 40    | 79                         | 54                  | 31                  | 48                  |
| Male     | 47    | 71                         | 33                  | 46                  | 27                  |
| Female   | 23    | 73                         | 33                  | 48                  | 21                  |
| Female   | 23    | 48                         | 15                  | 33                  | 06                  |
| Female   | 21    | 58                         | 71                  | 54                  | 38                  |
| Male     | 23    | 71                         | 27                  | 54                  | 21                  |
| Female   | 23    | 73                         | 31                  | 25                  | 19                  |
| Female   | 22    | 73                         | 15                  | 40                  | 10                  |
| Male     | 37    | 67                         | 35                  | 42                  | 19                  |
| Female   | 29    | 85                         | 35                  | 29                  | 15                  |
| Female   | 24    | 65                         | 77                  | 58                  | 31                  |
| Female   | 21    | 88                         | 38                  | 29                  | 06                  |
| Female   | 21    | 63                         | 29                  | 46                  | 19                  |
| Female   | 21    | 63                         | 38                  | 31                  | 23                  |
| Male     | 21    | 79                         | 35                  | 54                  | 33                  |
| Female   | 21    | 83                         | 29                  | 19                  | 17                  |
| Male     | 24    | 79                         | 44                  | 42                  | 17                  |
| Averages | 28.04 | 75.44                      | 37.22               | 41.11               | 21.70               |

## Chapter 6: Experimental Evaluation

*Table 6.3.3.2: Correct detections of emotional clusters using video clips.*

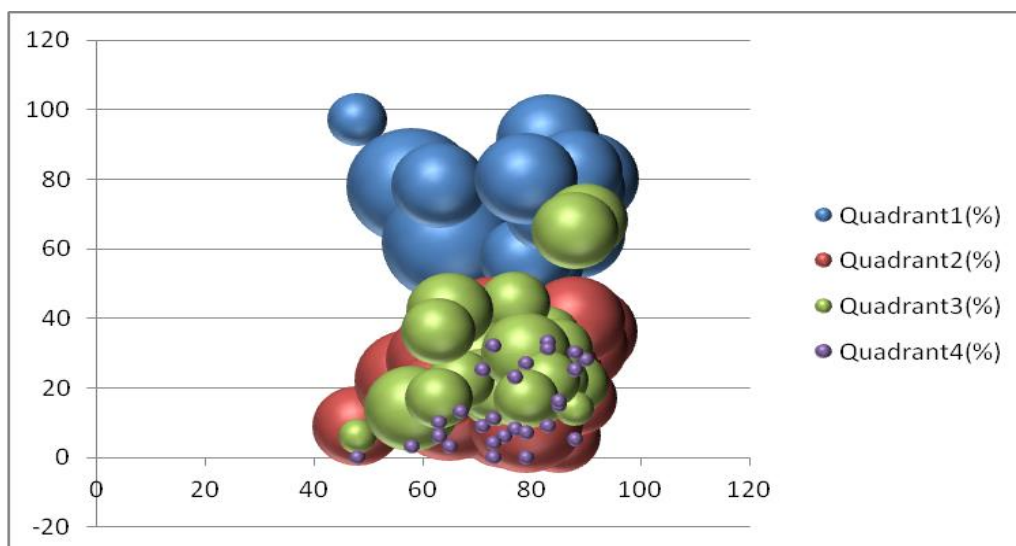
| Gender   | Age   | Absolute hits per quadrant |                     |                     |                     |
|----------|-------|----------------------------|---------------------|---------------------|---------------------|
|          |       | Q <sub>1V</sub> (%)        | Q <sub>2V</sub> (%) | Q <sub>3V</sub> (%) | Q <sub>4V</sub> (%) |
| Male     | 18    | 80                         | 36                  | 68                  | 28                  |
| Male     | 20    | 77                         | 32                  | 65                  | 25                  |
| Male     | 21    | 68                         | 23                  | 35                  | 33                  |
| Male     | 21    | 64                         | 39                  | 21                  | 30                  |
| Female   | 23    | 57                         | 14                  | 44                  | 08                  |
| Female   | 28    | 91                         | 12                  | 25                  | 31                  |
| Female   | 46    | 68                         | 25                  | 31                  | 15                  |
| Female   | 50    | 54                         | 39                  | 17                  | 32                  |
| Female   | 50    | 61                         | 09                  | 19                  | 06                  |
| Male     | 39    | 53                         | 24                  | 44                  | 23                  |
| Male     | 40    | 55                         | 25                  | 22                  | 27                  |
| Male     | 47    | 51                         | 21                  | 28                  | 25                  |
| Female   | 23    | 66                         | 28                  | 00                  | 04                  |
| Female   | 23    | 97                         | 09                  | 06                  | 00                  |
| Female   | 21    | 78                         | 22                  | 14                  | 03                  |
| Male     | 23    | 60                         | 18                  | 22                  | 09                  |
| Female   | 23    | 54                         | 17                  | 19                  | 00                  |
| Female   | 22    | 53                         | 19                  | 28                  | 11                  |
| Male     | 37    | 40                         | 25                  | 22                  | 13                  |
| Female   | 29    | 70                         | 06                  | 22                  | 16                  |
| Female   | 24    | 62                         | 14                  | 42                  | 03                  |
| Female   | 21    | 82                         | 17                  | 14                  | 05                  |
| Female   | 21    | 77                         | 28                  | 17                  | 06                  |
| Female   | 21    | 78                         | 19                  | 36                  | 10                  |
| Male     | 21    | 57                         | 10                  | 30                  | 00                  |
| Female   | 21    | 70                         | 22                  | 22                  | 09                  |
| Male     | 24    | 80                         | 28                  | 17                  | 07                  |
| Averages | 28.04 | 66.78                      | 21.52               | 27.04               | 14.04               |

It is worth commenting that all figures in the previous two tables denoting low engagement matching percentages (below 10%) were either because particular candidates had coincidentally watched the particular clip many times before. One female participant had problem handling

## Chapter 7: Discussion and Interpretation of Results

the apparatus indicating bad contact. Problems of contact with the sensing device were not encountered but partly in one section of the experiment caused by erroneous handling, otherwise data acquisition was obtained dependably and fully uninterrupted.

Data obtained from the psychosomatic response acquisition as successful allocations display optimally higher success rates as higher activation and more positive valence was expressed. A graphical representation indicating clearly this effect can be seen in Figure 6.3.3.2, presenting a percentage rate of successful measurements for IAPS pictures and video sessions cumulatively for each Quadrant.



**Figure 6.3.3.2: Correct detections of emotional clusters for IAPS pictures and videos.**

An important constraint in the study reported in this paper was related to the selection of the videos for the second part of the experimental procedure. Whereas the image sets consisted of images with pre-validated arousal and valence scores (IAPS), the videos were selected with practical and empirical criteria based upon the realistic expectations of the psychosomatic response they would elicit to the participants. However, the responses of the participants to the images for each quadrant (as detected by the system) were not correlated with the responses to the videos for that respective quadrant – in other words, no significant correlations were found

## Chapter 6: Experimental Evaluation

between image set 1 and video 1, image set 2 and video 2, image set 3 and video 3, and image set 4 and video 4 (as one would expect if the videos did indeed elicit the same affective reaction as their respective image sets). This was mainly because by design the picture assessment set was depicted with specific emotional allocation criteria, however, the video captions were actually causing a wider range of emotion much harder to concur.

Statistically significant correlations among image sets and videos may be seen on table 6.3.3.3.

**Table 6.3.3.3: Significant correlations between the accuracy of quadrant detection among image sets and videos.**

|     | Q4P    | Q1V    | Q3V   | Q4V    |
|-----|--------|--------|-------|--------|
| Q1P |        |        | .410* | .515** |
| Q2P | .652** |        |       |        |
| Q3P |        | -.329* |       |        |
| Q4P |        |        | .399* |        |
| Q2V |        |        |       | .522** |

\*  $p < 0.05$ ; \*\*  $p < 0.01$

(rows and columns that contain no significant correlations are omitted)

Furthermore, paired samples t-tests indicated that the differences in response accuracy for each image-video pair were significant:  $t_1(26) = 2.623$ ,  $p = 0.014$ ;  $t_2(26) = 5.400$ ,  $p = 0.001$ ;  $t_3(26) = 4.422$ ,  $p = 0.001$ ;  $t_4(26) = 3.029$ ,  $p = 0.005$  (Table 6.3.3.4).

**Table 6.3.3.4: Differences in participant response accuracy means for each image-video pair.**

| Pair    | Image | Video | $\Delta$ |
|---------|-------|-------|----------|
| Q1P-Q1V | 75.44 | 66.78 | 8.66*    |
| Q2P-Q2V | 37.22 | 21.52 | 15.70**  |
| Q3P-Q3V | 41.11 | 27.04 | 14.07**  |
| Q4P-Q4V | 21.70 | 14.04 | 7.66**   |

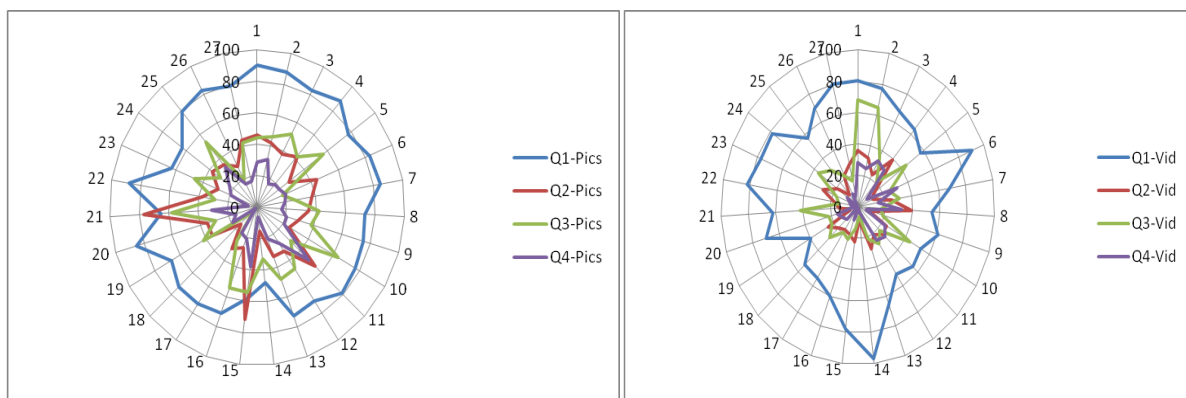
\*  $p < 0.05$ ; \*\*  $p < 0.01$

Given that the image set was already validated, the most obvious explanation for the discrepancy described above was that the videos, which were chosen for the experiments, did

## Chapter 7: Discussion and Interpretation of Results

not generate the emotion or emotional intensity exactly to the same level as that of the image set (IAPS). Another explanation, which was not necessarily contrary to the aforementioned one, was that cultural factors “offset” the expected affective response in one or both dimensions – i.e. an image that is expected to result in high valence and low arousal according to IAPS may, for reasons related to the Greek cultural context, not cause such a reaction to individuals who have primarily been exposed to that context. In light of the above, it is essential that validated video clip sets be used in future experiments so as to avoid any ambiguity in the interpretation of the results.

It is worth commenting that valid measurement rate was 98.2% for all participants with 1.8% “no-contact” errors occurring when the participants left the mouse during the test session. Percentage rates of valid measurements per quadrant can be seen in the cumulative radar plot in Figure 6.3.3.3, for IAPS pictures and videos respectively. The higher accuracy in quadrants indicating positive gradients of both valence and arousal, (i.e. high levels of engagement) can be seen clearly.

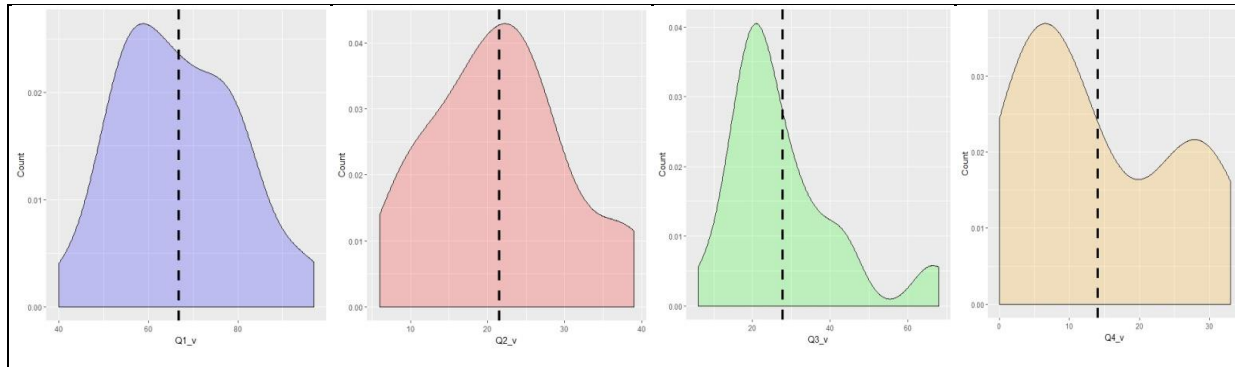


**Figure 6.3.3.3: Successful measurements per quadrant for IAPS pictures and videos respectively.**

Interesting observations can be made from studying the density functions of measured data for the responses of pictorial stimuli as well as the video sessions. The distribution of data in relation to the mean value representing the value of percent success rate, reveals behavioural patterns that can provide additional information if studied individually for each profile group. Clusters with HR/STC gradient data in quadrants 3 and 4, appear to have more distant

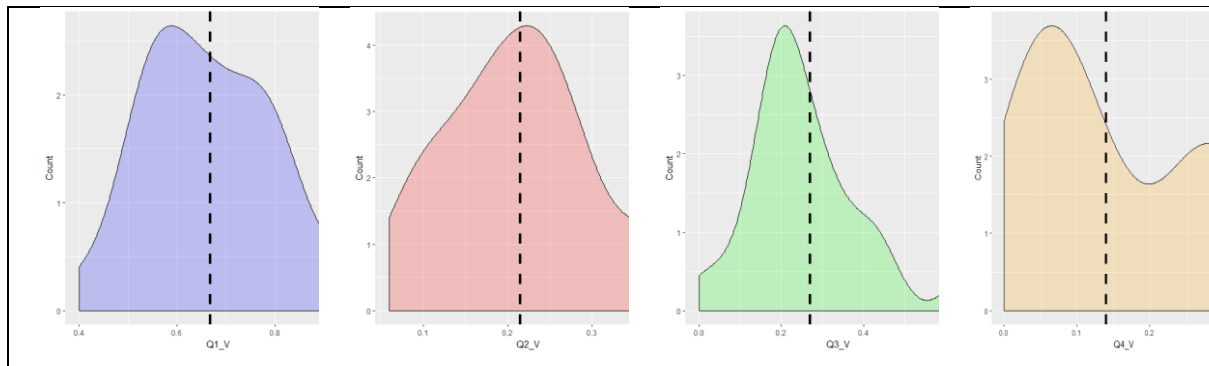
## Chapter 6: Experimental Evaluation

distributions from the mean value, showing tentative and rather volatile reading behaviour, (Figure 6.3.3.4).



***Figure 6.3.3.4: Density graph of allocated measurements per quadrant - IAPS***

For the video session the density functions expose a different pattern with salient responses only in the third quadrant as shown in Figure 6.3.3.5 below.



***Figure 6.3.3.5: Density graph of allocated measurements per quadrant - video***

Sharp responses indicated a more prominent transition in lesser space of time than those indicated by the wider distributions suggesting that impressions left in shorter amount of time (i.e. pictures) produced higher values of engagement than those produced by a smoother transition provided by consecutive impressions instigated by the video.

## Chapter 7: Discussion and Interpretation of Results

With the exception of the third video (Q<sub>3V</sub>), to which response accuracy was significantly higher for men compared to women (independent samples t-test:  $t(25) = -2.352$ ,  $p = 0.027$ ), no gender-related differences in accuracy were found to be statistically significant (see Table 6.3.3.5).

*Table 6.3.3.5: Comparison of response accuracy between male and female participants.*

|                 | MALE  |        | FEMALE |        |
|-----------------|-------|--------|--------|--------|
|                 | Mean  | s.d.   | mean   | s.d.   |
| Q <sub>1P</sub> | 38.18 | 4.285  | 34.81  | 4.956  |
| Q <sub>2P</sub> | 20.36 | 2.024  | 16.13  | 1.516  |
| Q <sub>3P</sub> | 22.09 | 4.969  | 18.06  | 5.836  |
| Q <sub>4P</sub> | 12.18 | 4.875  | 9.13   | 3.631  |
| Q <sub>1V</sub> | 67.55 | 10.093 | 66.25  | 15.597 |
| Q <sub>2V</sub> | 24.09 | 10.425 | 19.75  | 7.523  |
| Q <sub>3V</sub> | 34.82 | 18.236 | 21.69  | 10.806 |
| Q <sub>4V</sub> | 18.00 | 12.008 | 11.31  | 9.965  |

Compiled results from the psychometric tests were corresponded to those collected from the real time physiological response acquisition system. The participants' scores on the five factors of personality derived from the NEO-FFI and the categorization of these scores ("very low", "low", "average", "high", and "very high") may be seen on tables 5 and 6, respectively, separated by gender.

With respect to personality factors as numeric values, no gender differences were found to be statistically significant. However, significant associations between personality category and gender have been found in the case of Extraversion and Conscientiousness (Fisher's Exact Test,  $p = 0.023$  and  $p = 0.047$ , respectively).

## Chapter 6: Experimental Evaluation

*Table 6.3.3.6: Means and standard deviations of personality factors separated by gender.*

|          | MALE  |       | FEMALE |      | OVERALL |      |
|----------|-------|-------|--------|------|---------|------|
|          | mean  | s.d   | mean   | s.d. | mean    | s.d. |
| <b>N</b> | 20.00 | 8.44  | 25.06  | 8.23 | 23.00   | 8.54 |
| <b>E</b> | 35.27 | 8.27  | 30.50  | 4.37 | 32.44   | 6.56 |
| <b>O</b> | 27.73 | 7.24  | 32.13  | 6.73 | 30.33   | 7.15 |
| <b>A</b> | 29.00 | 7.25  | 29.69  | 3.77 | 29.41   | 5.34 |
| <b>C</b> | 31.55 | 10.55 | 33.63  | 4.90 | 32.78   | 7.59 |

## Chapter 7: Discussion and Interpretation of Results

*Table 6.3.3.7: Frequencies of the participants' personality factor categories separated by gender.*

|                                      | MALE | FEMALE | TOTAL |
|--------------------------------------|------|--------|-------|
| <b><i>Neuroticism</i></b>            |      |        |       |
| Low                                  | 2    | 3      | 5     |
| Average                              | 5    | 5      | 10    |
| High                                 | 3    | 5      | 8     |
| Very high                            | 1    | 3      | 4     |
| <b><i>Extraversion</i></b>           |      |        |       |
| Low                                  | 1    | 2      | 3     |
| Average                              | 2    | 7      | 9     |
| High                                 | 3    | 7      | 10    |
| Very high                            | 5    | -      | 5     |
| <b><i>Openness to experience</i></b> |      |        |       |
| Very low                             | 1    | -      | 1     |
| Low                                  | 2    | 1      | 3     |
| Average                              | 4    | 6      | 10    |
| High                                 | 4    | 4      | 8     |
| Very high                            | -    | 5      | 5     |
| <b><i>Agreeableness</i></b>          |      |        |       |
| Very low                             | 2    | 2      | 4     |
| Low                                  | 4    | 8      | 12    |
| Average                              | 2    | 5      | 7     |
| High                                 | 3    | 1      | 4     |
| <b><i>Conscientiousness</i></b>      |      |        |       |
| Very low                             | 4    | 1      | 5     |
| Low                                  | -    | 4      | 4     |
| Average                              | 4    | 8      | 12    |
| High                                 | 1    | 3      | 4     |
| Very high                            | 2    | -      | 2     |

## Chapter 6: Experimental Evaluation

Regarding Trait Anxiety, females (mean = 44.19, s.d. = 10.703) scored higher than men (mean = 35.18, s.d. = 9.622). A Mann Whitney test indicated that the difference in Trait Anxiety scores across genders was statistically significant ( $p = 0.024$ ). For the entire sample, the mean was 40.52 and the standard deviation was 11.061. As expected, Trait Anxiety correlates relatively highly with Neuroticism ( $r = 0.648$ ,  $p = 0.001$ ); Agreeableness ( $r = -0.423$ ,  $p = 0.028$ ). No correlation was found between Trait Anxiety and any of the other factors of personality.

The means and standard deviations of Emotion Regulation and its subscales may be seen on Table 6.3.3.8 (gender-based differences are not significant).

*Table 6.3.3.8: Means and standard deviations of Emotion Regulation and its subscales*

|                    | MALE |      | FEMALE |      | TOTAL |      |
|--------------------|------|------|--------|------|-------|------|
|                    | mean | s.d. | Mean   | s.d. | mean  | s.d. |
| <b>ER</b>          | 3.41 | 0.44 | 3.13   | 0.46 | 3.24  | 0.46 |
| <b>EI</b>          | 3.39 | 0.47 | 3.27   | 0.34 | 3.32  | 0.40 |
| <b>Ere</b>         | 3.91 | 0.56 | 3.98   | 0.48 | 3.95  | 0.50 |
| <b>EMa</b>         | 3.27 | 0.68 | 3.04   | 0.63 | 3.13  | 0.65 |
| <b>EMo</b>         | 3.09 | 0.77 | 2.91   | 0.53 | 2.98  | 0.63 |
| <b>Expe + Expr</b> | 3.19 | 0.60 | 2.81   | 0.58 | 2.96  | 0.60 |
| <b>Expe</b>        | 2.96 | 0.67 | 2.67   | 0.53 | 2.79  | 0.60 |
| <b>Expr</b>        | 3.42 | 0.75 | 2.94   | 0.80 | 3.13  | 0.80 |
| <b>SE</b>          | 3.87 | 0.70 | 3.52   | 0.64 | 3.66  | 0.68 |

*(ER: Emotion Regulation; EI: Emotional Intelligence; Ere: Emotion Recognition; EMa: Emotional Management; EMo: Emotional Motivation; Expe: Emotional Experience; Expr: Emotional Experience; SE: Self-Efficacy).*

Emotion Regulation significantly correlates with several factors and sub factors of the NEO-FFI model of personality. Specifically:

## Chapter 7: Discussion and Interpretation of Results

- Emotion Regulation (as a whole) correlates with Neuroticism ( $r = -0.718$ ,  $p = 0.001$ ) and both of its subscales, i.e. Negative Affect ( $r = -0.518$ ,  $p = 0.006$ ) and Self-Reproach ( $r = -0.724$ ,  $p = 0.001$ ).
- Emotional Intelligence correlates with Neuroticism ( $r = -0.560$ ,  $p = 0.002$ ) and one of its subscales, Self-Reproach ( $r = -0.594$ ,  $p = 0.001$ ). Additionally, it correlates with Conscientiousness ( $r = 0.472$ ,  $p = 0.013$ ) and two of its subscales, namely Orderliness ( $r = 0.400$ ,  $p = 0.039$ ) and Goal-Striving ( $r = 0.501$ ,  $p = 0.008$ ).
- Emotion Recognition correlates with the “Positive Affect” subscale of Extraversion ( $r = 0.446$ ,  $p = 0.02$ ), the “Intellectual Interests” subscale of Openness to Experience ( $r = 0.478$ ,  $p = 0.012$ ), and the “Dependability” subscale of Conscientiousness ( $r = 0.461$ ,  $p = 0.016$ ).
- Emotional Management correlates with Neuroticism ( $r = -0.495$ ,  $p = 0.009$ ) and both its subscales ( $r_1 = -0.491$ ,  $p = 0.009$ ;  $r_2 = -0.395$ ,  $p = 0.041$ ).
- Emotional Motivation correlates with the “Self-Reproach” subscale of Neuroticism ( $r = -0.457$ ,  $p = 0.016$ ), as well as with Conscientiousness ( $r = 0.57$ ,  $p = 0.002$ ) and all of its subscales, i.e. Orderliness ( $r = 0.481$ ,  $p = 0.011$ ), Goal-Striving ( $r = 0.545$ ,  $p = 0.003$ ), and Dependability ( $r = 0.398$ ,  $p = 0.04$ ).
- Emotional Experience and Expression, taken together as a single construct, correlate with Neuroticism ( $r = -0.565$ ,  $p = 0.002$ ) and both its subscales ( $r_1 = -0.416$ ,  $p = 0.031$ ;  $r_2 = -0.563$ ,  $p = 0.002$ ).
- Emotional Experience by itself correlates with Neuroticism ( $r = -0.513$ ,  $p = 0.006$ ) and its subscale “Self-Reproach” ( $r = -0.584$ ,  $p = 0.001$ ).
- Emotional Expression by itself correlates with Neuroticism ( $r = -0.468$ ,  $p = 0.14$ ) and both its subscales ( $r_1 = -0.415$ ,  $p = 0.031$ ;  $r_2 = -0.412$ ,  $p = 0.033$ ).
- Self-Efficacy correlates with Neuroticism ( $r = -0.789$ ,  $p = 0.001$ ) and both its subscales ( $r_1 = -0.598$ ,  $p = 0.001$ ;  $r_2 = -0.773$ ,  $p = 0.001$ ). It also correlates with Extraversion ( $r = 0.544$ ,  $p = 0.003$ ) and all its subscales, i.e. Positive Affect ( $r = 0.436$ ,  $p = 0.023$ ), Sociability ( $r = 0.439$ ,  $p = 0.022$ ), and Activity ( $r = 0.433$ ,  $p = 0.024$ ).

## Chapter 6: Experimental Evaluation

Trait Anxiety was found to correlate significantly with Emotion Regulation ( $r = -0.608$ ,  $p = 0.001$ ), Emotional Intelligence ( $r = -0.525$ ,  $p = 0.005$ ), Emotional Experience and Expression ( $-0.445$ ,  $p = 0.02$ ), Emotional Experience ( $r = -0.555$ ,  $p = 0.003$ ), and Self-Efficacy ( $r = -0.666$ ,  $p = 0.001$ ).

### 6.3.4 Analysis of the findings

Low to moderate correlations were found between two personality factors (Neuroticism and Openness to experience) and participant response accuracy for some of the videos. More specifically:

- Openness to experience was found to correlate with the response to Q<sub>2V</sub> ( $r = 0.384$ ,  $p = 0.048$ ) and Q<sub>4V</sub> ( $r = 0.493$ ,  $p = 0.009$ ).
- Neuroticism was found to correlate with the response to Q<sub>4V</sub> ( $r = 0.426$ ,  $p = 0.027$ ).
- Some factor subscales were found to correlate with participant response accuracy, namely:
  - Self-reproach (a subscale of Neuroticism) correlates with Q<sub>4P</sub> ( $r = -0.383$ ,  $p = 0.024$ ).
  - Aesthetic interests (a subscale of Openness to experience) correlates with Q<sub>1P</sub> ( $r = -0.391$ ,  $p = 0.022$ ) and Q<sub>4V</sub> ( $r = -0.472$ ,  $p = 0.006$ ).
  - Intellectual interests (a subscale of Openness to experience) correlates with Q<sub>1V</sub> ( $r = 0.44$ ,  $p = 0.011$ ), Q<sub>2V</sub> ( $-0.338$ ,  $p = 0.043$ ), and Q<sub>4V</sub> ( $r = -0.349$ ,  $p = 0.037$ ).
  - Dependability (a subscale of Conscientiousness) was found to correlate with Q<sub>2P</sub> ( $r = -0.404$ ,  $p = 0.018$ ).

When using a five-level personality factor category structure (“very low” to “very high”) instead of numerical scores, Conscientiousness was found to affect participant response to Q<sub>2P</sub> ( $F = 5.859$ ,  $p = 0.002$ ). When using a three-level category structure (“below average”, “average”, and “above average”) for the same factors, Neuroticism was found to influence participant response to Q<sub>3P</sub> ( $F = 3.463$ ,  $p = 0.048$ ).

The participants’ Trait Anxiety score only correlates significantly with response to Q<sub>4V</sub> ( $r = -0.460$ ,  $p = 0.016$ ).

## Chapter 7: Discussion and Interpretation of Results

Only response accuracy for the fourth image set ( $Q_{4P}$ ) correlates significantly with Emotion Regulation ( $r = 0.396$ ,  $p = 0.041$ ) and two subscales, “Emotional Experience” by itself ( $r = 0.428$ ,  $p = 0.026$ ) and Emotional Experience and Emotional Expression combined ( $r = 0.44$ ,  $p = 0.022$ ). However, a three-level categorization<sup>2</sup> of Emotional Management significantly affects  $Q_{2P}$  ( $F = 3.913$ ,  $p = 0.034$ ) and  $Q_{4P}$  ( $F = 3.527$ ,  $p = 0.045$ ). Likewise, a two-level categorization of the same construct significantly affects  $Q_{1V}$  ( $F = 5.743$ ,  $p = 0.024$ ).

