Participations
Journal of Audience & Reception Studies

Volume 11, Issue 1
May 2014

What does not happen: Quantifying embodied engagement using NIMI and self-adaptors

Harry J. Witchel,
Brighton and Sussex Medical School, UK

Carina E.I. Westling,
University of Sussex

Julian Tee,
Brighton and Sussex Medical School, UK

Aoife Healy,
Robert Needham,
Nachiappan Chockalingam,
CSHER, University of Staffordshire

Abstract:
Previous research into the quantification of embodied intellectual and emotional engagement using non-verbal movement parameters has not yielded consistent results across different studies. Our research introduces NIMI (Non-Instrumental Movement Inhibition) as an alternative parameter. We propose that the absence of certain types of possible movements can be a more holistic proxy for cognitive engagement with media (in seated persons) than searching for the presence of other movements. Rather than analyzing total movement as an indicator of engagement, our research team distinguishes between instrumental movements (i.e. physical movement serving a direct purpose in the given situation) and non-instrumental movements, and investigates them in the context of the narrative rhythm of the stimulus. We demonstrate that NIMI occurs by showing viewers' movement levels entrained (i.e. synchronised) to the repeating narrative rhythm of a timed computer-presented quiz. Finally, we discuss the role of objective metrics of engagement in future context-aware analysis of human behaviour in audience research, interactive media and responsive system and interface design.

Keywords: Digital representation, audience metrics, screen engagement, interactional synchrony, entrainment, affordances, motion capture, non-instrumental movement inhibition, NIMI

1. Summary

There is a complex relationship between engagement in its embodied intellectual and emotional manifestations and its associated non-instrumental physical responses. Here are two contrasting examples of engagement while watching television: a football fan watching his team score a goal will manifest engagement by raising his arms and shouting, while a child watching a fascinating cartoon will sit absolutely still and remain silent with his mouth hanging slightly open: both are engaged and entrained to the stimulus, but in the first case this is expressed with overt physical excitement, and in the second with rapt attention. In order to clarify the link between engagement and non-instrumental physical activity, we used a range of motion capture and video analysis techniques to characterise the movement patterns of seated experimental participants during interaction with a set of discrete screen-based, three-minute-long audio-visual stimuli with a matrix of divergent levels of A) interactivity, and B) intrinsic interest. Our data show that cognitive engagement as an embodied phenomenon can attenuate certain types of non-instrumental movements (larger postural movements and self-adaptors, i.e. limb or facial movements with no functional purpose). We propose that Non-Instrumental Movement Inhibition (NIMI) is an entrained instrumental response allowing for fuller engagement with audiovisual stimuli (especially when seated), and that non-instrumental movement disinhibition (e.g. 'break taking') can be either an indicator of engagement (when entrained, e.g. expressed during breaks between rounds or natural pauses) or disengagement (when disruptive and occurring during active parts of the presentation cycle).

Furthermore, we propose that when a viewer is entrained to the narrative rhythm of the stimulus, this is a proxy for media engagement (as explained in section 2.4.2). Intrinsic to these proposals is the relevance of context to attempts at quantifying human experience, in particular contextual awareness of the nature and narrative rhythms (e.g. active parts and pauses or call-and-response patterns) of the stimuli. We conclude that metrics for engagement should not only include actions, but also the lack of actions as important surrogates for emotional and intellectual aspects of engagement.

2. Introduction

2.1 Digital representation and audience behaviour research

Remote audience metrics based on data harvesting via distributed network technologies are currently developed in support of commercial and security interests (Berry, 2014). Analysis methods tend to promote reductive categorisation of users due to a combination of the procedural nature of computer processes (Bogost, 2007: 11-14) and the interests of the organisations that harvest and process such metrics. Of prime importance to commercial (and also security) interests are what people *do* (i.e. 'positive action'), e.g. purchase, share, register, travel in physical space: reflective absorption and stillness are of little interest, yet fundamental to human emotional and intellectual life and culture. As distributed networks become ever more pervasive in public and everyday life, the influence of analysis methods on how we are represented digitally, the types of behaviours that are included in the digital representation of human reality, and the potential impact of digital representation on cultural, political and economic development become more critical (Couldry, 2013). Our team endeavours to develop audience research and audience behaviour analysis methodologies to support more fine-grained, real-time appreciation of audience behaviours for the purpose of a) integration in responsive educational and assistive systems and b) responsive systems in art and design, and finally, c) contributing to more contextual and diverse digital representation beyond commercially driven objectives. In addition, we take the position that overt, rather than covert, audience research methodologies for the explicit purpose of enhancing audience/user engagement in these fields toward a more human-centred development can, in the longer term, contribute to broadening the discourse and awareness of both positive and negative potential in user data gathering and application.

In the education sector, automated capture of non-verbal learner responses offers new possibilities for real-time assessment of student engagement with interactive learning resources, for analysis against long-term retention of information (Bulger et al, 2008). Some researchers have built responsive learning systems, e.g. for the teaching of physics (Auto-tutor, D'Mello and Graesser, 2009). Such automated teaching systems 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.

In the cultural sector, we see an increase in collaboration between art practitioners and HCI researchers in order to integrate real-time audience behaviour data into responsive systems. This is well-established in installation and interactive art, where audience actions in response to the interface produce aesthetic and/or narrative changes. (Bilda et al., 2008; Edmonds et al., 2009; Gonsalves, 2009). More lately, producers of interactive art have shown interest in measuring internal states in the audience in search of a new intimate aesthetic through mapping the relationship between physiological processes, bodily sensations and perceived subjectivity (Loke and Poonkhin Khut, 2014). The turn toward incorporating cognitive audience responses in interactive art revitalises the concept of engagement as not just a physical, but also an intellectual and emotional phenomenon, where these three dimensions of engagement combine in sustained attention to, and interactions with, a target stimulus or activity. Thus, there is momentum for gathering objective audience response data for a deeper understanding of audience engagement in performance arts (Latulipe et al., 2011).

The measurement of engagement using sensors and motion analysis is gaining adherents as a potential technology for measuring deep engagement without the limitations of video-ethnography (Latulipe et al., 2011; Witchel et al., 2013b; D'Mello et al., 2012). The use of sensors and motion analysis is fundamentally an interdisciplinary topic of research: the cultural significance and narrative structure of the stimulus spans the arts and humanities, the complex nature of the human responses occupies the human sciences, and the possibilities and limitations presented by the sensor technologies enjoins engineering. As interdisciplinary research becomes more prevalent, the opportunities for making cultural meaning from these applied technologies proliferate.

2.2 Cognitive engagement with audio-visual stimuli: context and complexity

Here we propose a perspective on the quantification of non-verbal responses that allows for the assessment of embodied intellectual and emotional engagement, as well as physical engagement, in relation to the narrative rhythm of the stimulus. While a more comprehensive and context-sensitive approach to measuring audience engagement presents a complex challenge in terms of analysis and interpretation of data, it broadens the relevance of insights gained to include audience experiences that are not normally categorised as interactive, e.g. visual arts, music or non-interactive audio-visual stimuli (e.g. films and television) that require less physical interaction, but significant amounts of emotional and/or intellectual forms of engagement. One way to provide more fine-grained measures of embodied engagement is by attending to real-time or moment-to-moment (e.g. physiological) responses to audio-visual stimuli (Latulipe et al., 2011; Witchel et al., 2013a), and visual and audio stimuli (Witchel et al. 2013b). By approaching engagement as a moment-by-moment contextual phenomenon, it is hoped that the impact of cultural and media artefacts and practices can be more robustly determined and understood for the purpose of integration into interactive systems, and that more nuanced responsive interactive systems can be developed.

2.3 Boredom and interest: parameters explored in previous research

Movement parameters that have previously been explored as non-verbal proxies for engagement include proxemics and velocity, which we will discuss below. While both have yielded statistically significant results in specific experimental situations, neither has so far proved consistent as indicators of engagement, particularly cognitive engagement, when applied outside of the limitations of the experimental situations in which they were observed. In the concluding section of our introduction (2.4), we will introduce an alternative parameter that we propose, by virtue of movement being *absent* (Witchel, 2013c), as a more holistic proxy for cognitive engagement.

