Mean head and shoulder heights when seated: subconscious postural cycles during discrete, computerised stimuli

Harry J. Witchel
Brighton & Sussex Medical School
Brighton BN1 9PX UK
+44 1273 873549
h.witchel@bsms.ac.uk

Carina Westling
School of Media, Film & Music
University of Sussex
Brighton BN1 9RG UK
c.e.i.westling@sussex.ac.uk

Rob Needham
CSHER, Faculty of Health,
Staffordshire University
Stoke-on-Trent ST4 2DF UK
r.needham@staffs.ac.uk

Aoife Healy
CSHER, Faculty of Health,
Staffordshire University
Stoke-on-Trent ST4 2DF UK
a.healy@staffs.ac.uk

Nachiappan Chockalingam
CSHER, Faculty of Health,
Staffordshire University
Stoke-on-Trent ST4 2DF UK
+44 1782 295853
n.chockalingam@staffs.ac.uk

ABSTRACT
Discrete, three-minute, computer-presented stimuli (designed to range from engaging to incredibly boring) were used to elicit changes in cognitive/emotional states in seated, healthy volunteers. These stimuli did not require the use of a mouse, so movements were assumed to be non-instrumental. Stimuli included films, games, quizzes and music. Motion capture and video analysis were used to detect changes in head and shoulder position in response to the stimuli. Results include changes occurring between the first half and the second half of each of the main stimuli (i.e. arising in less than one minute as the volunteer "settles in"); in the second half of each stimulus, there were decreases in head height and shoulder height (i.e. position rather than movement). In conclusion, we speculate that non-instrumental changes in head height and shoulder height may suggest loss of vigilance or diminishing arousal in seated computer-users. Our unique contributions are: 1) discrete stimuli, were used on seated volunteers 2) without a mouse, to show that 3) modest (mm) head and shoulder movements in the vertical axis correlated with 4) subtle cyclical changes in boredom, not overall changes in fatigue. Future psychological validation of tutoring systems with discrete stimuli can use these postural parameters as part of a multimodal analysis of engagement.

Author Keywords
Motion capture; video analysis; fatigue; vigilance; arousal; boredom; head attitude; postural change; discrete stimuli.

ACM CLASSIFICATION KEYWORDS
HCI design and evaluation methods: Laboratory experiments.

General Terms
Human Factors; Affective computing; Measurement.

INTRODUCTION
The field of ergonomics, especially relating to affective computing in computer-presented learning, has an ongoing strand of research focused on the objective interpretation of posture and nonverbal human behaviour [21, 15, 8]. Automated teaching systems (e.g. auto-tutor [8]) are seen as needing a way to recognise when the human learner is bored, frustrated or confused, so that the teaching system might respond – by giving hints, presenting a more engaging problem, providing motivational encouragement, recommending a break etc. [10, 14]. Extensive human computer interaction (HCI) research to recognise mental states unconducive to learning has focused on facial expressions [14, 13]. In addition, the HCI literature also has tested systems based on recognising non-instrumental (i.e. non-purposeful and potentially subconscious) changes in seating posture (e.g. the chair-mat [21, 10]), which may detect putative disinterest indicators (e.g. fidgeting [6, 4]). Almost none of the measurements of these non-instrumental movements have been based around motion capture of individual parts of the body (e.g. the shoulders). Instead, postural measures of seated individuals has been limited to head position detection [1, 12, 15] or seat pressure mats, [6, 15, 21]; a corpus of positional data based on motion capture now exists, but that is based on standing, continuous video game play on the Wii [16].

The field of human factors also has an ongoing strand of research focused on the objective interpretation of posture and nonverbal human behaviour (e.g. MINDS [12]). In human factors, fatigue is considered a research priority because the majority of accidents made in trucking, (e.g. in the processes of mining during early morning hours) are made due to operator fatigue [19, 23, 25]. The importance of making non-invasive measures (e.g. postural measures, as opposed to subjective questionnaires) is that second-by-second mindset data is generated without interfering with the operator's task or attention; objective data is needed because in some cases simple subjective self-assessments of wakefulness/fatigue have been shown to be poorly correlated with compromised ability or with measured fatigue [19, 23, 25]. The putative postural and nonverbal measurements for fatigue are eye closures and large head pitch forward movements (i.e. "nodding off"), as well as fidgeting, eye rubbing and yawning.

