Proceedings of the Symposium

Symposium on Mental States, Emotions and their Embodiment

A symposium at the AISB 2009 Convention (6-9 April 2009)
Heriot-Watt University, Edinburgh, Scotland

Symposium Chairs
Nadia Barthouze
Marco Gillies
Aladdin Ayesh

Published by SSAISB:
The Society for the Study of Artificial Intelligence
and the Simulation of Behaviour
http://www.aisb.org.uk/

ISBN - 1902956788

Symposium on Mental States, Emotions and their Embodiment

A two-day symposium at AISB 2009 (6-9 April 2009).
http://www.doc.gold.ac.uk/~mas02mg/msee/

PROGRAMME CHAIRS

Dr Nadia Berthouze, University College London, UK
Dr Marco Gillies, Goldsmiths College London, UK
Dr Aladdin Ayesh, De Montfort University, UK

INTRODUCTION

An understanding of human mental states and emotions can have a strong impact on the design and creation of computational systems. This can be through simulations that model mental states; the automated recognition of mental states, or an understanding of the affective aspects of human computer interaction.

The study of affect in humans is a very diverse field. Emotions and mental states can be studied through cognitive or neuroscientific models or through their bodily expression. They can be studied in fields as diverse as the arts, psychology, neuroscience, computing and cultural studies.

This symposium demonstrates the interdisciplinarity and breadth of this field. It will focus on models of emotion and their bodily expression. Participants come from a wide range of disciplines including computing, psychology, fine arts, cultural studies and film.

The papers submitted also cover a wide range of topics ranging from cognitive models to embodied behaviour. Karg, Kühnlenz and Buss present a computation model of emotion. Beck, Stevens and Bard and Halovic and Kroos study different aspects of peoples perception of affective behaviour. Pan, Gillies and Slater, on the other hand use affective behaviour as an experimental measure for studying people's interaction with virtual reality. That perceptural studies can lead to computational models is demonstrated by Fridin et al. The study of affective behaviour is greatly facilitated by movement tracking technologies, Karanika presents a novel fabric for motion tracking and explores its artistic possibilities. Finally, the study of affect has many practical applications, and the symposium will include two novel and interesting one. Igareda and Maiche explore the difficult topic of describing affect in audio description of films, while Santos and Pitt discuss how ubiquitous computing can be used to embed social norms in the design of a workplace. This range of papers demonstrates how vibrant and exciting the field of affective computing can be and promising interesting research for years to come. Truly good research in any field relies on practical work but also a thorough exploration of the fundamentals of the field and a critical appraisal of the assumptions made. Sdrolia provides in her philosophical critique of the research area.

PROGRAMME COMMITTEE

Table of Contents

Facilitating the Perception of Anger and Fear in Male and Female Walkers

Shaun Halovic[1], Christian Kroos[1]

Abstract. Previous evidence has indicated that the perception of emotions in others is confounded by the gender of the person displaying the emotion. This confound may be conceptually based or structurally/kinematically based. Male and female full-light (conceptual and kinematic information) and point-light walkers (kinematic information) displaying either anger or fear were shown to perceivers. Perceiver identification times were recorded. Results indicate the perception of fear is facilitated if the walker displaying fear is female due to similar kinematics for fearful gait and female-specific gait.

1 INTRODUCTION

There is good evidence that specific emotions can be perceived through walking gait [1, 2, 3]. Furthermore, past studies have inferred that the speed in which perceivers can identify each emotion indicates the relative ease in identifying those emotions [4, 5, 6]. However, the gender of the walker likely influences the perception of different emotions due to gender-emotion stereotypic beliefs [7, 8] or due to overlapping kinematics between specific emotion expression and gender-specific gait [9, 10].

The relationship between emotionality and gender is inherent in our culture and represented in the majority of theories on emotional development [11, 12, 13]. In the classic study by Condry and Condry [14], an emotional display by an infant labelled 'girl' was rated by perceivers as less angry and more afraid than the same display labelled 'boy'. In addition to the stereotypic beliefs of each gender expressing specific emotions there is the belief that females experience and express emotions more frequently and with more intensity than males [7, 8]. The perception of emotions and the gender of the person displaying the emotion appear to be confounded on a conceptual level.