2.3.1 Proxemics

It is generally accepted that proximity of body position (and posture) can act as a fundamental indicator of dyadic engagement (Coan and Gottman, 2007). When

Volume 11, Issue 1
May 2014

characterizing engagement with an audiovisual stimulus, Bull (1987) found an association between engagement and viewers leaning forwards towards the stimulus. This association has however been disputed, with Mota and Picard (2003) asserting that position alone could not be used to distinguish between engagement and boredom. In our own research, comparing the average distance from the monitor during engaging vs. boring stimuli does not result in any statistically significant differences, possibly because the dual nature of boredom during seated interactions with a video monitor: there are two radically different boredom positions. Panels A and B of **Figure 1** show two images of the same experimental

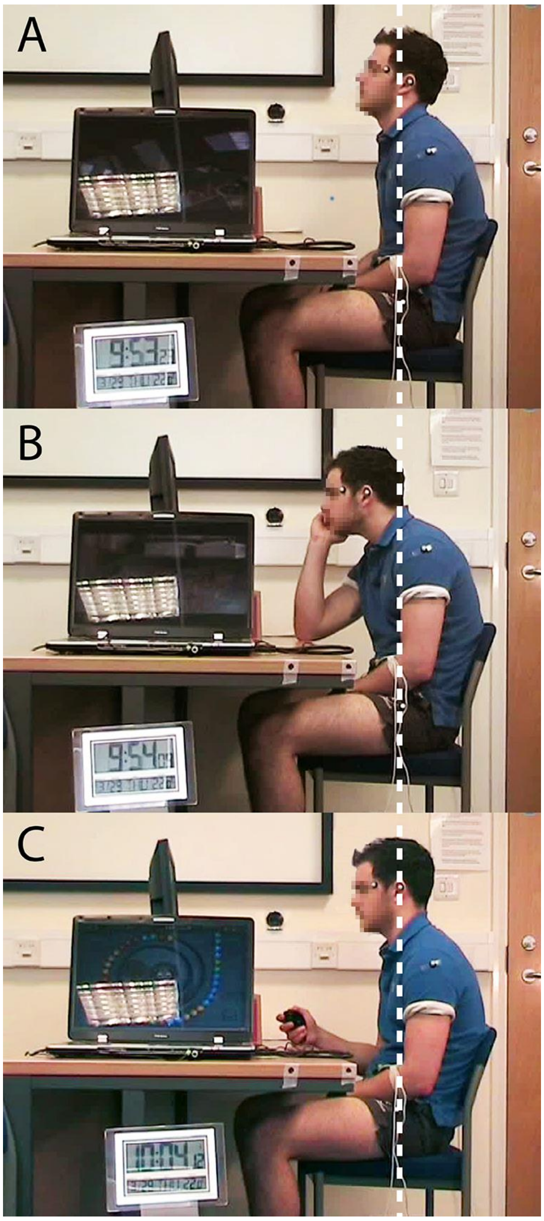

Figure 1: Boredom with a stimulus can elicit head movements either away from the monitor or towards the monitor. Panels A and B show the same volunteer interacting with the same boring stimulus (SPi), separated in time by less than one minute; the upper panel (A) is the earlier attitude. In panel C (lower) the same participant is shown sitting up straight (i.e. normally) while playing an interesting video game (ZU, Zuma). A white dashed line is shown for alignment. Reflective motion capture markers can be seen as described.[2]

participant spontaneously interacting with a very boring stimulus SPi (a film about the manufacture of sewage pipes in which the soundtrack is in Estonian); in the upper panel (A) participant Y022 is tilting his head back as if rejecting the stimulus, while the middle panel (B) shows the same participant during the same boring stimulus less than one minute later, balancing his chin on his palm ("load bearing" his head, a form of surrender), such that his head is much closer to the monitor than it would be when he is interested and sits up straight (panel C).

2.3.2 Velocity

Another suggested criterion for engagement has been increased movement velocity, with Bianchi-Berthouze et al. (2007) even going so far as to describe movement as the *essential* component of engagement. However, this conclusion too is problematic, for measurements of increased movement have been contrastingly associated with frustration, loss of interest and boredom (Kapoor et al., 2007; Mota and Picard, 2003; D'Mello et al., 2007), suggesting engagement can also correlate with episodes of low movement.

The fact that engagement with screen-based content can correlate with either high or low amounts of movement is intuitively clear, and can be highlighted by contrasting the rapt engagement of someone attending to a favourite television drama show (with a still body) compared to the very physical engagement of someone working out with an exercise video. We feel that this difference has not been sufficiently appreciated in recent attempts to measure engagement, and to the best of our knowledge, a detailed exploration of the way this dichotomy can be handled has not been put forward. If episodes of engagement can be contrastingly manifested as either large movements or inactivity, we must ask how we can understand engagement in terms of movement velocity at all. To do this, it is necessary to determine precisely the kinds of movements involved in these forms of engagement, and particularly the movements that are seen to diminish, which we shall focus on. It is only by accomplishing this, we propose, that reliable postural measures of engagement as a whole will be derived.

2.4 Non-instrumental movements: self-adaptors and postural micromovements

Rather than focusing exclusively on the presence of general movement as an indicator of engagement, our research team distinguishes between instrumental movements (serving a direct purpose in the given situation) and non-instrumental movements, and investigates them in the context of narrative rhythm (explained in section 2.4.2). Non-instrumental movements are those movements that are not instigated by deliberate activity resulting from the task or event at hand (i.e. engaging with the stimulus); examples of instrumental and non-instrumental activities are shown in **Table 1**. Self-adaptors (also known as manipulators) are a subset of body movements (especially arm/hand movements) that are non-instrumental, i.e. they emanate from cognitive states (e.g. emotions, boredom, fatigue) rather than from deliberate, goal-directed activity (Ekman and Friesen, 1969; Allen and

Honeycutt, 1997; Caso et al., 2006; Waxer, 1977). Self-adaptors include postural micromovements, fidgeting, and face touching, and are thought to be associated with cognitive and emotional states ranging from discomfort or anxiety (Ekman and Friesen, 1972; Ekman and Friesen, 1975; Allen and Honeycutt, 1997; McCroskey, 2009) and depressive states (Waxer, 1974; Waxer, 1976), to boredom (Niemeyer and Dirven, 1997, 290). Examples of self-adaptors include rubbing the back of one's neck when annoyed ("pain in the neck" gesture), rubbing one's eyes in response to an unpleasant thought ("I'd rather not see this in my mind's eye"), and thumb-sucking when anxious. They are called self-adaptors because they are proposed to allow the individual to adapt to internal affective states by making at least partial movements in response to real or imagined externalities. For example, Ekman and Friesen (1969) characterise leg fidgeting as a stunted "abortive flight movement", revealing a subject's stressful disposition in situations, particularly in social deception. Psychologists accept that relatively little is scientifically known about how self-adaptors (and related biomechanical parameters) are encoded, precisely because they are purposeless and outside of conscious awareness:

> Despite years of work, however, we still do not know how body movements are changed quantitatively when an individual experiences an emotion. (Gross et al., 2010)

The lack of progress in research of this nature may be the reason for the scarcity of more recent literature on the subject, and the focus on linking specific actions with internal states may be part of the reason that there has been little progress. As an alternative approach, we suggest that such non-instrumental behaviours more broadly are precipitated by, or associated with, a low engagement state (e.g. you cannot rub your eyes and be fully engaged with a video game at the same time), and propose that rapt engagement results in people diminishing their non-instrumental movements, but not their instrumental movements — a phenomenon we term Non-Instrumental Movement Inhibition (NIMI).

Instrumental Movements	Non-Instrumental Movements
Eye movements to see another part of the screen	Fidgeting
Rotating head to see another part of the screen	Break-taking (temporarily leaning back)
Leaning in to see a small object on the screen	Face touching, scratching and other self-adaptors
Operating mouse with hand	Facial expressions

Table 1: Examples of instrumental and non-instrumental movements that are performed by seated volunteers while interacting with computer-based stimuli.

2.4.1 Differentiation of instrumental and non-instrumental movements in context

To reliably measure this form of engagement requires that experimenters distinguish instrumental movements, resulting from intentional interaction with the stimuli, from non-instrumental movements, including self-adaptors. While previous empirical attempts to quantify engagement with interactive media by measuring embodied responses have been complicated by apparently idiosyncratic and inconsistent effects of engagement and boredom on net movement, we believe this is primarily a result of overlooking the above distinction.

As yet, machine recognition of nonverbal cues is not sufficiently advanced to reliably differentiate instrumental from non-instrumental movements, but progress on this front is being made. Mahmoud et al. (2013) suggest that a range of self-adaptor movements associated with psychological distress, including self-grooming, fidgeting gestures, and rocking motions, share a common repetitive rhythmic motion. By detecting rhythmic movements, they were thus able to make progress in detecting purely non-instrumental movements, but manual categorization of movements currently remains the most effective means of accomplishing this.

2.4.2 Interactional synchrony and entrainment to the rhythm of an external stimulus

When an organism synchronises itself to the external rhythm of a stimulus, consciously or unconsciously, it is said to be *entrained* to that target. Examples of deliberate entrainment include dancers who synchronise themselves to the music, as well as two people shaking hands with each other. Subconscious interactional synchrony with another person (e.g. when two people spontaneously make a gesture at the same time) manifests across broad biological and social levels, and has long been a source of fascination to both scientists and the general public (Chartrand and Bargh, 1999; Morris, 1977); other names and misnomers for this phenomenon include rapport, postural mirroring, and the chameleon effect, and it has been associated with liking, pro-social behaviour and affiliation with other persons (e.g. Chartrand and Bargh, 1999; Lakin et al., 2003).

Entrainment (and interactional synchrony) can encompass two people making very different actions occurring at the same times (e.g. the orchestra conductor waves his baton while the tuba player blows air into his instrument); the term interactional synchrony implies both agents have agency, while entrainment can imply an agent who follows a leader (e.g. a person seen on television) or a leading stimulus (e.g. recorded music) that is not responding to the follower. Entrainment to a stimulus (e.g. exercising to music) can be associated with improved performance in physical tasks (Thornby et al., 1995; Karageorghis et al., 2009). Entrainment to a stimulus that lacks a regular rhythm (e.g. to another person who is not dancing) has been difficult to describe mathematically; in psychology experiments interactional synchrony has been manufactured by having experiments with two volunteers, one of whom is secretly a confederate of the experimental team and has been instructed to copy the gestures and postures of the “real” experimental participant (Chartrand and Bargh, 1999).

Spontaneously-occurring subconscious synchronisation between people has been difficult to measure objectively; recently Ramseyer and Tschacher (2011) successfully employed an automated video analysis algorithm to capture interactional synchrony and demonstrate its correlation with relationship quality in psychotherapy situations.