There are two methodological aspects to the successful implementation of systems for postural recognition and interpretation in cognitive ergonomics and human factors: deployment of a range of sensors (with the ability for the signals to be collated continuously), and the analysis of the postural signals detected by those sensors. The present study focuses on the latter aspect: interpretation of postural signals for the potential detection of boredom, fatigue, vigilance and arousal. Several challenges exist in analysis of these postural signals. First, researchers in the field of the psychology of non-verbal behaviour have not established universally accepted criteria about how to classify postures [3, 17, 20], although an attempt has been made in HCI [11].

Second, despite extensive psychological research to analyse nonverbal postures based on manual coding of films, the nature of the relationship between posture and underlying mental/emotional states remains controversial. Ekman [9] proposed that body movements only carry information about the intensity of the emotion that is being experienced. In contrast, Bull [3] presented results showing that both body movements and specific postural positions can transmit information about boredom and interest. HCI researchers using sensors were surprised to find that those static postures of disinterest from manual coding of films were not sufficient for interpreting their sensor-based data. Rosalind Picard's group has specifically eschewed static postures in favour of dynamic postural changes:

> *"This work has never assumed that static postures reveal what a student is feeling inside; rather, observed patterns in the dynamics of the student's postures were found to disclose significant information related to the affective states of high interest, low interest and the related behaviour of taking a break" [21].*

While detection of mental states and emotions "in the wild" is a feature of our laboratory and of the work of others in this field [18, 22], a complementary and important approach is to use laboratory experiments to make clear and unequivocal links between specific, elicited mental states and their corresponding nonverbal correlates, so that sensor systems can be designed to make the appropriate measurements and analyses. HCI investigators have previously described emotions and their measurement as "murky" due to the individual differences and the subjective nature of emotions [6].

To make these links between posture and mental states clear, rather than using a continuous stimulus of a long, non-homogeneous task, we elicited highly divergent mental states using a collection of discrete, 3-minute internally homogeneous stimuli — some members of the collection are unequivocally interesting, others are unequivocally boring, and some are in between. The advantage of discrete stimuli is that **unexpurgated** time series data can be **objectively** associated to the mood of the stimulus.

Most previous studies measuring engagement/boredom have employed continuous stimuli [6, 21], and other psychological studies investigating emotional responses have used discrete, six-second stimuli (e.g. the International Affective Photographic System [2]), which are too short to make determinations about engagement. Hence there is a clear need for using discrete, longer stimuli specifically designed to elicit a range of engagement responses from extreme boredom to interest. This study has attempted to fill this gap by employing such stimuli to characterize the fine-structure of the postural responses to fatigue/boredom.

Although this manuscript focuses on establishing a methodological approach to quantify engagement to various stimuli, the observations and results from this study will have a clear impact on conducting further structured research into cognitive ergonomics and human computer interaction.

The results of the current paper demonstrate the utility of two sensor-detectable postural parameters, mean forehead height and mean shoulder height, which to the best of our knowledge have not been characterised before in the HCI literature.

METHODS

Experimental Volunteers

Fourteen healthy volunteers (2 female, age range 19-62, m ± sd: 29.4 ± 15.6) were recruited from the university community via advertisements and emails. Ethical approval was obtained from our local university ethics committees. The volunteers were seated in a standard four-legged, non-swivelling, armless "reception room" chair with cushioned and fabric-covered back and seat.

Protocol

After being briefed as to the nature of the study, participants were seated in a standard armless "reception room" chair at a desk with a 21.5 inch (diagonal) monitor. The monitor was set up such that the centre of the screen was at the eye level of the volunteer. Volunteers were allowed to adjust the seat position for comfort. After completing initial background questionnaires, participants experienced audiovisual stimuli, each lasting 170 seconds, and then rated the experience via a subjective questionnaire.