In contrast, this confounded relationship may be due to structural or kinematic similarities between specific emotion expression and the display of gender. The structural architecture of a gender-specific face has been shown to influence the speed in which specific emotions can be identified. Vaughn Becker, et al. [6] found that anger was identified faster and more accurately on male faces than female faces. Also, happiness was identified faster and more accurately on female faces. Male faces have several architectural structures that are different to female faces. Male faces tend to have a lower brow and a squarer jaw whilst female faces retain some of the neonatal features of children [15]. The structural architecture of a male face is similar to the facial expression of anger in which the lowered brow and squarer jaw mimics the furrowing eyebrows and clenching jaw of angry expressions. In contrast, the neonatal structure of a female face is similar to the facial expression of happiness in which the finer jaw line, less prominent brow and rounder cheeks mimic the upturned facial characteristics of the expression of happiness [16, 17, 18, 19]. It is therefore not surprising that perceivers are biased towards perceiving anger in male faces and happiness in female faces.

[1]MARCS Auditory Laboratories, Univ. of Western Sydney, Australia. Email: {s.halovic, c.kroos}@uws.edu.au

However, walking is clearly different to facial expressions and the perception of different emotions from facial expressions uses a different brain system to that used for the perception of emotions through bodily movement [20]. Thus previous findings may not generalise to cover the expression and perception of emotions through walking style. However, the gender of a walker is likely to still influence the perception of different emotions through walking style because males and females have distinct walking styles [10].

Any specific emotion that is expressed through gait movements that are similar to a gender specific walking style are likely to bias the perception of that emotion towards the gender of the walker. Troje [10] found that male-specific gait increased the walker's perceptual size (i.e. greater lateral sway of the body and with outward pointing elbows) whilst female-specific gait incorporated size reducing kinematics (i.e. smaller steps with backward pointing elbows). The size enhancing kinematics of male gait appears to make the walker look more dangerous therefore more likely to be perceived as displaying anger [9]. On the other hand the size reducing kinematics of female gait is similar to the cowering movements of fear [9] thus this is likely a bias towards perceiving fear when the walker is female. Therefore there is a likely confound between gender-specific gait and the perception of emotions.

The aim of this research is to investigate the influence of walker gender on the perceiver reaction time for identifying the emotions anger and fear. We therefore hypothesise: 1) Perceivers will identify anger faster when the walker is male compared to female. 2)

Perceivers will identify fear faster when the walker is female compared to male.

2 Stimulus Construction

2.1 Method

2.1.1 Participants

The walkers were 17 female and 19 male actors (Age: M = 32.5, SD = 9.8; years of acting training: M = 2.46, SD = 5.08; years of acting experience: M = 4.18, SD = 6.99). The recording of gait movements with real felt emotions (i.e. induced) using motion capture methods is methodologically difficult to impossible. However, actors are trained to reproduce a displayed emotion instantly on demand without the need of an external emotion inducing event. The actors were recruited from various acting agencies in the Sydney area and through recruitment flyers and were paid $50AUS for travel costs.

2.1.2 Apparatus

Walkers wore black spandex body suits (ankle to wrist length) to ensure that perceivers could adequately see the walker's gait movements. All walkers wore a black balaclava, leaving only the walker's eyes and lips visible to the camera. To create point-light (PL) displays of the walkers, 17 visible Vicon infrared reflective markers were attached with double sided tape to both sides of the walker's body on the ears, shoulders, elbows, wrists, hips, knees, ankles, insteps of the feet and a single marker on the manubrium (Figure 1). The reflective markers were tracked within a 3-dimensional space with a Vicon (Oxford Metrics Limited) motion capture system with 10 cameras. PL displays of the walkers were created by importing the motion capture data into Matlab (The Math Works, Inc.). Full-light (FL) displays of walkers were simultaneously recorded with a digital camera (Panasonic NV-MX300EN/A) from a frontal perspective.

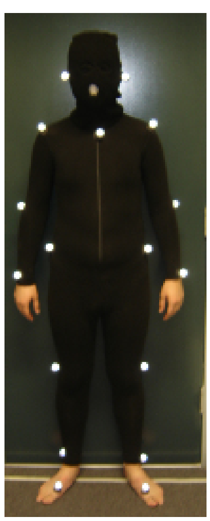


Figure 1. A FL walker shown with the corresponding Vicon markers used to create PL walkers.

2.1.3 Procedure

Prior to recordings the walkers were informed that they were to walk whilst displaying different emotions (i.e. anger and fear[2]) and were given time to think about how to express each emotion. For each emotion, the actors were asked to walk from one side of the room to the other. Both emotions were recorded at three levels of emotional intensity, starting with low, progressing to moderate and then high to assist in the expression of each emotion. At least two complete gait cycles was recorded for each emotion/intensity combination of each walker.