Entrainment to a televised stimulus might involve gazing at the television during the sitcom and leaving to make a coffee during commercial breaks; so long as the person keeps coming back for the show, they are entrained to the stimulus even when not in the room with the stimulus. This narrative structure of commercial television programming (e.g. sitcom, advertisements, sitcom, advertisements, next show) has a profound effect upon the gross observable behaviour of the television audience, and this gross behaviour is radically different from audience behaviour in a cinema. In a video game, the narrative structure might be activity/task, information about scoring, next level, activity/task, failure at task, "death", game over. If the activity segments of a video game have strict time controls (e.g. each activity segment lasts 60 seconds), the narrative structure of the game can be said to have a narrative rhythm, and this narrative rhythm provides a mathematically tractable framework for analysing entrainment by automated methods.

The fundamental problem with using measurements of activity as a metric of entrainment to a stimulus is that the experimenter measures total movement, including both instrumental and non-instrumental movements. This problem is even true for mouse actions; for example, we have observed a common behaviour for restless video game-players is to move the mouse around the screen when there are plainly no designed affordances to click on. Thus, there is a difference between the expected/designed narrative rhythm (based on affordance design), and the perceived narrative rhythm (evidenced by the actions of the game-player); the expected/designed narrative rhythm is the experimental variable that we have manipulated here, and we hypothesize that the expected narrative rhythm will influence the behavioural rhythm of the experimental participants, but at times their behaviour will inevitably depart from our designed narrative rhythm. One possible way to measure entrainment is how much the behavioural rhythm of our stimulus viewers departs from the designed narrative rhythm of the stimulus.

While entrainment is often conceived as mere synchronization of activity, a full apprehension of its properties requires the recognition that entrainment is not just about moving at the right time, but also about *not moving* at the right time, in order to be fully responsive to the narrative rhythm of the context. We would therefore expect the Non-Instrumental Movement Inhibition phenomenon to feature prominently during entrainment while seated, and that engagement with dynamic stimuli would manifest as a cycling between Non-Instrumental Movement Inhibition (rapt engagement) and non-instrumental activity (e.g. break-taking when changing levels during a video game). Breaking-taking during video games is a notable feature of game-player behaviour, and typically involves moving away from the screen by a few centimetres or tilting the head backwards temporarily (Mota and Picard, 2003; Balaban et al., 2004).

In order to develop our understanding of the complex relationship between engagement and non-instrumental embodied responses, we have thus used a range of motion capture and video analysis techniques along with manual categorisation of movements to analyse the movement patterns of experimental participants during interaction with a set of discrete audio-visual stimuli, in an attempt to verify whether engagement can diminish non-instrumental movements. We seek to test whether it is possible in principle to identify more reliable movement-based measures for engagement. Furthermore, we are testing whether Non-Instrumental Movement Inhibition, as an instantiation of embodied engagement, is not simply a standalone phenomenon, but also a crucial component of the larger phenomenon of *entrainment* that occurs when individuals engage cognitively with dynamic stimuli.

3. Methods

3.1 Experimental volunteers

Nineteen healthy volunteers (3 female, age range 19-62, m ± sd: 28.2 ± 14.0) 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.

3.2 Protocol

After being briefed as to the nature of the study, participants were seated 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 175 seconds, and then rated the experience via a subjective questionnaire. For interactive computer games and quizzes, participants used a handheld trackball that could be operated from any position, thus vouchsafing that no movements detected were instrumental movements required by operation of a mouse.

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.

3.3 Stimuli and subjective rating scales

Stimuli were a collection of interactive games and quizzes, and non-interactive film excerpts and musical excerpts as described by Witchel et al. (2012). 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. The stimuli and their features are listed in **Table 2**. 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/hear 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 distributions of subjective responses of this cohort is shown in **Figure 2**.

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

Table 2: 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. Mean subjective responses from the participants (see Figure 2) have supported the accuracy and success of these designs. “Expected Actions Per Minute” estimates the actions per minute (i.e. clicking of trackball/mouse) for a typical user; quizzes made by our team using Flash have precisely 3 APM by design[1].

3.4 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 Vicon motion capture system (at Staffordshire University). We have previously shown that these two technologies produce comparable results for head attitude and for small translational movements from the lateral aspect (Witchel et al., 2012). Passive reflective markers were positioned on the head, shoulder, and middle of the outer thigh (**Figure 3**), at standardised anatomical positions.[2] The outcome parameters were head pitch (relative to floor), front head marker from screen, front head marker from floor, shoulder marker from screen, shoulder 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.

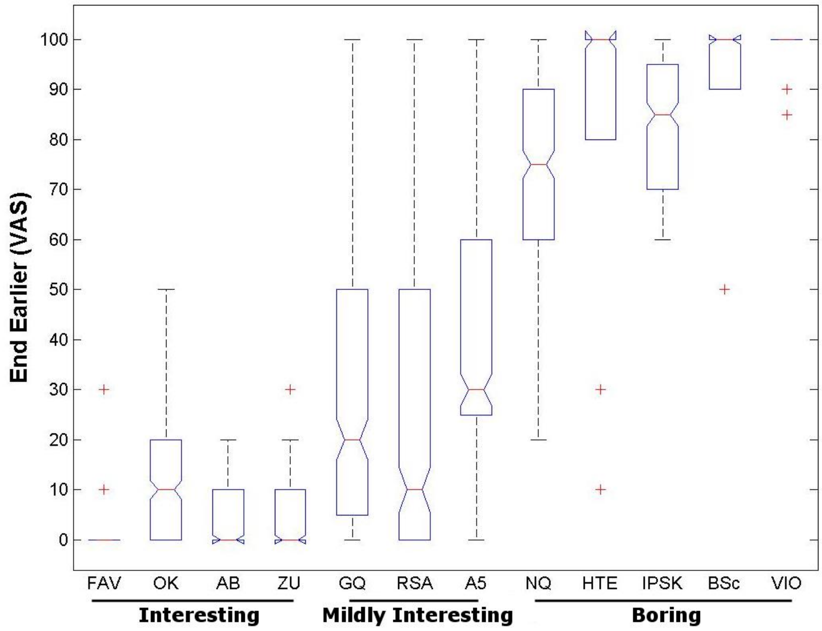

Figure 2: Subjective Visual Analogue Scale ratings for "I wanted it to end earlier" for stimuli in the "interesting", "mildly interesting" and "boring" groups of stimuli. VAS anchors are 0 = "not at all" and 100 = "extremely".[3]

Figure 3: Motion tracking of a participant during a spontaneous self-adaptor (face touch) seen from the lateral aspect. The participant has just answered a difficult question during a geography quiz (GQ), and is currently responding to the feedback to his answer. Tracking points for the front of the head, the ear, and the shoulder are shown as black/white circles, with one second of motion tracking shown in blue.

3.5 Statistics and analysis

Results for the manual self-adaptor count were gathered as follows. A standardised time period (175 seconds) for each stimulus was selected from videos of the volunteers, and these were inspected manually. Tallies were kept for observed self-adaptors that were coded as one of the following: total positional change, upper limbs, lower limbs, head and face. These movements were categorised again into large isolated movements, small isolated movements, and consistent movements (movements made repeatedly over a five-second period).

Average speed calculations for movements were calculated in Matlab as follows. Positions as 25 Hz time course data were low-pass filtered, and differences between consecutive time points were calculated (i.e. velocities in cm/frame). The absolute values of the velocities (i.e. moment-to-moment speeds) were averaged, and then standardised 80-second selections from the middle of the time course were selected by the algorithm listed previously (**Figure 2** from Witchel et al., 2013a), and from this selected time course a mean speed was calculated. Speed calculations resulted in very small values that were multiplied by 10,000 for ease of illustration; the units are cm per frame or degrees change per frame. Head pitch angle time courses were calculated using standard anatomical landmarks[4]

Mean normalised ranges were calculated in Matlab.[5] All statistics reported here are paired T tests calculated in Matlab. Time course graphs were created in Microsoft Excel, and paired plot graphs and box-and-whisker plots were made in Matlab.

4. Results

4.1 Engagement manifestations: representative examples of high and low movement

The diversity of stimuli we presented to participants ranged from highly engaging stimuli that offered many on-screen affordances (i.e. high expected actions per minute with exacting requirements for trackball activity) to experiences with no instrumental engagement with the screen (listening to one's favourite song while the screen remained blank). Although both engaging stimuli, they resulted in very different head movement responses (representative data shown in **Figure 4**); listening to one's favourite music with a strong beat (while the screen remained black, "FAV" in **Table 2** and **Figure 4A**) elicited in this participant many large regular movements of the head, which were in time with the music ("head-banging", see **Figure 5**). Note that when considering the entire cohort of participants, the head movements in response to music were not all of the sort seen in **Figure 4A** and **Figure 5**: some participants tapped their hands, others tapped their feet, and some did not move much at all.