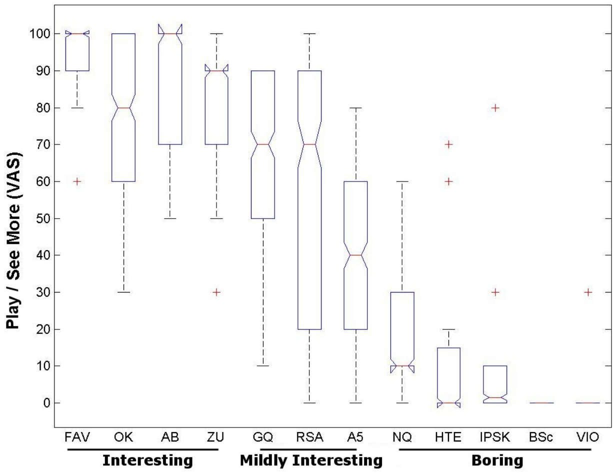

Figure 1. Subjective Visual Analogue Scale ratings for "I wanted to see / play more" for stimuli in the "interesting", "mildly interesting" and "boring" groups of stimuli. VAS anchors are 0 = "not at all" and 100 = "extremely". The box and whisker plots have boxes with lines at the lower quartile, median (red), and upper quartile values. The whiskers are lines extending from each end of the boxes to show the extent of the rest of the data (except for outliers). Outliers are data with values beyond the ends of the whiskers; the maximum whisker length is 1.5 x the inter-quartile range. The notches represent a robust estimate of the uncertainty about the medians for box-to-box comparison. Boxes whose notches do not overlap indicate that the medians of the two groups differ at the 5% significance level.

All experimental stimuli were presented in a counterbalanced order. All members of the scientific team left the room before each stimulus, such that the volunteer was alone in the room as they experienced the stimulus. At the beginning of the experiment, each participant was allowed to adjust the volume control of the sound system to a level they found comfortable, and they were encouraged to pick a level that was slightly quieter just for safety; participants were told that they could adjust the volume at any time if they found the sounds too loud.

Stimuli and Subjective Rating Scales

Stimuli were a collection of games, film excerpts, quizzes, and musical excerpts as described [24]. Each stimulus was preceded by 45 seconds of "television snow" plus white noise (to establish a baseline signal before each stimulus), followed by a brief synchronisation timing signal. Stimuli were rated by a questionnaire with 6 adjectives to be rated on a visual analogue scale (VAS). Each VAS was a 10 cm line with anchors at 0 (not at all) and 100 (extremely). The VAS statements were: I felt interested, I felt bored, I wanted to see/play more, I wanted it to end earlier, I was engrossed by the experience, I felt empathy or emotional attachment to what I saw. With a previous group of volunteers we verified that the subjective responses to our stimuli were as expected. An example of the subjective responses of this cohort is shown in Figure 1.

Motion Capture

Motion capture was performed by video analysis (Kinovea) of video from a lateral aspect (at BSMS) or by an 8-camera opto-electronic mocap system (at Staffordshire). We have previously shown that these two technologies produce comparable results for head attitude and for small translational movements in the sagittal plane [24]. Passive reflective markers were positioned on the head, badge of the deltoid, and middle of the outer thigh. Head markers were placed on the outer canthus of the eye and on the ear behind the tragus (Kinovea) or on a head band as a set of four (left front head, right front head, left back head, right back head); the Vicon movements were corrected for position and angle based on a frame at the beginning of the experiment for each volunteer. The outcome parameters were head pitch (relative to floor), front head marker from screen, front head marker from floor, deltoid marker from screen, deltoid marker from floor, thigh marker from screen, thigh marker from floor. The videos were made by a Canon MV890 mini-DV recorder and captured by Kinovea at 25 Hz. Vicon captured data at 50 Hz, which was down-sampled by Matlab to 25 Hz.

Statistics and analysis

All statistics reported here are paired T tests calculated in Matlab, or ANOVAs and linear regressions performed in Stata 7. Positions were calculated as the mean of each uni-dimensional parameter for the segment listed.

