ABSTRACT
Boredom and disengagement metrics are crucial to the correctly timed implementation of adaptive interventions in interactive systems. Psychological research suggests that boredom (which other HCI teams have been able to partially quantify with pressure-sensing chair mats) is actually a composite: lethargy and restlessness. Here we present an innovative approach to the measurement and recognition of these two kinds of boredom, based on motion capture and video analysis of changes in head and shoulder positions. Discrete, three-minute, computer-presented stimuli (games, quizzes, films and music) covering a spectrum from engaging to boring/disengaging were used to elicit changes in cognitive/emotional states in seated, healthy volunteers. Interaction with the stimuli occurred with a handheld trackball instead of a mouse, so movements were assumed to be non-instrumental. Our results include a feature (standard deviation of windowed ranges) that may be more specific to boredom than mean speed of head movement, and that could be implemented in computer vision algorithms for disengagement detection.

Author Keywords
Motion capture; video analysis; engagement; interest; boredom; head movements, postural change; discrete stimuli.

ACM CLASSIFICATION KEYWORDS
HCI design and evaluation methods: Laboratory
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 and engagement. This interpretation has thus far remained controversial due to the complexity of cognition [9, 24, 25].

While detection of mental states and emotions “in the wild” is a goal of our team and of the work of others in this field [16, 20], a complementary and important approach is to use laboratory experiments to make clear and unequivocal links between elicited engagement and its corresponding nonverbal correlates [21, 1, 9], 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 [7].

**Multiple Causes for Postural Movement**

Interpreting the affective underpinnings of postural movements during human-computer interaction is complex because not all movements are affective. Our stimuli (Table 1) cover a range of affects, but the impetus for postural movements can be instrumental (e.g. use of the mouse) as well as non-instrumental (e.g. boredom or screen

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Baseline?</th>
<th>Type</th>
<th>Expected Actions Per Minute</th>
<th>Interesting / Boring (expected design goal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAV (user selects their favourite music)</td>
<td>No</td>
<td>Music without video</td>
<td>Passive</td>
<td>Interesting</td>
</tr>
<tr>
<td>OK (music video by OK Go)</td>
<td>Yes</td>
<td>Video with music</td>
<td>Passive</td>
<td>Interesting</td>
</tr>
<tr>
<td>AB (Angry Birds)</td>
<td>No</td>
<td>Commercial Video Game</td>
<td>20-25</td>
<td>Interesting</td>
</tr>
<tr>
<td>ZU (Zuma)</td>
<td>No</td>
<td>Commercial Video Game</td>
<td>50-60</td>
<td>Interesting</td>
</tr>
<tr>
<td>GQ (Geography Quiz)</td>
<td>No</td>
<td>Quiz with sound</td>
<td>3</td>
<td>Fairly Interesting</td>
</tr>
<tr>
<td>RSA (Royal Society of Art animated lecture)</td>
<td>Yes</td>
<td>Video with voice narration</td>
<td>Passive</td>
<td>Fairly Interesting</td>
</tr>
<tr>
<td>A5 (20 positive photos each lasting 6 seconds)</td>
<td>Yes</td>
<td>Photo Montage (silent)</td>
<td>Passive</td>
<td>Mildly Interesting</td>
</tr>
<tr>
<td>NQ (Nutrition Quiz)</td>
<td>No</td>
<td>Quiz (silent)</td>
<td>3</td>
<td>Boring</td>
</tr>
<tr>
<td>HTE (soporific video on sewage pipes)</td>
<td>Yes</td>
<td>Video with voices</td>
<td>Passive</td>
<td>Boring</td>
</tr>
<tr>
<td>IPSK (arousing photo of ski jumper shown for 2 mins)</td>
<td>Yes</td>
<td>Still Photo (silent)</td>
<td>Passive</td>
<td>Boring</td>
</tr>
<tr>
<td>BSc (Black Screen lasting 2 mins)</td>
<td>Yes</td>
<td>Screen remains black</td>
<td>Passive</td>
<td>Boring</td>
</tr>
<tr>
<td>VIO (incompetent solo violin music)</td>
<td>Yes</td>
<td>Music without video</td>
<td>Passive</td>
<td>Dislike</td>
</tr>
</tbody>
</table>