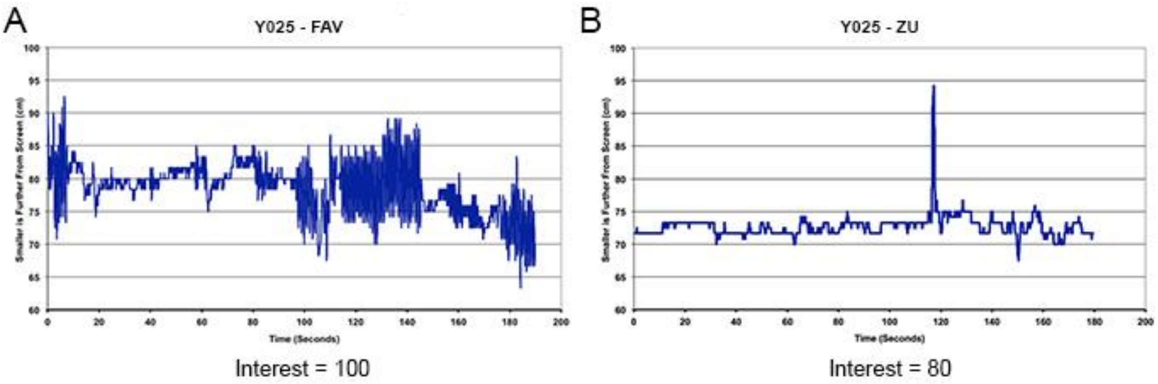

Figure 4: Representative motion tracking data during two engaging stimuli for one volunteer. The participant subjectively rated his interest level for both stimuli as 100 and 80 respectively on a 0-100 VAS scale, where 100 = "extremely". Panel A (left) shows the forehead marker distance from the screen (cm) while the participant was listening to his favourite music (stimulus "FAV"). The rapid fluctuations are sinusoidal movements of the head position in time with the music's beat. Panel B (right) shows similar measurements from the same participant while playing a computer game based on vigilance (Zuma, Popcap Games, "ZU" in Figure 2). During Zuma, the head position over the course of three minutes remained consistently within a 5 cm range (i.e. there was virtually no drift of the head's translational position), except for one "break-taking" episode during a natural pause in the game.

Figure 5: Photographic overlay of two frames from when a volunteer was nodding his head and tapping his hand to his favourite music. These photos are part of the response represented by the time series graph in Figure 4A; the frames are less than half a second apart.

4.2 Average values of movement for different stimuli

To determine whether the movements during the engaging stimuli were generally inhibited, we averaged the time course of head movements in response to an engaging film (OK Go's "This Too Shall Pass", Rube Goldberg version, "OK" in **Table 2**), resulting in the time course in **Figure 6A** (left panel). During the baseline fifty seconds (when the screen showed "television snow" while the audio track was of white noise), there was a tendency for volunteers to make sudden large movements. By contrast, during the music video itself, head movements were small and more akin to drift, with large, sudden movements being rare. A similar analysis of the participants' head movement responses to a boring stimulus (a single photograph remaining unchanged on the screen for two minutes, "IPSK" in **Figure 2**) resulted in a qualitatively different response (**Figure 6B**). While the range and mean of the overall head angle for the two stimuli over the course of two-minutes was broadly similar, we observed a general tendency during the boring stimulus for participants to make many sudden, large movements, which are manifest (although generated by single individuals) even when averaging the time courses over a number of participants.

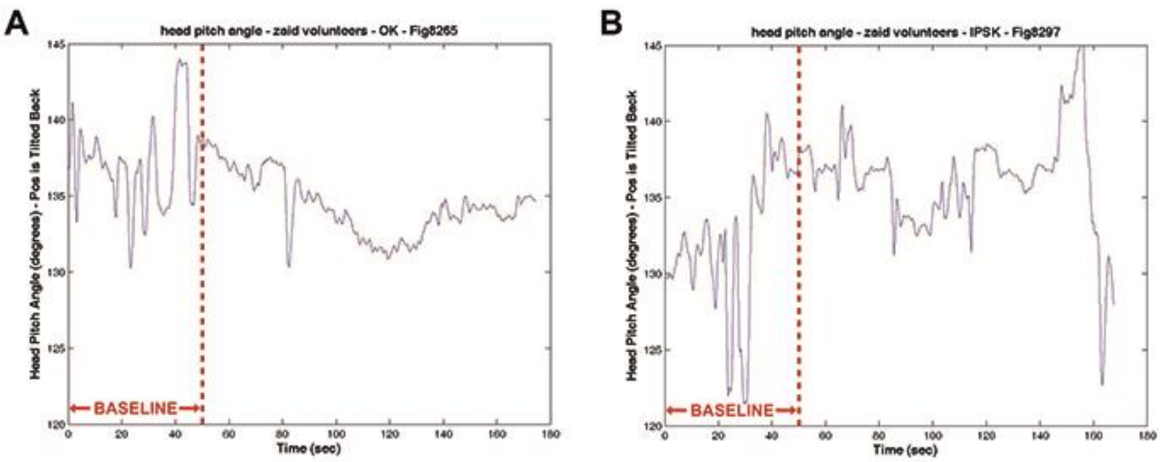

Figure 6: Time courses of head pitch angle for an engaging film stimulus (panel A, OK Go's music video "This Too Shall Pass") and a boring film stimulus (panel B, IPSK, a still photo shown for two minutes) averaged across the cohort. Larger values for head pitch angle correspond with the head being tilted back.

To make statistical comparisons of how these movement patterns compared in different stimuli, the mean movements during 80 seconds in the middle of each stimulus were compared as speeds (i.e. averaging of the time course the absolute values of the differences in angle (degrees) between adjacent time points of the low-pass filtered time course for each individual). **Figure 7A** shows a comparison of how head angle changed in two interesting stimuli (left panel), one of which was accompanied by video and the other had no video accompaniment (i.e. the screen remained black throughout); **Figure 7B** (right panel) makes the same statistical comparison in two boring (and silent) stimuli, one of which was accompanied by video and the other was not. In both cases, there was a statistically significant diminution of head angle speed when the stimulus was accompanied by video, presumably due to the fact that the participants needed to keep their heads steadier to see what was on the screen.

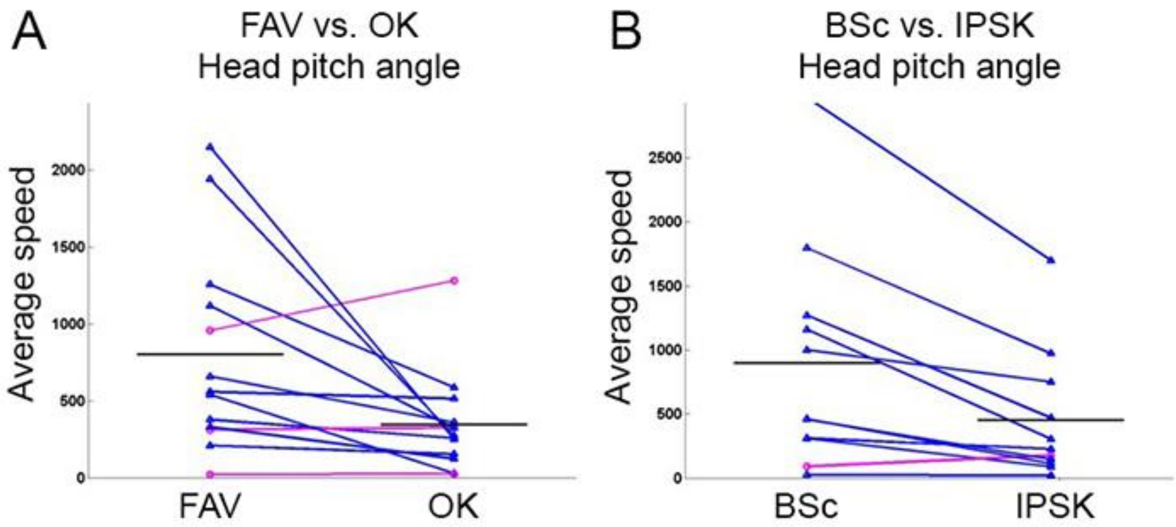

Figure 7: Non-Instrumental Movement Inhibition of head pitch movements due to gaze. Both panels demonstrate that when a stimulus includes on-screen visual cues, it results in diminished head movement (due to the head being kept steadier in order to look at the images). Panel A shows a comparison of the speed of head angle movements for a musical stimulus without visuals (FAV) and a music video (OK). Both stimuli are rated as very interesting (i.e. the participants do not want the stimulus to end sooner, see Figure 2). In the paired plot, each line represents one volunteer; blue lines are when the average head angle speed during FAV is higher than the average head angle speed for OK, while pink lines are where head speeds during OK are greater than during FAV. The black horizontal lines are the mean values, which are significantly different (Paired t Test, $P < 0.05$). Average speeds are calculated as listed in the methods. In panel B a similar comparison is made for two boring stimuli: BSc (a black screen) and IPSK (a still photo shown for two minutes). Although the photograph in IPSK is very interesting, seeing it for two minutes is both too long (see Figure 2) and boring — mean VAS rating for boring = 70.0 ± 6.6). During the stimulus, there is more movement during BSc (paired t Test, $P < 0.01$). Speed units shown are 10,000 X degrees change per frame (25 Hz).

To determine if the lack of visual accompaniment was generally associated with more movement, we pooled data from two different musical excerpts in our stimulus set (FAV, and VIO, a piece of aversive solo violin music) and compared these to pooled data from films (OK and IPSK). When non-interactive (i.e. without mouse activity) films used in this study were compared to the non-visual (and non-interactive) music excerpts used in this study, the films have an apparent Non-Instrumental Movement Inhibition effect on head pitch angle movements (Paired t Test, $P < 0.01$, see **Figure 8A**, left panel); this difference is paired for the interesting film compared to the interesting music excerpt, and the boring film compared to boring music; however, the same result is statistically significant for each film compared to each musical excerpt ($P < 0.05$ for all, not shown). Because our selection of films and the selection of music are not comprehensive, it is not possible to prove that films (including photo montages and still photographs) always diminish total head movement compared to music alone – presumably due to the act of gazing at the screen causing Non-Instrumental Movement Inhibition. However, it is interesting that the film stimuli (compared to music only) are also associated with reduced movement (Non-Instrumental

Movement Inhibition) of the thigh (Paired t Test, P < 0.01, see **Figure 8B**, right panel); while it is possible to argue that gaze would limit head movements, the rationale for gaze leading to a reduction in thigh movement is less clear.