3 Experiment 1: The Influence of Walker Gender on the Reaction Time for Identifying Specific Emotions in FL Walkers

3.1 Method

3.1.1 Participants

The sample of 'perceivers' comprised 38 (34 female) first year psychology students from the University of Western Sydney. The sample had a mean age of 22.61 years (SD = 8.88). All perceivers had normal or corrected to normal vision and were given course credit for their participation in the experiment.

3.1.2 Materials

Due to technical faults during stimulus construction, some walkers failed to be recorded thus were not shown to perceivers. Perceivers were shown the remaining 187 FL walker stimulus items (anger = 93, fear = 94) with a minimum 30 examples of each emotion/intensity combination.

The experiment-control software Alvin (version 1.19, [21]) was used to show perceivers the FL walkers with an inter-stimulus interval of two seconds between walkers. The program displayed video clips of a walker walking towards the perceiver. It required the perceiver to click on 1 of 5 categorical buttons corresponding to the five displayed emotions (happy, sad, anger, fear, and neutral[2]) and automatically measured the time between the initial display of the FL walker and the mouse click on one of the categorical buttons.

3.1.3 Procedure

Perceivers were told they would view a series of walkers shown in full video display format. It was explained that the walkers would be walking with one of five emotional states: happiness, sadness, anger,

[2]Five different emotions (i.e. happy, sad, anger, fear and neutral) were collected and tested in these experiments for other research purposes which are irrelevant to this paper. We will thus only report the details for the emotions anger and fear due to their relevance for the research questions investigated here.

fear or neutral and that they were required to identify which emotion the walker was feeling by clicking the button on the computer screen that corresponded to the emotion they perceived. Perceivers were also told to make their judgments as quickly and as accurately as possible and they need not wait until the end of the video clip to make their judgement.

3.2 Results

Only correctly identified emotions were included in these analyses thus ensuring any emotion/gender gait similarities, whether stereotyped or not, were valid displays of that individual emotion. Paired-sample t-tests were conducted comparing the perceiver reaction times of correctly identified emotions between both genders. The mean reaction times for each perceiver were used as the data scores for this analysis in order to reduce the high degrees of freedom. A Bonferroni corrected alpha of .025 was used for both t-tests. Fear was identified significantly faster in female walkers compared to male walkers, $t(37) = 4.6$, $p = .000$. There was no significant difference in the reaction time for identifying anger between male and female walkers, $t(37) = 2.02$, $p = .050$. Descriptive statistics are shown in figure 2.

Figure 2. Fear was identified significantly faster in female FL walkers than male FL walkers. Anger was identified in equivalent time in both female and male FL walkers.

3.3 Discussion

We found that perceivers identified fear faster in female walkers than in male walkers. There are two possible explanations for this finding. The first of which is that the size reducing kinematics of female-specific gait [10] is similar to the expression of fear [9] in walkers. The similarity between female-specific gait and the expression of fear likely facilitated the perception of fear in female walkers leading to faster reaction times for female walkers but not male walkers.

The second possible explanation for this finding is that the concept of a fearful individual is confounded with the perception of females. Females, due to their reduced stature, are physically less able to defend themselves from assault than males, possibly explaining why females are more susceptible to experiencing fear than men [22, 8]. The perception of fear is therefore likely facilitated when the perceiver can identify the walker as female.

To distinguish between these two possible explanations we replicated this first experiment with point-light (PL) walkers. The gender of FL walkers is easily apparent to the perceivers. PL walkers, on the other hand, deny perceivers of most structural information thus making the gender of the walker increasingly difficult to perceive [23, 24]. The gender-specific kinematics of the walker however, are left undiminished [10]. If the concept of fear is confounded with the concept of female, then the findings of this experiment would not be reproduced in the next experiment with PL walkers. However, if the kinematics of fearful gait is confounded with the kinematics of female-specific gait then we would find similar results with PL walker stimuli.

4 Experiment 2: The Influence of Walker Gender on the Reaction Time for Identifying Specific Emotions in PL Walkers

4.1 Method

4.1.1 Participants

The participant requirements were the same as in experiment 1. The sample comprised of 34 perceivers (19 female) first year and postgraduate psychology students and had a mean age of 26.71 years (SD = 5.61).