RESULTS

Human postural positions, even when seated, are affected by a wide range of known and unknown influences. To allow for comparisons between one stimulus and the next, each stimulus was preceded by a 45-second baseline period (white noise, WN, see methods). Analyses of stimuli were performed by breaking each stimulus into time segments (Figure 2). The start (St), end (E) and transition (Tr) segments were isolated from the rest of the stimulus because they represented large-scale changes in how or where the experimental participant is expected to focus their attention; in psychophysiology, often the strongest responses occur when the experimenter says, "The experiment will begin… now." We also found that comparatively large postural changes often occur during these segments (data not shown), justifying the automated removal of this part of the data from the analyses. Thus, analysis focused on the main stimuli, but not at its ends.

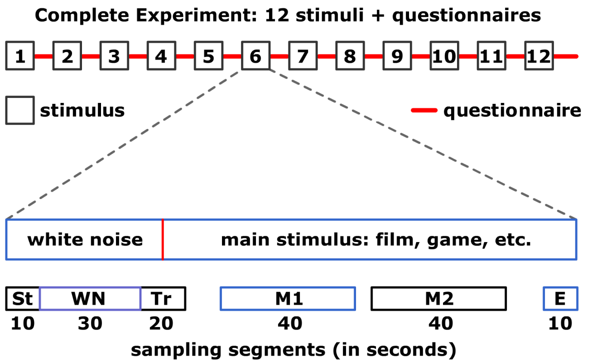

Figure 2. Organization of one entire experiment (above) and micro-structure of a single stimulus (below). The sampling windows within each 170 second stimulus are St = start of stimulus, WN = white noise, Tr = transition period, M1 = the first half of the main stimulus, M2 = the second half of the main stimulus, E = end of main stimulus. The main stimulus is a game, quiz or film; the white noise is used to generate a baseline. The red line between white noise and the main stimulus is the timing signal.

The start of each stimulus begins with the experimenters leaving the room and the participant coming to terms with the initiation of a stimulus (beginning with white noise); we expect this to be a period going from high to lower arousal. The central period of white noise is expected to be low in vigilance, arousal and attention. The transition period would include a sudden increase in attention and arousal. The main stimuli are all homogeneous, so Main Part 1 and Main Part 2 should be identical except for effects of "settling in" and/or becoming bored with the stimulus.

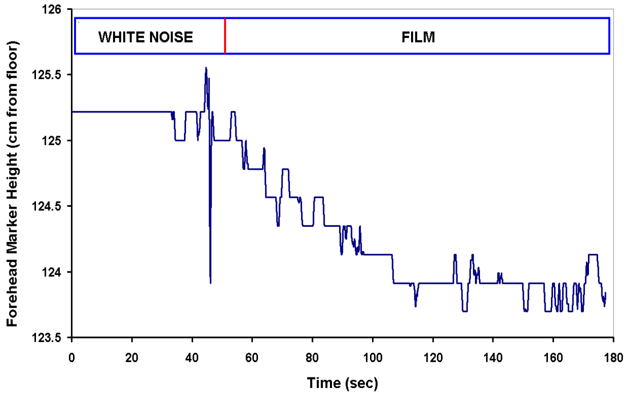

Figure 3. Representative data showing the change in height of the forehead marker of one volunteer during the course of watching the stimulus "OK".

Postural Changes during the Stimuli

Although it is axiomatic in human-to-human dyadic communication that increased interest and engagement is accompanied by proximity and approach [5], we found no statistical evidence for this in this experimental protocol (either testing for distance from monitor or head pitch angle). However, we found clear evidence that the forehead marker height was statistically related to a range of engagement interactions (a representative trace is shown in Figure 3). Forehead height varied in the different segments (ANOVA, $P < 0.05$), and post hoc tests showed that there were statistically different mean forehead heights between the first half of the main stimulus and the second half of the main stimulus ($P < 0.05$, see Figure 4).