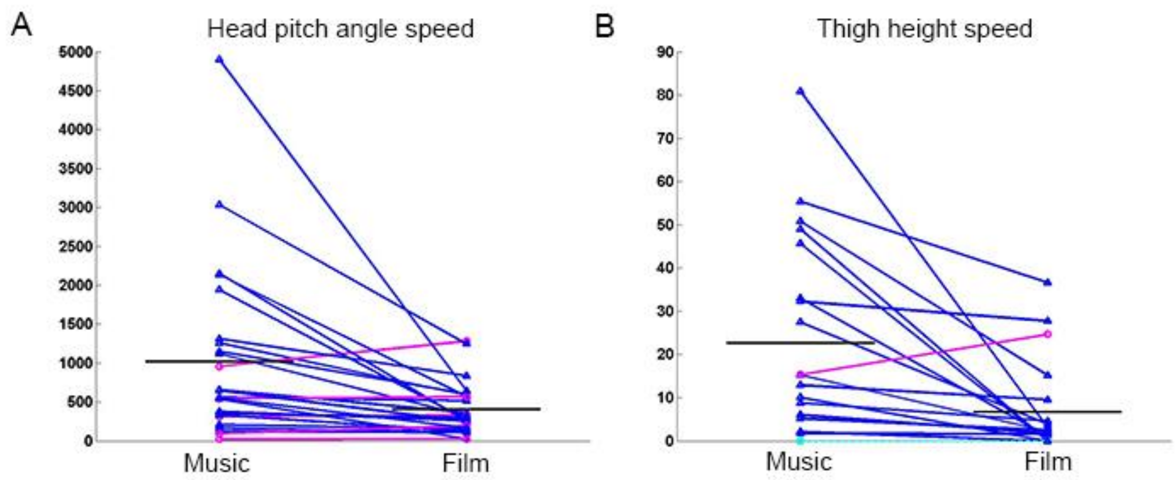

Figure 8: Compared to music alone, film causes Non-Instrumental Movement Inhibition in both the head and the thigh. Generalising from Figure 7A (left panel), not only do our non-interactive visual stimuli reduce head movement compared to non-visual stimuli, but they do so in thigh height movements as well (Figure 8B). Speed units shown are 10,000 X degrees change per frame (panel A) and 10,000 X cm/frame at 25 Hz (panel B).

4.3 Non-Instrumental Movement Inhibition as a manifestation of entrainment

Given that some engaging stimuli result in non-instrumental movement (i.e. preferred music can lead to head banging and foot tapping), while boring stimuli are associated with restless movements (Ekman's "abortive flight movements"), there is some ambiguity in interpreting when non-instrumental movements are a manifestation of engagement or disengagement. We propose the model in **Figure 9**, in which the cognitive state can be determined by the relationship between the timing of the movements and the narrative rhythm of the stimulus: entrained movements will imply engagement, while disruptive (out of synch) movements will imply disengagement.

To test whether entrained movements were more often associated with engagement, while disruptive movements (i.e. isolated self-adaptors) were associated with disengagement and restlessness, we manually scored films (standardised periods lasting for 175 seconds) for movements of the head, face, upper limb, lower limb and entire posture change, and compared preferred music to aversive music (**Figure 10**). While isolated arm movements (**Figure 10A**, left panel) were significantly more common during the aversive music (VIO), consistent movements (**Figure 10B**, right panel) were more common during the preferred music (FAV). While this result is provocative, it is based on manual scoring, which is open to criticisms of interpretation (i.e. whether a manual coder can consistently differentiate without bias what is a consistent/entrained movement vs. an isolated movement).

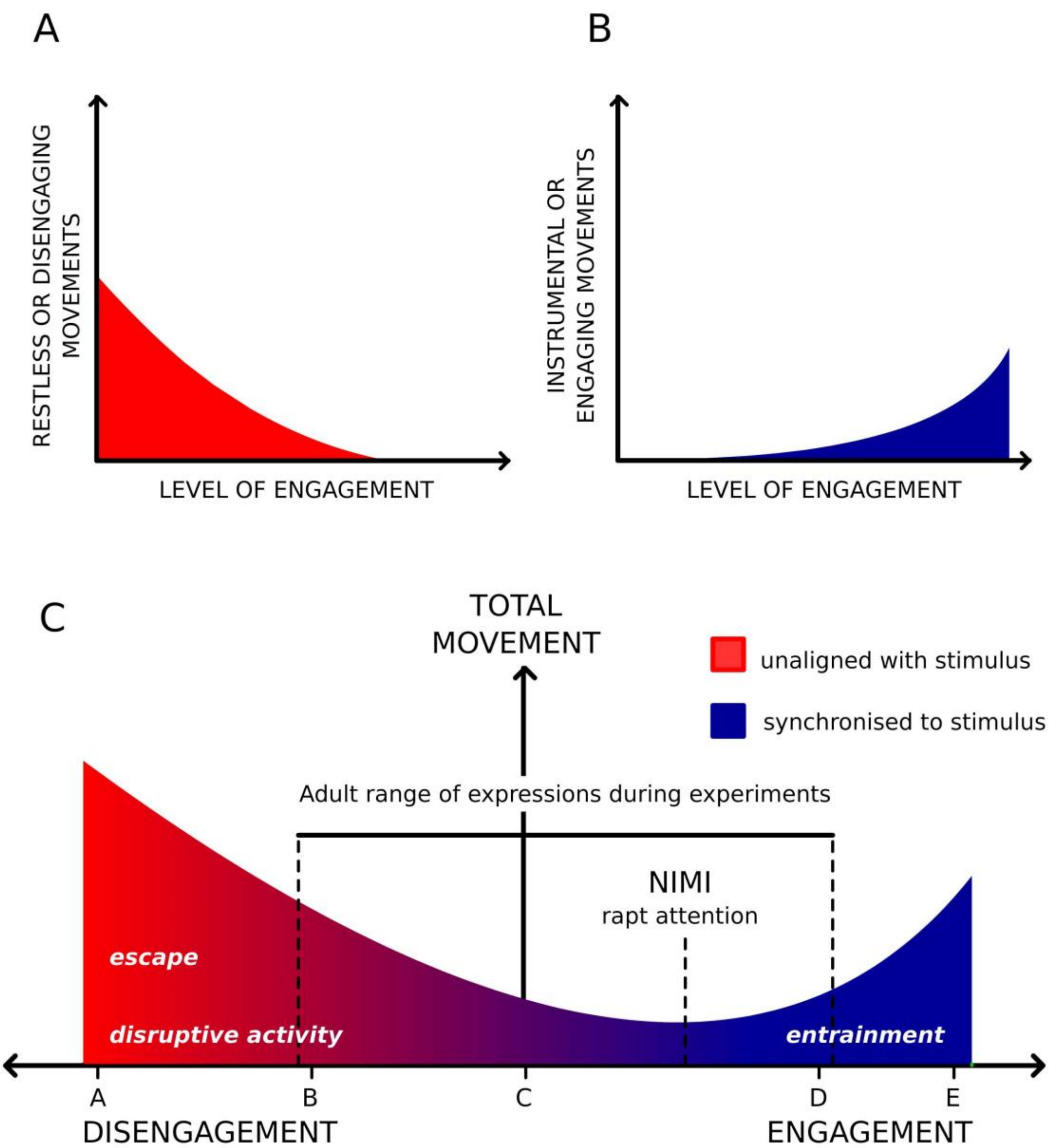

Figure 9: Proposed schematic description of the relationship between engagement and the amount of measurable physical activity. In all panels the X axis represents the response of the audience, which runs from disengagement on the left to engagement on the right. The Y axis represents the amount of movement made by the audience member (where up is more movement). The movements in panel A would be disruptive movements (e.g. running away or squirming), while the movements in panel B would be entrained to the stimulus (e.g. dancing or energetically turning the steering wheel during a driving game). In panel C, which is the summation of panels A and B, high levels of total movement/activity can result from either engagement or disengagement, and can be differentiated according to whether the activity is entrained (interactional synchrony) or disruptive (suppressed escape) to the stimulus. The tick marks on the X axis of panel C represent examples of audience responses: A is running away (which never happens during our experiments), B is constantly squirming in one's seat, C is when a bored person lets their head droop forward, D is vigorously activating the controls on a video game (or a football fan raising his hands in the air after a goal), and E is dancing to music. "NIMI" marks where cognitive engagement, expressed as focused attention, causes a distinct reduction in total physical activity — for example, when a still child watches a cartoon.