4.1.2 Materials

Due to technical faults during stimulus construction, some walkers failed to be completely recorded (i.e. missing markers) thus were not shown to perceivers. Perceivers were shown the remaining 184 PL walker stimuli (anger = 100, fear = 84) with a minimum 27 examples of each emotion/intensity combination. For every walker the common translatory component of each marker was subtracted by performing a principal components analysis [25] in Matlab. The first principal component always represented the translation of the walker across the designated movement space. Therefore this component was excluded from a reconstruction of the remaining data based on all remaining components thus creating stationary walkers with the ecologically valid kinematics of a locomotive walker.

The procedure was identical to that used in the FL walker experiment.

4.2 Results

Two paired-sample t-tests were conducted in the same way as in the previous experiment. The mean reaction times for each perceiver were again used as the data scores for this analysis. A Bonferroni corrected alpha of .025 was used for both t-tests. Both fear, $t(32) = 4.7$, $p = .000$, and anger, $t(33) = 3.0$, $p = .005$, were identified significantly faster in female walkers compared to male walkers. Descriptive statistics are shown in figure 3.

4.3 Discussion

Our finding that fear was identified faster in female walkers than in male walkers is congruent with the results of the previous experiment with FL walkers. This lends support to the argument that the kinematics of fearful gait is similar to the kinematics of female-specific gait thus facilitating the perception of fear in female walkers. The perception of fear in walkers is therefore confounded with the gender of the walker.

Figure 3. Fear and anger were both identified significantly faster in female PL walkers than in male PL walkers.

However, our additional finding that anger was also identified faster in female PL walkers is surprising. There is no kinematic similarity between anger-specific gait and female-specific gait [3, 9, 10]. One possible explanation for this finding is that the female walkers were more skilled at displaying both anger and fear through gait. Females experience and express their emotions more frequently and with more intensity than males [22] thus are likely to be more skilled at expressing their emotions through gait movements. The greater expressive skill of female walkers likely compounded the kinematic similarities between female-specific gait and fearful gait, further facilitating the perception of fear in female walkers in both FL and PL conditions. However, this argument does not explain why anger was identified faster in female walkers in the PL condition but equivalently the FL condition.

Conceivably, our initial argument for the confound between the perception of anger and male walkers had merit but was counteracted in the FL condition by the additional skill that female walkers had at expressing their emotions. However, unlike the kinematic similarities between fearful gait and female-specific gait, the influence of male gender on the perception of anger in walkers appears to be conceptually based. That is, when perceivers could easily identify the gender of a male walker in the FL condition, they interpreted the emotion displayed by the walker in a stereotype-congruent manner [8]. This stereotype-congruent information facilitated the perception of anger in male walkers but this advantage was counteracted in the FL condition by the increased expressive skill of the female walkers.

However, when the gender of the walker was no longer so apparent in the PL condition, perceivers were denied the additional cue of stereotype-congruent information. The perception of anger in male PL walkers was therefore not facilitated. However, the greater expressive skill of female walkers did facilitate the identification of anger in the PL condition. Consequently, anger was identified faster in female walkers than male walkers in the PL condition but not in the FL condition.

5 Conclusion

We have provided support that the kinematics of fearful gait is similar to the kinematics of female-specific gait thereby facilitating the perception of fear in female walkers. We have also argued that male gender and anger are confounded on a conceptual level therefore when the gender of the walker is easily perceived as male then the perception of anger is facilitated. However, stronger conclusions could be made if these experiments were replicated with walker stimuli that was normalised for the level of expression. These findings have important implications for the development of anti-victimisation programs whereby women may be taught how to walk in a way that may discourage potential attackers.

REFERENCES

[1] Chouchourelou, A., Matsuka, T., Harber, K., & Shiffrar, M. The visual analysis of emotional actions. *Social Neuroscience,* 1(1): 63-74, (2006).
[2] Heberlein, A. S., Adolphs, R., Tranel, D., & Damasio, H. Cortical regions for judgments of emotions and personality traits from point-light walkers. *Journal of Cognitive Neuroscience,* 16(7): 1143-1158, (2004).
[3] Montepare, J. M., Goldstein, S. B., & Clausen, A. The identification of emotions from gait information. *Journal of Nonverbal Behavior,* 11(1): 33-42, (1987).
[4] De Sonneville, L. M., Verschoor, C. A., Njiokiktjien, C., Op het Veld, V., Toorenaar, N., & Vranken, M. Facial identity and facial emotions: Speed, accuracy, and processing strategies in children and adults. *Journal of Clinical and Experimental Neuropsychology,* 24(2): 200-213, (2002).
[5] Eastwood, J. D., Smilek, D., & Merikle, P. M. Negative facial expression captures attention and disrupts performance. *Perception & Psychophysics,* 65(3): 352-358, (2003).
[6] Vaughn Becker, D., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., & Smith, D. M. The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology,* 92(2): 179-190, (2007).