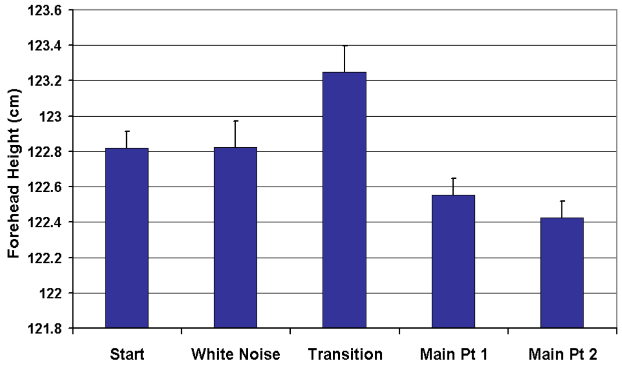

Figure 4. Comparison of mean forehead marker height during the different segments of all stimuli. St, WN and Tr are not significantly different from one another, but all three are significantly different from M1 & M2, and M1 and M2 are significantly different from each other.

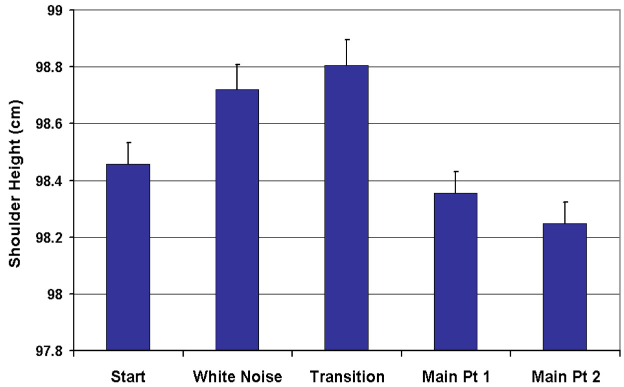

Figure 5. Comparison of mean shoulder (deltoid) marker height during the different segments of all stimuli. Significant differences are as per figure 2.

A similar (but not identical) relationship was observed for shoulder height (Figure 5).

Postural Changes during the Length of the Experiment

The known changes in the head and shoulder positions during long tasks (e.g. driving) associated with fatigue might explain the observations here – so we tested whether our experiments might elicit fatigue, and whether the main stimulus may elicit "micro-fatigue" between M1 and M2. We performed linear regressions of head (and shoulder) height based on the independent variables stimulus number and volunteer (because each volunteer is a different height to begin with). See Table 1.

Model				
	F	**d.f.**	**Prob > F**	**R^2**
Forehead	46.6	14,147	0.0000	0.82
Shoulder	74.4	14,147	0.0000	0.88
Stimulus Number as Variable in Model				
	βcoef	t	P > |t|	95% CI
Forehead	-0.063	-1.08	0.28	-.1 / .05
Shoulder	-0.051	-1.35	0.18	-.1 / .02

Table 1. Linear regressions for forehead height and shoulder height as a function of stimulus number and volunteer.

Although the regression models had high R-squared values (showing the importance of volunteer), the results showed only a very weak decrease of these parameters with stimulus number (i.e. with time), which did not reach statistical significance; presumably this is because fatigue did not increase greatly during the 70 minutes of the experiments, which took place during normal daylight hours. The effect of stimulus number on forehead height and on shoulder height can be seen in figures 6 and 7, respectively. These graphs show that these parameters, when split into M1 and M2, consistently exhibit the slumping down between M1 and M2.

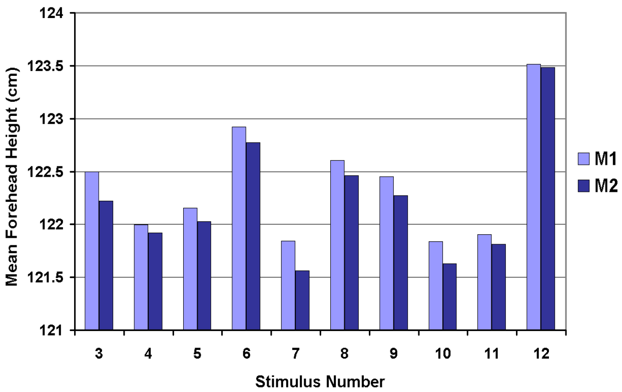

Figure 6. Comparison of mean forehead marker height in the first half of the main stimulus (M1) to the second half of the main stimulus (M2), by stimulus number.