### 6.3.5 Discussion

Intensified or reduced cardiac rhythm and psychological stress levels were observed in mutually coinciding patterns derived in emotion stimulating environments. The above physiological classifiers were studied in order to evaluate their merits for use as a tool for detecting and categorizing emotional activation denoting levels of engagement. A purpose built real time acquisition system that has been validated and optimised for detecting focused attention and user engagement was used for this experimental investigation. Fundamentally the system has shown a good potential by obtaining high rate of measurements with inconspicuous operation, fulfilling its primary design objective. No optimization algorithms whatsoever were applied in the above system or modification of the generic measurement responses in order to improve data allocation on the real time display model. This was elected by the authors as they intended to validate the raw capabilities of the two physiological quantities to infer emotional, affective and cognitive activation. Results have shown that the output of the system was biased towards the emotional groups with higher positive activation and disposition, as expected due to the design of the primary use of the system, (i.e. measurements have a rectified projection towards positive and active). As a result the accuracy was very good at higher levels of activation caused by emotional affective or cognitive workload, and shown diminishing performance hedonic and activation levels.

---

<sup>2</sup> We arrived at this categorization by using the z-score as a cut-off point between categories; the “Average” category contains scores that deviate less than one standard deviation from the mean. All other values are categorized as either above or below average, based on the direction of the deviation.

## **Chapter 6: Experimental Evaluation**

The capabilities of the system to detect involvement and active engagement in absolute values without any correction or optimization algorithms was optimistically high for the first quadrant (Male:75%, Female:66%) and below 50% for the other three quadrants. The latter verified the author's views that engagement could be detected more clearly when measurements produced higher activation levels of arousal and valence, as the two physiological parameters measured were not sufficient to deduce in detail the constituents of emotional elements characterised by lower activation. As shown in the results outlined in previous sections, identifying emotion was more difficult than simply measuring cumulative physiological responses interpreted as states of engagement, as reactions to identical stimuli varied among people with differences in emotional regulation and personality profile.

### **6.4 Phase Four - Comparison with commercial products**

The experiment used in this part of the system approval resembled a real life application of educational learning-while-playing environment. The operational outcome of the experiment was to test the hypothesis that levels of mental workload, engagement, positive emotions and self-reported game play experience would differ enormously while participants play free-form (FF) versus formally structured (FS) digital games. The incentive of this research was twofold, first to demonstrate that FF and FS digital games elicit different kind of game play experiences and that it can be clearly revealed by psycho-physiological measurements and secondly to compare data produced between the system developed in our laboratory and those of a commercially proven product employing facial expression in order to derive the same outcome. Accomplishment of this study necessitated simultaneous recordings of physiological measurements of skin conductance (STC), heart rate (HR) and emotion detection data derived from facial expression (Noldus). A Self-Assessment Manikin (SAM) appraisal was also included in order to assess how the players perceived their psychosomatic state while playing Minecraft which was classified as a FF game and Subway Surfers which was considered as an FS game.

A number of dependent and independent variables defined analytically in the next chapters have been derived using the STC-HR and Noldus systems in lieu of exploring the correlations

## Chapter 7: Discussion and Interpretation of Results

between game structure and player engagement, however, a very significant parameter indicating how results from both systems agree and to what extent entails an empirical assessment derived by visual observation of each particular measurement taken for each participant. The above parameter provided a very important variable for the statistical analysis and was defined as gradient and entailed the sign of the angle of the vectorised form of data measured from both systems. Possible values would be “+” for rising and “-“ for falling gradient and was used in direct comparison as it was derived by both testing systems on each instance of measurement. Exploring also the purpose to find optimal relationships between FF and FS with view to integrate characteristics of digital games in educational settings, an effort was made to identify significant correlations between the quality of players’ engagement and the style of play that a digital game offered, depending on whether the latter focused more on the game dimension (formally structured games) or the play dimension (free-from games) of gameplay. Putting the above plan in context within a broader research plan for using digital games (adventure games, in particular) as learning frameworks, it was aimed to:

- a) evaluate results produced during the experiments by two systems operating and deriving conclusions from completely different principles and identify similarities in commonly identified variables. One system was using biofeedback principles effectively measuring STC and HR, while the other system was deriving its conclusions by employing interpretation of facial expression. Comparison of conclusively derived quantities was also accomplished when and where that was possible. Finally SAM responses from the participants were also included and correlated.
- b) formalize a concept of game structure, and prove the ability to evaluate its presence/absence in different digital games; more specifically, that was an attempt to achieve the formalization and, eventually the quantitative assessment of a concept of game structure in FF and FS structured digital games, through representing the gameplay flow and states on diagrammatic formalisms such as State Transition Diagrams, and trying to arrive at typologies that would allow to place different digital games at various points on an axis between free creativity and rule-bound complexity.
- c) formalise a concept of player engagement in digital games and making possible to size the mental engagement of the user while involved in a digital game or learning

## Chapter 6: Experimental Evaluation

environment. In this respect recognising the difficulties involved when it comes to distinguishing between focusing attention, mental, cognitive, affective or psychological engagement and particularly due to the similar responses of the above psychological states when it comes to either physiological or facial expression, we aimed at measuring player engagement as a compact unit conclusively by employing a variety of available tools such as observation, questionnaires and emerging approaches such as facial expression recognition and bio-feedback measurements.

- d) explore the elements of game mechanics that digital games should have in order to exhibit a game structure that leads to optimal player engagement. Longer term intention was to develop the capability to design and validate a heuristic framework for designing formally structured, and thus more engaging, digital games.
- e) provide essential guidelines in order to select and/or design digital games for language learning for preschool and primary school children that will exhibit an optimal game structure. The capability to formulate and evaluate fundamental guidelines for selecting and/or designing digital games for learning verified by quantified biofeedback user responses and facial expression in order to provide an optimised structure and thus create maximum engagement was not in existence up to the date of this study.
- f) study the learning outcomes of preschool and primary school children on language learning while playing digital games and being assessed by our systems, with view to exploring and maximising the effectiveness of FF and FS digital games for language learning.

An analysis on how was engagement derived by measuring STC, HR and facial expression during freeform and formally structured game play and what was the role of the terms valence and arousal in the composition of the psychosomatic analysis of the user can be found further on in this text. An overview of the theoretical framework in the field of freedom of play and existing methods of evaluating game play experience was following. Next, was found a description of the experimental design and the data collection methods used in this research. Finally, the main research findings and a short discussion with concluding remarks and suggestions for future research in this area were drawn in the last section.

## **Chapter 7: Discussion and Interpretation of Results**

### **6.4.1 Free-form and formally structured digital games**

Game play thought to be by definition in game studies a special kind of a formalized subset of action, within various ‘ludic activities’, which in turn were placed within the even more general category of ‘being playful’ (Salen & Zimmerman 2004).

Fundamental to the discussion of rules and freedom in play was the distinction Roger Caillois (2001) introduced by identifying two forms of play, *paidia* and *ludus*. According to Caillois, it was possible to approach each kind of game or play form with structures, rules, goals and limits (*ludus*), or with a more unstructured, spontaneous and free act of play (*paidia*).

Gonzalo Frasca (2003) has developed Caillois’ theory further and noted how certain games were more clearly designed for a formally defined *ludus* gameplay – they were typically focused on winning and losing, the counting of points, and they generally provided players with clearly defined goals during gameplay. In contrast, games designed with more open goals, exploration, experimentation and improvisation in mind were more likely to invite a *paidia* style of free, playful behaviors. As McGregor (2008) designated, using Frasca’s understanding of Caillois terms, the spectrum between *paidia* and *ludus* in relation to digital games operates between *ludus* as a structured goal-driven type of activity, with clearly defined or formalized rules, and *paidia* as a freeform type of activity, with undefined goals and undefined but implicit or informal cultural rules.

Konstantin Mitgutsch (2008) recently noted that while a number of digital games focus on the game-dimension of game play, such as rules, goals and structures (referred to as *ludus* (Caillois, 2001)) there are other digital games that focus on the play-dimension of game play, such as unstructured, spontaneous and free act of play (referred to by Caillois (2001) as *paidia*). On this basis it may be concluded that digital games categories can be considered to lie at various points on an axis of play between totally FF activity and FS goal-driven activity:

## Chapter 6: Experimental Evaluation

The “formally *structured*” end of the axis focused on the game-dimension of game play, such as rules, goals and structures. Games that would lie at this end comprise “formally structured” digital games which are more strictly defined since

- (a) they have an explicit set of winning and losing rules which force players to take specific paths to reach goals and ensure that all players follow the same paths;
- (b) they have pre-determined and clear goals whether these are to beat players’ high score, progress to the next level or complete the game in full, etc.;
- (c) they provide immediate feedback which lets players know immediately whether what they have done is positive or negative for them in the game, whether they are staying within or breaking the rules, moving closer to the goal or further away and how they are doing versus the competition;
- (d) they have structured designer-generated activities with linear game play which confront players with a fixed sequence of challenges; and
- (e) they have a defined space and time which include any narrative or story elements in the game.

The “*free-form*” end of the axis focused on the play-dimension of game play, such as unstructured, spontaneous and free acts of play. Games that would lie at this end comprised “free-form” digital games that

- (a) had no set of rules pre-determined by the game designers or if there were rules they were informal and flexible and the player had the freedom to use them or not;
- (b) exhibited no pre-defined goals but accommodated players’ goals that were entirely intrinsic and personal;
- (c) had no “winning plot”, as they were more open-ended; and
- (d) included nonlinear game play which allowed for greater player freedom than games with linear game play.

In free-form games the players had more freedom to decide what to do, to set their own individual goals and determine their rules, path, scenarios and plots. As a system of rules that defined a victory or a defeat, a gain or a loss was not in place, the players’ aim was not to win

## Chapter 7: Discussion and Interpretation of Results

but rather to learn the game through exploration and discovery play. The players were in effect the authors of the play and the goals in digital games of this category, and this was the only decision made by the game designers in regard to the players.

### 6.4.2 Psycho-physiological metrics

Recent methods of evaluating game play experience embraced both objective and subjective measurements. Objective methods entail tangible non-invasive measurements that attempt to identify attention, mental effort, engagement and emotional responses typically expressed in two constituent components of active or passive state of arousal and positive or negative emotional valence. Physiological measurements providing data as indicators of player's emotional and psychosomatic condition included:

- a) Electro-dermal activity (EDA), one of the most commonly acquired direct and reliable physiological quantities (also referred to as skin conductance, SC and galvanic skin response, GSR). The quantity measured was directly related to psychological stress expressed as dilatation of sweat gland secretion, activity that was directly related to physical reactions to activated mental activity and positive arousal.
- b) Cardiovascular activity readings (heart rate, HR; pulse rate, BPM; heart rate variability, HRV) have been among the most widely used physiological indicators in many research areas, but because the heart and circulatory system could be affected and regulated by many different neurological, hormonal or adrenal bodily processes, interpreting the signal's relevance to the game context can be challenging.
- c) Muscular activity measurement (electromyography, EMG), a technology providing for readings of the electrical activity of muscles. Mental engagement, cognitive workload, affect or basic emotion cause involuntary muscular contractions in the intestinal tract as well as skeletal movements in response to excessive disconcertion. (Tortora & Grabowski 2009).
- d) Facial expression assessment (emotional facial recognition, EFR), a measurement technology using visual analysis of facial expression through emotional reaction. Basic emotion have a well reflected signature in facial expressions. Subsequently, this

## Chapter 6: Experimental Evaluation

allowed for a mapping of expressions to emotion and successively in the classification model of affect (Russell, 1980).

- e) Electroencephalography (EEG), the measurement of electrical potentials of brain activity in the form of frequency wave patterns. That measurement required the participant to be kept immovable and wear precisely located scalp electrodes, making this measurement appropriate strictly for laboratory use. Brain waves were usually described in terms of frequency bands, such as alpha, theta etc. representing empirical allocation to mental activities. (e.g. Theta waves correspond to sleeping etc.).
- f) Electro-oculogram (EOG) is the measurement measuring the corneo-retinal standing potential that exists between the front and the back of the human eye hence activity of the visual cortex. Strictly for laboratory use and advanced diagnosis of mental awareness such as in aeroplane pilots and medical pathology.
- g) Eye pupil size measurement (Pupilometry PM), visual detection of changes in participants' pupillary diameter due to dilations or contractions caused when focusing with varied magnitude of attentiveness. Since the optical nerve of the human eye is anatomically connected to amygdala substantia innominata (Jainta & Baccino, 2010) it has been recognised as a dependable indicator of emotional and affective activity (Lewis, Critchley, Rotshtein, Dolan. 2007).

Although scientific literature regarding physiological game studies is growing, there have been few attempts at validating proven psycho-physiological results in the context of digital games. Mandryk (2008) supports the use of psycho-physiological data in game research, although utilizes small sample sizes. Others researches support the use of EDA and HR (Drachen et al. 2010) or EDA and EFR (Nacke & Lindley 2009) and concluded with a recommendation for the methodology in a game context. Mandryk and Atkins (2007) modelled five emotions using an input of EDA, EFR and a cardiac measurement, to predict self-reports with tentative success. Finally, Yannakakis and Hallam (2008) successfully used a similar approach to create a model of children's entertainment preferences, measuring cardiac indices and EDA.

## **Chapter 7: Discussion and Interpretation of Results**

Psycho-physiological measurements were optimally beneficial because they could provide accurate and involuntary hence objective data recordings, unaffected by misleading behavioural or other intentional distractions introduced by the participant (Cacioppo, Tassinary & Berntson, 2007). Furthermore, the above measurements when in an appropriate experimental setup such as the one described herein, could operate and be recorded in real-time, without disturbing or affecting by any means the participant's natural behaviour. Another important benefit when employing psycho-physiological measurement methods was the adaptable sensitivity: 'measures are sensitive enough to pick up responses smaller than what the human eye can detect' (Kivikangas et al., 2011). Adding to the above, levels of arousal and responsiveness can differ among individuals and even between situations (Glynn, Christenfeld & Gerin, 2002). As far as accuracy of measurements goes, external interference factors such as unrelated physical activity or motion could alter responses - as it would be more obvious in systems measuring muscular activity that would in such cases produce misleading or hard to interpret results - however, in the system and experimental setup used for this reading the effect from the above factors was minimal, as muscular activity was not included.

According to Ravaja (2004), physiological measurements do not constitute self-reliant indicators of emotion or affect, therefore, they are most commonly applied together with supporting tools and questionnaires in order to contextualize the measurements (Nacke 2009). This process included the very important task to identify patterns in the physiological responses of users that reflected provoked emotional responses. Therefore, the correlation of physiological measurement with self-reported responses was receiving more attention in game research (Nacke, 2009; Nacke & Lindley. 2009). Knowledge about how well objective and subjective measures of gameplay experiences would complement or correlate is relatively limited to date. Mandryk, Inkepn, & Calvert, (2006), Yannakakis & Hallam, (2008), Ravaja et al. (2008) and Drachen et al. (2010) used both psycho-physiological and subjective measures in the study of emotional components of gameplay experience, however, amid those, the studies of Yannakakis & Hallam (2008) and Drachen et al., (2010) found statistical correlations among these measures. Mandryk, Inkepn, & Calvert, (2006) reported correlates between

## **Chapter 6: Experimental Evaluation**

psycho-physiological measurements and specific self-reported measures of boredom, challenge, frustration and fun.

The two systems used for the experiments served a complementary purpose, deducing from one side natural physiological responses and on the other side emotional status through analysis of facial expressions. There was however at least one measured quantity that was common to both systems, namely valence. In the SCL/HR system, valence was represented as one variable on a biaxial representation while in Noldus system valence was represented as a line in a graph. During instances that HR and STC were both producing an increased or decreased quantity compared to their previous value of measurement (slope technique), this condition was defined as positive or negative gradient respectively. Gradient responses from the HR/STC system plotted and compared to the corresponding graph of NOLDUS system for identical instances of measurement have shown a high degree of agreeing values ranging from 63.48 – 100.00%, mean 89.60% and sd=9.63%. It is worth noticing that although valence was deduced by two systems based on two completely different principles, resulting measurements were nearly identical. Since player's engagement can greatly impact learning, psycho-physiological data can indeed be useful for the assessment of player's affective performance in digital games, especially when correlated with subjective data. Until now only findings for First Person Shooting games have been reported (Nacke, Grimshaw & Lindley 2010; Drachen et al. 2010). Would the psycho-physiological and subjective measurement prove to be informative for other game genres, including adventure games? This study was endeavoured to fill the gap in the literature by examining the usefulness of psycho-physiological try-outs in conjunction with subjective methods for adventure digital games.

### **6.4.3 Research objectives and hypotheses**

The scientific objective of this research beyond the cross-validation of the two independent systems employed for the experiments was the validation of the effectiveness of a Freedom of Choice (FoC) model for the quantitative and qualitative assessment in adventure digital games based on the structural characteristics of games and the perceptions of students. The development of FoC model has been based on a systematic literature review and has been

## **Chapter 7: Discussion and Interpretation of Results**

grounded empirically in our previous studies (Kirginas & Gouscos, 2015a; 2015b). More specifically the conception of this study was to answer the question of whether psycho-physiological reactions measured by facial analysis software, biofeedback measuring devices, and self-reported experiences, represented in a two dimensional rating system based on mood (valence) and excitement (arousal), differ between digital games focused on play-dimension versus game-dimension of game play.

During an experimental session, the trait of successive STC data produced the tonic response while instantaneous measuring values produced the phasic responses, namely a discrete response to a stimulus. By design the accuracy required that psycho-physiological data were examined for individual game events, in addition to comparing the mean values scored in both of the games. Thus, five distinct game events were defined characterized by discreet and specific affective and emotional characteristics. Events were falling into one of the following headings: win, loss, reward, control and move. This classification applied to both digital adventure games (Minecraft and Subway Surfers). The game events distinguished above could be regarded as elementary choices of games, and although digital games could differ as a whole or may be quite analogous to each other, many of them include very comparable game events (thus, these game events owed to be applied with some notion of generality). Given the fact that there was minimum material from prior studies, no firm hypotheses could be framed about relationships between emotional responses and game events.

More specifically, this study aimed to (a) to measure the overall physiological responses, facial expressions (i.e. HR, STC, derived state of engagement, affect derived from facial expressions) and self-reported experiences to different kinds of digital games, (b) to examine physiological responses and facial expressions to different digital game events and (c) to cross validate measurements derived from two fundamentally different platforms of affective detection.

## Chapter 6: Experimental Evaluation

Based on the above objectives the following experimental hypotheses were formulated:

**H1:** Participants should be reporting higher levels of pleasure and excitement when playing freeform digital games (Minecraft) than formally structured ones (Subway Surfers).

**H2:** Participants should be experiencing lower values of physiological stress when playing freeform digital games (Minecraft) than those measured when playing formally structured ones (Subway Surfers).

**H3:** Participants should be experiencing in average lower HR values and lower HRV when playing freeform digital games than formally structured ones.

**H4:** Participants should express higher levels of positive engagement when playing freeform digital games than formally structured ones.

**H5:** Participants measured data should deduce higher positive experience values when playing freeform digital games than formally structured ones.

### 6.4.4 Methodology

The experiments were conducted in a classroom environment in the premises of the Laboratory of New Technologies in Communication, Education and Mass Media at the University of Athens. Physiological data acquired by using an HR and STC biofeedback system developed in house including electronic devices and computer software. Simultaneously, a typical computer camera captured players' facial expressions using FaceReader 4 Software developed by Noldus™. The screen visualization of the game, the camera recording of the facial activity content, and the screens containing the physiological data, graphs and real-time predictive parameters were synchronized into a single video display, recorded onto hard disc. After a brief description of the experiment, all participants filled out a background questionnaire. The questionnaire was used to gather personal information such as age, experience with digital games, frequency of digital gameplay and preferences of freedom of choice. The playing order was randomized between participants. Participants played each of the two games for 10 minutes following an introduction to the game controls. After each play session, participants performed a valence and arousal rating for each of the two games via the Self-Assessment Manikin (SAM) scale.

## Chapter 7: Discussion and Interpretation of Results

### 6.4.5 Participants

A total of 25 students (18 male and 7 female), aged between 9 and 12 years old (mean=10.77, SD=0.74) took place in the experiments. Three of the above students were not included in the final assessment as they refrained from proper use of the system during the session, producing unnecessary discontinuities either leaving the computer mouse or causing inconsistencies in facial video recordings. An overwhelming majority of the participants consisted of players classifying themselves as “more frequent” players. When users were asked to rate how often they played digital games, 13 subjects stated that they played games every day, 7 twice a week and 2 subjects stated that they played rather rarely (once or twice a month). A substantial part of the participants consisted of more experienced players as 20 of them used to play digital games in the past for more than 4 years.

The criteria for selecting the games used in this research were distinguished a game according to pre-designated goals and pre-designed game play paths and a game that, having no pre-designated goals, neither a single “winning plot”, could give room to many different player-generated gameplay paths. Games in the first category were considered as “*formally structured*” (FS) games (Subway Surfers), while the ones in the second category were considered as “*free-form*” (FF) games (Minecraft).

Since there are numerous FS and FF digital games, the following set of criteria were used to further select the most suitable game formation:

### 6.4.6. Nonlinear gameplay vs. linear gameplay

In non-linear game play it was implied that games have formed choices available to the players; different paths that the players could select arbitrary in order to explore the game scenario from one point to another throughout the game. Linear game play on the contrary, was characterised by the sentiment that players had to follow a single path to only one objective with no options for alternate paths or methods for accomplishing this objective. Players in linear game play could learn new experiences in a predetermined manner and developers utilise this form ensuring that players would follow a pre-designed game path exclusively.

## **Chapter 6: Experimental Evaluation**

### **6.4.7. Multiple solutions vs. unique solution to challenges**

FF games present to the player multiple solutions and challenges during the unfolding scenario and authorise many different player-generated paths for game play. Not every player will go about solving a situation in the same way and, given that these alternate solutions are reasonable, almost any challenge must have multiple ways for players to overcome it. Formally structured games offer unique solutions to challenges (every player overcomes a challenge in the same way) and this does not allow players to come up with different ways to proceed in the game.

### **6.4.8. Free vs. fixed sequence of challenges**

In FF games players have the ability to choose the order in which they face challenges. Giving players choices of different challenges to overcome allows them to put aside a difficult challenge and occupy themselves with another one for a while. After completing the second challenge, players may return to the first gratified and less stressed/better concentrated, and thereby stand a better chance of overcoming it. On the contrary, FS games have structured designer-generated activities which confront players with a fixed sequence of challenges. Games with a fixed sequence of challenges allow players to approach only one challenge at a time. In order to even attempt a second challenge, players must complete the first one. This is especially frustrating when players cannot overcome a particular challenge and can do nothing else until that challenge is successfully met.

### **6.4.9. Selection vs. Pre-Defined Sets of Challenges**

FF games allow players to choose which challenges they want to overcome or not. From the perspective of games for learning in particular, such games can result in improving students' learning by increasing their interest due to the pleasant and attractive environment that they offer, often contrary to conventional learning environments. On the contrary, FS games have structured activities which confront players with a pre-defined set of challenges, that all need to be overcome.

## Chapter 7: Discussion and Interpretation of Results

### 6.4.10 Criteria for Selection of Digital Games

Applying the above criteria, two digital games with different characteristics and gameplay types were selected and used in this research:

- 1) Minecraft (<https://minecraft.net/>) and
- 2) Subway Surfers (<http://www.kiloo.com/games/subway-surfers/>).

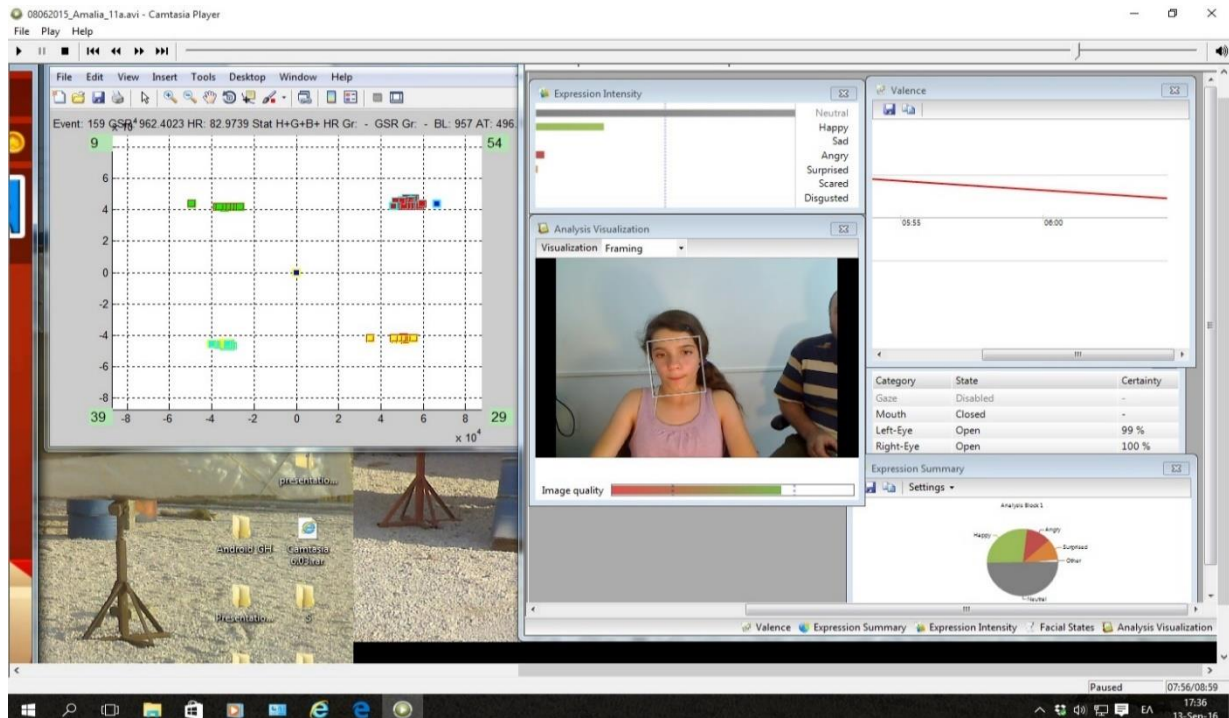
A brief presentation of these games follows. Minecraft (Figure 6.4.10.1) is an open environment game that has no specific goals for the player to accomplish, offering to players a large amount of freedom in choosing how to play the game. The core game play revolves around breaking and placing blocks. The game world is composed of rough 3D objects -mainly cubes- arranged in a fixed grid pattern and representing different materials, such as dirt, stone, various ores, water, lava, tree trunks, etc. While players can move freely across the world, objects can only be placed at fixed locations on the grid. Players can remove these material blocks by destroying them and place new ones elsewhere, thus allowing for various constructions. The game primarily consists of four game modes: survival, creative, adventure, and spectator. It also has a changeable difficulty system of four levels; the easiest difficulty (peaceful) removes any hostile creatures that spawn.

## Chapter 6: Experimental Evaluation



*Figure 6.4.10.1: Screenshot from Minecraft*

A screenshot of data capture during the experiment is shown in Figure 6.4.9.1 and a screen sample of the experimental setting in Figure 6.4.10.2 respectively.



*Figure 6.4.10.2: Sample experiment screenshot*

## Chapter 7: Discussion and Interpretation of Results

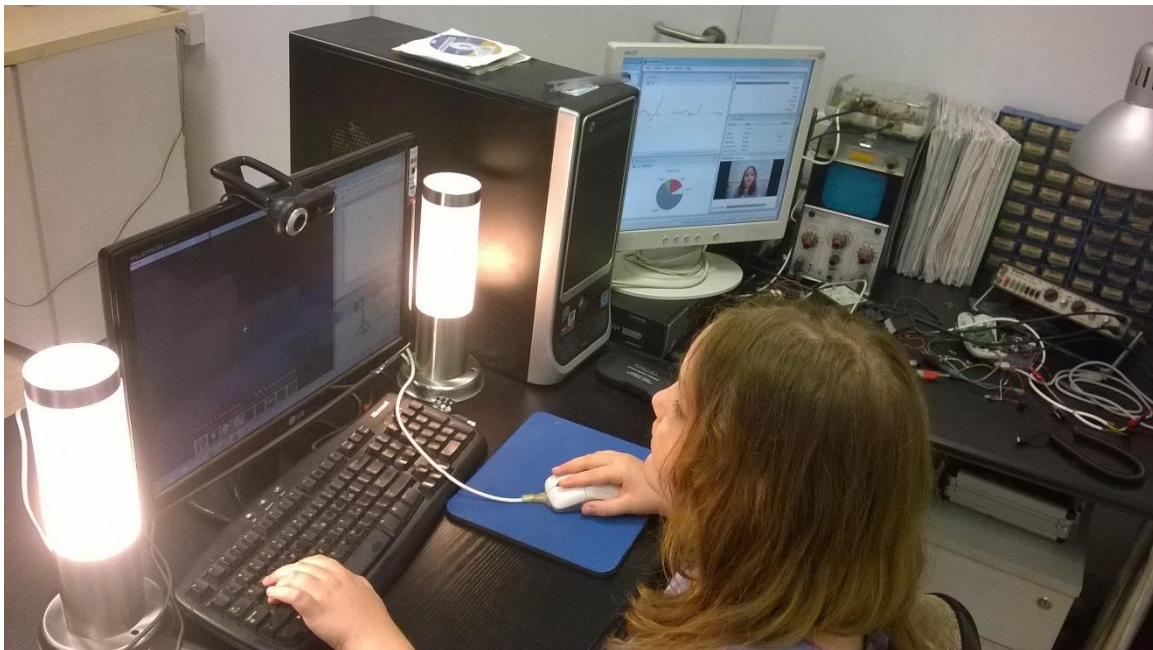
Subway Surfers (Figure 6.4.10.3) is an "endless running" digital game. Players of the game take the role of teenage hooligans who, upon being caught in the act of applying graffiti to (tagging) a metro railway site, run down the tracks to escape the Inspector and his dog. The objective of the game is to collect coins and other rewards while running through an endless game world. Trains and other obstacles must be avoided by performing well-timed jumps (swipe up), rolls (swipe down) and sideways moves (swipe left and right). Occasionally the characters surf on boards, soar over the train tracks and even run along overhead wires. Special missions reward players with bonuses for accomplishing specific tasks. The game ends when the surfer either stumbles and is apprehended or crashes directly into various obstacles. In this experiment we used a customized version of the game for PC.