Table 1. Table of stimuli. Each stimulus lasts 175 seconds, and begins with 50 seconds of baseline “white noise and television snow” (except for stimuli where that is not possible, as shown in column “baseline”). Listings under “Interesting/Boring” represent the scientific team’s design target for the response of the participants.
disengagement). Causes for screen disengagement are listed in Table 2. The most obvious cause would be when no information is presented visually on the monitor; a subclass of minimal visual information would be when very little new information is presented visually (such as when a static image remains on screen for a long time). Break-taking is when the user acts as if the current information on the monitor temporarily requires less watchfulness; in a video game this may be during level changes (or listing of scores) between playing episodes, but it may also occur as time passes while watching a static photograph. Boredom or other negative affective states (e.g. hopelessness, fright, disgust) may also lead to monitor disengagement. Note that multiple causes for monitor disengagement and watchfulness may occur simultaneously.

<table>
<thead>
<tr>
<th>Disengagement</th>
<th>Watchfulness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-visual stimulus</td>
<td>Visual stimulus</td>
</tr>
<tr>
<td>Internal mentation</td>
<td>High content rate</td>
</tr>
<tr>
<td>Break-taking</td>
<td>Persistent new content</td>
</tr>
<tr>
<td>Boredom</td>
<td>Interest</td>
</tr>
<tr>
<td>Negative Affect</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Potential causes for monitor disengagement compared to causes of watchfulness.

Complexity of Boredom: Restless vs. Lethargic

Even if one restricts oneself to affective causes for postural changes, the ability to discriminate interest from boredom remains problematic, mostly because the definitions of boredom are somewhat conflicting [22]. Mikulas and Vodanovich [18] have defined boredom as ‘a state of relatively low arousal and dissatisfaction, which is attributed to an inadequately stimulating situation’, whereas for Barbalet [2] boredom is a state of high arousal: ‘Boredom, in its irritability and restlessness . . . is not a feeling of acceptance of or resignation towards a state of indifference’. According to one qualitative study of the phenomenon of boredom, “Feelings comprising the experience of boredom were almost consistently those of restlessness combined with lethargy.” [17] This implies boredom could be a high activity or low activity state (see Figure 1). Restless activity would include fidgeting or stunted efforts to get out of the chair. Lethargic boredom might be the viewer balancing his/her head sideways on the palm of their hand, such that if the hand was removed, the head would fall. A similar argument holds for engagement. Dynamic engagement could be a football fan raising his arms in celebration of a goal. Rapt engagement might be a child watching a favourite cartoon while sitting still with his/her mouth slightly open; rapt engagement often inhibits non-instrumental movements (NIMI) because large movements can interfere with focusing of gaze [25].

Measuring Two Kinds of Boredom And Two Kinds of Engagement

| Dependent on Stimulus (And Audience Preferences + Knowledge) |
|-------------|-------------|-------------|
| Interested | Bored       |
| Physically Active | Dynamic Engagement | Restless |
| Physically Still | Rapt Engagement | Lethargic |

Figure 1. Schematic simplification of how both engagement and boredom can be characterised as either physically active or physically still.

Our research question is: if boredom can be physically active or physically still, and engagement can be physically active or still, how might it be possible to use movement parameters as an indicator of boredom or engagement?

METHODS

Experimental Volunteers

Twenty-nine healthy volunteers (4 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.

Protocol

The complete methodological description can be found in [24, 25]. 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. Participants experienced audiovisual stimuli, each lasting 170 seconds, and then rated the experience via a subjective questionnaire.

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.

Stimuli and Subjective Rating Scales

Stimuli were a collection of games, film excerpts, quizzes, and musical excerpts as described (see Table 1 and [23]).
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. An example of the subjective responses of this cohort is shown in Figure 2.

### Statistics and analysis

Analyses of stimuli were performed by breaking each stimulus into time segments and removing the transitions at the start, the end and the baseline white noise, with automated collection of 80-second segments [24]. Positions were calculated as the mean of each uni-dimensional parameter for the segment listed. Speeds (in cm/min) were calculated as the absolute value of the difference between two adjacent time points; all speeds are reported with respect to sampling frequency (here 25 Hz), as increasing sampling will increase the apparent speed.