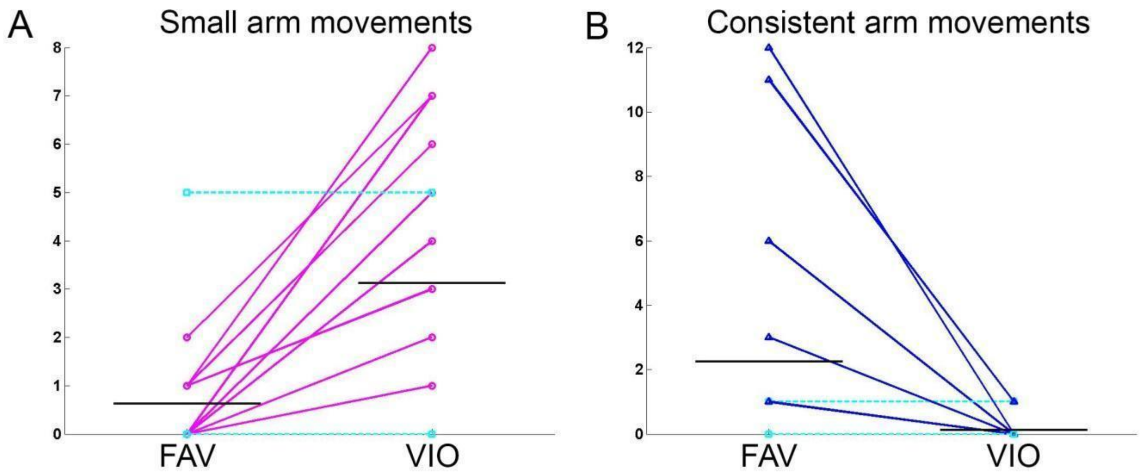

Figure 10: Manually scored self-adaptors during liked music (FAV) are more consistent (i.e. repeated, see panel B) while self-adaptors in response to disliked music (VIO) are more isolated and disruptive (see panel A). Films of volunteers experiencing music (without visual accompaniment) were scored manually for ~3 minutes for self-adaptors that were isolated and small (e.g. hand or foot only), large (e.g. movement of the whole arm or whole leg) or consistent (a repetitive movement happening at least 5 times over a 5 second period).

4.4 Measuring Non-Instrumental Movement Inhibition vs. Narrative Rhythm

To make fully objective measurements of entrainment, we compared movements within a stimulus, where the cycle time was long. In the geography quiz stimulus (GQ), the participant is shown 3-way multiple choice questions in cycles of 20 seconds (i.e. 3 actions per minute); when the participant selects a choice with the handheld trackball, they are shown whether the answer is correct or incorrect for 2 seconds, and then in any remaining time before the twenty-second cycle ends, they are given interesting information explaining the answer. The new question always appears after twenty-seconds, whether the participant has supplied an answer or not.

In a cyclic activity such as this, we would expect engagement and attention to be greatest when the question first appears and the participant reads the question and formulates an answer; after supplying an answer or while receiving the interesting information, we would expect non-instrumental movements to be more prevalent (‘break-taking’ see Kapoor et al. 2007), similar to the “back bracing” response described in Balaban et al (2004). By binning low-pass filtered head position data for each question and estimating movement activity as the range in each bin, it was possible to determine an average of movement for each bin (normalised for each person, as some people move much more than others) and then to compare the distributions of normalised ranges for the twenty-seconds that this cycling of questions lasts (**Figure 11**). As expected, during the geography quiz (panel A, left), the participants were entrained to the stimulus, as evidenced by non-instrumental movement inhibition during the first six seconds of questioning; during the rest of the question cycle, movement was comparably higher, presumably due to self-adaptors and postural movements during break-taking. When a similar analysis was

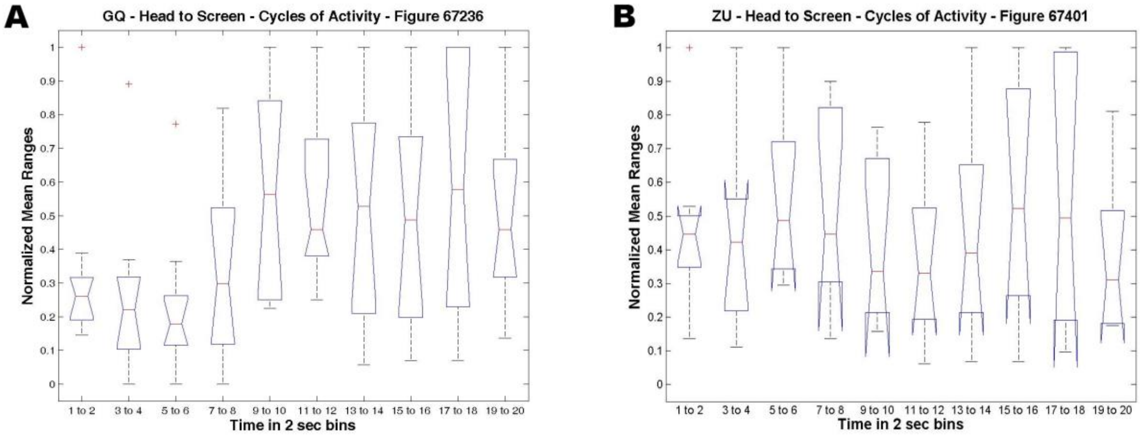

Figure 11: Entrainment of head to screen translational movement during a regularly repeating quiz. In panel A (left) are the distributions of mean normalized ranges for head translational position for the consecutive two-second bins of time during the 20-second geography quiz cycles. At time zero a new question appears, and as can be seen the mean normalized ranges of movement are comparatively small (< 0.3) during the first six seconds of these questions compared to the latter twelve seconds (when the participant is more likely to receive feedback on their selection or take a break while receiving an explanation to the answer). Panel B (right) shows ZU, a stimulus that does not have 20-second cycles or a regular narrative rhythm.

performed on Zuma (which does not have 20-second cycles), no apparent entrainment was seen (panel B, right).

There is a question as to whether this entrainment (as detected in regularly repeating stimuli) is associated more with engaging stimuli than with disengaging stimuli; a similar entrainment analysis is shown in **Figure 12** for a boring quiz (NQ). Compared to the clear inhibition of non-instrumental movement during the geography quiz, the duration of time where NIMI occurs in the nutrition quiz is limited. More importantly, when the question first appears (time zero to four seconds), participants make comparatively large movements, as if break-taking while the question arrives because they know there are no affordances for interaction yet. Thus, if we expected the actions and behaviours of the experimental participants during the boring nutrition quiz to be similar to the interesting geography quiz, instead we find that by making the quiz boring (and by making NQ reject button presses early in the narrative cycle), the actual behaviour is changed. That is, there is a difference between the observed cycles of behaviour in the interesting quiz (GQ) compared to the observed cycles of behaviour during the boring quiz (NQ) — despite the fact that both quizzes have three expected actions per minute.

5. Discussion

5.1 Novel contributions

This set of studies has made three new contributions: 1) Non-Instrumental Movement Inhibition (NIMI) has been defined as a manifestation of engagement, and it has been possible to numerically observe it by characterizing self-adaptors or by measuring postural

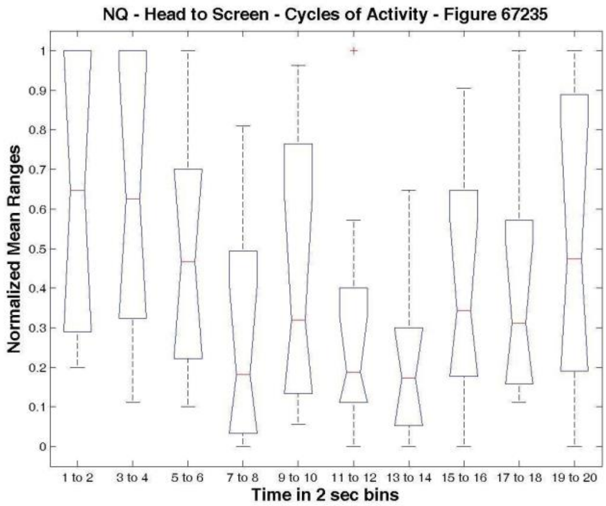

Figure 12: Weak entrainment of head to screen translational movement during a regular repeating quiz that is boring. The distribution of mean ranges for head translational position for the consecutive two-second bins of time during the 20-second nutrition quiz cycles (NQ in Figure 2) are shown. At time zero a new question appears, but the options do not appear immediately. To make the quiz more boring, the five options appear one at a time every two seconds, and the participant is not allowed to answer until all five options are on the screen. As can be seen, the normalized ranges of movement are comparatively large until all the options are on the screen and the participant can make a choice (seconds 11-14 of the question cycle).

micromovements. 2) Non-instrumental movement inhibition is elicited by visual engagement with a monitor, and it can be seen not only in head movements, but also in thigh movements. 3) Entrainment to stimuli can be identified as periods of non-instrumental movement inhibition, which are interspersed with “break-taking” during less demanding periods of stimulus interaction. That is, the experiment in **Figure 11A** provides the first controlled, objective evidence for the theoretical model in **Figure 9C**.

5.2 Ramifications for research on media

The description and ability to objectively observe Non-Instrumental Movement Inhibition will greatly benefit more holistic attempts to use postural cues and self-adaptors to assess of how audiences engage with screen-based and potentially other types of media. Although manual scoring of recorded nonverbal behaviours has been practised for many decades now (e.g. Bull, 1987), the fact that this could be subjective and biased (as well as extremely time consuming) has meant that progress in the nonverbal behaviour field has been slow (Gross et al., 2010). Improved analysis methods taking context and narrative rhythm into account could be applied to automated, real-time machine vision within responsive systems that incorporate functionality dependent on non-verbal signals of engagement and/or disengagement from the audience (D’Mello and Graesser, 2009). Such measures of emotional and intellectual audience engagement are potentially relevant to audience research, performance art, supportive technologies, and emergent responsive systems where the cognitive processes of the participants are of primary interest (Bilda et al., 2008; Edmonds et al., 2009; Gonsalves, 2009).

Evolving quantitative audience metrics toward affective and intellectual dimensions is an important contribution toward a step change in perspectives on interaction design in digital media and beyond. Digital production economies are primarily driven by quantitative factors acquired by remote metrics, e.g. page views, footfall, and conversions as measures of successful design. Such audience research methodologies shape the industry definitions of engagement (Adobe, 2013), and promote a procedural design rhetoric that is geared toward persuasion (Bogost, 2007: 28-29). By defining cognitive (intellectual and emotional) engagement as a quantifiable phenomenon, and therefore 'real' in terms of digital representation, there is increased scope for interface design objectives that are decoupled from simplistic actions and end results.