Subanalysis comparing Interesting to Boring Stimuli

Sub-analyses were performed to determine whether the slump between M1 and M2 was caused by the boring stimuli or by the more engaging stimuli. In the sub-analysis with only the engaging stimuli, which compared the parameters during M1 to M2, forehead height was significantly affected ($P < 0.05$, paired T test) and shoulder height was highly significantly affected ($P < 0.01$); however, with the boring stimuli, only forehead height was affected ($P < 0.05$), and shoulder height was not ($P > 0.2$).

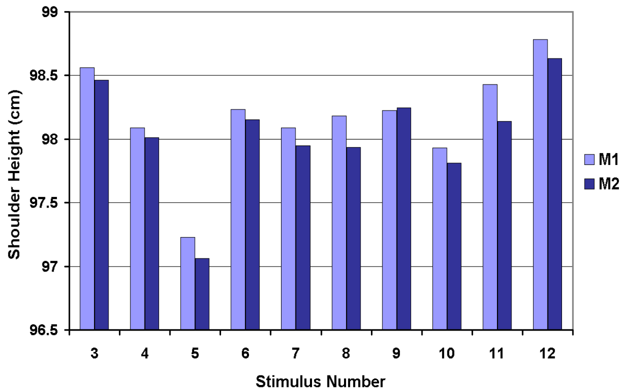

Figure 7. Comparison of mean shoulder (deltoid) marker height during the first half of the main stimulus (M1) to the second half of the main stimulus (M2).

CONCLUSIONS

Shoulder height and forehead height decrease during discrete, three-minute computer-presented stimuli. Discrete, homogeneous stimuli are of fundamental importance when designing experiments testing psychological issues, as they provide a foundation for validating systems. These slumping-related changes are statistically associated more with interesting than with boring stimuli; so, the slumping changes may be less connected with fatigue, and more related to "settling in" to a stimulus, indicating reduced arousal or vigilance. Discrete stimuli reveal the cyclical effects of boredom arising from discrete, homogeneous activities.

Although this research focused on establishing a procedural approach for quantifying engagement of seated volunteers with computer-presented stimuli, the results from this study will have clear repercussions on how researchers conduct further structured research into cognitive ergonomics and human computer interaction. These observations are likely to aid systems that use continuous postural monitoring to determine mental states such as boredom, relaxation and fatigue. Although initial research examining seated posture during human-computer-interaction explicitly measured head attitude as well as seat pressure [1, 21], subsequent studies have been focused on the more easily deployed chair-mat [7]. The current study presents highly significant statistical evidence that two parameters, mean head height and mean shoulder height, diminish subtly as the volunteer settles in to interesting stimuli. These parameters are unlikely to be redundant information when used in combination with a chair-mat, which is particularly insensitive to small head movements in the vertical axis.

ACKNOWLEDGMENTS

We gratefully acknowledge Dr. Paul O'Dette for the original idea about changing fatigue states with music.