### RESULTS

The postural changes elicited by two passive stimuli (i.e. films that did not involve interaction) were compared in terms of position and movement of the head and shoulders. One stimulus was boring and one was not boring: IPSK (a photograph that is left unchanged on the monitor for 2 minutes) and OK (an internet classic of OK Go’s “This Too Shall Pass, Rube Goldberg version” music video). Although the photograph in IPSK is very interesting (it is photo 8030 from the International Affective Photographic System [3], which has the highest mean rating for arousal in the entire IAPS), seeing it for two minutes is too long and ultimately boring (mean VAS rating for boring = 77.4 ± 5.9).

It is axiomatic in human-to-human dyadic communication that increased interest and engagement is accompanied by proximity and approach [6]. However, in this human-computer interaction, engagement was not sufficient to change the mean positions of the head to monitor, shoulder to monitor, and shoulder height – all of which were not statistically different during engagement and boredom (P > 0.5 for all); these results for spatial positions are concordant with previous HCI measures [19]. The mean positions of head height were borderline significant (P < 0.1); previous studies have not agreed on whether head height position changes between interesting and boring films: an HCI study of head position found no difference [19], but a film study with manual coding did find that head height dropped during boring films [4]. Average speeds of the movements were also compared (see Table 3); the shoulder speeds were significantly different, but the head speeds were not.

### 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 [23]. 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 headband as a set of four (left front head, right front head, left back head, right back head). 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.

![Boxplot of VAS ratings for I was bored](image)

**Figure 2.** Subjective Visual Analogue Scale ratings for “I was bored” 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 (red plus signs) 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.

### Table 3. Mean speed of body markers comparing an engaging (OK) to a boring (IPSK) stimulus. NS, not significant (P > 0.1).

<table>
<thead>
<tr>
<th>Movement Type</th>
<th>P</th>
<th>OK (cm/min)</th>
<th>IPSK (cm/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forehead to Monitor</td>
<td>NS</td>
<td>23.2</td>
<td>30.8</td>
</tr>
<tr>
<td>Shoulder to Monitor</td>
<td>&lt; 0.01</td>
<td>10.8</td>
<td>20.6</td>
</tr>
<tr>
<td>Forehead Height</td>
<td>NS</td>
<td>15.7</td>
<td>19.5</td>
</tr>
<tr>
<td>Shoulder Height</td>
<td>&lt; 0.01</td>
<td>10.9</td>
<td>13.4</td>
</tr>
</tbody>
</table>
Figure 3. Representative motion tracking data. The left panel shows forehead marker distance from the screen (mm) (sampled at 25 Hz) for an interesting passive stimulus (OK, left) vs. a boring passive stimulus (IPSK, right) for one volunteer (Y023). Subjective (VAS) ratings for this volunteer are shown at the lower left of each trace. We propose that the spike-and-flat morphology of the trace for IPSK is typical of restlessness punctuating lethargy that occurs during many boredom episodes, while the slow downward ramp during OK is not relevant for either type of boredom.

The Structure of Movements during Boredom
To better understand the lack of statistical difference in head to monitor movement between the interesting and boring film, a pair of representative traces for an engaging passive stimulus and a boring passive stimulus are shown in Figure 3. Both time series have a similar total amount of movement (which would result in similar mean speed measurements), but the structures of the movements are different. The range of the movement during boredom is larger, and the movements tend to be large sudden movements interspersed with long (> 5 seconds) periods of stillness; by contrast, movements during interest are smaller and less spiky, but they are more pervasive. We suggest that the spike-and-flat morphology seen during boredom may relate to the mental states of restlessness and lethargy. During interest, by contrast, the small movements may be instrumental movements required by gaze, while larger movements (bodily adjustments for comfort) may be prevented by Non-Instrumental Movement Inhibition (NIMI) [25].

Non-Instrumental Movement Inhibition by Gaze
Another potential problem with using head speed as an indicator for disengagement is that head speed may be strongly influenced by the visual import of the stimuli. When a stimulus includes visual cues on the computer screen, it results in diminished head movement (due to the head being kept steadier in order to look at the images). We demonstrated this by comparing head movements elicited by non-visual musical stimuli to filmed stimuli. We found that our non-visual stimuli elicited significantly more (> double the amount of) forehead-to-monitor movement (i.e. average speed = 52.1 cm/min) compared to visual films of similar interest value (speed = 24.7 cm/min, P < 0.01, N = 33). Likewise, for forehead height speed, music nearly doubled the speed compared to films (see Figure 4, P < 0.001, N = 33). In the paired plot, each line represents one volunteer; blue lines (triangles) are when the average head height speed during non-visual music is higher (for that volunteer) than the average head height speed for the matched film, while pink lines are where head speeds during the film are greater than during non-visual music. The black horizontal lines are the group mean values.