The existence of NIMI (see **Figure 11A**) suggests that engagement, as is tacitly understood within the arts, the humanities, and the education sector, is a phenomenon that incorporates perseverance and challenge, and is not solely a 'positive action' outcome. While the definition of engagement is controversial, some authors have included a positive mental state as part of the definition. For example, Skinner and Belmont (1993) posited that school children who are engaged 'show sustained behavioural involvement in learning activities accompanied by *positive emotional tone*' [emphasis added]. Likewise, the Advertising Research Foundation (2006) in one of their white papers on engagement stated, 'Consumer engagement is *a positive consumer attitude* resulting from the communication of (a) a given brand, (b) a given category (product/service/etc.)….' [emphasis added].
In contrast to this positive perspective, we observe that in response to some of our more challenging stimuli (e.g. GQ and SPi), half of our participants engage with cognitive states that include only negative (e.g. frustration, disappointment, low self-esteem) or a mixture of both negative and positive states. In fact, although participants numerically rated the stimulus GQ as significantly more engaging than the stimulus A5 (see **Figure 2**), in free text feedback participants provided much more negative descriptions of GQ than of A5 (**Table 3**). In the wake of the big data trend and the proliferation of remote audience metrics, such a complex understanding of engagement is the exception. Our longer-term research interests include making objective measurements to contribute to a differentiated definition of engagement that incorporates more fully perspectives that are known, but currently statistically invisible, within the arts, the humanities, and the education sector.

5.3 Limitations

There are many limitations to the conclusions and applicability for this study, whose function was to delineate and encourage further research using these technologies and analyses, rather than to propose a complete solution. 1) This research is done in a laboratory with experimental participants being aware that they are being filmed by multiple cameras. 2) The presence of cameras alone can potentially change people's behaviour by motivating special behaviours for the camera (Laurier and Philo, 2006) or by suppressing actions as if being scrutinized socially. However, in these experiments we

PARTICIPANT/STIMULUS/FREE TEXT			PARTICIPANT/STIMULUS/FREE TEXT		
Y013	GQ	annoyed, disappointed	Y013	A5	slightly interested then bored
Y015	GQ	frustrated, annoyed	Y015	A5	calm, cheerful
Y016	GQ	interested	Y016	A5	bored
Y017	GQ	good, interesting	Y017	A5	interesting
Y019	GQ	informative, difficult	Y019	A5	confused, puzzled
Y020	GQ	interesting, educated	Y020	A5	intrigued, anticipation
Y021	GQ	annoying	Y021	A5	interested
Y022	GQ	annoyed, upset, stupid, insulted	Y022	A5	interested
Y023	GQ	under pressure, felt a bit stupid	Y023	A5	some of them were interesting
Y024	GQ	interested, engaged, enjoyable	Y024	A5	reflective, alive, uplifted
Y025	GQ	interested	Y025	A5	it was nice
Y026	GQ	interested, frustrated, stupid	Y026	A5	mild interest
Y027	GQ	engaged	Y027	A5	attentive
Y028	GQ	challenged	Y028	A5	happy, longing, then a bit sad
Y029	GQ	frustrated, ignorant, humiliated	Y029	A5	happy, relaxed, in awe, amazed
Y030	GQ	quite stupid	Y030	A5	cheerful, random, bored
Y032	GQ	mentally active, enjoyment, interest	Y032	A5	mixed
Y033	GQ	frustrated	Y033	A5	intrigued
Y034	GQ	interested	Y034	A5	happiness, amazement, fortunate

Table 3: Free text responses to the question, "While you were watching/experiencing the previous stimulus, what did you feel?" This question was always the first question asked after a stimulus, to avoid letting the VAS descriptors influence the participant's first instinctive response. The left three columns refer to GQ, the right three columns refer to A5. Although the median rating to "I wanted it to end earlier" (Figure 2) was significantly higher for A5 than for GQ (and GQ was significantly more "engrossing", data not shown), this table shows that the free text responses for GQ were much more negative: 8 out of 19 responses were purely negative, while for A5 only 1 out of 19 was completely negative.

followed the recommendations of Heath and Hindmarsh (2002) of leaving the camera unattended on a tripod and having it set for a wide angle. 3) These experiments were performed in a chair. 4) These experiments used brief (two-minute), emotionally-homogenous, discrete stimuli rather than more "life-like" continuous stimuli. 5) No claims are made as to how engagement — in the here and now — influences thoughts, emotions and behaviours in the future. Our research programme is still at the early stages of demonstrating that people exhibit consistent non-instrumental movements when exposed to boring or engaging stimuli, and that these movements can be detected objectively by motion capture and video analysis.

5.4 Future Directions

While this study indicates that engagement, in some cases, inhibits non-instrumental movements, ambiguity remains in the case of quantifying engagement by using movement and velocity. While as of yet unable to remove this uncertainty, our research demonstrates that it is critical to differentiate the non-instrumental restless movements from the instrumental movements a person makes while interacting with a stimulus in order to make progress with affective analysis based on movement and velocity. This may be possible by detailed analysis of the structure and frequency spectrum of movements (e.g. D'Mello et al., 2012), or it may involve a strategy of associating each cognitive state with the movements of particular body parts (and particular directions of movement) (e.g. Bull, 1987). Finally, generalising that there is an association between non-instrumental movements and negative states (e.g. boredom) ignores positive non-instrumental movements, such as the raised arms of the football fan in celebration of a goal. Taking the cultural context and narrative rhythm into account also in the analysis of positive non-instrumental movements could yield significant progress, especially in situations engendering high arousal. Future research will need to differentiate negative non-instrumental boredom movements from non-instrumental expressions of positive emotion and from gestures of consideration (e.g. chin stroking), although for now, our experiments are not burdened with this problem because the range of positive movements we see in the lab is greatly reduced compared to what we know occurs "in the wild" (see **Figure 9**). We also wish to pursue a definition of engagement that is inclusive of composite cognitive states such as enjoyment and frustration, or confusion and interest (see **Table 3**).

Biographical notes:

Dr. Harry Witchel is Discipline Leader in Physiology at Brighton and Sussex Medical School. His research interest is in psychobiological metrics of engagement and emotions based on time series analysis. Contact: h.witchel(at sign)bsms.ac.uk.

Carina Westling is a Ph.D. candidate at the School of Media and Film at the University of Sussex. Her background is in performance and digital media, and her thesis interests are digital media engagement and mediated spaces.

Julian Tee is a final year medical student at Brighton and Sussex Medical School. This paper includes work performed as part of his fourth year Independent Research Project for the degree of Bachelor of Medicine, Bachelor of Surgery (BM BS).

Aoife Healy is a Senior Research Officer within the CSHER at Staffordshire University. She currently manages the Gait Analysis Laboratory and is writing up her PhD on Biomechanics.

Robert Needham is a Lecturer in Biomechanics at Staffordshire University. His research focuses on Gait and Postural analysis with an emphasis on Spinal mechanics.

Professor Nachiappan Chockalingam is Professor of Clinical Biomechanics at Staffordshire University, and is the Director of the CSHER, where he directs the Movement Analysis Laboratory and leads the Biomechanics team. He is also an Affiliate Professor at the Faculty of Health Sciences, University of Malta and a Visiting Professor at the Department of Sports Medicine and Arthroscopy, Sri Ramachandra University, Chennai. His research interests

centre around musculo-skeletal biomechanics and gait analysis with special interests in scoliosis, foot and footwear biomechanics and motion capture.

Acknowledgments:

We greatly acknowledge Michael Roberts and Amy Moffat for manually scoring the movement films for self-adaptors, and for help with formulating the manuscript. We also acknowledge Brian Wilson for the initial idea of combining music with self-adaptors.

References:

Adobe. 'Achieving a Single Marketing View of the Customer.' *Adobe Campaign White Paper*, 2013.

Advertising Research Foundation. 'Engagement: Definitions and Anatomy,' *ARF White Paper* (Laborie), J. Plummer (ed), 2006.

Allen, Terre H., and James M. Honeycutt. 'Planning, Imagined Interaction, and the Nonverbal Display of Anxiety.' *Communication Research,* 24.1, 1997: 64-82.

Balaban, Carey D., et al. 'Postural Control as a Probe for Cognitive State: Exploiting Human Information Processing to Enhance Performance.' *International Journal of Human-Computer Interaction,* 17.2, 2004: 275-286.

Berry, David. *Critical Theory and the Digital.* London: Bloomsbury, 2014: 36-39.

Bianchi-Berthouze, Nadia, Whan Woong Kim, and Darshak Patel. 'Does Body Movement Engage You More in Digital Game Play? And Why?' *Affective Computing and Intelligent Interaction,* 102-113.

Bilda, Zafer, Ernest Edmonds, and Linda Candy. 'Designing for Creative Engagement.' *Design Studies,* 29.6, 2008: 525-540.

Bogost, Ian. *Persuasive Games: The Expressive Power of Videogames.* Cambridge, Mass: MIT Press, 2007: 11-29.

Bradley, Margaret M., and Peter J. Lang. '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: Oxford University Press, 2007: 29-46.

Bulger, Monica E., Richard E. Mayer, Kevin C. Almeroth, and Sheridan D. Blau. 'Measuring Learner Engagement in Computer-Equipped College Classrooms.' *Journal of Educational Multimedia and Hypermedia,* 17 (2), 2008: 129-143.