*Figure 6.4.10.3: Screenshot from Subway Surfers*

## Chapter 6: Experimental Evaluation

A screenshot of the complete setup for the experiment is shown in Figure 6.4.10.4.



*Figure 6.4.10.4: Experimental setup screenshot*

### 6.4.11 Data collection methods

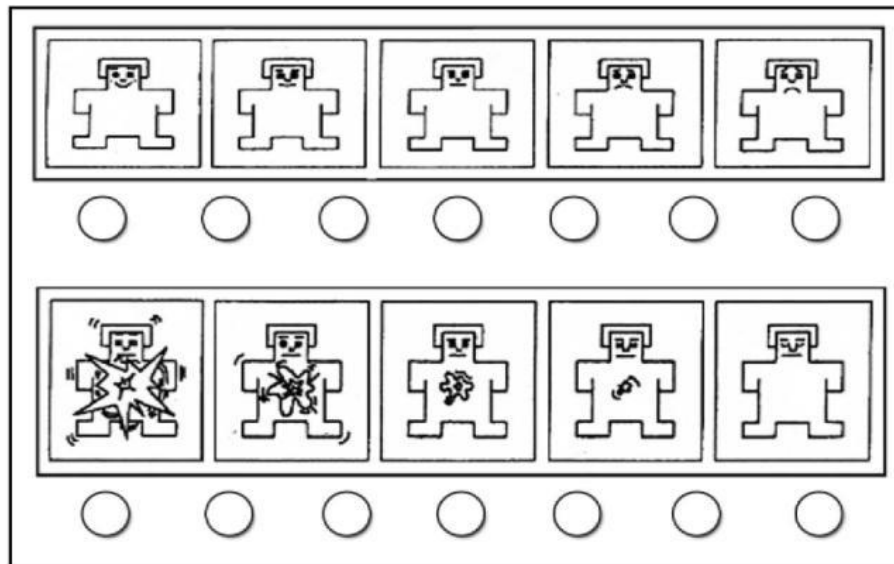
The study was designed to measure physiological responses, emotional reactions and self-reported experiences during game play. Consequently, during the study we measured both students' self-reported experience of each digital game played and measured their physiological responses and emotional reactions using a range of measurement tools. The measures are described in more detail below.

#### 6.4.11.1 Self Assessment Manikin

The Self-Assessment Manikin (SAM) was originally a three dimensional rating system based on pleasure (valence), activation (arousal) and superimposition (dominance) (Lang, 1985). In this experiment only the two dimensions for valence and arousal were utilised. Also, a seven point scale was used instead of the original nine point scale to reduce the decision time for the subjects. Figure 6.4.11.1.1 shows the graphical representation of SAM. The upper scale evaluates the valence, while the lower evaluates the arousal. Both of the scales have seven

## Chapter 7: Discussion and Interpretation of Results

different values. For the evaluation the different values got numbers from -3 to 3, in order to map them into the valence-arousal space.



*Figure 6.4.11.1.1: The Self-Assessment Manikin (SAM) used to rate the affective dimensions of valence (top panel) and arousal (bottom panel)*

### 6.4.11.2 Biofeedback Measuring Device

A device designed and built in the Laboratories of New technologies of the department was used for the experiments (Psaltis & Mourlas, 2011), featuring the advantage of a completely unobtrusive measurement and seamless operation from the part of the user. The complete system consisted of: a sensing part that was accommodated onto a typical computer mouse, an analogue electronic circuit that processed and fed the two modals acquired through to a typical home computer and finally a software component that transformed the measured quantities into an appropriate meaningful form to accommodate display and recording. STC was detected by direct skin contact of the thumb and ring fingers with contact ring shape sensors located on the left and right sides of the computer mouse respectively. The areas of the epidermis (stratum corneum) of the fingertips reputedly the densest parts of the human body in sweat glands and nerve endings (Gray, 1977) conveniently provide a highly reliable source for our instrument with optimal response to stress type stimuli. HR was detected by infrared sensors located in the centre of the ring shaped STC sensors, based on the principle of reflective Near Infra Red

## Chapter 6: Experimental Evaluation

Absorption (NIRA) occurring during the changes of the colour absorption of the skin caused by the pulsation of the blood content in the tissue. A dual sensing circuit of the HR, (one at each finger) minimised interference and reading errors and helped to eliminate movement artefacts. Further analogue filtering and scaling ensured reliable pulse detection while Software algorithms compensated for missing pulse detection, ectopic beats, offset pulse elimination, as well as inconsistencies between measurements improving further the accuracy of the system. Skin Conductance and Heart Rate values, after an appropriate value scaling were mapped in real time onto a biaxial arrangement and were treated as values corresponding to those of arousal and valence respectively. Coinciding increases and reductions of STC and HR simultaneously generated a corresponding positive or negative gradient (slope technique). The four combinations derived from the two modals (i.e.  $G_{HR}+G_{SC+}$ ,  $G_{HR}+G_{SC-}$ ,  $G_{HR}-G_{SC-}$ ,  $G_{HR}-G_{SC+}$ ) were treated as indicators, classifying thus the user state into a four quadrant representation of clusters of emotion, analogous to circumplex of emotion (Russell, 1980).

### 6.4.11.3 FaceReader Emotion Detection Software

Face-Reader version 4.0, a commercial product developed by Noldus Information Technology, was used to identify and classify facial expressions. FaceReader, using a computer camera and appropriate software program components and algorithms, evaluated on a video frame-by-frame basis, facial expressions in terms of seven emotional states – happy, sad, angry, surprised, scared, disgusted, and “neutral”. The values of those variables reflect a measure of the magnitude of the specific emotion being detected in a scale from 0 (not at all) to 100 (perfect match). The emotional state “Surprise” was excluded from the present study, given its ambiguity with respect to valence component. Behavioural problems resulting difficulties in capturing expressions from two subjects and impediments in conformance with speechlessness and breathing made data being doubtful end therefore classified unusable. While data for these participants were excluded from all analyses, it is worth noticing that the hitch was purely caused by erroneous neglect from the part of the user and not because of downsides of the systems. A caption of Face-Reader 4.0 during an experiment is shown in Figure 6.4.11.3.1.

## Chapter 7: Discussion and Interpretation of Results



*Figure 6.4.11.3.1: The interface of FaceReader 4.0*

### 7. Discussion and Interpretation of Results

Primarily the distinction between essential forms of data involved in this effort was deemed necessary since the nature of data was ranging from permeability current signals up to pure parametric derivatives and quantified states of engagement. In one end, input data were actually generated from biological electrical activity and for that reason they could not be simulated by devices or simulators. To illustrate this further, considering the skin trans-conductance measurement where initially no signal of any kind exists in the human body. If an instantaneous minute current was transmitted through the skin, it would have been dissipated and absorbed by the subcutaneous levels of the skin due to its physical electrolytic characteristics spreading the electrons to a full attenuation of the signal, thus producing no output. Repetitively nourishing the small current by generating a feedback loop, the circuit obtains intermittently the instantaneous output corresponding to the instantaneous input produced. Since the instantaneous input is stable, invariable and known, any changes in the output are entirely depended upon alterations in the characteristics of the skin permeability. Subsequently, according to studies mentioned in the literature review documented thoroughly, the measured quantity derived from the above measurement is the expression of the excitation of the autonomic nervous system.

As far as the HR measurement is concerned, the input signal is obtained by amplified optical signals detecting the change in coloration of the skin. The output of the HR circuit fed to the PC is an analogue voltage wave and therefore for testing purposes, a simulated electrical signal was used producing a pre-programmed signal identical to a perfect HR wave. The simulated input signal made possible to verify the precision and accuracy of reproduction of the shape and the count of the pulses.

Input signals as they were produced and conditioned by the electronics were effectively transformed into numerical values representing a frequency signal for the STC and a pure analogue pulse for the HR, which was then processed by software in order to derive the rate of heart beats per second. STC, HR and HRV data were then used to derive gradients of each

## **Chapter 7: Discussion and Interpretation of Results**

channel of input and then forwarded to the data display allocation algorithm on each cycle of measurement and in real time. It is apparent from the above illustration that a method to correspond the interpreted values to those expected had to be devised. The four phases of the study were undertaken to reach the point of validation, verification, repeatability and consistency of the system for approximations of evoked affective conditions.

### **7.1 Results: Phase One – Device and Systems Evaluation**

Testing procedures of the HR-STC system were designed to validate seamless operation, timely response and certainly accuracy, of measured quantities, based on:

- bench tests, (ie. spectrum analysis, frequency response stability, pulse count burst analysis),
- electronic stress tests (i.e. responses to noisy input signals, interference tests and signals below or above the allowed quantities) and
- pilot tests with actual participants and strong emotional case scenarios.

Although the input signals to this phase were essentially translating unstructured informal physiological data, it was of particular importance the fact that the above tests produced successfully a number of positive results. Firstly they proved that the seamless operation of the system can be fulfilled. Secondly they have shown a high tolerance to movement of the fingers from the part of the participant and thirdly and most importantly proved the feasibility to obtain reliably and consistently the required physiological measurements, a fact that was anticipated initially to be a very optimistic concept. The output data derived from this phase were recorded as numerical data and real time processing allowed for visual presentation. Data included raw numerical input, STC and HR graphs, trait values in charts of continuous measurements, and a real time mapping as a visual allocation data on Russell's circumplex model of affect.

Results from the validation stage in phase one have been satisfactory as the recordings and system operation was verified to operate correctly. Moreover, one of the most crucial elements derived from the tests was the reliability of signal acquisition and HR in particular. To explain this further, initially, it was anticipated that because the bio-sensing elements were on the

## **Chapter 7: Discussion and Interpretation of Results**

mouse which requires to be moved in any direction, movement of the fingers about the sensing rings could cause inconsistencies. That was proved wrong firstly because the speed of the dual analogue circuit was able to compensate and produce a dependable stabilised output, and secondly because the concentration of the participants on the screen and the scenarios proved to be the major factor so there was no erratic movement to cause inconsistencies from the first place.

### **7.2 Results: Phase Two – Systems operational evaluation**

The second phase involved experimental scenarios aiming to cause predictable affective conditions characterised by all possible combinations of valence and arousal levels as for example sorrow (caused by moving emotional scenes) or satisfaction (caused by successful accomplishment of tasks). Data produced included both, raw data recordings and display recordings allowing thus to derive the success rate of response data by direct comparison between threshold scaled inputs and visual allocation data generated by the system.

Experiment data recordings went through a visual evaluation and an exploratory data analysis with view to investigate if the system could detect consistently, physiological activities when the participant experienced strong psychological events. Inferences between responses expressed as allocation onto the model of affect were sought against the pre-valued estimates. Cognitive models employed for estimating mental workload were based on the GOMS model (Helander, 1988:135-188). In our experiments, just as GOMS batteries use a disrupter task in order to allow for working memory usage between tasks, a blank screen was intervened to reinstate the confidence interval for a few seconds, as an evacuating period from the mental activities generated by the previous task. Evaluation of the vigour of the tasks has been performed using methods taken from clinical practice as the tasks had been scaled according to behaviours that correspond to estimated mental workload. In pictorial evaluation for mental workload, marking was used in particularly important designations that at the same time related designations to tasks with severity indexes. A relationship between the physiological reactions of the participants to the presentation of the content and the categories of the explicit emotions portrayed was then assessed.

## Chapter 7: Discussion and Interpretation of Results

### 7.2.1 Findings: Phase Two – Systems operational evaluation

Correlations between scaled values of expected markings and measured responses was made to observe both system efficacy and also to identify significant differences between associated segments of quadrants of Russell's circumplex model of affect. Output metrics have shown discrete detection of participant's mental engagement while individuals experienced transition between tasks with increasing demands of intellectual activity.

Studies of data derived from the experiments indicated a high degree of proximity (91.35%) between optimally assigned values and those produced by the system with negligible instances of No-Reading errors. Participants described their psychosomatic state by selecting manually one of the states displayed on the bottom right corner of the screen as follows:

- '*Focused Involvement / Engagement*' 71%,
- '*Contentment*', 9%
- '*Perceived Difficulty*' 13%,
- '*Non-involvement*' 7%

The percent symbol corresponds to the percentage of the time of the experiment. The deviation between the values obtained through self-report and those obtained through measurements was approximately 17%, although it was worth noting that users were not inclined to use the manual selection feature very often. Positive user involvement during the experiments was represented in our model by the states of Focused Involvement and Contentment. The opposite states were actually mapping either the state where users were changing level of difficulty between scenarios or while they spent time waiting for the next event with diminished or diminishing involvement. Overall results classified all users in the above four cases at, Focused Involvement: 44.2%, Contentment: 12.6%, Perceived Difficulty: 31.9%, and Non Involvement / Apathy: 11.3% of the time of the experiment. Detailed data regarding allocation per quadrant and gender is shown in Table 7.2.1.1.

## Chapter 7: Discussion and Interpretation of Results

*Table 7.2.1.1: Mapping distributions per quadrant and gender*

|               | Emotional Mapping Distribution (%) |      |      |      |
|---------------|------------------------------------|------|------|------|
|               | AI                                 | C    | PD   | NI   |
| <b>Male</b>   | 44.6                               | 11.6 | 32.2 | 11.1 |
| <b>Female</b> | 43.9                               | 11.0 | 31.6 | 14.2 |
| <b>Median</b> | 44.2                               | 11.3 | 31.9 | 12.6 |

*(AI = Active Involvement, C = Contentment,  
PD = Perceived Difficulty, NI = Non-involvement)*

The difference in responses between male and female participants was trivial ( $< 1\%$ ), although event-to-response evaluation has shown coincidence between 87.76% for female and 98.13% for male users. Coincidences between predicted and measured values in the game task were similar for both male and female participants, however, differences were observed in the video session. In our view, this small difference has its origins in the profiles of male participants being affected more by aversive driving experiences (age of 32-47 with driving experience) than female participants (age 18-27 with little or no driving experience whatsoever). Differences in both male and female participants between gaming and video sessions indicated similarly accurate event / response matching during the gaming session data and that during visual observation, which was an indication that the system could perform equally well in a variety of types of tasks involving greater or lesser use of mouse movement.

### 7.2.2. Discussion: Phase Two - Systems operational evaluation

State of engagement was the most important psycho-physiological condition we attempted to identify and quantify in this research as it was highly regarded in Affective Computing research (e.g. Peter & Herbon, 2006) as well as other areas of application like virtual environment control.

Results have verified our views that user engagement transitions between affective activation states can be detected consistently by identifying simultaneous changes in HR and STC gradients. Indication was immediate as expressed by previous research work assessing similar aspects by means of fMRI and STC (Tsatsou, 2006). It may be concluded that the bio-sensing

## Chapter 7: Discussion and Interpretation of Results

system produced a valid and to a high degree accurate snapshot of the user's state of engagement because, as reported earlier, whenever the users looked away or otherwise disengaged their attention from the screen, the system immediately detected this shift of focus with high accuracy, by labelling their state of attention as either "Perceived Difficulty" or "Non-involvement" – both of which entail negative valence when represented as coordinates for the dimensions of Russell's circumplex model of affect.

Throughout the entire experimental process, a transition from the state of Active Involvement to that of Contentment immediately after stressful stimuli was most frequently observed. This transition was interpreted as the effect of the slow decay of the stress level, which produced a negative STC gradient, effectively indicating a reduction of arousal.

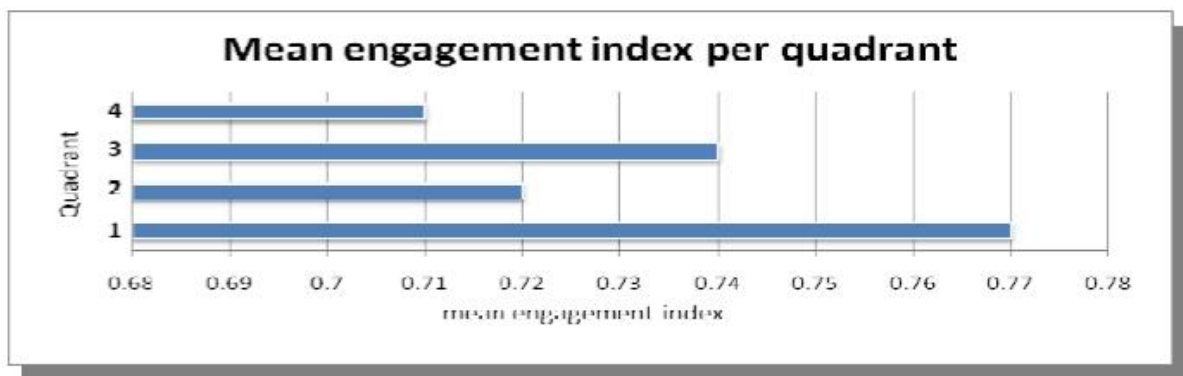
Taken together, the aforementioned points indicate that a mapping of the users' state of engagement onto Russell's circumplex model of affect was reassuringly correct, not as far as a precise emotion area goes but at least with respect to placing the identified state in the correct quadrant (i.e. positive / negative valence, active / passive arousal).

From the above, as well as the overall results of the experiments, we have indications from this system that the correlation of common gradients of present and past values of HR and STC reveal a high probability of success in determining various states of engagement. The time interval for each measurement was crucial for the accuracy of the system and optimized accordingly in order to provide enough time for the subjects' physiological response to settle into a detectable timeframe. Additionally, this timeframe allowed for the normalization of artefacts that could have been misleadingly accounted for as spontaneous reactions of the participant.

The time frame was chosen following optimization deduced by the assessment of data produced by adding data frames of measurements into a larger buffer of data. Time intervals between measurements for over 2 seconds had shown a smoother transition between states but also slowed down the detection of user response and therefore were not adopted.

## Chapter 7: Discussion and Interpretation of Results

With reference to results produced in recent studies assessing affect and mental engagement (Chaouachi, Chalfoun, Jraidi & Frasson, 2010) entailing multi-modal physiological assessment including STC, HR, EEG, EMG and audio, remarkable similarities could be seen, although the system used in this study only employed STC and HR and was certainly holding an advantage as its use was not restricted for laboratory use only as was the case in the above study, Figure 7.2.2.1.



**Figure 7.2.2.1: Mean engagement index per quadrant (Chaouachi, Chalfoun, Jraidi, & Frasson, 2010)**

Success rate in all four levels of engagement ( i.e. ‘*Focused Involvement / Engagement*’, ‘*Contentment*’, ‘*Perceived Difficulty*’, ‘*Non-involvement*’) was unexpectedly high as derived from the analysis, although it was observed that accuracy was higher towards levels with positive valence and higher intensity of arousal. This was considered to be an important finding as studies employing fMRI support that brain activity is higher at levels of negative to neutral valence levels and lower intensity of arousal (ref.) In our view, this happens because the aforementioned system measures the physiological expression resulting from the brain activities involved in affective, emotional and cognitive effort while an fMRI system detects pure neuronal activity that may at the same time correspond to other than the somatic system. Effectively, our system has shown capabilities to identify scaled magnitude of state of engagement with growing accuracy from smaller affective activation states to higher activation levels. Also it was observed that a transition from the state of Active Involvement (i.e. positive valence and high arousal) was frequently reduced to that of Contentment (i.e. positive valence

## Chapter 7: Discussion and Interpretation of Results

and lowering arousal) immediately after stressful stimuli. This transition was frequent and was caused due to the effect that psychological stress response shows a swift ascend subsequent to a stressful event, followed by a slow decay while the parasympathetic system intervenes to establish the initial state of equilibrium; as shown in Figure 7.2.2.2.



*Figure 7.2.2.2: Typical skin conductance response exhibiting the immediate overshoot and slow decay.*

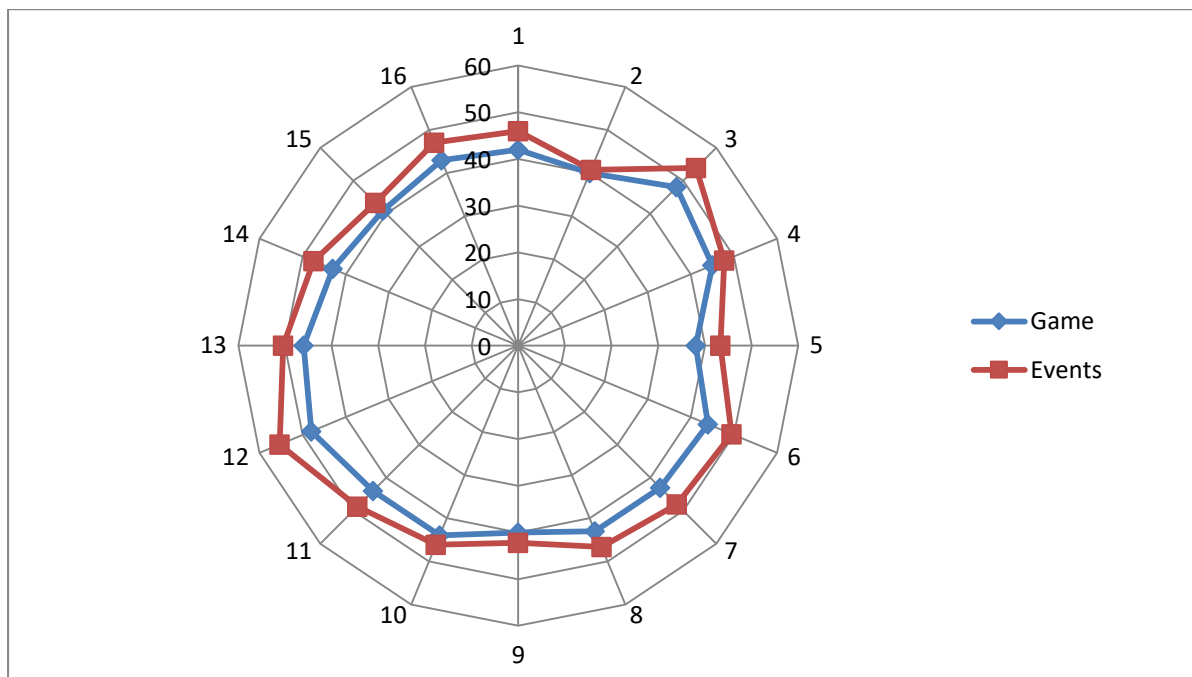
The physiological interpretation of the slow decay is self-explanatory since the sweat congregated by the sweat glands which is responsible for the increased permeability of the skin takes a while to evaporate, proportionally longer than it takes to develop.

Another very important finding was that the system could maintain readings of positive involvement during tasks keeping the candidates in such a state for a while. One could argue that a state of rising STC and HR could not be maintained for long, while a person was continuously in a state of high involvement reaching a saturation state; however, the fact that during such a condition the STC was maintained above the baseline combined with the factor derived by the HRV, even if the increase was minimal it could be projected correctly with no corrective algorithm whatsoever. When the candidate was exceeding the point of saturation where HR was high but began diminishing and STC was not rising any more the system

## Chapter 7: Discussion and Interpretation of Results

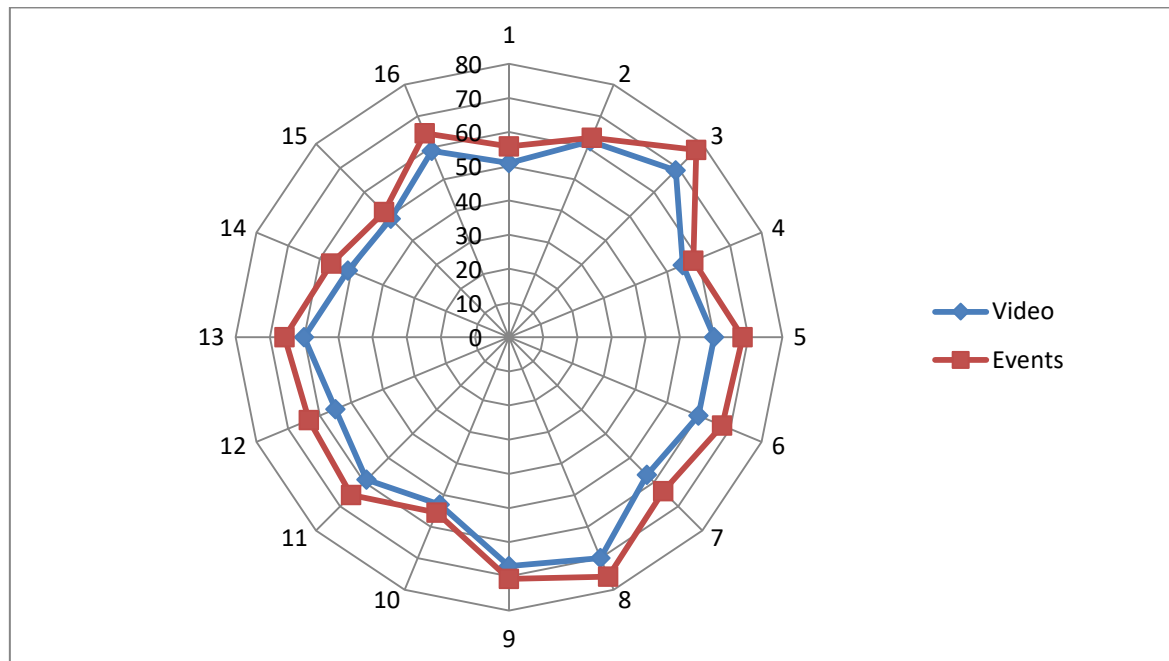
response was less accurate. The above discrepancy could be corrected if a predictive algorithm was incorporated to compensate for sustained conditions.

On the whole, values derived from the GOMS estimation of event allocation plotted against the measured responses in the radar plots is shown in Figure 7.2.2.3 for the game and Figure 7.2.2.4 for the video assessment respectively.



*Figure 7.2.2.3: Estimated vs measured values – game assessment*

## Chapter 7: Discussion and Interpretation of Results

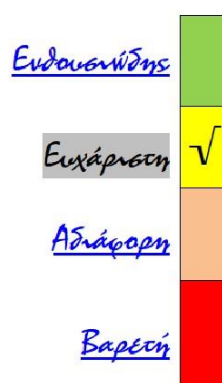


*Figure 7.2.2.4: Estimated vs measured values – video assessment*

Although accuracy maintained a deficit from 0 - 10%, overall consistency was apparent.

Real-time self-assessment was carried out during tests and was also recorded. User responses were rated in four levels expressing the emotional clusters presented in each quadrant of the allocation model, (Figure 7.2.2.5).

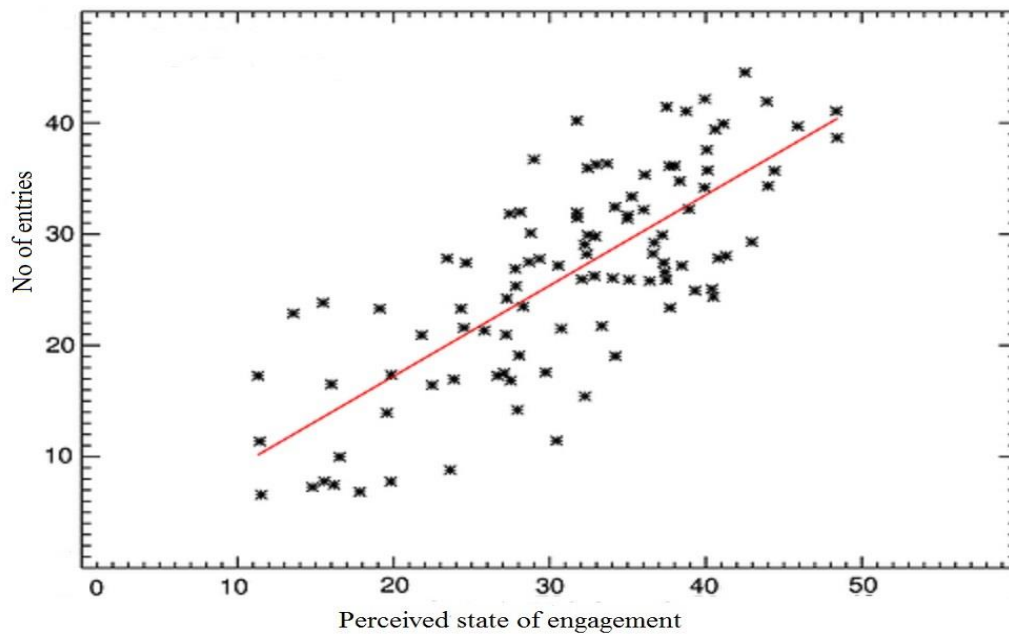
Πως αντιλαμβάνεσθε την  
ψυχολογική σας διάθεση;



*Figure 7.2.2.5: Estimated affective condition as it was perceived by the participant*

## Chapter 7: Discussion and Interpretation of Results

Data produced from the above input during tests differed in frequency to those consistently measured by the system every two seconds because the participants seemed to update the above table infrequently or simply because they felt comfortable with the same condition for long periods of time. A scatter graph representing how close the judgement of the participants was to the real time measurements of the mental workload is presented in Figure 7.2.2.6.