Figure 4. Head height mean speed elicited by music (non-visual) vs. film. An interesting musical stimulus without visual accompaniment (FAV — the participant listens to their favourite music with a strong beat) is compared to an interesting film (OK Go’s music video) and a boring/irritating musical stimulus (VIO) is compared to a dull film (HTE). The mean values are significantly different (Paired t Test, P < 0.001). Speed units shown are cm/min at 25 Hz.
The Effects of Break-Taking on Head Movements

Another limitation of using head to monitor speed as an indicator of engagement/boredom is that break-taking has a profound influence on head movement. Break-taking is one of the most obvious postural responses seen during human-computer interaction, and it was described as one of the six basic “affective states” that can be recognised by human coders watching video-gameplay [19]. From the perspective of machine measurement of posture, head distance from the monitor during the transition from break to gameplay was one of the two postural movement patterns that were identified by the DARPA Augmented Cognition Technical Integration Experiment [1, 21].

We found strong evidence that one interesting stimulus (GQ), which is a quiz, has a mean speed that is comparable to boring or non-visual stimuli (FAV). This is likely to be due to break-taking, which occurs every 20 seconds on this timed quiz [25].

\[
R_t = \max(X_1, X_2, ... X_t) - \min(X_1, X_2, ... X_t)
\]
for \( t = 1, 2, ... n \)

\[
\sigma = \sqrt{\frac{1}{N} \sum_{t=1}^{N} \left( R_t - \left( \frac{\sum R_i}{n} \right) \right)^2}
\]

As an example, when the two head-to-screen time courses in panels A and B from Figure 3 are compared, the speeds are 17.2 (OK) and 26.2 (IPSK) cm/min, respectively (i.e. the speed feature for the boring IPSK is 50% faster than for the interesting OK); the standard deviation of ranges feature is 0.56 for OK and 1.36 for IPSK (i.e. the boring IPSK is 2.4 × OK).

Using this feature on all the volunteers, a significant difference in head to monitor standard deviation of 2-second ranges can be detected between the passive stimuli IPSK and OK (Figure 6, \( P < 0.05, N = 25 \)).

Figure 5. Mean speed of forehead to monitor speed elicited by each stimulus. Box plot symbols as per Table 1. High speeds are elicited not only by boring stimuli, but also by non-visual stimuli (FAV) and stimuli with many break-taking episodes (GQ). There is a dashed line at 25 cm/min for comparison.

A Feature to Detect Occasional Large Movements such as Those Elicited by Boredom

The challenge in detecting rare large movements interspersed with inactivity is that two factors are being sought: spikes and inactivity. Measuring speed is limited because it is equally sensitive to ramps (Figure 3, OK, left panel) and spikes (Figure 3, IPSK, right panel), so long as they are the same height. One approach for a new feature to detect this is to break the time series into windows relevant to human movements (2 seconds) and to determine the standard deviation of the range of movement in each window:

Figure 6. Standard deviation of 2-second windows of ranges of head to monitor distance elicited by passive stimuli OK vs. IPSK.

A similar result can be found when comparing engaging vs. boring interactive stimuli (see Figure 7). This analysis shows that people playing Zuma (ZU), a high actions per minute commercial video game based on vigilance, move less than those interacting with a slow (3 actions per minute), boring nutrition quiz (NQ). Thus, despite requiring more player interactions, the commercial video game elicits less variation in postural movement; this is a clear example of Non-Instrumental Movement Inhibition (NIMI). Note that while the standard deviations of ranges are significantly different, the speeds are not significantly different (\( P > 0.1 \)). For ZU the mean speed was 25.9 cm/min, while for NQ it was 32.1 cm/min; compare this to the means of the standard deviation of 2 second windows of head-to-monitor distance, which were: 0.39 (ZU) vs. 0.54 (NQ).
3. We gratefully acknowledge Thomas Montgomery Newman for the original idea about different kinds of music.

REFERENCES