REFERENCES

1. Balaban, C.D., Cohn, J., Redfern, M.S., Prinkey, J., Stripling, R., Hoffer, M. Postural control as a probe for cognitive state: exploiting human information processing to enhance performance. *International Journal of Human-Computer Interaction*, 17 (2004), 275-286.
2. Bradley, M. M. & Lang, P. J. The International Affective Picture System (IAPS) in the study of emotion and attention. In J.A. Coan and J. J. B. Allen (Eds.). *Handbook of emotion elicitation and assessment*. Oxford University Press, Oxford. (2005), 29-46.
3. Bull E. P. *Posture and Gesture*, Pergamon Press, Oxford, 1987.
4. Calvo, R. A. & D'Mello, S. Affect Detection: An Interdisciplinary Review of Models, Methods. *IEEE Trans. on Affective Computing*, 1 (2010), 18-37.
5. Coan, J. A., Gottman, J.A. The specific affect coding system (SPAFF). In J.A. Coan and J. J. B. Allen (Eds.). *Handbook of emotion elicitation and assessment*. Oxford University Press, Oxford. (2005), 267-285.
6. D'Mello, S. & Graesser, A. Automatic Detection of Learner's Affect From Gross Body Language. *Applied Artificial Intelligence*, 23 (2009), 123-150.
7. D'Mello, S., Dale, R., & Graesser, A. Disequilibrium in the mind, disharmony in the body. *Cognition & Emotion*, 26, 2, (2012), 362-374.
8. D'Mello, S., Graesser, A., & Picard, R. W. Toward an affect-sensitive AutoTutor. *Intelligent Systems* 22, 4, (2007), 53-61.
9. Ekman, P. The differential communication of affect by head and body cues. *Journal of Personality and Social Psychology*, 2, (1965), 726-735
10. Graesser, A. C., McDaniel, B., Chipman, P., Witherspoon, A., D'Mello, S., & Gholson, B. Detection of emotions during learning with AutoTutor. In *Proceedings of the 28th Annual Meetings of the Cognitive Science Society*, (2006), 285-290.
11. Gunes, H., & Piccardi, M. Creating and annotating affect databases from face and body display: A contemporary survey. In *Systems, Man and Cybernetics, 2006. SMC'06*. 3, (2006), 2426-2433.
12. Heitmann, A., Guttkuhn, R., Aguirre, A., Truschel, U. & Moore-Ede, M. Technologies for the Monitoring and Prevention of Driver Fatigue. *Proc. First Int. Driving Symp. on Human Factors in Driver Assessment Training and Vehicle Design*, (2001), 81-86.
13. Hernandez, J., Hoque, M. E., & Picard, R. W. Mood Meter: large-scale and long-term smile monitoring system. In *ACM SIGGRAPH 2012 Emerging Technologies* ACM (2012), 15.
14. Hoque, M. E., McDuff, D. J., & Picard, R. W. Exploring Temporal Patterns in Classifying Frustrated and Delighted Smiles. *IEEE Transactions on Affective Computing*, 3, 3 (2012), 323 - 334.
15. Kapoor, A., Burleson, W. & Picard, R. Automatic prediction of frustration. *International Journal of Human-Computer Studies*, 65 (2007), 724-736.
16. Kleinsmith, A., Bianchi-Berthouze, N., & Steed, A. Automatic recognition of non-acted affective postures. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, 41 (2011), 1027-1038.
17. LaFrance, M. Posture Mirroring and Rapport. In M. Davis (ed.), *Interaction Rhythms: Periodicity in Communicative Behavior*. Human Sciences Press, New York, (1982), 279-298.
18. Lavia, L., Easteal, M., Close, D., Witchel, H., Axelsson, O., Ware, M., & Dixon, M. Sounding Brighton: Practical approaches towards better soundscapes. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 2012, 11. Institute of Noise Control Engineering, Washington DC, USA (2012), 436-444.
19. Mackie, R.R. & Miller, J.C. Truck and Bus Driver Fatigue. US Department of Transport, Report HS 803 799, 1978.
20. Mehrabian A., Friar J. T. Encoding of attitude by a seated communicator via posture and position cues, *Journal of Consulting and Clinical Psychology*, 5, (1969), 330-336.
21. Mota, S., Picard, R. W. Automated Posture Analysis for Detecting Learner's Interest Level. Conf. on Computer Vision and Pattern Recognition, 5 (2003), 49-49.
22. Picard, R. W. Measuring affect in the wild. In *Affective Computing and Intelligent Interaction*. Springer, Berlin, (2011), 3.
23. Williamson, A. M., Feyer, A-M., Friswell, R. & Leslie, D. *Strategies to Combat Fatigue in the Long Distance Road Transport Industry. Stage 2: Evaluation of Alternative Work Practices*. Federal Office of Road Safety, Report CR 144, Australian Government Publishing Service, Canberra, Australia, 1994.
24. Witchel, H.J., Healy, A., Needham, R., Westling, C., & Chockalingam, N. Comparing four technologies for measuring postural micromovements during monitor engagement. *Eur. Conf. Cognitive Ergonomics*, 2012.
25. Wylie, C. D., Shultz, T., Miller, J. C., Mitler, M. M. & Mackie, R. R. *Commercial Motor Vehicle Driver Fatigue and Alertness Study*. Report No. FHWA-MC-97-002, Federal Highway Administration, US, 1996.