***Figure 7.2.2.6: Scatter plot of affective engagement as perceived by the participants***

The red line in Figure 7.2.2.6 represents the actual instantaneous state of engagement as it was measured by the system. Dispersion was high with distant responses from the centreline as expected because the precision and frequency of perceived mental condition by the participant could not be exact and contiguous as in the subconscious measurements produced by the system. This was primarily because the participants were selecting one of the four states very infrequently and secondary because their perceived state was rather vague comparing to those derived by the actual physiological responses measured by the HR-STC system.

## **Chapter 7: Discussion and Interpretation of Results**

### **7.3 Results: Phase Three - Application oriented evaluation**

Phase three was designed to take on the mission to substantiate the quantity derived by combining the two aforementioned modals. This was done by using pre-validated emotional constructs to aliment scenarios used to evaluate the output of the system. Input data obtained from the International Affective Picture System (IAPS), developed to be used exclusively in scientific research, expressing three norms, (namely, valence, arousal and domination), which have been verified by multiple validation studies, methodologies and measurements, (Bradley & Lang, 2008). Selection of pictures was of particular importance and was optimised for particular clusters of basic emotion, distinctly classified in one of the four quadrants of Russell's model, with significantly high values of both valence and arousal. This was chosen in order to produce the most vivid stimulations possible to the participants. Categorised sentimental quantification of selected IAPS subsets were correlated with the real time output of the system with main objective to produce a sincere justification of the validity of output classification.

#### **7.3.1 Findings: Phase Three - Application oriented evaluation**

In the following analysis wherever Shapiro-Wilk test for normality has shown that the data distribution is normal, parametric tests were performed. Similarly for non-normal distributions non parametric tests were performed accordingly.

For the third quadrant of the video tests ( $Q_{3V}$ ), response accuracy was significantly higher for men compared to women (independent samples t-test:  $t(25) = -2.352$ ,  $p = 0.027$ ), (see Table 7.3.1.1).

## Chapter 7: Discussion and Interpretation of Results

*Table 7.3.1.1: Comparison of response accuracy between male and female participants.*

|                 | MALE  |        | FEMALE |        |
|-----------------|-------|--------|--------|--------|
|                 | mean  | s.d.   | mean   | s.d.   |
| Q <sub>1P</sub> | 38.18 | 4.285  | 34.81  | 4.956  |
| Q <sub>2P</sub> | 20.36 | 2.024  | 16.13  | 1.516  |
| Q <sub>3P</sub> | 22.09 | 4.969  | 18.06  | 5.836  |
| Q <sub>4P</sub> | 12.18 | 4.875  | 9.13   | 3.631  |
| Q <sub>1V</sub> | 67.55 | 10.093 | 66.25  | 15.597 |
| Q <sub>2V</sub> | 24.09 | 10.425 | 19.75  | 7.523  |
| Q <sub>3V</sub> | 34.82 | 18.236 | 21.69  | 10.806 |
| Q <sub>4V</sub> | 18.00 | 12.008 | 11.31  | 9.965  |

Compiled results from the psychometric tests were corresponded to those collected from the real time physiological response acquisition system. The participants' scores on the five factors of personality derived from the NEO-FFI and the categorization of these scores ("very low", "low", "average", "high", and "very high") may be seen on Tables 7.3.1.2 and 7.3.1.3 respectively, separated by gender.

*Table 7.3.1.2: Extraversion vs. Gender*

|        | EXTRAVERSION |     |         |      |           |
|--------|--------------|-----|---------|------|-----------|
|        | Very low     | Low | Average | High | Very high |
| Female | -            | 2   | 7       | 7    | -*        |
| Male   | -            | 1   | 2       | 3    | 5*        |

\*  $|z| > 1.96$  ( $p = 0.05$ )

*Table 7.3.1.3: Conscientiousness vs. Gender*

|        | CONSCIENTIOUSNESS |     |         |      |           |
|--------|-------------------|-----|---------|------|-----------|
|        | Very low          | Low | Average | High | Very high |
| Female | 1*                | 4   | 8       | 3    | -         |
| Male   | 4*                | -   | 4       | 1    | 2         |

\*  $|z| > 1.96$  ( $p = 0.05$ )

## **Chapter 7: Discussion and Interpretation of Results**

With respect to personality factors as numeric values, no gender differences were found to be statistically significant. However, significant associations between personality category and gender have been found in the case of Extraversion and Conscientiousness (Fisher's Exact Test = 8.665,  $p = 0.023$  and FET = 8.435,  $p = 0.047$ , respectively). The associations are observed in the case of categorizing personality across five levels (very low, low, average, high, and very high). However, no significant interactions are observed when using a three-level categorization (below average, average, above average), Table 7.3.1.4.

## Chapter 7: Discussion and Interpretation of Results

*Table 7.3.1.4: Means and standard deviations of personality factors categorised by gender.*

|   | MALE  |       | FEMALE |      | OVERALL |      |
|---|-------|-------|--------|------|---------|------|
|   | mean  | s.d   | mean   | s.d. | mean    | s.d. |
| N | 20.00 | 8.44  | 25.06  | 8.23 | 23.00   | 8.54 |
| E | 35.27 | 8.27  | 30.50  | 4.37 | 32.44   | 6.56 |
| O | 27.73 | 7.24  | 32.13  | 6.73 | 30.33   | 7.15 |
| A | 29.00 | 7.25  | 29.69  | 3.77 | 29.41   | 5.34 |
| C | 31.55 | 10.55 | 33.63  | 4.90 | 32.78   | 7.59 |

The means and standard deviations of emotion regulation and its subscales may be seen on Table 7.3.1.5 (gender-based differences were not statistically significant).

## Chapter 7: Discussion and Interpretation of Results

*Table 7.3.1.5: Frequencies of the participants' personality factor categories separated by gender.*

|                               | MALE | FEMALE | TOTAL |
|-------------------------------|------|--------|-------|
| <b>Neuroticism</b>            |      |        |       |
| Low                           | 2    | 3      | 5     |
| Average                       | 5    | 5      | 10    |
| High                          | 3    | 5      | 8     |
| Very high                     | 1    | 3      | 4     |
| <b>Extraversion</b>           |      |        |       |
| Low                           | 1    | 2      | 3     |
| Average                       | 2    | 7      | 9     |
| High                          | 3    | 7      | 10    |
| Very high                     | 5    | -      | 5     |
| <b>Openness to experience</b> |      |        |       |
| Very low                      | 1    | -      | 1     |
| Low                           | 2    | 1      | 3     |
| Average                       | 4    | 6      | 10    |
| High                          | 4    | 4      | 8     |
| Very high                     | -    | 5      | 5     |
| <b>Agreeableness</b>          |      |        |       |
| Very low                      | 2    | 2      | 4     |
| Low                           | 4    | 8      | 12    |
| Average                       | 2    | 5      | 7     |
| High                          | 3    | 1      | 4     |
| <b>Conscientiousness</b>      |      |        |       |
| Very low                      | 4    | 1      | 5     |
| Low                           | -    | 4      | 4     |
| Average                       | 4    | 8      | 12    |
| High                          | 1    | 3      | 4     |
| Very high                     | 2    | -      | 2     |

Regarding Trait Anxiety, females (mean = 44.19, s.d. = 10.703) scored higher than men (mean = 35.18, s.d. = 9.622). An independent samples t-test indicated that the difference in Trait Anxiety scores across genders was statistically significant ( $t = 2.232$ ,  $p = 0.027$ , 95% BCa CI [1.651, 15.879]). For the entire sample, the mean was 40.52 and the standard deviation was 11.061. As expected, Trait Anxiety correlated relatively highly with Neuroticism ( $r = 0.647$ ,  $p$

## Chapter 7: Discussion and Interpretation of Results

= 0.001) and its two constituent elements, Negative Affect ( $r = 0.548$ ,  $p = 0.003$ ) and Self-Reproach ( $r = 0.590$ ,  $p = 0.001$ ). Positive Affect, a component of Extraversion, is inversely correlated with Trait Anxiety ( $r = -0.405$ ,  $p = 0.036$ ). A similar correlation was observed between the factor of Agreeableness ( $r = -0.423$ ,  $p = 0.028$ ) and its component, Non-Antagonistic Orientation ( $r = -0.392$ ,  $p = 0.043$ ).

**Table 7.3.1.6: Means and standard deviations of Emotion Regulation and its subscales**

|            | MALE |      | FEMALE |      | TOTAL |      |
|------------|------|------|--------|------|-------|------|
|            | mean | s.d. | mean   | s.d. | mean  | s.d. |
| ER         | 3.41 | 0.44 | 3.13   | 0.46 | 3.24  | 0.46 |
| EI         | 3.39 | 0.47 | 3.27   | 0.34 | 3.32  | 0.40 |
| Ere        | 3.91 | 0.56 | 3.98   | 0.48 | 3.95  | 0.50 |
| EMa        | 3.27 | 0.68 | 3.04   | 0.63 | 3.13  | 0.65 |
| EMo        | 3.09 | 0.77 | 2.91   | 0.53 | 2.98  | 0.63 |
| Expe +Expr | 3.19 | 0.60 | 2.81   | 0.58 | 2.96  | 0.60 |
| Expe       | 2.96 | 0.67 | 2.67   | 0.53 | 2.79  | 0.60 |
| Expr       | 3.42 | 0.75 | 2.94   | 0.80 | 3.13  | 0.80 |
| SE         | 3.87 | 0.70 | 3.52   | 0.64 | 3.66  | 0.68 |

**(ER: Emotion Regulation; EI: Emotional Intelligence; Ere: Emotion Recognition; EMa: Emotional Management; EMo: Emotional Motivation; Expe: Emotional Experience; Expr: Emotional Experience; SE: Self-Efficacy).**

Emotion Regulation significantly correlated with several factors and sub-factors of the NEO-FFI model of personality. Specifically:

- Emotion Regulation (as a whole) correlated with Neuroticism ( $r = -0.718$ ,  $p = 0.001$ ) and both of its subscales, i.e. Negative Affect ( $r = -0.518$ ,  $p = 0.006$ ) and Self-Reproach ( $r = -0.724$ ,  $p = 0.001$ ).
- Emotional Intelligence correlated with Neuroticism ( $r = -0.560$ ,  $p = 0.002$ ) and one of its subscales, Self-Reproach ( $r = -0.594$ ,  $p = 0.001$ ). Additionally, it correlated with Conscientiousness ( $r = 0.472$ ,  $p = 0.013$ ) and two of its subscales, namely Orderliness ( $r = 0.400$ ,  $p = 0.039$ ) and Goal-Striving ( $r = 0.501$ ,  $p = 0.008$ ).

## Chapter 7: Discussion and Interpretation of Results

- Emotion Recognition correlated with the “Positive Affect” subscale of Extraversion ( $r = 0.446$ ,  $p = 0.02$ ), the “Intellectual Interests” subscale of Openness to Experience ( $r = 0.478$ ,  $p = 0.012$ ), and the “Dependability” subscale of Conscientiousness ( $r = 0.461$ ,  $p = 0.016$ ).
- Emotional Management correlated with Neuroticism ( $r = -0.495$ ,  $p = 0.009$ ) and both its subscales ( $r_1 = -0.491$ ,  $p = 0.009$ ;  $r_2 = -0.395$ ,  $p = 0.041$ ).
- Emotional Motivation correlated with the “Self-Reproach” subscale of Neuroticism ( $r = -0.457$ ,  $p = 0.016$ ), as well as with Conscientiousness ( $r = 0.57$ ,  $p = 0.002$ ) and all of its subscales, i.e. Orderliness ( $r = 0.481$ ,  $p = 0.011$ ), Goal-Striving ( $r = 0.545$ ,  $p = 0.003$ ), and Dependability ( $r = 0.398$ ,  $p = 0.04$ ).
- Emotional Experience and Expression, taken together as a single construct, correlated with Neuroticism ( $r = -0.565$ ,  $p = 0.002$ ) and both its subscales ( $r_1 = -0.416$ ,  $p_1 = 0.031$ ;  $r_2 = -0.563$ ,  $p_2 = 0.002$ ).
- Emotional Experience by itself correlated with Neuroticism ( $r = -0.513$ ,  $p = 0.006$ ) and its subscale “Self-Reproach” ( $r = -0.584$ ,  $p = 0.001$ ).
- Emotional Expression by itself correlated with Neuroticism ( $r = -0.468$ ,  $p = 0.14$ ) and both its subscales ( $r_1 = -0.415$ ,  $p_1 = 0.031$ ;  $r_2 = -0.412$ ,  $p_2 = 0.033$ ).
- Self-Efficacy correlated with Neuroticism ( $r = -0.789$ ,  $p = 0.001$ ) and both its subscales ( $r_1 = -0.598$ ,  $p_1 = 0.001$ ;  $r_2 = -0.773$ ,  $p_2 = 0.001$ ). It also correlated with Extraversion ( $r = 0.544$ ,  $p = 0.003$ ) and all its subscales, i.e. Positive Affect ( $r = 0.436$ ,  $p = 0.023$ ), Sociability ( $r = 0.439$ ,  $p = 0.022$ ), and Activity ( $r = 0.433$ ,  $p = 0.024$ ).
- Trait Anxiety was found to correlate significantly with Emotion Regulation ( $r = -0.608$ ,  $p = 0.001$ ), Emotional Intelligence ( $r = -0.525$ ,  $p = 0.005$ ), Emotional Experience and Expression ( $r = -0.445$ ,  $p = 0.02$ ), Emotional Experience ( $r = -0.555$ ,  $p = 0.003$ ), and Self-Efficacy ( $r = -0.666$ ,  $p = 0.001$ ).

Low to moderate correlations were found between two personality factors (Neuroticism and Openness to experience) and participant response accuracy for some of the videos. More specifically:

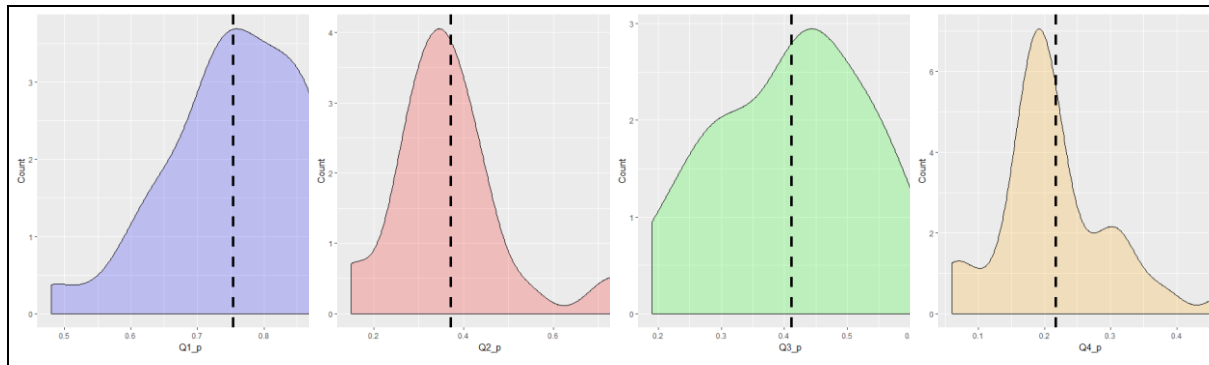
## Chapter 7: Discussion and Interpretation of Results

- Openness to experience was found to correlate with the response to Q<sub>2V</sub> ( $r = 0.384$ ,  $p = 0.048$ ) and Q<sub>4V</sub> ( $r = 0.493$ ,  $p = 0.009$ ).
- Neuroticism was found to correlate with the response to Q<sub>4V</sub> ( $r = 0.426$ ,  $p = 0.027$ ).
- Some factor subscales were found to correlate with participant response accuracy, namely:
  - Self-reproach (a subscale of Neuroticism) correlated with Q<sub>4P</sub> ( $r = -0.383$ ,  $p = 0.024$ ).
  - Aesthetic interests (a subscale of Openness to experience) correlated with Q<sub>1P</sub> ( $r = -0.391$ ,  $p = 0.022$ ) and Q<sub>4V</sub> ( $r = -0.472$ ,  $p = 0.006$ ).
  - Intellectual interests (a subscale of Openness to experience) correlated with Q<sub>1V</sub> ( $r = 0.44$ ,  $p = 0.011$ ), Q<sub>2V</sub> ( $r = -0.338$ ,  $p = 0.043$ ), and Q<sub>4V</sub> ( $r = -0.349$ ,  $p = 0.037$ ).
  - Dependability (a subscale of Conscientiousness) was found to correlate with Q<sub>2P</sub> ( $r = -0.404$ ,  $p = 0.018$ ).

When using a three-level category structure (“below average”, “average”, and “above average”) for the personality factors, Neuroticism was found to influence participant response to Q<sub>3P</sub> ( $F(2,24) = 3.500$ ,  $p = 0.046$ , Tukey HSD: Average – Below Average = 0.159,  $p = 0.037$ ). Openness to Experience was found to affect participant response to Q<sub>4V</sub> (Kruskal-Wallis  $H(2) = 7.246$ ,  $p = 0.023$ , 95% CI[0.020, 0.026]), but pair wise comparisons did not highlight any significant differences.

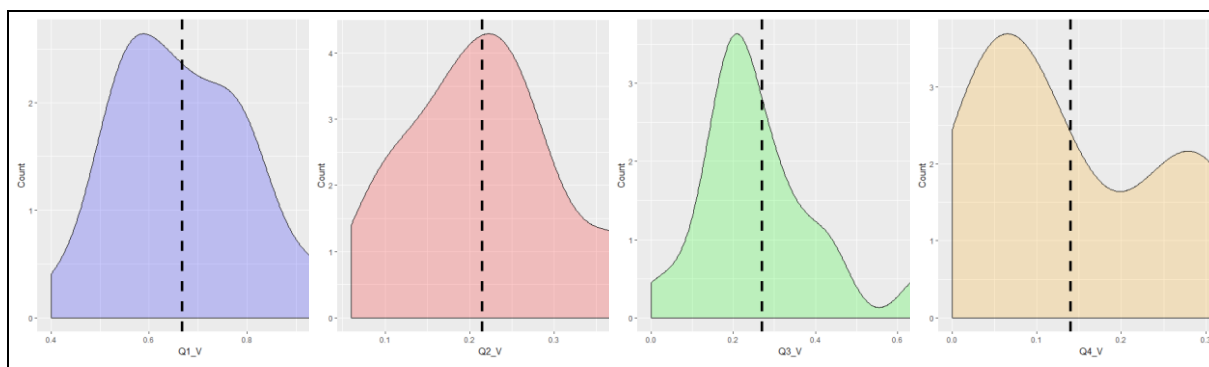
Density functions of measured data for the responses of pictorial stimuli show the distribution of data in relation to the mean value with data in quadrants 2 and 4 appearing more vivid in expression while the remaining two quadrants show a wider spreading, (Figure 7.3.1.1).

## Chapter 7: Discussion and Interpretation of Results



***Figure 7.3.1.1: Density graph of allocated measurements per quadrant - IAPS***

For the video session the density functions expose a different pattern with salient responses only in the third quadrant as shown in Figure 7.3.1.2.



***Figure 7.3.1.2: Density graph of allocated measurements per quadrant - video***

Sharp responses indicate a more prominent transition in lesser space of time than those indicated by the wider distributions suggesting that impressions left in shorter amount of time (i.e. pictures) produce higher values of engagement than those produced by a smoother transition provided by consecutive impressions instigated by the video.

In post processing, correlations between pre-validated data and measured responses exhibit a bias in accuracy towards more positive states of engagement. It is worth noting that although measurements have shown a higher rate of allocation on Russell's model of affect for emotional clusters characterized by more positive valence responses and more active levels of arousal, a

## Chapter 7: Discussion and Interpretation of Results

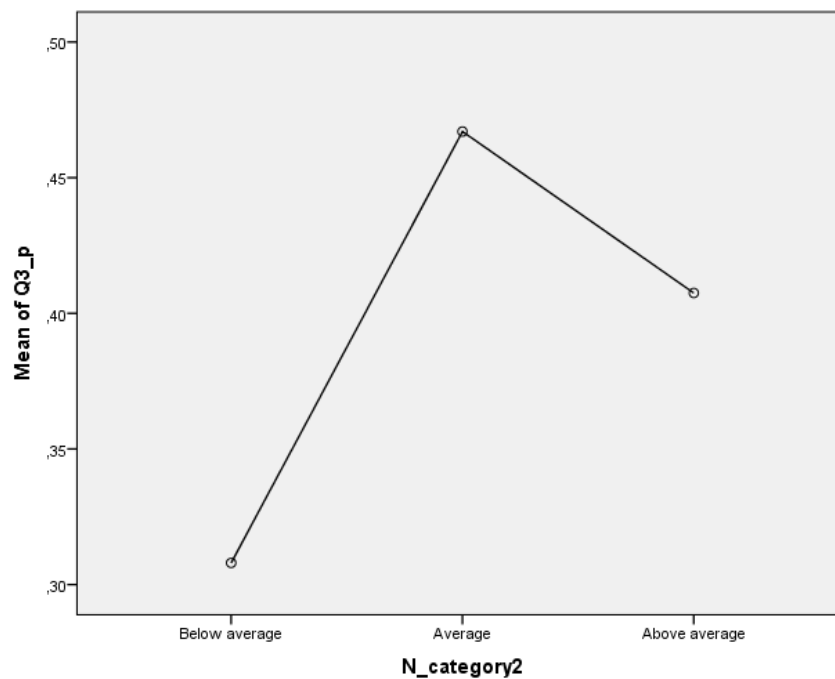
particularly intriguing effect was observed. An unexpectedly higher rate of allocation to the quadrant indicating '*Non-involvement*' was observed, suggesting that emotional constituents were not conclusively expressing cerebral involvement neither that the participant was highly engaged. Hypothesizing that the above case was correct and the system response was accepted as valid, the outcome was then consistent with views that cerebral activity may be high while a person is unimpressed or experiences aversive emotional events (Lewis et al., 2000:742-748). Effectively, comparing results from phases two and three, one can suggest that tasks based purely on emotional constituents seem to be allocated differently from the system compared to cognitive hungry and memory engaging tasks. This finding could be very important, however, its credibility may be hampered by the complexity involved when one attempts to distinguish amongst emotional or affective involvement and that produced by cognitive demand, memory activity or combinations thereof.

Self-assessment data included; statement of estimated demand and effort exerted by the candidate, personality profiles and trait anxiety. Personal estimates of vivacity during the tests meant to identify how the participant perceived the following categories of demand in mental conditions:

- Knowledge Demand: How much previous knowledge was required for the task?
- Mental Demand: How mentally demanding was the task (cognitive or affective)?
- Memory Demand: How much memory dependent or demanding was the task?
- Temporal Demand: How hurried or rushed was the pace of the task?
- Frustration: How insecure, discouraged, irritated, stressed, or annoyed felt during the task?
- Performance: How successful were in accomplishing what they were asked to do?
- Effort: How hard did they have to work to accomplish their level of performance?

Statistical analysis of data produced by SAM's using repeated measures Analysis of Variance (ANOVA) have produced some tendencies as follows. No statistical significance was found between groups except for the third quadrant  $F(2,24) = 3.500$ ,  $p = 0.42$  (Figure 7.3.1.3).

## Chapter 7: Discussion and Interpretation of Results

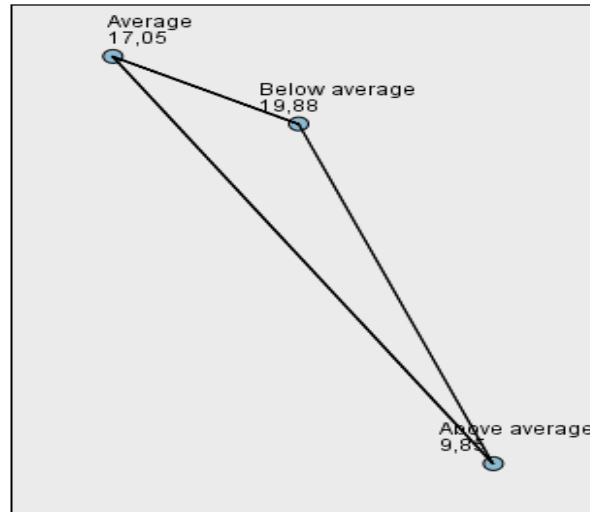


*Figure 7.3.1.3: Mean values of one-way analysis for the third quadrant*

A five-level category structure for the personality factor of Conscientiousness was found to significantly affect participant response to Q<sub>4v</sub>: Welch  $F(4,22) = 7.645$ ,  $p = 0.009$ , Games-Howell: Very high – Low = 0.1875,  $p = 0.025$ ).

## Chapter 7: Discussion and Interpretation of Results

**Pairwise Comparisons of O\_category2**



Each node shows the sample average rank of O\_category2.

| Sample1-Sample2             | Test Statistic | Std. Error | Std. Test Statistic | Sig. | Adj.Sig. |
|-----------------------------|----------------|------------|---------------------|------|----------|
| Above average-Average       | 7,204          | 3,335      | 2,160               | ,031 | ,092     |
| Above average-Below average | 10,029         | 4,533      | 2,213               | ,027 | ,081     |
| Average-Below average       | 2,825          | 4,690      | ,602                | ,547 | 1,000    |

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is ,05.

The participants' Trait Anxiety score only correlated significantly with response to Q<sub>4v</sub> ( $r = -0.460$ ,  $p = 0.016$ ).

Only response accuracy for the fourth image set (Q<sub>4p</sub>) correlated significantly with Emotion Regulation ( $r = 0.396$ ,  $p = 0.041$ ) and two subscales, "Emotional Experience" by itself ( $r = 0.428$ ,  $p = 0.026$ ) and Emotional Experience and Emotional Expression combined ( $r = 0.44$ ,  $p = 0.022$ ). However, a three-level categorization<sup>3</sup> of Emotional Management significantly affected Q<sub>2p</sub> ( $F = 3.913$ ,  $p = 0.034$ ) and Q<sub>4p</sub> ( $F = 3.527$ ,  $p = 0.045$ ). Likewise, a two-level categorization of the same construct significantly affected Q<sub>1v</sub> ( $F = 5.743$ ,  $p = 0.024$ ).

<sup>3</sup> We arrived at this categorization by using the z-score as a cut-off point between categories; the "Average" category contains scores that deviate less than one standard deviation from the mean. All other values are categorized as either above or below average, based on the direction of the deviation.

## **Chapter 7: Discussion and Interpretation of Results**

### **7.3.2 Discussion: Phase Three - Application oriented evaluation**

Intensified or reduced cardiac and stress variability observed in mutually coinciding patterns derived in emotion stimulating environments, was evaluated for use as a tool for detecting and categorizing emotional activation. A real time acquisition system that was previously validated and optimised for detecting focused attention and user engagement (Psaltis & Mourlas 2011) was used for this experimental investigation. No optimization algorithms whatsoever were applied in the above system or modification of the generic measurement responses in order to improve the data projection process as we intended to validate the raw capabilities of the system to allocate emotional data, being aware of the risk that it could fail. Results have shown that the output of the system was biased towards the emotional groups with positive activation and disposition, as expected due to the aims set in the initial design of the system, (i.e. measurements had to have a rectified projection towards positive and active).

The capabilities of the system to detect emotion in absolute values without any software predictive or optimization algorithms was enthusiastically high for the first quadrant (Male:75%, Female:66%) and below 50% for the other three quadrants. That was verifying our views that firstly the emotional aspect was in many cases overlaid by focusing attention, engagement, cognitive and emotional workload while at the same time confirming that the two physiological parameters measured (i.e. STC and HR) were not sufficient in order to devise the constituents required in order to explain the mechanisms of emotion and moreover determine precise emotional elements. A notable conclusion was that effectiveness of the system seemed to have been greater for emotional constructs expressed with higher values of valence and arousal as opposed to those with lower values respectively. As shown in the results outlined in previous sections, identifying emotion was more demanding than simply measuring physiological responses, as reactions to identical stimuli varied among people with differences in emotional regulation and personality profiles.