Bull, Peter E. *Posture and Gesture*. Oxford: Pergamon, 1987.

Caso, Letizia, et al. 'The Impact of Deception and Suspicion on Different Hand Movements.' *Journal of Nonverbal Behavior,* 30.1, 2006: 1-19.

Chartrand, Tanya L., and John A. Bargh. 'The Chameleon Effect: The Perception–Behavior Link and Social Interaction.' *Journal of Personality and Social Psychology,* 76.6, 1999: 893-910.

Coan, James A., and John M. Gottman. 'The Specific Affect Coding System (SPAFF)' in J. A. Coan and J. J. B. Allen (eds) *Handbook of Emotion Elicitation and Assessment*. Oxford: Oxford University Press, 2007: 267-285.

Couldry, Nick. *Media, Society, World*. Cambridge: Polity, 2013: 30-31.

D'Mello, Sidney, Patrick Chipman, and A. C. Graesser. 'Posture as a Predictor of Learner's Affective Engagement.' *Proceedings of the 29th Annual Cognitive Science Society*. Austin, TX: Cognitive Science Society, 2007.

D'Mello, Sidney and A. C. Graesser. 'Automatic Detection of Learner's Affect from Gross Body

Language.' *Applied Artificial Intelligence*, 23, 2009: 123-150.
D'Mello, Sidney, Rick Dale, and Art Graesser. 'Disequilibrium in the Mind, Disharmony in the Body.' *Cognition & Emotion,* 26.2, 2012: 362-374.
Edmonds, Ernest, Zafer Bilda, and Lizzie Muller. 'Artist, Evaluator and Curator: Three Viewpoints on Interactive Art, Evaluation and Audience Experience.' *Digital Creativity,* 20.3, 2009: 141-151.
Ekman, Paul and Wallace V. Friesen. 'Nonverbal Leakage and Clues to Deception.' *Psychiatry,* 32 (1), 1969: 88-106.
Ekman, Paul, and Wallace V. Friesen. 'Hand Movements.' *Journal of Communication,* 22.4, 1972: 353-374.
Ekman, Paul, and Wallace V. Friesen. *Unmasking the Face: A Guide to Recognizing Emotions From Facial Clues*. Upper Saddle River, NJ: Prentice-Hall, 1975.
Gonsalves, Tina, et al. 'The Chameleon Project: An Art Installation Exploring Emotional Contagion.' *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*. IEEE, 2009.
Gross, M. Melissa, Elizabeth A. Crane, and Barbara L. Fredrickson. 'Methodology for Assessing Bodily Expression of Emotion.' *Journal of Nonverbal Behavior,* 34.4, 2010: 223-248.
Heath, C. and J. Hindmarsh. 'Analyzing Interaction: Video, Ethnography and Situated Conduct.' In T. May (ed) *Qualitative Research in Action*. London: Sage, 2002: 99-121
Kapoor, Ashish, Winslow Burleson, and Rosalind W. Picard. 'Automatic Prediction of Frustration.' *International Journal of Human-Computer Studies,* 65.8, 2007: 724-736.
Karageorghis, Costas I., et al. 'Psychophysical and Ergogenic Effects of Synchronous Music During Treadmill Walking.' *Journal of Sport & Exercise Psychology,* 31, 2009: 18-36.
Lakin, Jessica L., et al. 'The Chameleon Effect as Social Glue: Evidence for the Evolutionary Significance of Nonconscious Mimicry.' *Journal of Nonverbal Behavior,* 27.3, 2003: 145-162.
Latulipe, Celine, Erin A. Carroll, and Danielle Lottridge. 'Love, Hate, Arousal and Engagement: Exploring Audience Responses to Performing Arts.' *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 2011.
Laurier, Eric, and Chris Philo. 'Natural Problems of Naturalistic Video Data.' In H. Knoblauch, B. Schnettler, J. Raab, H.-G. Soeffner (eds) *Video Analysis: Methodology and Methods. Qualitative Audiovisual Data Analysis in Sociology*. Frankfurt am Main: Peter Lang, 2006.
Loke, Liam, and George Poonkhin Khut. 'Intimate Aesthetics and Facilitated Interaction.' In Linda Candy and Sam Ferguson (eds) *Interactive Experience in the Digital Age*. Springer, 2014: 91-108.
Mahmoud, Marwa, Louis-Philippe Morency, and Peter Robinson. 'Automatic Multimodal Descriptors of Rhythmic Body Movement.' *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*. ACM, 2013.
McCroskey, James C. 'The Communication Apprehension Approach.' In J. A. Daly (ed) *Avoiding Communication*. Beverly Hills, CA: Sage, 1984: 13-38
Morris, Desmond. *Manwatching: A Field Guide to Human Behaviour*. G. Desebrock (ed). New York: HN Abrams, 1977.
Mota, Selene, and Rosalind W. Picard. 'Automated Posture Analysis for Detecting Learner's Interest Level.' *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on*. Vol. 5. IEEE, 2003.
Niemeyer, Suzanne, and René Dirven (eds). *The Language of Emotions: Conceptualization, Expression, and Theoretical Foundation*. Amsterdam, NL: J. Benjamins, 1997.

Ramseyer, Fabian, and Wolfgang Tschacher. 'Nonverbal Synchrony in Psychotherapy: Coordinated Body Movement Reflects Relationship Quality and Outcome.' *Journal of Consulting and Clinical Psychology,* 79.3, 2011: 284-295.

Skinner, Ellen A., & Michael J. Belmont. 'Motivation in the Classroom: Reciprocal Effects of Teacher Behavior and Student Engagement Across the School Year.' *Journal of Educational Psychology,* 85, 1993: 571 – 581.

Thornby, Maureen A., François Haas, and Kenneth Axen. 'Effect of Distractive Auditory Stimuli on Exercise Tolerance in Patients with COPD.' *CHEST Journal,* 107.5, 1995: 1213-1217.

Waxer, Peter H. 'Nonverbal Cues for Depression.' *Journal of Abnormal Psychology,* 83*,* 1974: 319-322.

Waxer, Peter H. 'Nonverbal Cues for Depth of Depression: Set Versus No Set.' *Journal of Consulting and Clinical Psychology,* 44*,* 1976: 493.

Waxer, Peter H. 'Nonverbal Cues for Anxiety: an Examination of Emotional Leakage.' *Journal of Abnormal Psychology,* 86.3, 1977: 306.

Witchel, Harry J., et al. 'Comparing Four Technologies for Measuring Postural Micromovements During Monitor Engagement.' *Proceedings of the 30th European Conference on Cognitive Ergonomics*. ACM, 2012.

Witchel, Harry J., et al. 'Mean Head and Shoulder Heights When Seated: Subconscious Postural Cycles During Discrete Computerised Stimuli.' *Proceedings of the 31st European Conference on Cognitive Ergonomics*. ACM, 2013a.

Witchel, Harry J., et al. 'Using Body Language Indicators for Assessing the Effects of Soundscape Quality on Individuals.' *AIA-DAGA Conference on Acoustics*. Berlin: Deutschen Gesellschaft für Akustik, 2013b.

Witchel, Harry J. 'Engagement: The Inputs and the Outputs.' *Proceedings of Inputs-Outputs.* ACM, 2013c.

Notes:

[1] Origin of stimulus set. All photographs (including those in the montage A5) are from the International Affective Photographic System (IAPS). Videos were harvested from the internet; the original soundtrack for SPi was replaced by the scientific team with a conversation in Estonian that was meant to be incomprehensible to our non-Estonian volunteers. FAV is a stimulus where the participant listens to their favourite music with a strong beat; in advance of the experiment, participants are asked to select their favourite music with a strong beat, which the scientific team plays via an iPhone through the experimental sound system, such that participants remain facing the black computer monitor. VIO was harvested from the internet; it was originally 24 seconds, and was repeated to make a 2 minute stimulus. IPSK is photo 8030 (a skier from the top of a ski jump) from the International Affective Photographic System (Bradley and Lang, 2007); this photograph has the highest mean arousal ratings of all the IAPS photos. BSc (black screen) is two minutes in which the participant is alone in the room while the computer only shows a blank, black computer screen after the pre-stimulus of television snow and white noise The black screen stimulus is not explained, nor are the volunteers forewarned; it begins with the white noise baseline, and then the screen goes completely black for two minutes while the participant is alone.

[2] Reflective markers for motion capture. 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. A marker for shoulder movement was placed on the badge of the deltoid, and another marker was on the thigh.
[3] Box and whisker plots. 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.
[4] Head Pitch Angle is modelled as being only at the atlanto-occipital joint, where larger angles correspond with the head being tilted back. This angle is measured as the angle between a line running from the outer canthus of the eye to the mastoid process (crossing the tragus of the ear), and a conjoined line running from that point on the mastoid process to the C7 vertebra (spinous process).
[5] Mean normalized ranges were calculated in Matlab as follows: the stimulus was divided into 20-second segments (each representing one question on a quiz), and the first 16 seconds and the final 4 seconds (of the entire 175 second stimulus) were ignored (to avoid movement artefacts associated with the beginnings and ends of stimuli (see Witchel et al. 2012 and Witchel et al. 2013a)), resulting in data for eight quiz questions per person (i.e. per stimulus). The 20-second time segments were divided into 2 second bins, and for each question the absolute value of the range in each two-second bin was calculated (as a surrogate for self-adaptors, i.e. large movements), and for a given person/stimulus these were normalised to the largest range in any bin for that person/stimulus. The normalised values for all the questions for that person/stimulus were averaged, resulting in one normalised value for each of the ten two-second periods in an average question.