The discrepancy observed involving the accurate detection of emotion which was more profound in the case of images and video clips associated as a valid response to the fourth quadrant, could be explained as a result of the nature of engagement and focused attention,

## **Chapter 7: Discussion and Interpretation of Results**

which was inclined to be accompanied by an intensification of cognitive processes. As such, the act of focusing attention was capable of influencing the physiological quantities measured by the system, thus effectively masking emotions not consistent with high arousal and valence.

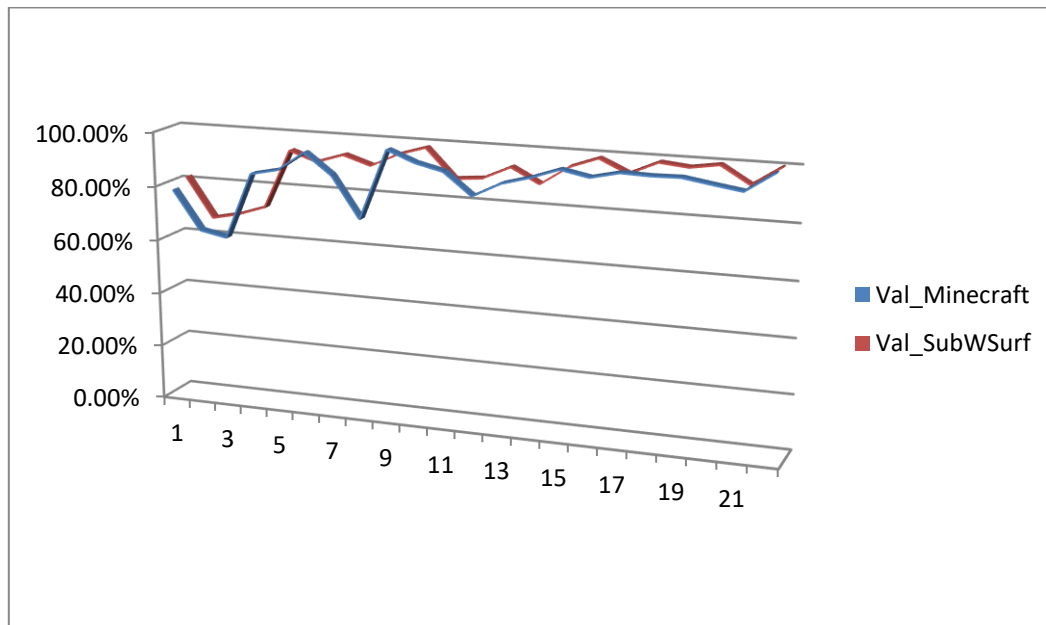
### **7.4 Results: Phase Four - Comparison with commercial products**

Phase four was effectively deliberated to use the aforementioned system of HR-STC along with a credible commercial product and verify two major capabilities of the system as follows:

- the precise and timely response when compared in sync with the output of another system
- the convergence or deviation from data coming from the platform approved for commercial use and
- evaluate common tendencies or indications expressed on both systems.

Analysis of the direct comparison between real-time responses of the STC-HR system in parallel with the simultaneously produced output from Noldus Face-Reader has been presented in the next section. It should be noted that although the two systems have attempted to deduce similar constituents of emotional and affective states they were effectively doing so implementing algorithms from totally different approaches. In essence the HR-STC system was operating via the detection of physiological signals while Noldus operated by evaluating facial expression. Both systems deducing similar quantities employing completely different methods and processes, have shown remarkably close indication of gradients of valence as seen in Figure 7.4.1 below for the 23 candidates.

## Chapter 7: Discussion and Interpretation of Results



*Figure 7.4.1: Matching gradients of valence between HR-STC and Noldus systems*

### 7.4.1 Findings: Phase Four - Comparison with commercial products

Using the Kolmogorov-Smirnov test of normality it was determined that physiological and subjective data were not normally distributed and therefore nonparametric tests were used to examine the difference between the two related samples. On the contrary, using the Kolmogorov-Smirnov test of normality it was determined that the psychological data were normally distributed and therefore parametric tests were used to test the difference between the two related samples.

#### 7.4.1.1 Impact of FF and FS structured digital games on satisfaction and activation

Wilcoxon analysis of the Self-Assessment Manikin (SAM) scores showed that although the average values for satisfaction (valence) and activation (arousal) were higher for Minecraft as compared to Subway Surfers, these differences were not statistically significant.

In addition, a Spearman's correlation coefficient was used to examine the relationship between the variables of player's satisfaction (coded as *Pleasure*) and players' arousal (coded as *Arousal*) for both games.

## Chapter 7: Discussion and Interpretation of Results

The Spearman's correlation coefficient showed that there was a strong positive correlation between the variables of pleasure and arousal in Minecraft ( $r_s=0.769$ ,  $p=0.05$ ), as well as between the variables of pleasure and arousal in Subway Surfers ( $r_s=0.758$ ,  $p=0.05$ ). This finding indicates that increases/decreases in the level of pleasure are correlated with increases/decreases in the level of arousal in both.

Furthermore, analysis of the SAM scores across gender, age, frequency of digital game use and prior gaming experience showed that there was a significant difference only for the variable of gaming experience. Subway Surfers was rated highest by the novice players and Minecraft by the experienced players. This finding strongly suggests that experienced players derive pleasure and excitement from the freeform game play to a higher degree than novice players.

### 7.4.1.2 Impact of FF and FS digital games on stress level

Wilcoxon analysis showed that there was a significant main effect of game for the STC ( $Z = -2.127$ ,  $p = 0.033$ ). Participants who played Subway Surfers had a significant higher level of STC (954.10) than did participants who had played Minecraft (933.94). This pattern was consistent for 15 out of the 22 subjects.

The comparison between the means for two digital games provides a good basis for using physiological measures as an objective indicator of experience with games. Direct comparison of STC raw data have shown that the tonic level was 13-28% higher in the FS games comparing to those of the FF games. Since the number of responses were different for each user as the games produced more randomly distributed stimuli, a comparison between commonly grouped game events was employed. There were five game events (move, loss, win, control, reward) in both games. All of five events, participants experienced a significantly higher STC to move game event in Subway Surfers versus move game event in Minecraft ( $Z=-2.354$ ,  $p=0.019$ ). All other game events comparisons apart of move event were not significant.

## **Chapter 7: Discussion and Interpretation of Results**

In addition, the Kruskal-Wallis test showed that there were no statistically significant differences in STC across gender, age, frequency of digital game use and gaming experience. Based on the above results, hypothesis H1 has to be accepted.

### **7.4.1.3 Impact of FF and FS digital games on HR**

Even though the mean value of heart rate when playing Minecraft was lower (96.76) as compared to that when playing Subway Surfers, (101.40) this difference was not significant ( $Z = -1.09$ ,  $p = 0.277$ ).

Furthermore, HR when examined for individual game events was found to not represent statistically significant differences. Finally, the Kruskal-Wallis test showed that there was no statistically significant difference across gender, age, frequency of digital game use and gaming experience. Therefore, although the statistical findings show results less vital to those expected in H2, it was assumed that HR had no statistically significant value neither on Minecraft nor Subway Surfers.

Although the statistical outcome has shown literally minute influence of FF and FS game play on HR (and vice-versa), from physiological point of view the findings show exactly the opposite, as an average elevation of heart beats per minute causes a significant increase in cardiac output, indicating an increased awareness, mental activity, mobility and eventually after effects such as fatigue and psychological pacing. In this particular assessment, the measurement indicating differences in activation between FF and FS game play was predominantly the derivative of HR trait values, indicating the Heart Rate Variability (HRV) that was also found higher in FS rather than FF games from raw data recordings. HRV has been subject to a debate as it could be affected greatly by physical fitness, age and ambient conditions, however, since during the sessions of the experiments environmental conditions were similar and also age of candidates was pretty close too, the findings from HRV measurements was considered very reliable indicators for the purpose of game play comparison.

### 7.4.1.4 Impact of FF and FS structured digital games on engagement

According to the model for the measurement of users' engagement derived by the HR/STC physiological measurements, coinciding elevations and reductions of STC and HR were treated as indicators, classifying thus the user state into a four quadrant circumplex of engagement as defined below:

- State of Focused Involvement / Engagement (positive arousal and valence), where the user was happily engaged with game play or satisfied for fulfilling the task successfully.
- State of Contentment (negative arousal, positive valence), where the user was finding it difficult or unable to fulfil a task while maintained a high level of activation.
- State of Perceived Difficulty (positive arousal, negative valence), where the subject's focus on a task did not change significantly, while satisfaction diminishes.
- State of Non-involvement / Apathy (negative arousal and valence), where an uninterested and inattentive person performed a task in negative disposition and disengaged arousal levels.

Overall, average of active involvement was significantly higher when playing Minecraft ( $Z=$ ) as compared to playing Subway Surfers (54.5), ( $Z = -1.997$ ,  $p = 0.046$ ). This pattern was consistent for 16 of the 22 subjects. A Kruskal-Wallis test was used to analyse engagement across gender, age, frequency of digital game use and gaming experience. The Kruskal-Wallis test showed that there was a statistically significant difference only at the state of active involvement for Minecraft (chi-square = 12.167,  $p = 0.029$ ). Experienced participants were more engaged players as compared to novice participants. Based on the above results, hypothesis H3 has to be accepted.

## Chapter 7: Discussion and Interpretation of Results

### 7.4.1.5 Impact of freeform and formally structured digital games on positive experience

Paired-samples t-tests were used to compare the impact of Minecraft and Subway Surfers on participants' basic emotions as indicated by their facial expressions. These tests showed that there was a significant main effect of game on participants' happiness ( $t(22)=2.164$  and  $p=0.043$ ). Participants' happiness while playing Minecraft had significantly higher mean strength scores than participants' happiness while playing Subway Surfers (for Minecraft,  $M=41.28$ ,  $SD=23.55$ ; for Subway Surfers,  $M=28.87$ ,  $SD=18.13$ ). This pattern was consistent for 15 of the 22 participants. The global mean average score of facial expressions of all other basic emotions were not different in any of the digital games, as summarized in Table 7.4.1.5.1.

**Table 7.4.1.5.1: A paired-t test analysis of basic emotions**

| Emotion   | Games     |                | T      | df | Sig. (2-tailed) |
|-----------|-----------|----------------|--------|----|-----------------|
|           | Minecraft | Subway Surfers |        |    |                 |
| Neutral   | 63.71     | 62.31          | 0.280  | 20 | 0.783           |
| Happy     | 41.28     | 28.87          | 2.164  | 19 | 0.043           |
| Sad       | 32.98     | 35.72          | -0.482 | 18 | 0.635           |
| Angry     | 54.43     | 31.97          | 0.791  | 18 | 0.439           |
| Surprised | 32.91     | 30.13          | 0.682  | 20 | 0.503           |
| Scared    | 35.27     | 39.22          | -0.356 | 13 | 0.728           |
| Disgusted | 16.78     | 31.51          | -1.801 | 10 | 0.102           |

Furthermore, an analysis of the impact of Minecraft and Subway Surfers on participants' basic emotions across game events showed that there were no significant differences which might have affected the subsequent analysis of the data. Analyzing the five game events it was found that participants' happiness to the move game event while playing Minecraft had significantly higher mean strength scores ( $t(22)=2.062$  and  $p=0.035$ ), than participants' happiness while playing Subway Surfers (Minecraft,  $M=47.44$ ,  $SD=30.15$ ; for Subway Surfers,  $M=27.31$ ,  $SD=24.25$ ). Analysis also showed that participants' emotion of sadness to the reward game event while playing Minecraft had significantly lower mean strength scores ( $t(22)=-3.938$  and

## **Chapter 7: Discussion and Interpretation of Results**

$p=0.008$ ), while playing Subway Surfers (Minecraft,  $M=19.89$ ,  $SD=15.44$ ; for Subway Surfers,  $M=40.67$ ,  $SD=26.20$ ).

Repeated measures using ANOVA across gender, age, frequency of digital game use and prior gaming experience showed that there were no statistically significant differences between these variables. Therefore, hypothesis H4 had to be accepted.

## **Chapter 7: Discussion and Interpretation of Results**

### **7.5 Discussion: Phase Four - Comparison with commercial products**

Essential objectives of phase four concerned the quantitative and qualitative assessment of freedom of choice in adventure digital games based on the structural characteristics of games and the perceptions of students, to examine whether free-form and formally structured digital games elicit different kind of game play experiences, the validation of the model for the above assessment, and a cross validation of two systems based on different principles of operation for data acquisition.

Regarding the first objective of the study the findings could validate the effectiveness of FoC model for the quantitative and qualitative assessment of freedom of choice in adventure digital games; Minecraft, which was considered as FF digital game, elicited different kind of game play experiences than the Subway Surfers, which was considered as FS digital game. Specifically, results showed that participants felt less stressed, happier and more engaged when playing freeform digital game, and found this kind of games more pleasant and exciting than the formally structured digital game.

Concerning the specific objective of the study, playing Minecraft was, as expected, more pleasant and exciting for the participants (H1). It was hypothesized that both STC and HR would increase with the more stressful game (namely Subway Surfers). However, this was only true for the skin conductance (H2). Contrary to our expectations we could not find significant differences on HR values (H3). Furthermore, it was found that the state of engagement was higher when playing Minecraft as compared to playing Subway Surfers (H4). Finally, the results showed that while playing the Minecraft participants showed significantly higher values of happiness than they did when playing the Subway Surfers (H5).

The increase of stress level indicated by STC measurements while participants played the more tense game was in agreement with the results reported by Drachen, Nacke, Yannakakis, & Pedersen, (2010) and Cusveller, Gerritsen, & Man, (2014). Drachen found that skin conductance correlates with negative affect, namely frustration while Cusveller noted that skin conductance would increase with the more hectic, hence more stressful games. The lack of significant differences on HR values is in contrary to our expectations and to previous findings

## Chapter 7: Discussion and Interpretation of Results

(Drachen, Nacke, Yannakakis, & Pedersen, 2010; Cusveller, Gerritsen, & Man, 2014). There was no simple explanation for this result, but this finding is in line with Ravaja et al. (2005) who supports that HR alone may not be an optimal measure of arousal as carries information on both sympathetic and parasympathetic activity, which may entail interpretative difficulties. Increased cardiac sympathetic activity is related to emotional arousal and causes the heart to speed up, whereas increased cardiac parasympathetic activity is related to information intake and attention or engagement and causes the heart to slow down (Turpin, 1986). The higher level of participants' engagement while playing Minecraft was in agreement with the results reported by Kirginas & Gouscos (2016) who found that FF digital games lead students towards, higher engagement and more positive game play experience. Finally, the significantly higher values of happiness while playing the Minecraft was consistent with those expressed by Kirginas & Gouscos (2016) who found that FF digital games lead students towards more positive emotions as well as Wiklund et al. (2015) who found that strong emotions, whether negative or positive in valence, have been shown to influence motivation. This in turn, may lead to rich learning performance, as well as a high degree of game enjoyment. This notion is further supported by Moreno (2006) argued that learning with multimedia is affected by emotional and motivational factors. This finding is significant as positive emotions, such as happiness may result in higher player involvement.

Combining all these findings together it is purported that there is a direct correlation between the quality of players' experience and the kind of play that a digital game offers, depending on whether the latter focuses on the game dimension (structured games) or the play dimension (FF games) of game play. We would like to note at this point that the concept of freedom of choice substantially concerns serious games, in the sense that these games have an inherent learning purpose. Given that psycho-physiological measurements and subjective self-reported experience could be used as objective indicators for the assessment of learning experience it can be concluded that players feel less stressed when playing freeform digital game, they feel happier and more engaged, and find this kind of games more pleasant and exciting.

## Chapter 7: Discussion and Interpretation of Results

Stemmed from the results it was stipulated that a key factor for the design of effective digital games for learning seems to be the development of games that:

- a) focus on the play dimension of game play, allowing for free, unstructured activities without pre-defined rules and pre-designed goals;
- b) create non-linear game play;
- c) allow multiple solutions to the game challenges and enable many different player-generated game play paths;
- d) give players the ability to choose the order in which they face game challenges; and
- e) allow players to choose which challenges they want to overcome or not,
- f) give players the ability have control over the game character,
- g) allow players to follow their pace, and (h) have no win or lose states.

Furthermore, it was believed that the concept of freedom of choice substantially concerns educational digital games. If educational games uptake characteristics of free-form play they may be better suited to lead to more engaging and more FF learning processes, which may in turn be able to better support innovative approaches to learning and, more generally, to the acquisition of 21st century transferable skills.

Although our current results were promising, it should be cited that the research conducted thus far has faced a number of limitations. Firstly, there are other variables that we could not account for and could possibly affect the correlations found. For example, game reviews, predispositions, expectations, needs, motivation, mood, personal interests, and social setting were expected to have an impact on what kind of games players prefer and play. Secondly, given the time intensive nature of our study, our sample size was rather small. Although we attempted for variation in age, gender, gaming experience and gaming frequencies, we cannot claim representativeness for the general gaming population. Finally, there is a lack of an established theory of gameplay experiences that could provide a frame for deeper analysis and relatively limited experimental research studies about game play experience in digital games to which these findings could compare.

## **Chapter 7: Discussion and Interpretation of Results**

As far as the cross validation study between Noldus Facereader 4.0 and the HR-STC data acquisition system goes, the two systems performed properly executed together in the same computer producing synchronized appraisal that proved to be most valuable in assessing data produced. A common variable deduced by both systems separately (namely Valence) was found to be closely the same, although the first system derived this DV from emotional facial recognition while the second from pattern recognition of physiological responses. Studies of raw data recordings have shown extensively consistencies between responses and empirical interpretations of expressions or physiological reactions.

## **Chapter 7: Discussion and Interpretation of Results**

### 8. Conclusions

Recognising the need to improve responsive adaptation in learning and personalised environments, it was thought that measuring user's physiological expressions would provide the additional components required for real time identification of subconscious psychosomatic expressions. For this purpose, an optimised integrated bio-sensing system has been developed, that can deduce seamlessly the heart rate and physiological stress reactions of a learner. Effective implementation and adept usability concepts have been employed, successfully eliminating side effects and errors introduced by distraction imposed on the user during the measurements. Based on a long standing hypothesis that coinciding intensification of HR and STC may express some relation to an emotional state, it was decided to assess the possibilities to detect accumulated emotional, affective and cognitive involvement which have been proven to be indicative of underlying psychosomatic condition such as focused attention and engagement.

A major debate concerning the underlying constituents expressed through the psychophysiological signals has been investigated using latest technologies and advanced operational concepts. Meticulously examined experimentation methodologies and mathematical modelling techniques have been employed towards the accomplishment of this study, fulfilling the major obstacle of the final implementation, that of seamless operation of the system.

Results from especially designed experiments have shown that simultaneous excitations of STC and HR correlate to underlying levels of emotional, cognitive or affective excitation in a cumulative fashion. On the other hand, precision obtained was not sufficient to deduce exact emotional or affective constituents. As expected, this was due to the fact that the excitation state detected and interpreted by the system as a state of engagement could represent one of many variant combinations of emotional states, producing similar physiological responses and measurements that coincide with the interpretation given by the converging gradient algorithm.

Effectively, conclusive findings show that precision of the HR-STC system varied according to activation levels. System measurements verify the variable intensity of emotional

## Chapter 8: Conclusions

expressions and the need for further investigation regarding vigour, onset time and duration of major affective and emotional components, not available to date. In such scientific fields of research, the HR-STC system could prove to be a valuable tool for assessing specific subconscious conditions, in conjunction with pre-validated scenarios.

The HR-STC system has demonstrated its abilities to detect cumulative mental activation with distinct non linearity, showing higher accuracy when the two converging physiological modalities express high activation levels, diminishing as the opposite occurs. Taken into account that no particular optimisation algorithm has been incorporated entailing predictive or pattern analysis, it can be claimed that the system presents only a portion of its capabilities. According to initial specification, design prerequisites included detection of subconscious physiological quantities serving as classifiers of manifested states of engagement. Detection of precise affective or cognitive constituents or distinction of combinations of the above components has not been attempted or claimed possible by using the HR-STC system. At this stage of development, the intensity of activation and hedonic state constituents can be considered sufficient for use in research and most applications regarding learning and performance assessment.

The HR-STC system has been revealed as rudimentary and rather imprecise for detailed mapping of particularly low activation emotional states such as frustration, sadness, depression etc. In order to improve on that part, further development may incorporate baseline deduction and also intelligent algorithms assessing trends of values over a longer time interval than looking only at instantaneous measurements like the system in its present version. For instance, averaging of measurements looking at the last three or five measurements together have been expected to improve the accuracy particularly in minimally fluctuating measurement conditions.

Further research that would evaluate affective and emotional onset would also provide a significant component for improving system performance. Despite that imprecision in psychosomatic states with low intensity, it was thought that the system could be used efficiently

as a component of intelligent interfaces for detecting the user's high degrees of involvement in various applications (e.g. AmI, UC, distance learning etc.). It is worth noting that the capability to obtain such information was not available to date before the development of the HR-STC system.

Achieving the essential objective of minimally obstructive and effective acquisition of two vital physiological quantities for determining states of engagement of a user, the HR-STC system presents a tool for development of innovative applications in human-centric attention monitoring and interactive technologies. The effectiveness of the algorithm exploring the direction of excitation of the two basic physiological quantities has been proven meaningful, supporting the initial hypothesis and also verifying previous findings. Further research employing the HR-STC system in experiments related to different models of psychosomatic response allocation may prove more accurate thus enhancing its capabilities. The system development has met the target to obtain the physiological signals by simply maintaining contact with two relatively small areas instead of special electrode probes. This latter feature provides the flexibility to incorporate the system and sensing elements literally on any handheld device that may be applicable to making use of the psychosomatic responses as an enhancement for adaptive, responsive or simply monitoring systems. Consequently, several application areas beyond education and learning can utilise instantaneous human responses for interactive environments such as gaming or attention assessment.

As explored in the following chapters, a range of possible improvements in conjunction with future research may lead to further enhancements in both the accuracy of the system as well as precision in the identification of more detailed psychosomatic conditions. Also improved systems performance indeed can be achieved by adding functional components to the system as explained further in this text. Applications of the HR-STC system in new areas of research and evaluation studies of different psycho-physiological models may reveal new findings and concepts as new technologies and innovative ideas emerge steadily.

## Chapter 8: Conclusions

### 9. Applications

The HR-STC system not seen only as a development tool for this research but rather as a concept that can add the capabilities to deduce specific attributes of the inner self of a human in a notoriously unobtrusive fashion holds a revolutionary potential in many areas of application. Existing devices deducing HR have found new grounds in fitness monitoring applications providing information on cardiovascular effort exerted and useful derivative information like for instance fitness coefficients derived from the speed that HR takes to return to normal levels. Devices measuring psychological stress have found applications in approved methods for deducing truthfulness of responses or in stress therapy and relaxation.

Combined measurements attempting to derive psycho-physiological markers in such a simplified manner so that they can be incorporated in any handheld device present a momentous option seen for the first time. As a starting point, the system has shown that simultaneous elevation of the above two physiological markers reveal psychosomatic excitation in consistency with previous studies. The value of an index providing levels of excitation in real time can be essential for monitoring as well as designing novel responsive applications as illustrated further below. Improvements on system precision and efficiency may reveal additional capabilities opening doors to even more advanced application areas such as predictive assessment or traits helping to automatically deducing personal profiles.

The implementation of the HR-STC system by using analogue electronic components - as opposed to embedded processor implementation - provides a great flexibility for miniaturisation and most importantly eliminates program preload requirements that are substantially higher when looking at volume production.

The major drawback for scientists when they attempted to use bio-feedback devices for psychosomatic assessment in applications targeting natural environments is the need for initial calibration. All known devices set that prerequisite in order to determine the range of values

## **Chapter 9: Applications**

that can be given by each person individually so they can scale their measurements. Instant measurement interpretation without the above prerequisite presents a unique feature of the HR-STC system, removing that long standing obstacle. This important feature is opening doors to wide fields and areas of application ranging from education and learning to gaming and interactive communication.

Users do not require prior knowledge of the system or specific training before they use the HR-STC system. Even at frequently interrupted operation, responses of the system remain equally dependable and valid.

A battery operated version of the HR-STC system could provide the flexibility required to be easily fitted as an integral part of handheld devices following the tendencies of the marketplace for the present as well as the near future.

### **9.1 School personalised learning**

Being inconspicuous, discreetly hidden in peripheral computer devices such as the computer mouse, the HR-STC system has the upper hand over any other existing device for effective application in class like environments. Teachers could have displayed in their screen information like:

- Colour coded levels of attention for each student in the class
- Information regarding user activity on a percentage scaled vertical bar next to each user screen
- Timers for inactivity of each computer station indicating possible need for assistance
- Progress analysis data associated to completion targets set by the tutor leading to automated scoring information instantaneously as well as at the end of the session
- The teacher could obtain information regarding completion of tasks from each individual student, so he can reward them, creating thus new instruments for motivation.

In more advanced teaching applications, adaptive scenarios can increase or reduce levels of difficulty either automatically depending on intelligent software algorithms processing psychosomatic responses of the user or by intervention from the educator.

### **9.2 Cloud based educational personalisation**

Usefulness of the HR-STC system can be even more prominent in applications where visual or other form of communication between tutor and learner is absent as is the case in cloud educational environments. Learners can enter the educational platform at any time and carry out instructive tasks with no specific supervision by their tutor.

A cloud application utilising the HR-STC system could obtain useful traits of psychosomatic expressions of the learner regarding information related to perceived difficulty derived from levels of attention and completion time. Similarly, information about user data in conjunction with the content may help in evaluation and classification of content in ways not possible before the HR-STC system. The abovementioned data may produce content analytics and new correlates for effectiveness, efficiency, preferences, timing information, levels of difficulty and more.

In dynamically adaptive cloud environments, process automation enhanced by the added features provided by the HR-STC system entailing psychosomatic responses, attendance data and preferred format of content could gain enormously in interface optimisation (HCI) and efficiency of content development. Further statistical analysis of usage attributes and data mining methodologies may lead to new functional components related to psychosomatic responses and their applications.

## **Chapter 9: Applications**

### **9.3 Gaming interaction**

Applications incorporating psycho-physiological assessment in gaming could find a wide field for development. Primarily, multiplayer configurations could obtain indexes that may be used in gaming in various ways such as varying speed or options provided to the player. Also the environment of some games may become adaptable according to responses obtained. From the part of the player, the HR-STC system could represent a separate entity promoting or demoting their gameplay accordingly. For instance, in a gaming scenario where levels of attention of the player have been higher than average, a number of additional game options may be presented making the gameplay more interesting according to criteria set by developers. In the opposite case where players have shown reduced interest and focused involvement, the game could adaptively present easier challenges.

In competition games, the added features made available by the HR-STC system would provide added functionality to developers to incorporate say colour or elevation characteristics in their game or new ideas entailing responsive environments.

### **9.4 e-learning adaptive content**

Internet based education incorporating e-learning applications constitute a part of the emerging new technologies. As an educational environment, e-learning retains most characteristics of the classroom environment. In contrast with cloud or distant learning approaches, e-learning entails real-time communication between the tutor and the student. An HR-STC system applied in such technologies could provide additional information to the tutor similar to those of student attendance in the classroom as described above regarding educational applications.

The value of physiological responses in e-learning environments can be considered even more valuable than in the class, since in the former educational schemes there is a distance between tutors and students. Information about involvement and indications of personal motivation of the student can be considered substantially important, not possible before the development of the HR-STC system or with any other existing methods even those employing semantic analysis and AI.

### 9.5 Communication and social media

The exponential growth of media sharing technologies has introduced many new styles in communication between known and unknown recipients. A rapid increase in membership counts have already placed social media providers as the highest traffic producers of the Internet generating substantial amount of communication and media content. Affective information has been a concern in this recent and still rapidly developing means of communication. It should be noted that the Open Systems Interconnection (OSI) pyramid defining the Internet and network communication model has provision for affective interfaces providing specification of relevant protocols.

Attempts to incorporate AI and semantic analysis or other instruments in social media environments have been made and being in progress in order to achieve communication of affective and emotional conditions between members, (Grassi, Cambria, Hussain, Piazza, 2011). With further research investigating appropriate affective models the HR-STC system could provide optimized markers correlating physiological responses to affective conditions.

## Chapter 9: Applications

### 10. Limitations and Future Developments

Efferent physiological reactions of computer users were literally expanded in this study outside the human body, establishing a form of communication between instantaneous mental activation and a typical computer. A study in this direction required:

- a profound understanding of the biological activity employed while psychological, affective, cognitive or emotional operations take place in the human brain,
- a thorough acquaintance with the electrical properties of the above as they are being expressed by the nervous system,
- a proficiency of an approved method to distinguish HR and STC from other similar biological signals
- a knowledge of electronic design and development of a device capable to acquire the bio-electrical signals derived from the evoked expressions in an accurate and dependable fashion,
- the ability to translate those generic quantities into a cognizant form of data, prove their potentials and interpret those values according to their significance.

Individual component design was employed entailing sensing interface, timing, signal conditioning electronics and a suite of software components for configuration, display and recording of data, providing active responses to any platform that would take advantage of psychosomatic indicators of a computer user. Thus experimental work in this particular study was accountable not only to evaluate data and their respective inferences and statistical correlations but also to prove beyond doubt the operational validity of the system developed from the first principle of operation. The primary hypothesis of the study as it was analysed in the preceding chapters was implemented partly in hardware, where the HR signal was convoluted filtered and conditioned, and partly in software where algorithms translated the behavioural interpretation of the combined biofeedback acquisition of both HR and STC quantities.

## **Chapter 10: Limitations and Future Developments**

The major constraint of this study was to produce a system that would work seamlessly with no obstructions or restrictions of any kind from the side of the user. Uses of assorted medical devices in experimentation studies before this development have imposed biasing factors on the participants. The second limitation was that even if a system capable to identify the aforementioned quantities was developed, a proper area of the human body had to be found that would give a decent signal for the HR and STC acquisition. The next hurdle was that even if all the above were made possible, artefacts and movement related noise and interferences had to be counteracted.

A difficult task also was to verify that the quantities measured by the system were what it was aiming for and formulate those quantities in an unambiguous format in order to produce psycho-physiological parameters in a convincing manner. For this purpose the experimental procedure was effectively organized in four discreet phases fulfilling specific accomplishments in an ordered fashion as follows. In the first experimental setting, system validation and functional approval was endeavoured, fulfilling the requirements successfully as analysed in stage one of this chapter. The essential objective was to prove the proper operation of the system and detection of the actual bioelectrical signals in a timely and dependable manner. The experiment included laboratory techniques, electronic workbench testing and human control group experiments verifying the validity, correctness and responsiveness of the system using quantitative and empirical methods of assessment. The system characteristics satisfied the operational requirements for the real time detection and process of psychosomatic condition of a user. In stage two, an experimental procedure was designed so that the actual capabilities of the system to detect and present psychosomatic activity could be demonstrated in real time. Signals obtained from the system were combined and treated according to their physiological respective behavioural interpretation, subsequently classified into meaningful sets according to their associated affective significance. Classified results were displayed as they occurred and following a visual assessment it was found that they actually represented the expected responses with performance accuracy better than anticipated initially.

## Chapter 10: Limitations and Future Developments

Assessment was based on an unconfined predictive interpretation of empirically estimated emotional stimulations evaluated according to their corresponding constituent values of valence and arousal as they were judged for each evoked psychosomatic condition respectively. Results were better than anticipated initially, indicating that the assessment process conceivably was more lenient than it should by design, acceptable presumption for an initial validation as the accuracy of the system was to be assessed again in later stages.

Effectively, a limitation was presented in the process of assessment of the results in phase two. The derivation of states of engagement had to be moderated by subjective evaluation based on visual observation and judgmental interpretation of the instantaneous display as ruled by GOMS. For example, when a participant was playing a game involving shooting at coloured balls flowing through a pipeline, aiming to destroy clusters of three or more of them before they reach the end of the pipe, it was assumed that the player was highly involved and tentatively focused or stressfully loaded when the balls were closing dangerously to the end of the pipe as the game was going to be lost. This process, although sensible, incorporated a possible imprecision as affective combinations could vary in strength between people. Obviously there was no instrument to facilitate the element of strength of the emotion or affect at that instant and the quantity derived by the system could not be verified quantitatively with high precision. Testing of the system was therefore reduced to the precision of the coordinate values allocating the emotional or activity construct to a distinct quadrant in the projected allocation model.

The above limitation introduced by the difficulties in estimating potency of emotional construct as test data could produce slightly misleading results in the allocation model for obvious reasons; however, in worst case scenario where data with questionably deduced allocation were excluded, accuracy was reduced to average 81.6%. Also it was found that although the accuracy of detected responses was remarkably high, a bias towards positive values existed, resulting in an overall shift of displayed values towards the positively active engagement. This was expected because emotional stimulation was physiologically expressed by more vivid bioelectrical quantities in the affective grouping of active engagement and

## **Chapter 10: Limitations and Future Developments**

gradually less active reactions as the emotional, affective or other psychosomatic condition was approaching negatively characterized affective groups such as boredom and apathy.

It was derived from the quantified results of this test that the two constituents of the affective model referred to as valence and arousal, were expressed with both more vivid and higher absolute values when pertained to positive engagement characteristics and less intense when expressing negative affect and impassive levels of interest. Accuracy improvements and possible correction algorithms were considered in the corresponding analysis of the stage one conclusion and recommendations, however, none of those suggestions were applied in the next phases of the experimental procedure, as for the purpose of system verification and data validation, the primal system capabilities had to be identified.

In phase three, system dependability and accuracy were tested against pre-validated affect evoking scenarios. Visual content with quantified levels of predisposition (valence) and activation (arousal) by means of a greater variety of biofeedback, biometric, vocal and combined scientific methods, provided a basis for comparison between widely accepted affect causing values and those produced by the system. Results have shown that dependable and convincingly accurate readings were produced by the system accomplishing the aims of the experiment successfully.

In phase four, experimentation in natural user setting has proved the capabilities of the system to perform in a non-laboratory like environment while at the same time system performance was comparing closely with results obtained from another commercially approved product such as the Noldus Facial Recognition suite.

In summary, the chapter of experimental methodology laid out the process followed to investigate whether the integrated hardware and software system developed specifically for biofeedback measurements of the two aforesaid major physiological activities could correlate to identifiable psychosomatic conditions in real world applications. Multiple avenues of

## Chapter 10: Limitations and Future Developments

engaging audio-visual content were explored and utilized in conjunction with most established methodologies, indicating effectively that an inconceivably plain presumption that simultaneous increase of HR and STC are indicative to psychological or affective condition with higher precision as the requirements for higher mental engagement intensify.

Encapsulating conclusions from all four phases of the experimental processes, the system has been proven successful in acquiring and processing the biofeedback markers in a dependable fashion and seamlessly in a typical computer user environment with no particular restrictions whatsoever. Accuracy of the system was not proven sufficient in all four segments of allocation.

With the system in its primary operational form with raw data as they were formed by bioelectrical responses and without correction algorithms employed, the system performance was considered very decent. Bearing in mind that the system developed was generic and there was no such a system commercially available for an analogy and given that there is plenty of room for improvement and optimization, the system was considered to be a successful development introducing innovative concepts.

Precision of system measurements was considered satisfactory after a series of bench and application tests carried out but the overall system allocation was degraded by a number of issues as explained below:

- duration of expression of emotion has not been given an estimate through scientific investigation
- analysis of emotional constructs is vague lacking precise coordinate values of valence and arousal, for example anger can be represented by say Valence=-3, Arousal=-5 but also by Valence = -4, Arousal=-4 and boredom also by Valence=-4, Arousal=-4.

## Chapter 10: Limitations and Future Developments

- precision provided by the two quantities of valence and arousal in deducing psychosomatic condition proved insufficient. Additional classifiers expressing say intensity would be required.
- there is no previous evaluation of distinct physiological expressions of cerebral processes to clearly define affect, cognitive, emotional or memory activity and distinguish them from other cerebral activities.
- research in combining HR and STC responses employing the model of gradients in order to identify mechanisms to deduce psychosomatic condition through emotional and affective composition has not been found to date. In this regard, to verify accuracy, information can only be investigated from first principles.

Additional components incorporated into the system such as gaze analysis and pupil size estimation could improve validity and usability of the system dramatically in many more areas of application as described in the relevant chapter of this text.

Right from the beginning of the concept analysis of this study, a system was planned with a leg on each side of two equally essential objectives. First objective was certainly the perception and development of a novel approach to develop an integrated device with established capabilities for application in educational environments. The second objective was to provide a high quality of measurement and produce a tool not existing before that would reveal physiological responses as identifiable values corresponding to specific mental states of engagement and effectively levels of attention. It is clinically proven that physiological measurements of STC and HR can provide valid information on the actual psychophysiological state of a human and they can project an accurate indication of their psychosomatic condition. In clinical environments and in order to achieve a more precise evaluation, the participant has to be connected physically with gel electrode and binding sensors and be subject to kinetic and postural restrictions during the measurement.

## Chapter 10: Limitations and Future Developments

The first major limitation to overcome was effectively to devise an approach that would eliminate the need for gel electrodes and straps causing the above limitations. Even though experimentation using physiological data acquisition and bio-sensing has been the core engine in many scientific fields to date, the fundamental concept of this study establishes a solid foundation of seamlessness in application. This alone presents a major leap against any previous claims for impervious experimentation as this is the first system available that has removed the psychological impact caused by feeling that pervasive measurements are actually taking place. The system described above employs a user friendly approach, eliminating requirements for kinetic and other restrictions caused by attached electrodes so that it can be applied to any typical computer user who simply uses a computer mouse; similarly applicable to mobile or other devices. Clearly this setup based on optical HR detection and trans-conductance instrumentation allows for a very realistic psychosomatic state detection as the subject is entirely free not only from sensors and attachments but more significantly from the overwhelming pressure caused by the sense of being under some invasive observation.

Previous attempts to achieve meaningful results by utilising the computer mouse have been abandoned either because the quality of measurement was computer platforms (gaming and such) allowed other auxiliaries to be used (helmets, garments, glasses etc), or because the application requirements were such that the user did not have to use continuously the mouse during a particular session. Adding to that rejection of the mouse based physiological measurement systems are undoubtedly the operational challenges facing the realisation of a robust system, that would be unaffected by unreliable reading and artefacts produced by undesirable movements in the point of contact of the user with the sensing elements.

The system portrayed in this literature has been designed for the assessment of the psychosomatic state of a computer user, mainly in educational environments, with only prerequisite that the user is using a computer mouse or a specifically adapted case of a mobile device as the input peripheral device and in specifically designed learning or assessment educational environments (Lekkas, Tsianos, Germanakos, & Mourlas, 2007, 2008; Schunk,

## **Chapter 10: Limitations and Future Developments**

1989). In those environments, the composition, appearance and completion time requirements during the assessment are actively changing according to subconscious responses of the user.

A major advantage of the HR-STC system comparing to other systems used for User Behavioural and Psychosomatic Detection in use to date is that data is being processed in real time and utilised in real time, as opposed to others systems usually producing recorded data for post processing. Conducting further tests with numerous subjects in the areas of educational assessment for validation, verification of existing hypotheses and optimisation of the significance of the combined parameters derived from measurements, this system is candidate to become the first system that can produce in real time a display of the Affective Computing Circumplex for research organisations and actively adapted HCI platforms. This characteristic of the system could find application in broader areas of behavioural or psychosomatically affected adaptive platforms beyond those dedicated to education. Also such a system can be applicable for longer term psychosomatic profiling, providing a user classification index prompting the user to understand when they reach or exceed a threshold of stress and discomfort.

### 10.1 Improvements

Beginning with the assumption that physiological measurements of HR and STC portray a true relation between the psychological reactions and realistic experiences, results from our experiments have proven that the system could be used as an objective indicator for the assessment of learning experience. Confirmation comes also from the findings of other studies supporting validity of the above (Drachen, Nacke, Yannakakis, Pedersen, 2010; Cusveller, Gerritsen & Man, 2014; Nacke, Grimshaw & Lindley, 2010). Notwithstanding the fact that the HR-STC system has proven competent effectively providing reliable measurements, projections of objective psycho-physiological conditions derived through measurements have shown increased accuracy in conditions where learners expressed higher states of alertness, diminishing as they become less involved. Subsequently, areas of improvement have to be channelled towards the allocation algorithms and additional processing modules aiming to increase the precision of the system.

An improvement on the accuracy of detection of the correct quadrant and consequently the correct cluster presenting the focused attention could be achieved by optimizing the system in the following aspects:

- A baseline value derived from the auto-calibration algorithm of the system needs to be deducted from the measured value thus subtracting the offset value that has proved to bias the allocation to the quadrants towards the more positive/active representation. This would improve the accuracy of allocation dramatically.
- The present algorithm for allocating the combined measurement values to the corresponding quadrant was not representing the no-change values. Representing those values as “retained last valid representation” would give an improvement of 18-23% in all four quadrants.
- A predictive model algorithm adapted to the profile of the participant taking into consideration behavioural, personal and emotional tendencies and trait anxiety could be incorporated and this was expected to have improved system accuracy even further.

## Chapter 10: Limitations and Future Developments

- Extracting more experimental data and studying multivariate correlations from larger samples of participants could lead to new algorithms using fuzzy logic or neural network architectures that would use the two measured quantities more intelligently for further improvement in accuracy.
- Respiration rate could be deduced from the heart rate with advanced signal processing and that would have been a very significant quantity contributing to higher accuracy of detection of emotion and mental effort.
- Finally recognising the volatility of instant measurements that may be prone to large fluctuations of readings, a significant step expected to improve dramatically the precision and even add predictive capabilities to the HR-STC system would be achieved if intelligent algorithms were incorporated to the existing algorithm. For example, curve fitting applied to trend data in time intervals of say one minute, would help to produce a more stable projection of psychosomatic states particularly useful in applications that require longer term information rather than instant measurements such as cloud or distant learning monitoring applications.

Further research in educational environments may address issues such as:

- a. selecting and/or designing applications specifically for curriculum learning of preschool and primary school children that may exhibit optimised structures, could achieve enhanced students' learning performance, and
- b. studying the learning outcomes of preschool and primary school children on adaptive learning scenarios, with a view to exploring and maximising the effectiveness of free-form and formally structured applications for curriculum programs and examination material.

Finally the forthcoming stages of our research include the development of a pupil size variability detection system that will be incorporated into our system and the development of specifically designed interfaces for interactive applications in mobile telephony and pioneering educational technologies.

### 10.2 Future work

Following the endorsement received during the process of validation of the system when compared the system responses with those derived from commercially approved emotional recognition software (i.e. Noldus facial expression analysis), diverse avenues considered to be explored. Some thoughts were to incorporate other commercial products in future implementations provided that the extensions did not violate the requirement for restrictions posed by additional obstructive wearable sensing elements. Optimisation of commercial products with the potential to be integrated with the HR-STC system has led to promising choices as investigated in the following pages. Essentially, in order to maintain system flexibility and minimise the use of additional sensors, commercial products extracting information from visual assessment were explored such as those based on small cameras and additional software processing modules. The above optimisation has revealed a number of products that could be integrated easily inside the HR-STC system as explored in the following text.

The HR-STC system may be applicable to fields of application beyond education and learning, where the human psychosomatic state needs to be detected and responded to accordingly such as airport security control, interactive games, aeroplane pilot warning systems, clinical monitoring of longer term data etc. Given the fact that the aforementioned system could be developed in small sizes and be adapted appropriately to portable devices, wider application areas would be in mobile communication devices, AmI systems, e-learning and educational assessment, virtual environments, and further areas requiring remote detection of a user's psychosomatic condition. With further study on improving precision of the descriptors, the system would present an innovative and at the same time discreet solution in areas of application such as security systems assessing severe psychosomatic condition of people like terrorists in airports and other important check points.

The reason that similar systems have not been utilised in the above areas of application was that EDR devices developed in the past have not been able to provide a continuous reading as

## **Chapter 10: Limitations and Future Developments**

they did not employ Trans-Conductance but rather inferior methods of measurement based on voltage measurements. As a result the human skin could compensate for the electron build-up of the measuring devices in the same way it compensates for static electricity, producing thus a decaying attenuation compromising the EDR signal after a short while. Since the HR-TSC system can produce continuous and uncompromised measurements for long time intervals presents new opportunities for monitoring applications in the abovementioned areas and possibly fresh ideas spawn by emerging technologies.

### **10.2.1 Pupil size detection**

The autonomic nervous system in humans is one of the two components comprising the peripheral nervous system (the other being the somatic), that is responsible for keeping the homeostatic balance (i.e. optimal physiological status of equilibrium) in the human body (Tortora & Grabowski, 2006). When a balanced state (homeostasis) experiences a disturbance caused by external factors like a sudden impact or a very loud noise, the parasympathetic and sympathetic systems react in tandem in order to compensate for maintaining the initial homeostatic balance. Correcting actions induced by the autonomic nervous system take place in various parts of the body and in various ways (e.g. neurotransmitters, hormonal secretion etc.) One important instantaneous reaction particularly sensitive to emotional responses initiated by the autonomic nervous system is the dilation of the pupil of the eye. It is known to be dilated by the sympathetic system when a stimulus causes emotional instigation and constricted (or returning to normal size) by the parasympathetic system when the effect of the stimulus diminishes (Gray's, 2008). A small camera based pupil size detection component could be incorporated in the system; expected to endorse the responses of the system with respect to emotional reactions revealed by the other two physiological quantities already utilised.

In future versions of the system, the user's pupil size could be assessed also in real time and simultaneously with the other two physiological measurements. This improvement was considered greatly valuable since could be implemented unobtrusively, hence maintaining the initial concept of the design and fundamental objective of the system. Requirements for the

## **Chapter 10: Limitations and Future Developments**

implementation of the Pupil Size Detection (PSD) and the integration with the HR-STC system necessitate top-end computer processing power (that is available today) and software modules for face detection and stabilisation, eye feature extraction and pupil size deduction and assessment respectively. The only additional prerequisite to the existing contact of the fingers to the sensing elements is that the user needs to be in viewing range of a camera. It was expected that a relative size variation in relation to the iris would definitely complement system accuracy and provide additional information on focus attention and emotional loading.

### **10.2.2 Virtual and augmented reality**

The usefulness of the biofeedback acquisition system portrayed in this development would present a significant benefit if used in virtual reality systems, where the concepts of user immersion and presence were of primary importance. In brief, presence could be seen as a multidimensional construct encompassing various factors that pertain to the system, the user, or both, and potentially even other users. Although numerous definitions of presence have been articulated (e.g. Jennett, et al., 2008, Kuniavsky, 2010, Lombard, Ditton, 2013, Lykken & Venables, 1971, Strongman, 2003, Witmer, Singer, 1998, Witmer, Jerome, Singer, 2005), the common thread among them is the users' impression of being present not in their physical surroundings, but in an alternate, computer-generated world. Immersion was often regarded as a vital component of presence, nowadays increasingly studied in the context of computer games (e.g. Damasio, 2006) – the latter being the primary manifestation of commercially available and widely accessible virtual environments.

The strong point of the HR-STC system over self-report methods is the elimination of the significant problem of disrupting the flow of interactive communication in order to assess various aspects of the users' psychological state, resulting in a greater decrease in the overall perceived quality of the interaction. Therefore, a way of objectively ascertaining user state can result on the one hand in a more truthful and accurate measurement of user psychosomatic states and on the other hand in an improved adaptation of the virtual environment to the users' characteristics. The utilization of the HR-STC system described herein as a tool for measuring the users' engagement and evaluation of various aspects of virtual environments was planned

## **Chapter 10: Limitations and Future Developments**

for the near future. More specifically, a series of virtual environments were being designed with the aim of being used in order to investigate the impact of various environmental parameters on user experience. Bio-sensing would be used as a measurement instrument that would allow for the effective adaptation of environmental stimuli to the users' aims and characteristics.

### **10.2.3 Mobile devices and applications**

Another great advantage of the system developed for the purposes of this research was its versatility when supplied by auxiliary power so that it could nurse portable applications. Provision was made in the design of the electronics for low voltage and low power operation, so the system could operate on batteries and therefore could communicate with Bluetooth-Low-Energy (BLE) devices opening doors to developments for IoT (Internet of Things) applications. The value of the system as an aid in user performance and experience evaluation in mobile contexts of use could be very promising. Sensing components could quite easily be fitted onto a case of a mobile device, while miniaturization of electronic parts and minute power requirements make the infrastructure easily within reach of development. A small size BLE version of the HR-STC system would provide instantaneous readings to responsive game platforms and apps as well as long term data records for user trends and statistical performance analysis. Comparable wearable devices measuring HR and muscular activity have found application on fitness and calorie expenditure estimation. Indication of psychosomatic condition could be a valuable quantity in addition.

The context of mobile usage of the system (in both structured and unstructured activities) was feasibility and operationally tested. Mobile applications differ from ordinary desktop applications in requirements, largely due to the fact that the use of mobile devices occurs “on the go”, which elevates the importance of how engaged users are with the task at hand or where they focus their attention (in this particular case, excessive or minimal engagement may have to be prevented in the interest of safety). Communication may be enhanced if incorporated aspects of psychosomatic state revealing affective conditions between the remote speakers.

## **Chapter 10: Limitations and Future Developments**

In education, as the usage of mobile devices has been increasing, an interface that for first time would deduce psychosomatic user response could provide an additional dimension used in varying difficulty levels according to rules influenced by the above responses. Effectively, the HR-STC system could conversely be used in time varying context environments such as in educational assessment, user performance, context valuation and statistical analysis.

An immensely interesting area of application was also that of the next generation interactive gaming. There, if psycho-physiological condition was respected as a major contributor to making progress through gaming scenarios, that could create a new exciting multiuser competing game experience. For example, participants from the Internet or other network connections could compete in a racing game with their vehicle making progress as their attention level remains high. Also in other scenarios the psychosomatic condition could be facilitated seeking a balance between contentment and discomfort or any other pair of basic emotional constituents producing a tool for users to develop new tactics for emotional regulation.

### **10.2.4 Gaze analysis**

Development of a gaze analysis software module or the incorporation of an existing commercial product could be integrated in the system providing reinforcement of information on valuable parameters of specific usability characteristics. Focusing duration of specific display areas was up to now assessed without additional classifiers deducing say intensity or vigour. Educational scenarios may gain additional information in evaluating pleasing aspects of presentation layout improving usability and communication with the learners.

### **10.2.5 Cloud computing**

The notion of cloud computing presented numerous advantages to computer or mobile device users by unlocking their dependence to local processing and storage limitations. Communication has been improved as users can use, store, share or promote information of literally unlimited size and at the same time avoid congestion and storage bottleneck problems

## **Chapter 10: Limitations and Future Developments**

in their local computer or mobile devices. Widely available applications have been made readily available to users which no longer require local installation. The impact of cloud computing has not yet been fully estimated as new application areas continuously step forward. Latest versions of software can be incorporated as soon as they have been released by the companies and the users do not need to renew their versions of software. Users of cloud environments can access their data from any place or device safely and with higher security than that of a local system that may be impaired by hardware failure.

Sensing of engagement or derivative classifiers as the HR-STC could provide if made available to cloud engines could augment responsive personalisation and adaptive capabilities. Cloud engines may then be able to evaluate and improve content, assess preferential attitudes of the users and levels of satisfaction as they occur.

Seen from a different angle, development of cloud based applications can respond to a tutor or manager in a bidirectional fashion presenting psychosomatic information regarding specific users.

### **11. Contribution to Science and Technology**

The HR-STC system development has produced a number of possible improvements and in diverse disciplines as this research study stretches into different domains of scientific fields. Primarily the aforementioned system, as an instrument, provided the means to measure reliably two important physiological quantities achieving significant improvements over existing devices and also introduced some particularly valuable concepts. With electronics and software development as its core discipline, HR-STC has shown an integrated implementation able to obtain the physiological quantities of Heat Rate and Skin Trans-Conductance with dependability over long term measurement. Invaluable concepts were also the synchronisation and homogeneity of data measurements that have proven to be an improvement comparing to similar real time systems composed by separate laboratory devices.

Very important novelties introduced that made this undertaking possible were based on original ideas that may be used in future scientific implementations. More specifically the implementation based on latest technologies of Trans-Conductance instead of Resistive measurements constitutes an improvement over previous measurement techniques. The compact design and elimination of electrodes and attachments has improved system's usability enabling the use of the system in a variety of mobile and desktop applications. Instantaneous measurements and post processing that produced conclusive interpretations of each valid measurement, eliminated problems of discontinuity that would otherwise affect adversely the output of the system in many applications.

A major obstacle for research affiliates and scientists when considered using physiological stress measurements in studies and applications of user activation and engagement was the need for initial calibration of the device used for each specific user. Vector facilitation of measurements used in the HR-STC system has been another crucial concept as it can be applied in similar cases where initial calibration and familiarisation phase needs to be avoided. In this conception, vector information of measurements has been used rather than the actual quantities measured which may vary significantly between users. Results can thus be conclusively accurate while at the same time maintain an independence of the actual quantitative values

## **Chapter 11: Contribution to Science and Technology**

measured from each individual. As a result, the system produces accurate results even for different users during the same session of measurements.

Flexibility in adaptation of the HR-STC system either as an accessory or as an embedded component that may be incorporated into any applicable device that needs to acquire HR and STC signals exhibits another important asset.

Provisions for low cost design and miniaturisation have also been important, extending the possibilities for volume production of the HR-STC system for use in widespread educational and gaming environments.

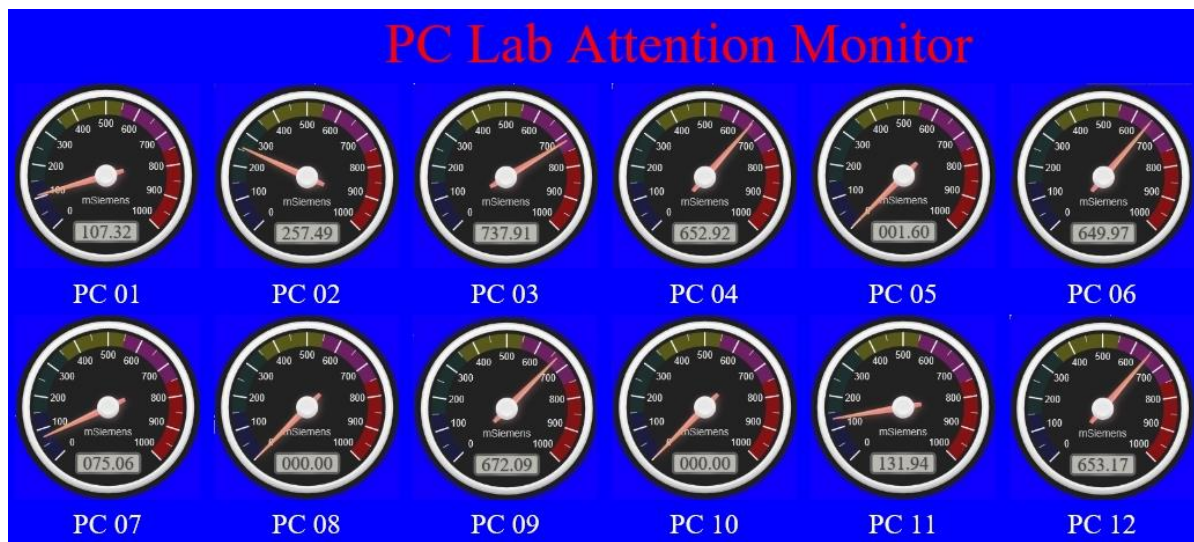
Beyond the fields of electronics and user interface usability concepts the impact of the effectiveness of the HR-STC system has obvious improvements in the scientific fields of psychophysiology as it constitutes a generic instrument that can be used for verification and study of various models of psychosomatic expressions. For instance, derivative classifiers produced by the system may be used either standalone or in conjunction with additional inputs such as gaze or iris dilation detection systems for study and verification of various interpretations in psychology, education and other fields.

Results from the present study of the converging excitation model have produced firstly a justification that previous suggestions of this algorithm have genuine foundations and credible expression of mental activation and secondly that accuracy of the algorithm is diminishing as the activation levels decline. This finding constitutes a scientific conclusion useful for many relevant scientific fields such as psychology where physiological expressions of emotional and affective constituents are studied extensively in search for classifiers pertaining to attentiveness and focusing awareness.

The possibilities presented by the HR-STC system to operate during any form of communication (e.g. in mobile devices, internet and cloud configurations) provide a valuable asset for future application development in education, gaming and responsive environments. Examples can be ranging from purely learning applications, to gaming and user satisfaction response logging for the film industries.

## Chapter 11: Contribution to Science and Technology

Specifically for applications in education, the HR-STC provides a practical tool that can produce a realistic detection of scaled projection of student levels of engagement. Even at this stage that detection accuracy has been unconvincingly high in high activation levels and even more questionable at lower levels respectively, the HR-STC represents a pioneer implementation that may find applications in monitoring where a margin of error is acceptable. As an example, an application that can approximate levels of attention in a school lab or in distant learning environment could prove very useful even with low accuracy (Figure 11.1).



*Figure 11.1 Example application- Attention monitor*

In educational settings where the contact between the tutor and the student is absent such as distant learning, knowing that a student is simply attending the task is a type of information not up to now available.

## References





### A

Adams, T.K. (2001). *Future warfare and the decline of human decision-making Parameters* U.S. Army War College Quarterly (Winter 2001–2002) 57–71.

Adler R.B. & Rodman G. (2006). *Understanding Human Communication* (3rd edition). Volume 10. New York and Oxford: Oxford University Press.

Ark W.S., Dryer D.C. & Lu D.J. (1999). *The Emotion Mouse*, in: Proceedings of HCI International (the 8<sup>th</sup> International Conference on Human-Computer Interaction) on Human-Computer Interaction: Ergonomics and User Interfaces – Volume I, 818-823.

Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., & Christopherson, R. (2009, July). Emotion Sensors Go To School. In *AIED*, Volume 200, 17-24.

### B

Backs, R., & Walrath, L. (1992). Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. *Applied Ergonomics*, 23(4), 243-254.

## References

Backs, R., & Boucsein W. (Eds.). (2000). *Engineering Psychophysiology: Issues and applications*. Mahwah, NJ: Erlbaum.

Bard, P., & Macht, M. B. (1958). The behaviour of chronically decerebrate cats. In *Ciba Foundation Symposium-Neurological Basis of Behaviour* 55-75. John Wiley & Sons, Ltd.

Beatty, J. (1982). *Task-evoked pupillary responses, processing load, and the structure of processing resources*. *Psychological Bulletin*, 91, 276-292.

Bechara, A., Damasio, H., & Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral cortex*, 10(3), 295-307.

Bechara, A. (2004). The role of emotion in decision-making: evidence from neurological patients with orbitofrontal damage. *Brain and cognition*, 55(1), 30-40.

Benetti, C., & Kambouropoulos, N. (2006). Affect-regulated indirect effects of trait anxiety and trait resilience on self-esteem. *Personality and individual differences*, 41(2), 341-352.

Benjamin, A. (2014). *Factors Influencing Learning*. Internal Psychology, University of Illinois. R. Biswas-Diener & E. Diener (Eds), Noba Textbook Series: Psychology

## References

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T. & Craven, P. L. (2007). *EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks*. *Aviation, space, and environmental medicine*, 78(5), B231-B244.

Bernat, E., Bunce, S., & Shevrin, H. (2001). *Event-related brain potentials differentiate positive and negative mood adjectives during both supraliminal and subliminal visual processing*. *International Journal of Psychophysiology*, 42(1), 11-34.

Boucsein, W. (2012). *Electrodermal activity*. Springer Science & Business Media.

Bradley M.M. and Lang P.J. (1994). Measuring emotion: *the Self Assessment Manikin and the Semantic Differential*, *Journal of Behavior Therapy and Experimental Psychiatry*, 25, 49-59.

Bradley M.M. & Lang P.J. (2007). Emotion and Motivation. In Cacioppo, J.T., Tassinary, L.G., Berntson, G. (eds.), *Handbook of Psychophysiology* (3rd edition), Cambridge: Cambridge University Press, 581-607.

Bradley M.M. and Lang P.J. (2007). The International Picture Affective System (IAPS) in the study of emotion and attention, in: *Handbook of Emotion Elicitation and Assessment*, J.A. Coan and J.J.B. Allen (eds.), Oxford University Press, Oxford and New York, 29-46.

## References

Brame, C. J., & Biel, R. (2015). Test-enhanced learning: *the potential for testing to promote greater learning in undergraduate science courses*. CBE-Life Sciences Education, 14(2), es4.

Brave, S., & Nass, C. (2003). Emotion in human–computer interaction. *Human-Computer Interaction*, 53.

Brave S. & Nass C. (2008). *Emotion in Human-Computer Interaction*, in: The Human-Computer Interaction Handbook: *Fundamentals, Evolving Technologies, and Emerging Applications* (2nd edition), A. Sears and J.A. Jacko (eds), Lawrence Erlbaum Associates, New York and London, 77-92.

Brouwer, A. M., Hogervorst, M. A., Van Erp, J. B., Heffelaar, T., Zimmerman, P. H., & Oostenveld, R. (2012). *Estimating workload using EEG spectral power and ERPs in the n-back task*. Journal of neural engineering, 9(4), 045008.

## C

Cacioppo, J. T., & Tassinary, L. G. (1990). *Principles of psychophysiology: Physical, social, and inferential elements*. Cambridge University Press.

Cacioppo, J., Tassinary, L., Berntson, G. (2007). "25". *Handbook of Psychophysiology* (3rd ed.). Cambridge University Press. (581–607).

## References

- Caillois, R (2001) *Man, Play, Game*, [1958]. Chicago, IL: University of Illinois Press. (Original work published 1958).
- Cain, B. (2007). *A review of the mental workload literature*. Defence Research And Development Toronto (Canada).
- Calleja G. (2011). *In-game: From immersion to incorporation*, MIT Press, Cambridge, MA.
- Cannon, W. B. (1927) The James-Lange theory of emotion: *A critical examination and an alternative theory*. American Journal of Psychology, 39, 10-124.
- Chaouachi M. Chalfoun P. Jraidi I. and Frasson C. (2010). “*Affect and Mental Engagement – Towards Adaptability for Intelligent Systems*”, 3-6.
- Comerchero, M. D., & Polich, J. (1999). *P3a and P3b from typical auditory and visual stimuli*. Clinical Neurophysiology, 110(1), 24-30.
- Costa P.T. & McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO-PI-R Five-Factor Inventory (NEO-PI-R-FFI) Professional Manual*. Odessa, FL: Psychological Assessment Resources.

## References

Craik, F. I. M., & Jennings, J. M. (1992). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition*. 51-110. Hillsdale, NJ: Lawrence Erlbaum Associates.

1.1.1.1 Cusveller, J., Gerritsen, C. & Man, J. de (2014). *Evoking and Measuring Arousal in Game Settings*. In S. Gobel & J. Wiemeyer (Eds.), *Games for Training, Education, Health and Sports*, 165-174. Springer International Publishing.

## D

Damasio A.R. (1994/2006). *Descartes' Error: Emotion, Reason, and the Human Brain*, Vintage, London.

Damasio, H. (1995). *Human brain anatomy in computerized images*. Oxford university press.

Darrow, C.W. (1937). Neural Mechanisms Controlling the Palmar Galvanic Skin Reflex and Palmar Sweating: A Consideration of Available Literature. *Archives of Neurology & Psychiatry*, 37(3), 641-663.

Davou, B. (2002). Unconscious processes influencing learning. *Psychodynamic Practice*, 8(3), 277-294.

## References

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual review of psychology*, 41(1), 417-440.

Drachen, A., Nacke, L.E., Yannakakis, G., Lee Pedersen, A. (2010). *Correlation between heart rate, electrodermal activity and player experience in First-Person Shooter games*. In Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games, Los Angeles, CA. 49-54.

Dimitriadou D., & Stalikas A. (2012). (In Greek) Ερωτηματολόγιο Πέντε Παραγόντων Προσωπικότητας NEO, Τύπος: Αυτοαναφορά (NEO Five- Factor Inventory, NEO-FFI, Form S). Σε Σταλίκας, Α., Τριλίβα, Σ., Ρούσση, Π. (επιμ.), *Τα Ψυχομετρικά Εργαλεία στην Ελλάδα* (2η έκδοση), Αθήνα: Εκδ. Πεδίο.

## E

Ebling, F., Eady, R., & Leigh, I. (1992). *Anatomy and organization of human skin, Textbook of Dermatology*, RH Champion, JL Burrington, FJG Ebling, eds.

Ekman P. (1999). Basic Emotions, in: *Handbook of Cognition and Emotion*, T. Dalgleish and M. Power, eds., John Wiley and Sons, Sussex, UK, 45-60.

Eysenck M. (1982). *Attention and Arousal*, Springer-Verlag, New York.

## References

Eysenck M.W. (2006). *Fundamentals of Cognition*, Psychology Press, Hove and New York.

Eysenck, M. W., & Keane, M. T. (2005). *Cognitive Psychology: A Student's Handbook*. 115-128.

## F

Fairclough S.H. (2009). Fundamentals of Physiological Computing. *International Journal of Human-Computer Interaction* 21, 133-145.

Fairclough, S. H., Moores, L. J., Ewing, K. C., & Roberts, J. (2009, September). *Measuring task engagement as an input to physiological computing*. Proceedings of 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops 1-9. IEEE.

Fontaine, J. R., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). *The world of emotions is not two-dimensional*. Psychological science, 18(12), 1050-1057.

Fowles, D. C. (2007). 10 The Measurement of Electrodermal Activity in Children. *Developmental psychophysiology: Theory, systems, and methods*, 286.

## References

Frasca, G. (2003) "Simulation Versus Narrative: Introduction to Ludology", Retrieved January 23, 2016 from: [http://www.ludology.org/articles/VGT\\_final.pdf](http://www.ludology.org/articles/VGT_final.pdf).

Frijda N.H. (1986). *The emotions*, Cambridge University Press, Cambridge.

## G

Gaillard, A., & Wientjes, C. (1994). Mental load and work stress as two types of energy mobilization. *Work and Stress*, 8, 141-152.

Glynn, L., Christenfeld, N., Gerin, W. (2002). "The Role of Rumination in Recovery From Reactivity; Cardiovascular Consequences of Emotional States". *Psychosomatic Medicine*. 64 (5), 714–726.

Gonsalvez, C. J., & Polich, J. (2002). *P300 amplitude is determined by target-to-target interval*. *Psychophysiology*, 39(3), 388-396.

Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance*, Vol. 2. *Cognitive processes and performance* 1-49. Oxford, England: John Wiley.

## References

Gray F.R.S. (1977). *Gray's Anatomy: The Unabridged Running Press Edition of the American Classic*.

Gray H., Carter H. (2008). *Gray's Anatomy*.

Grassi, M., Cambria, E., Hussain, A., & Piazza, F. (2011). Sentic web: A new paradigm for managing social media affective information. *Cognitive Computation*, 3(3), 480-489.

Greenberg, L. M., Kindschi, C. L., Dupuy, T. R., & Hughes, S. J. (1994). Test of Variables of Attention continuous performance test. *Los Alamitos, CA: Universal Attention Disorders*.

Gross J.J. and Levenson, R.W. (1995). Emotion elicitation using films. *Cognition & emotion*, 9 (1), 87–108.

Grossberg, S. (1987). A psychophysiological theory of reinforcement, drive, motivation, and attention. In *Advances in psychology*, (42), 3-81.

Groth-Marnat, G. (2009). *Handbook of psychological assessment*. John Wiley & Sons.

Guyton, A. (1977). *Basic human physiology: Normal function and mechanisms of disease*. Philadelphia: Saunders.

### H

Haapalainen, E., Kim, S., Forlizzi, J. F., & Dey, A. K. (2010, September). Psycho-physiological measures for assessing cognitive load. In *Proceedings of the 12th ACM international conference on Ubiquitous computing* 301-310. ACM.

Hancock P.A. & Caird J.K. (1993). “*Experimental Evaluation of a Model of Mental Workload*”. Human Factors 35(3), 413-430.

Harting, K., & Erthal, M. J. (2005). History of distance learning. *Information technology, learning, and performance journal*, 23(1), 35.

Hassett J. (1978), *A primer of psychophysiology*, W. H. Freeman, San Francisco.

Helander, M. (1988). Towards a practical GOMS model methodology for user interface design. In *Handbook of human computer interaction* 135-158. North-Holland.

Hess, E. & Polt, J. (1964). Pupil size in relation to mental activity during simple problem solving. *Science*, 143(3611), 1190-1192.

## References

Hockey, G. (1986). A state control theory of adaptation to stress and individual differences in stress management. In G. R. J. Hockey, A. W. K. Gaillard, & M. G. H. Coles (Eds.), *Energies and human information processing* 285-298.

Hodgkin, A. L., & Huxley, A. F. (1939). *Action potentials recorded from inside a nerve fibre*. Nature, 144 (3651), 710-711.

Horowitz, P., & Hill, W. (1989). *The art of electronics*. Cambridge Univ. Press.

Hunt, E., & Lansman, M. (1975). *Cognitive theory applied to individual differences*. Handbook of learning and cognitive processes, (1), 81-110.

## I

Insko B.E. (2003). *Measuring Presence: Subjective, Behavioral and Physiological Methods*, in: *Being There: Concepts, Effects and Measurement of User Presence in Synthetic Environments*, G. Riva, F. Davide, and W.A. IJsselsteijn (eds.), IOS Press, Amsterdam, the Netherlands, 109-119.

### J

Jainta S. Baccino T. (2010). *“Analysing pupillary response due to increased Pupillary demand”*, Elsevier.

James, W. (1884). What is emotion?, *Mind*, 9(34), 188-205.

James, W. (1890). The consciousness of self. *The principles of psychology*, 8.

Jennett C., Cox A.L., Cairns P., Dhoparee S., Epps A., Tijts T., and Walton A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies* 66(9), 641-661.

### K

Kim, U., Yang, K. S., & Hwang, K. K. (2006). Contributions to indigenous and cultural psychology. *Indigenous and Cultural Psychology*, 3-25.

Kirginas, S., & Gouscos, D. (2016). Exploring the Impact of Free-Form and Structured Digital Games on the Player Experience of Kindergarten and Primary School Students. In *Handbook of Research on Gaming Trends in P-12 Education* 394-420. IGI Global.

## References

Kirginas, S., & Gouscos, D. (2015a). A Study for the Quantitative Assessment of Freedom of Choice in Adventure Games for Learning. Manuscript submitted for publication.

Kirginas, S., & Gouscos, D. (2015b). Development and Validation of a Questionnaire to Measure Perceptions of Freedom of Choice in Digital Games. Manuscript submitted for publication.

Kivikangas, J. M., Ekman, I., Chanel, G., Järvelä, S., Cowley, B., Salminen, M., Henttonen, P., and Ravaja, N. (2010). Review on psychophysiological methods in game research. Proceedings of the 1st Nordic DiGRA Confernece, Stockholm, Sweden.

Koedinger, K. R., Alevan, V., Heffernan, N., McLaren, B., & Hockenberry, M. (2004, August). *Opening the door to non-programmers: Authoring intelligent tutor behaviour by demonstration*. In *Intelligent tutoring systems* Vol. 3220, 7-10.

Kort, B., & Reilly, R. (2002). An affective module for an intelligent tutoring system. In *Intelligent Tutoring Systems* 955-962. Springer Berlin/Heidelberg.

Kramer, A. (1991). Physiological metrics of mental workload: A review of recent progress. In D. Damos (Ed.), *Multiple task performance* 279-328. London:

## References

Kuniavsky M. (2010). *Smart Things: Ubiquitous computing user experience design*. Morgan Kaufmann, Burlington, MA.

## L

Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American psychologist*, 50(5), 372

Lazarus, R. S. (1982). Thoughts on the relations between emotion and cognition. *American psychologist*, 37(9), 1019.

Law E.L.C. and Sun X. (2012). Evaluating user experience of adaptive digital educational games with activity theory. *International Journal of Human-Computer Studies*, 70(7), 478-497.

Lessiter J., Freeman J., Keogh E. and Davidoff J.D. (2001). A cross-media presence questionnaire: the ITC Sense of Presence Inventory. *Presence: Teleoperators and Virtual Environments* 10(3), 282-297.

Lewis M., Haviland-Jones J.M. & Feldman-Barrett L. (2000). *Handbook of Emotions*, Third Edition.

## References

Lewis, P.A. Critchley H.D. Rotshtein P. and Dolan R.J. (2007). “*Neural correlates of processing valence and arousal in affective worlds*”, Journal: Cerebral Cortex, March.

Lombard D.M. and Ditton T., At the Heart of It All: The Concept of Telepresence. *Journal of Computer Mediated Communication* 3 (1997) available at:  
<http://jcmc.indiana.edu/vol3/issue2/lombard.html> [accessed 22 June 2013].

Lowenstein, O., & Loewenfeld, I. (1962). The sleep-waking cycle and pupillary activity. *Annals of the New York Academy of Sciences*, 117(1), 147-156.

Lubar, J. F., Swartwood, M. O., Swartwood, J. N., & O'Donnell, P. H. (1995). Evaluation of the effectiveness of EEG neurofeedback training for ADHD in a clinical setting as measured by changes in TOVA scores, behavioural ratings, and WISC-R performance. *Applied Psychophysiology and Biofeedback*, 20(1), 83-99.

Lykken D.T. and Venables P.H. (1971). Direct measurement of skin conductance: a proposal for standardization, *Psychophysiology* 8, 656-672.

Lysaght, R. J., Hill, S. G., Dick, A. O., Plamondon, B. D., & Linton, P. M. (1989). *Operator workload: Comprehensive review and evaluation of operator workload methodologies* (No. TR-2075-3). ANALYTICS INC WILLOW GROVE PA.

### M

Maarten A. Hogervorst, Anne-Marie Brouwer, Jan B. F. van Erp. (2014). *Frontiers in Neuroscience*. 2014; 8: 322. Published online Oct 14.

Malmivuo, J., & Plonsey, R. (1995). *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields*. Oxford University Press, USA.

Mandryk, R. (2008). Physiological measures for game evaluation. In: K. Isbister and N. Schaffer (Eds.), *Game Usability: Advancing the player experience*. Burlington, MA: Morgan Kaufmann Publishers.

Mandryk, R. L., Inkepn, K. M., and Calvert, W. (2006). Using psychophysiological techniques to measure user experience with entertainment technologies. In *Behaviour & Information Technology*. 25(2), 141-158.

Mandryk, R. & Atkins, M. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play environments. *International Journal of Human-Computer Studies*. 65(4), 329-347.

## References

McCrae, R. R., & Costa Jr, P. T. (1999). A five-factor theory of personality. *Handbook of personality: Theory and research*, 2, 139-153.

McCrickard, D. S., & Chewar, C. M. (2006). Designing attention-centric notification systems: five HCI challenges. *Cognitive Systems: Human Cognitive Models in Systems Design*, 3, 67-89.

McGregor, G. (2008) "Terra ludus, terra paidia, terra prefab: spatialization of play in videogames & virtual worlds". In *Proceedings of the 5th Australasian Conference on Interactive Entertainment*, 5, ACM.

McQuiggan, S. W., & Lester, J. C. (2009). Modelling affect expression and recognition in an interactive learning environment. *International Journal of Learning Technology*, 4(3-4), 216-233.

Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning*, 22(3), 149-158.

Mourlas, C. (Ed.). (2009). *Cognitive and Emotional Processes in Web-Based Education: Integrating Human Factors and Personalization: Integrating Human Factors and Personalization*. IGI Global.

## References

Mourlas, C., Tsianos N., and Germanakos P. (2009). *Cognitive and Emotional Processes in Web-Based Education: Integrating Human Factors and Personalisation*. IGI Global.

Mourlas, C., Tsianos N., and Germanakos P. (2008). *Intelligent User Interfaces: Adaptation and Personalization Systems and Technologies*. IGI Global.

Murugappan, M., Rizon, M., Nagarajan, R., Yaacob, S., Zunaidi, I., & Hazry, D. (2007). EEG feature extraction for classifying emotions using FCM and FKM. *International journal of Computers and Communications*, 1(2), 21-25.

## N

Nacke, L. E. (2009). *Affective Ludology: Scientific Measurement of User Experience in Interactive Entertainment*. Ph.D. Thesis. Blekinge Institute of Technology.

Nacke, L., & Lindley, C. (2009). *Affective Ludology, Flow and Immersion in a First- Person Shooter: Measurement of Player Experience*. Loading. 3, 5.

Nacke, L. E., Grimshaw, M. N., Lindley, C. A. (2010). More Than a Feeling: Measurement of Sonic User Experience and Psychophysiology in a First-Person Shooter Game. *Interacting with Computers*, 22(5), 336-343.

## References

Naglieri, J. A., & Das, J. P. (1997). *Cognitive Assessment System*.

Norman D.A. (2004). *Emotional Design: Why We Love (or Hate) Everyday Things*. Basic Books, New York.

## O

Oatley, K., Keltner, D., & Jenkins, J. M. (2006). *Understanding emotions*. Blackwell publishing, 23, 421-441.

Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in cognitive sciences*, 9(5), 242-249.

## P

Palomba C.A. & Banta T.W. (1999). *Assessment Essentials: Planning, Implementing and Improving Assessment in Higher Education*. San Francisco: Jossey-Bass. [405]. A general guide to assessment theory and practice for institutions and their programs.

## References

- Parasuraman, R., & Rovira, E. (2005). *Workload modeling and workload management: Recent theoretical developments* (No. ARL-CR-0562). George Mason Univ Fairfax VA Dept of Psychology.
- Parasuraman, R., & Riley, V.A. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230–253.
- Parasuraman, R., Sheridan, T. B., & Wickens, C.D. (2000). A model of types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans* 30(3), 286–297.
- Pessoa, L. (2013). *The cognitive-emotional brain: From interactions to integration*. MIT press.
- Peter C. & Herbon A. (2006). Emotion representation and physiology assignments in digital systems. *Interacting with computers* 18(2), 139-170.
- Phan, K. L., Taylor, S. F., Welsh, R. C., Decker, L. R., Noll, D. C., Nichols, T. E., ... & Liberzon, I. (2003). Activation of the medial prefrontal cortex and extended amygdala by individual ratings of emotional arousal: a fMRI study. *Biological psychiatry*, 53(3), 211-215.
- Picard, R. W. (2003). Affective computing: challenges. *International Journal of Human-Computer Studies*, 59(1), 55-64.

## References

Prinzel, L., Pope A., & Freeman, F. (2001). Application of physiological self-regulation and adaptive task allocation techniques for controlling operator hazardous states of awareness. NASA/TM-2001-211015.

Prinzel, L., Pope, A., Freeman, F., Scerbo, M., & Mikulka, P. (2001). Empirical analysis of EEG and ERPs for psychophysiological adaptive task allocation. (Tech. Rep. Nr. TM-2001-211016). Hampton, VA: Langley Research Center: NASA.

Prinzel, L., Freeman, F., Scerbo, M., Mikulka, P., & Pope A. (2000). A closed-loop system for examining psychophysiological measures for adaptive task allocation. *The International Journal of Aviation Psychology*, 10(4), 393-410.

Psaltis, A. & Mourlas C. (2011). A real time attachment-free, psychophysiological stress and heart rate measurement system, *International Journal of Measurement Technologies and Instrumentation Engineering* 1, 1-13.

Psaltis, A., Rizopoulos, C., Lekkas, Z., Mourlas, C. (2014). Deducing User States of Engagement in Real Time by Using a Purpose Built Unobtrusive Physiological Measurement Device: An Empirical Study and HCI Design Challenges. In Schmorow, D.D. & Fidopiastis, C.M. (eds.). *Foundations of Augmented Cognition: Advancing Human Performance and Decision-Making through Adaptive Systems – Proceedings of the 8<sup>th</sup> International Conference on Augmented Cognition* (AC 2014), Heraklion, Greece, 22-27 June 2014, LNAI vol. 8534, Springer, 55-66.

Psaltis, A., Rizopoulos, C., Mourlas, C. (2014). Mapping User Engagement to States of Affect via an Unobtrusive Biofeedback Device: A Dynamic Visualization of Real-Time Assessment. In Proceedings of the 2014 International Conference on Physiological Computing (PhyCS'14), Lisbon, Portugal, 7-9 January 2014.

### Q

Quartz, S. R. 2009. Reason, emotion and decision-making: risk and reward computation with feeling. *Trends in Cognitive Sciences*. 13(5), 209-215.

### R

Ravaja, N. (2004). Contributions of psychophysiology to media research: Review and recommendations. *Media Psychology* 6(2), 193-235.

Ravaja, N., Saari, T., Laarni, J., Kallinen, K., Salminen, M., Holopainen, J. & Järvinen, A. (2005). The Psychophysiology of Video Gaming: Phasic Emotional Responses to Game Events, Proceedings of DiGRA 2005. Changing Views - Worlds in Play, Retrieved December 19, 2015 from: <http://www.digra.org/dl/db/06278.36196.pdf> [10.3.2016].

Ravaja, N., Turpeinen, M., Saari, T., Puttonen, S. and Keltikangas-Jarvinen, L. (2008). The psychophysiology of James Bond: Phasic emotional responses to violent video game events. *Emotion* 8(1), 114–120.

## References

Reiser, R. A. (2002). A history of instructional design and technology. A. R. Reiser (Ed). *Trends and issues in instructional design and technology*, 3, (26-53), Florida State University

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome Jr, E. D., & Beck, L. H. (1956). A continuous performance test of brain damage. *Journal of consulting psychology*, 20(5), 343.

Rovira, E., K. McGarry, & Parasuraman, R. (2007). Effects of imperfect automation on decision making in a simulated command and control task. *Human Factors* 49(1), 76–87.

Russell, J.A. (1980). *A circumplex model of affect*. *Journal of Personality and Social Psychology*: 39, 1161–1178.

Russell J.A. (2003) *Core Affect and the Psychological Construction of Emotion*. *Psychological Review*, 110, 145-172.

Russell J.A. and Pratt G. (1980). A description of the affective quality attributed to environments, *Journal of Personality and Social Psychology* 38, 311-322.

## References

Russell J.A. and Snodgrass J. (1987). *Emotion and the Environment*, in: Handbook of Environmental Psychology (1<sup>st</sup> edition), D. Stokols and I. Altman (eds.), New York: John Wiley and Sons, 245-280.

Russell, & J. Laffey (Eds.) (2015). *Handbook of Research on Gaming Trends in P-12 Education* 394-420. Hershey, PA: Information Science Reference.

## S

Salen, K. and Zimmerman, E. (2004). *Rules of play: Game design fundamentals*, Cambridge, MA, MIT Press.

Salovey, P., & Mayer, J.D., Emotional Intelligence. *Imagination, Cognition and Personality*. 1990, Baywood Publishing Company Inc.

Scerbo, M. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* 37- 63. Mahwah, NJ: Erlbaum.

Scerbo, M., Freeman F., & Mikulka P. (2000). A biocybernetic system for adaptive automation. In R. W. Backs & W. Boucsein (Eds.), *Engineering psychophysiology* 241-253. Mahwah, NJ: Erlbaum.

## References

Scerbo, M., Freeman F., Mikulka P., Parasuraman R., Di Nocero F., & Prinzel III L. (2001). The efficacy of psychophysiological measures for implementing adaptive technology. NASA / TP-2001-211018.

Schachter, S. & Singer, J. E. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, 69, 379-399.

Scherer, K. R. (2005). What are emotions? And how can they be measured?. *Social science information*, 44(4), 695-729.

Schunk, D. H. (1989). Social cognitive theory and self-regulated learning. In *Self-regulated learning and academic achievement*, 83-110. Springer, New York, NY.

Schupp, H. T., Flaisch, T., Stockburger, J., & Junghöfer, M. (2006). Emotion and attention: event-related brain potential studies. *Progress in brain research*, 156, 31-51.

Sherry, D. F., & Schacter, D. L. (1987). The evolution of multiple memory systems. *Psychological review*, 94(4), 439.

## References

Snow, R. E., Corno, L. and Jackson, D. 1996. “*Individual differences in affective and cognitive functions*”. In Handbook of Educational Psychology. Prentice Hall.

Solomon R.L. & Corbit J.D. (1974). An opponent-process theory of motivation: I. *Temporal dynamics of affect*. Psychological Review, 81, 119-145.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). Manual for the state-trait anxiety inventory (Palo Alto, CA, Consulting Psychologists Press). *Inc.*

Sterman, M. B., Kaiser, D. A., Mann, C. A., Suyenobu, B. Y., Beyma, D. C., & Francis, J. R. (1993, October). Application of quantitative EEG analysis to workload assessment in an advanced aircraft simulator. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 37, No. 1, 118-121. Sage CA: Los Angeles, CA: SAGE Publications.

Stevens, R., Galloway, T., & Berka, C. (2006). Integrating EEG models of cognitive load with machine learning models of scientific problem solving. *Augmented Cognition: Past, Present and Future*, 2, 55-65.

Strongman K.T. (2003). *The psychology of emotion: from everyday life to theory* (5<sup>th</sup> edition), John Wiley and Sons, Chichester, UK.

## T

## References

Tecce, J. (1992). Psychology, physiology, and experimental psychology. In Ed, McGrawHill Yearbook of science & technology, 375-377. NY: McGraw-Hill.

Tortora G.J, Grabowski S.R. (2006) *Principles of Physiology and Anatomy*, 9th Edition, (248-255). Publisher: John Wiley & Sons ISBN-13: 978-0471366928

Tsatsou, K. (2006). *A Study of Electrodermal Response to Bodily Evoked Potentials* (Doctoral dissertation, Masters Dissertation, University of Patras, Greece) [In Greek].

Tsianos, N., Germanakos, P., Lekkas, Z., Mourlas, C., Samaras, G., & Belk, M. (2009). Working Memory Differences in e-Learning Environments: Optimization of Learners' Performance through Personalization. *User Modeling, Adaptation, and Personalization*, 385-390.

Turpin, G. (1986). Effects of stimulus intensity on autonomic responding: The problem of differentiating orienting and defense reflexes. *Psychophysiology*, 23(1), 1-14.

## References

Venables P.H. and Christie M.J. (1980). *Electrodermal activity*, in: Techniques in psychophysiology, I. Martin and P.H. Venables (eds), New York, John Wiley and Sons, 2-67.

## W

Welford, N. T. (1952). An electronic digital recording machine-the SETAR. *Journal of Scientific Instruments*, 29(1), 1.

Whitton, M. C. (2003). Making virtual environments compelling. *Communications of the ACM*, 46(7), 40-47.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human factors*, 50(3), 449-455.

Wickens, C. D. (2012). Workload assessment and prediction. *MANPRINT: an approach to systems integration*, 257.

Wilson, G. F., & O'Donnell, R. D. (1988). Measurement of operator workload with the neuropsychological workload test battery. *Advances in Psychology*, 52, 63-100.

## References

Wiklund, M., Rudenmalm, W., Norberg, L., Westin, T., Mozelius, P. (2015). Evaluating Educational Games Using Facial Expression Recognition Software: Measurement of Gaming Emotion. In *Proceedings of the 9th European conference on game-based learning*, Nord-Trondelag University College, Steinkjer, Norway, Reading: Academic Conferences Publishing,

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3), 225-240.

Witmer, B. G., Jerome, C. J., & Singer, M. J. (2005). The factor structure of the presence questionnaire. *Presence: Teleoperators & Virtual Environments*, 14(3), 298-312.

## Y

Yannakakis, G. N., & Hallam, J. (2008). Entertainment modeling through physiology in physical play. *International Journal of Human-Computer Studies*, 66(10), 741-755.

Yannakakis, G. N., Hallam, J., & Lund, H. H. (2008). Entertainment capture through heart rate activity in physical interactive playgrounds. *User Modeling and User-Adapted Interaction*, 18(1-2), 207-243.

## Appendices

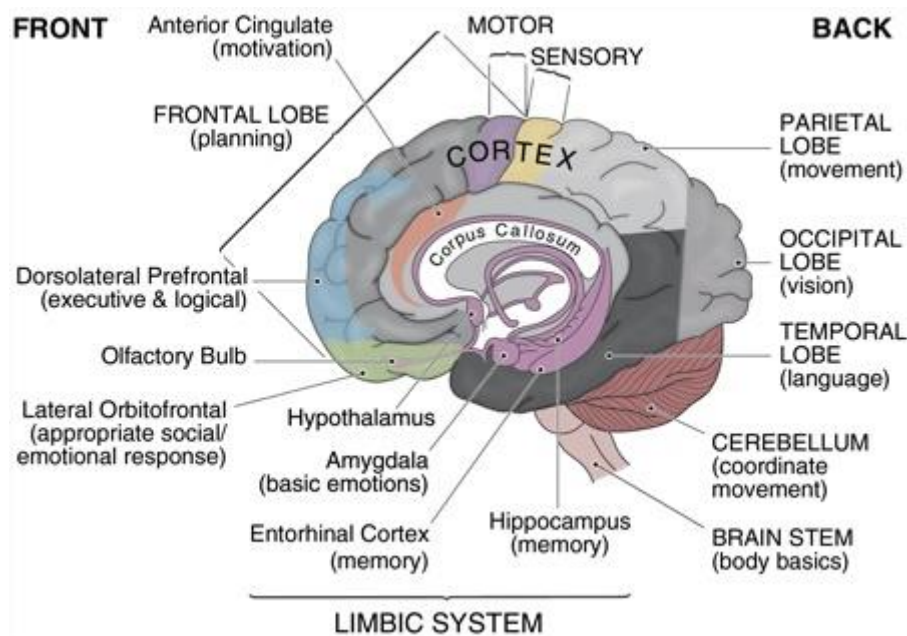






### Appendix A: Functional topology of the human brain

Extensive research in medical sciences, biology and psychology have gone a long way to explore neural activity aspects and pinpoint brain areas involved in preset tasks involving cognitive, emotional and behavioural physiological processes. Figure A.1 shows a sagittal view of the topological allocation of the main areas of the human brain with their associated activity area as they have been traced by means of multimodal assessment while under the instigation of specific evoked potentials.



*Figure A.1 Human brain areas and their associated functions*

Emotional, affective and cognitive functions have been identified to take place in the sub-area known as limbic. Vast literature in valuation studies of somatic markers (Bechara, Damasio H & A.R., 2000) has identified that emotional status influences the way people react to events and verified that functioning emotions rationalise decision making. From the side of human physiology, findings concluded that once an emotional stimulus occurs, the brain initiates a response via neuronal, hormonal and frequently kinetic reaction thus instructing the autonomic nervous system to cope with the urgency and severity of the felt event while that is under

## Appendices

cognitive identification. The hormonal activation includes the secretion of adrenaline in order to increase the heart and respiration rate, however, neuronal mediation causes skin tentative cautiousness by promoting sweat gland dilatation, in order to counteract for an aggressive stimulus activating the sympathetic nervous system (fight state). As soon as the severity of the event has been evaluated by the brain and classified as low or the intensity of the event has eased away, the parasympathetic nervous system alleviates the imbalances introduced in the previous time interval, thus returning the hormonal and neuronal stimulations back to ordinary levels (flight state) and subsequently reinstating the equilibrium of homeostatic balance of the human body. Apparently, as the above processes take place subconsciously, if this course was somehow identified and ascertained then it could be considered as an optimised indicator of human involvement in a mental task. Additionally, as these modus operandi of physiology are common to all humans with similar reactions to a great extent, they provide a reliable method for assessing psychosomatic arousal and active engagement by measuring mental activation and its acuity levels. Accurate selection of physiological signals is deemed essential as detection of homeostatic disturbance has to be complemented by quantified values of severity and duration so that a classification of tentativeness and excitement can be achieved.

### Appendix B: Physiological process of emotion

Although some considered emotion as being only psychologically based, scientific discoveries have shown emotion are based upon physiological processes as: *“A complex reaction pattern of changes in nervous, visceral, and skeletal muscle tissues in response to a stimulus. As a strong feeling, emotion is usually directed toward a specific person or event and involves widespread physiological changes, such as increased heart rate and inhibition of peristalsis.”* (The Longman Dictionary of Psychology and Psychiatry). Recent advancements in neuroscience demonstrate that emotion is an interaction between chains of amino acids that form neuropeptides and receptors. Emotions are normal physiological (organic) processes in the body, some of which are pleasant and others that are quite unpleasant. Although the primary locations for identifying the physiology of emotions are in the brain, spine and autonomic nervous system, emotion affect any and all parts of the body in a physiological way. Research has demonstrated that the biochemical substances produced as a result of an emotion travel to almost every cell in the body in various ways such as hormonal; through the endocrine glands secreting directly into the bloodstream, or neurologically transmitted; either adrenergic, activating / deactivating neurotransmitters such as epinephrine norepinephrine, or directly transmitted to induce muscular reactions caused by fear, startle response or other abrupt events. Analysing the mechanism of operation of neurotransmitter interaction would provide a comprehensive clarification and understanding of the principles employed, as these will build the foundation for our approach on a specific design for the purpose of this study. Neurotransmitters comprise a range of small-molecule chemical substances released by a process - called exocytosis - from a nerve ending on the arrival of a nerve impulse, implying that neurotransmitters are specific to particular neurons. They interact with receptors on adjacent structures to trigger off a response, either excitatory or inhibitory. The adjacent structure may be another nerve, a muscle fiber or a gland. The main neurotransmitters are acetylcholine, glycine, glutamate, dopamine, noradrenaline, adrenaline, serotonin, histamine and gamma-amino-butyric acid (GABA), producing radical and quick establishment of their effective presence. Just as feelings cause for the most part unknown combinations of biological interactions, the task to identify affect or emotional components becomes more complex

## Appendices

limiting prospective ambitions to interpret the brain activity with high precision, however, as biological interactions produce immediate responses that can be identified either as simple or congregated reactions, the possibility to study behavioural patterns is made possible. Bio-sensing is the use of chemically sensitive biological detectors utilising electronic analytical procedures aiming to identify biomolecular reactions or concentrations thus revealing details concerning specific biological processes. Various bio-sensing instruments have been developed with commercially available sensors or transducers that can transform a biochemical event or state into an analogous electrical, electronically or sometimes chromatic scale thus distinctly identifiable form. Examples of commonly used biosensors are electrochemical biosensors like skin PH sweat patches, continuous monitoring wearable sensors and combined metal based reactive sensors used for detecting conductive or electrical activity such as Electro-Cardio-Gram (ECG) Electro-Encephalo-Gram (EEG) etc.

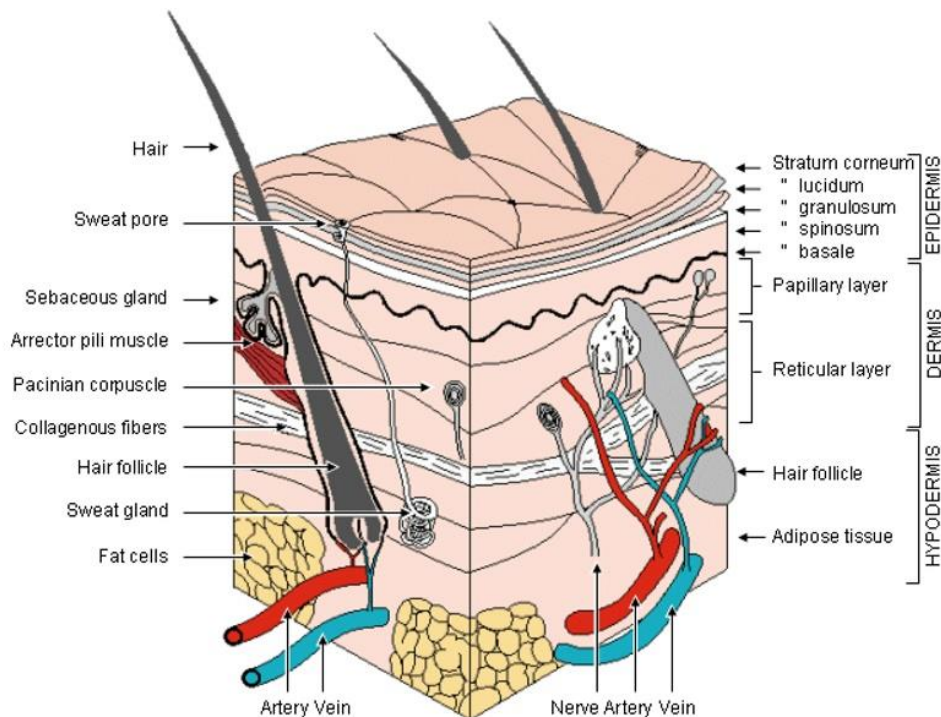
### Appendix C: Psychophysiology of Electro Dermal Response (EDR)

Although an important function of the skin is to protect the organism it covers by keeping bacteria, parasites, and noxious chemicals out and keeping vital fluids in, it also has a role in thermoregulatory activities. The thermoregulatory contribution is produced by dilation of blood vessels in the skin and increased sweating; both of which result in decreased skin and body temperature. Blood vessel dilation increases blood flow in surface areas to enhance cooling. The increased sweat on the skin surface produces cooling through evaporation. Humans are remarkable for their relatively hairless skin and high density of innervated thermoregulatory sweat glands. The density of sweat glands in human skin varies from around 50-200 glands/cm<sup>2</sup> on the trunk and limbs, to around 200-400 glands/cm<sup>2</sup> on the face and the dorsal sides of the hands and feet, with 600 to perhaps more than 2000 glands/cm<sup>2</sup> on the palms of the hands and soles of the feet. An interesting aspect of sweating is that it is not merely thermoregulatory (cooling). This fact forms the basis for many behavioural studies. Sweating or sweat gland activity, is reflected as changes in skin potential (SP) and skin conductance (SC) in a variety of situations, including those that are emotionally arousing. For example, eccrine glands of the palms and fingers of the hand respond only weakly at certain levels of heat and strongly to psychological and sensory stimuli. One may have noticed wet or "clammy" palms in situations causing fear or anxiety. Sweating from psychological stimuli has been considered by researchers to have adaptive value. It has suggested that the sweating of palms and soles may be an adaptive response that persisted over the course of evolution, because it aids in grasping objects. Darrow, as early as 1937, suggested that palmar sweating served to increase grip strength and tactile acuity, reminding the typical scene of a labourer spitting on his hands before grasping a pickaxe or shovel.

The interpretation of skin conductance requires some understanding about the structure of tissues at and beneath the skin surface. Figure C.1 shows the main features of the skin. The most superficial layer is called the epidermis and consists of the stratum corneum, the stratum lucidum (seen only on "frictional surfaces"), the granular layer, the prickle cell layer, and the basal or germinating layer. The surface of the skin is composed of dead cells, while at its base

## Appendices

one finds healthy, living cells. Between these two sites there are transitional cells that constitute the horny layer. Blood vessels are found in the dermis whereas secretory cells of the eccrine sweat gland are found at the boundary between the dermis and the panniculus adiposus, also referred to as hypodermis and superficial fascia. The eccrine sweat glands consists of an excretory duct which is a simple tube made up of a single or double layer of epithelial cells. This duct ascends up to the surface of the skin. It is undulating in the dermis and follows a spiral and inverted conical path through the epidermis to terminate in a pore on the skin surface. Cholinergic stimulation via fibres from the sympathetic nervous system comprises the major influence that elicits the production of sweat by these eccrine glands. From an observation of Figure C.1 one can understand that the epidermis has a high electrical resistance due to the thick layer of dead cells with thickened keratin membranes. This characteristic serves the function of the skin to provide a barrier and protection against abrasion, mechanical assaults, and so on. The entire epidermis (with the exception of the desquamating cells) constitutes the barrier layer), a permeability barrier to flow. Experiments show its behaviour to be that of a passive membrane, however, the corneum is penetrated by the aforementioned sweat ducts from underlying cells and while these ducts fill, a relatively good conductor emerges, and results in many low-resistance parallel pathways. A further increase in conductance results from the hydration of the corneum due to the flow of sweat across the duct walls. Sweat can be considered as a weak electrolyte equivalent to a 0.3% NaCl salt solution. As a consequence to the flow of sweat the resulting skin conductance can vary accordingly, depending on variations of the eccrine stimulation. The aforementioned process is vivid particularly in the palmar and plantar regions because at these areas whilst the epidermis is very thick, at the same time the eccrine glands are immensely dense. A major function of the skin is to protect the body from the environment and also to prevent the loss of body water, however, the evaporation of water at the same time works as a means of regulating body temperature. These two functions of a barrier layer that prevents the loss of water to the outside except through the sweat glands are mediated by the sympathetic nervous system. So effectively the measurement of the volumetric capacity of the sweat glands identified by Electro Dermal Response (EDR), provides a comparable estimate of the level and extent of sympathetic activity.



**Figure C.1** Section of smooth skin taken from the sole of the foot. (Redrawn from Ebling, Eady, & Leigh, 1992.)

It has been established scientifically that EDR is associated with sweat gland activity from experiments which prove a direct correlation with stimulated sweat gland activity. Also extensive literature exists (Fowles, 2007) showing that when sweat gland activity is eliminated, then there is an absence of EDR signals. There are two major types of measuring EDR. One involving exosomatic measurement of voltage potential applied to the skin and the other endosomatic, where the source of voltage is internal. Obviously the exosomatic type has the upper hand being non-intrusive.

## **Appendices**

### **Appendix D: Literature on psychometric tests employed**

Test sample pages to be placed here copyright permission allowing.

#### **NEO-FFI Item Clusters:**

The personality inventory was designed to assess the five factors or dimensions of the five factor model (FFM) of personality that is explained in detail in the next chapter. Looking back and examining the inventory itself, there are 60 items, 12 for each scale. Every fifth item is from the same scale. Items 1, 6, 11, and so on assess Emotionality or Neuroticism (N). Items 2, 7, 12, and so on assess Extraversion (E). Items 3, 8, 13, and so on assess Openness to Experience (O). Items 4, 9, 14, and so on assess Agreeableness (A). Items 5, 10, 15, and so on assess Conscientiousness (C). A number of the items are reverse scored. For example, item 1 ("I am not a worrier") is reverse scored. The more you disagree, the higher your score on N. Item 27 ("I usually prefer to do things alone") is also reverse scored. The more you agree, the lower your score on E.

In order to explain the mechanics of the test we'll look at some score calculations.

First, compute your score for N.

These are items: 1, 6, 11, 16, 21, 26, 31, 36, 41, 46, 51, 56.

Mark an "R" next to the reverse scored items: 1, 16, 31, 46.

For the non-reversed-scored items, SD=0, D=1, N=2, A=3, SA=4.

For the reversed-scored items, SD=4, D=3, N=2, A=1, SA=0.

The sum of all 12 items is your score for N.

Now, use the same procedure to compute your scores for E, O, A, and C. The following are the items for each scale, with the reverse-scored items in bold-faced type.

N (rev) 1, 11, 16, 31, 46, 51, 56.

E: 2, 7, 12, 17, 22, 27, 32, 37, 42, 47, 52, 57.

O: 3, 8, 13, 18, 23, 28, 33, 38, 43, 48, 53, 58.

A: 4, 9, 14, 19, 24, 29, 34, 39, 44, 49, 54, 59.

C: 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60.

One common way to interpret test scores is to compare them to the scores of other people who have already a test. The other people who have taken the test are called a normative sample. The developers of the NEO have given the test to a large sample of adults of all ages in the United States. The following allows you to compare your scores to that normative sample.

Compared to males in the normative sample:

N: If you scored below 13 you are low (and below 6 very low) on N. If you scored above 21 you are high (and above 29 very high) on N. Otherwise, you scored in the average range.

E: If you scored below 24 you are low (and below 18 very low) on E. If you scored above 30 you are high (and above 36 very high) on E. Otherwise, you scored in the average range.

## Appendices

O: If you scored below 23 you are low (and below 18 very low) on O. If you scored above 30 you are high (and above 36 very high) on O. Otherwise, you scored in the average range.

A: If you scored below 29 you are low (and below 24 very low) on A. If you scored above 35 you are high (and above 40 very high) on A. Otherwise, you scored in the average range.

C: If you scored below 30 you are low (and below 25 very low) on C. If you scored above 37 you are high (and above 43 very high) on C. Otherwise, you scored in the average range.

Compared to females in the normative sample:

N: If you scored below 16 you are low (and below 8 very low) on N. If you scored above 25 you are high (and above 32 very high) on N. Otherwise, you scored in the average range.

E: If you scored below 25 you are low (and below 19 very low) on E. If you scored above 31 you are high (and above 37 very high) on E. Otherwise, you scored in the average range.

O: If you scored below 23 you are low (and below 18 very low) on O. If you scored above 30 you are high (and above 36 very high) on O. Otherwise, you scored in the average range.

A: If you scored below 31 you are low (and below 26 very low) on A. If you scored above 36 you are high (and above 41 very high) on A. Otherwise, you scored in the average range.

C: If you scored below 32 you are low (and below 26 very low) on C. If you scored above 38 you are high (and above 44 very high) on C. Otherwise, you scored in the average range.

The NEO inventory measures differences among normal individuals. It is not a test of intelligence or ability, and it is not intended to diagnose problems of mental health or

adjustment. It does, however, give you some idea about what makes you unique in your ways of thinking feeling, and interacting with others.

### N

- 🇺🇸 People who score in the average range tend to be generally calm and able to deal with stress, but sometimes experience feelings of guilt, anger, and sadness
- 🇺🇸 People who score above average tend to be sensitive, emotional, and prone to experience feelings
- 🇺🇸 People who score below average tend to be secure, hardy, and generally relaxed even under stressful conditions

### E

- 🇺🇸 People who score in the average range tend to be moderate in activity and enthusiasm. Enjoy the company of others but also value privacy.
- 🇺🇸 People who score above average tend to be extraverted, outgoing, active, and high-spirited; prefer to be around people most of the time.
- 🇺🇸 People who score below average tend to be introverted, reserved, serious; prefer to be alone or with a few close friends.

### O

## Appendices

🌍 People who score in the average range tend to be practical but willing to consider new ways of doing things; seek a balance between the old and the new.

🌍 People who score above average tend to be open to new experiences; have broad interests and very imaginative.

🌍 People who score below average tend to be down-to-earth, practical, traditional, and pretty much set in your ways.

### A

🌍 People who score in the average range tend to be generally warm, trusting, and agreeable, but you can sometimes be stubborn and competitive.

🌍 People who score above average tend to be compassionate, good-natured, and eager to cooperate and avoid conflict.

🌍 People who score below average tend to be hardheaded, skeptical, proud, and competitive; tend to express anger directly.

### C

🌍 People who score in the average range tend to be dependable, moderately well-organized; generally have clear goals but are able to set you work aside.

🌍 People who score above average tend to be conscientious and well-organized; have high standards and always strive to achieve your goals.

- People who score below average tend to be easygoing, not very well-organized, sometimes careless; prefer not to make plans.

**NEO-FFI Reliability, Means, Standard Deviations & Constituent Items**

| Personality Factor       | Sub-Category                       | Reliability | Mean r | Mean  | S.D. | Constituent Items                    |
|--------------------------|------------------------------------|-------------|--------|-------|------|--------------------------------------|
| <b>Neuroticism</b>       | <i>Negative Affect</i>             | .80, (.75)  | .44    | 8.91  | 4.24 | 1*, 11, 16*, 31*, 46*                |
|                          | <i>Self-Reproach</i>               | .82, (.83)  | .40    | 9.11  | 5.06 | 6, 21, 26, 36, 41, 51, 56            |
| <b>Extraversion</b>      | <i>Positive Affect</i>             | .76, (.72)  | .44    | 9.94  | 2.90 | 7, 12*, 37, 42*                      |
|                          | <i>Sociability</i>                 | .67, (.57)  | .34    | 8.09  | 2.78 | 2, 17, 27*, 57*                      |
|                          | <i>Activity</i>                    | .69, (.52)  | .36    | 8.68  | 2.91 | 22, 32, 47, 52                       |
| <b>Openness</b>          | <i>Aesthetic Interests</i>         | .75, (.81)  | .49    | 7.24  | 2.71 | 13, 23*, 43, 33*                     |
|                          | <i>Intellectual Interests</i>      | .74, (.64)  | .50    | 7.73  | 2.57 | 48*, 53, 58                          |
|                          | <i>Unconventionality</i>           | .56, (.40)  | .24    | 8.85  | 2.62 | 3*, 8*, 18*, 38*                     |
| <b>Agreeableness</b>     | <i>Nonantagonistic Orientation</i> | .71, (.72)  | .24    | 21.03 | 4.42 | 9*, 14*, 19, 24*, 29*, 44*, 54*, 59* |
|                          | <i>Prosocial Orientation</i>       | .58, (.58)  | .28    | 12.72 | 1.73 | 4, 34, 39*, 49                       |
| <b>Conscientiousness</b> | <i>Orderliness</i>                 | .74, (.70)  | .37    | 13.52 | 3.37 | 5, 10, 15*, 30*, 55*                 |
|                          | <i>Goal-Striving</i>               | .68, (.69)  | .42    | 8.04  | 2.07 | 25, 35, 60                           |
|                          | <i>Dependability</i>               | .66, (.62)  | .35    | 12.41 | 2.19 | 20, 40, 45*, 50                      |

Note. N = 732, with exception of alpha coefficients in parentheses, which are from the additional sample (N = 253). Mean r = mean interitem correlation. Alpha coefficient under name of domain is reliability of NEO-FFI scale in full sample. \* indicates reverse-keyed item.

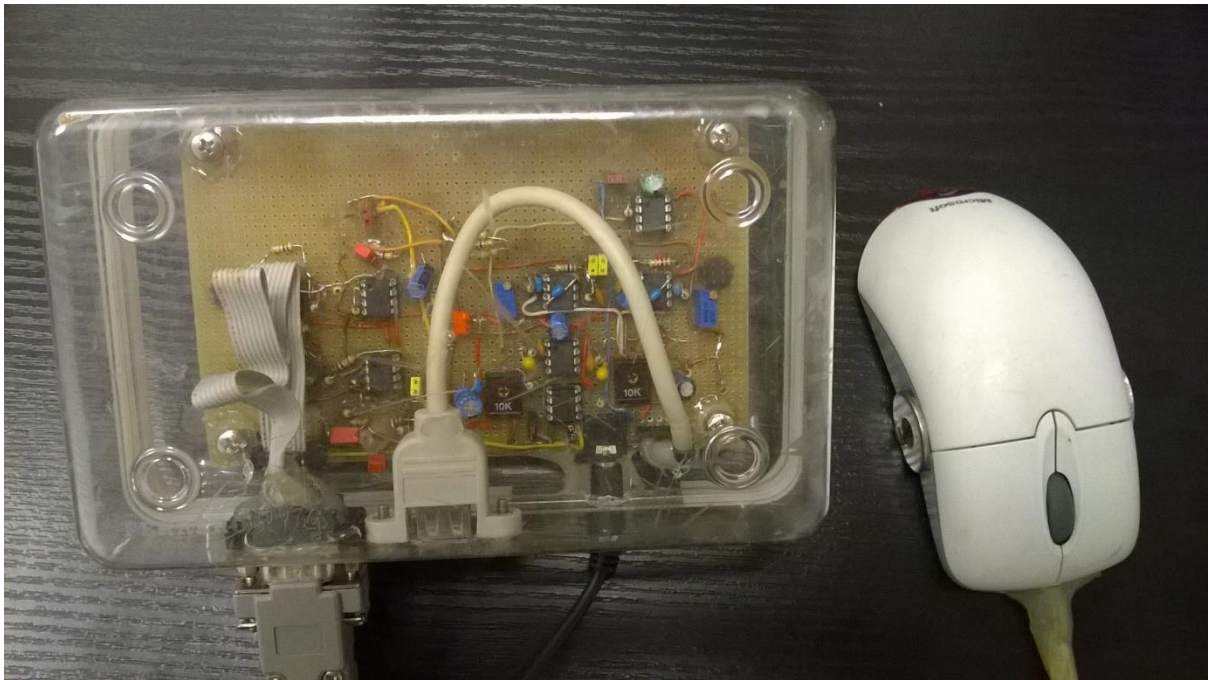
## Appendices

### Alternative Rationally-derived Decomposition of Neuroticism Domain:

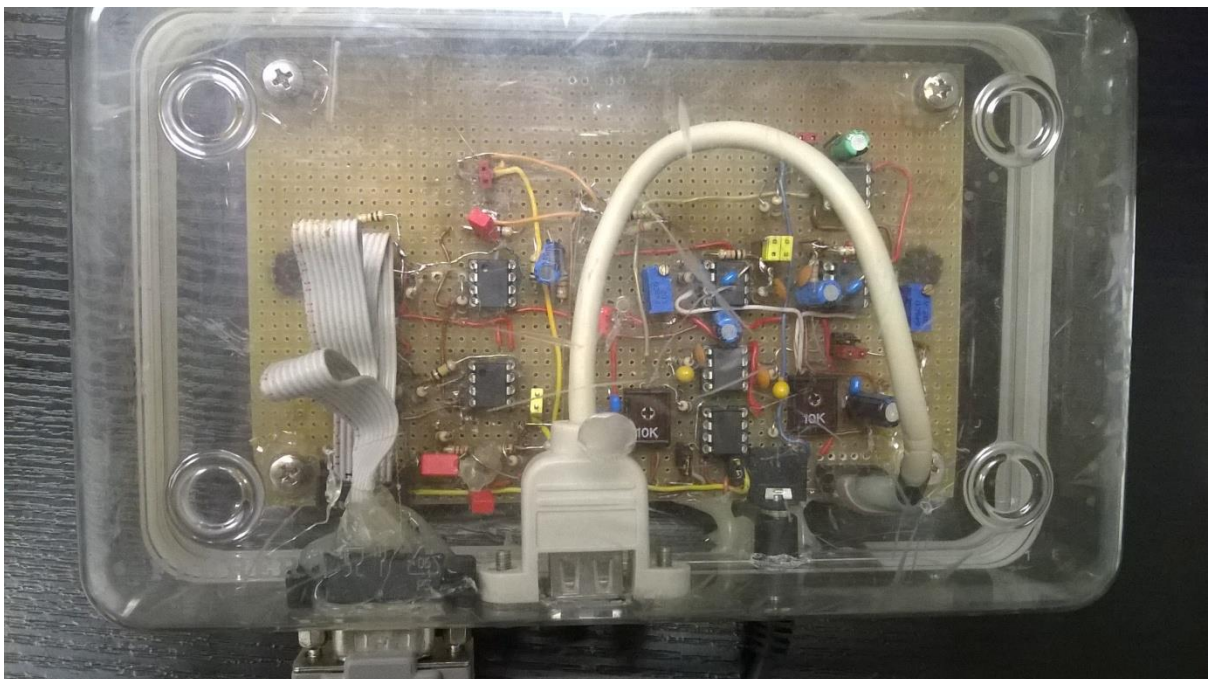
| Personality Factor | Sub-Category         | Reliability | Mean r | Mean | S.D. | Constituent Items |
|--------------------|----------------------|-------------|--------|------|------|-------------------|
| <b>Neuroticism</b> | <i>Anxiety</i>       | .70         | .45    | 5.37 | 2.64 | 1*, 21, 31*       |
|                    | <i>Depression</i>    | .76         | .51    | 4.56 | 2.64 | 16*, 41, 46*      |
|                    | <i>Self-Reproach</i> | .78         | .42    | 6.83 | 4.01 | 6, 11, 26, 51, 56 |

Note. N = 732. \* indicates reverse-keyed item.

**Appendix E: Pictures of the electronic prototype**



Main electronics prototype unit with mouse



Main unit with connectors to USB and mouse respectively.

## Appendices



Mouse with details of embedded sensors



Sample of a bead implementation of the circuit.