[7] Fabes, R. A., & Martin, C. L. Gender and age stereotypes of emotionality *Personality and Social Psychology Bulletin,* 17(5): 532-540, (1991).

[8] Plant, E. A., Hyde, J. S., Keltner, D., & Devine, P. G. The gender sterotyping of emotions. *Psychology of Women Quarterly,* 24: 81-92, (2000).

[9] Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception, 33*, 717-746, (2004).

[10] Troje, N. F. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision,* 2: 371-387, (2002).

[11] Buck, R. Emotional development and emotional education. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research and experience (*Vol. 2, pp 259-292). Academic Press, Orlando, FL, (1983).

[12] Chodorow, N. *The reproduction of mothering*. University of California Press, Berkely, (1978).

[13] Kemper, T. D. *A social interactional theory of emotions*. Wiley, New York, (1978).

[14] Condry, J. & Condry, S. Sex differences: A study of the eye of the beholder. *Child Development,* 47: 812-819, (1976).

[15] Bruce, V., Burton, A. M., Hanna, E., Healey, P., Mason, O., Coombes, A., et al. Sex discrimination: How do we tell the differences between male and female faces? *Perception,* 22(2): 131-152, (1993).

[16] Darwin, C. *The expression of the emotions in man and animals* (3rd ed.). HarperCollins, London, (1872/1999).

[17] Ekman, P., & Friesen, W. V. Measuring facial movement. *Environmental Psychology and Nonverbal Behavior,* 1(1): 56-75, (1976).

[18] Massaro, D. W., & Egan, P. B. Perceiving affect from the voice and face. *Psychonomic Bulletin & Review*, 3(2): 215-221, (1996).

[19] Padgett, C., Cottrell, G. W., & Adolphs, R. *Proceedings of the Cognitive Science Conference*. (1996).

[20] Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., et al. Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *European Journal of Neuroscience,* 13: 400-404, (2001).

[21] Hillenbrand, J., & Gayvert, R. T. Open source software for experiment design and control. *Journal of Speech, Language, and Hearing Research,* 48: 45-60, (2005).

[22] Grossman, M., & Wood, W. Sex differences in intensity of emotional experience: A social role interpretation. *Journal of Personality and Social Psychology,* 65(5): 1010-1022, (1993).

[23] Cutting, J. E., Proffitt, D. R., & Kozlowski, L. T. A biomechanical invariant for gait perception. *Journal of Experimental Psychology: Human Perception and Performance,* 4(3): 357-372, (1978).

[24] Johansson, G. Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics,* 14(2): 201-211, (1973).

[25] Jackson, J. E. *A user's guide to principal components.* John Wiley & Sons, New York, (1991).

Male Bodily Responses towards a Virtual Woman

Xueni Pan [1] and **Marco Gillies**[2] and **Mel Slater**[1, 3]

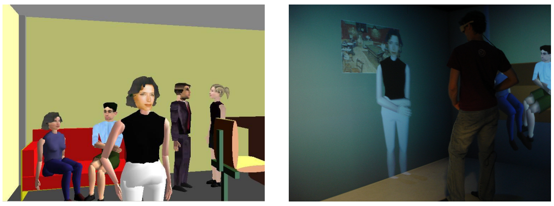
Abstract.

Current research measure participants' responses in a Virtual Environment through questionnaires, physiological responses, and post-experiment interviews. However, questionnaires and interview results are subjective. Physiological data, although objective, provide only evidence of a person's autonomic system responses rather than higher level behavioural responses. Different from the above methods, bodily movement is a gross overall indicator of a person's state, and is relatively easily observable. It therefore could offer us an additional method to assess the realism of people's responses within a virtual environment. This work presents the analysis of the bodily movement displayed by male participants, while talking with a life-size virtual woman in a virtual social encounter within a CAVE-like system. People tend to respond realistically to virtually generated data, both emotionally and physically. While interacting with a virtual character, their responses should be similar to how they would respond to a real person. Here we concentrate on their body movements as the virtual character speaks to and approaches them.

This is part of an ongoing study of whether people with social phobia react with appropriate anxiety to virtual reality depictions of social encounters. People who suffer from social phobia experience strong fear in one or more social situations, where they are afraid of acting in a way which is humiliating or embarrassing that other people would judge them in a negative way. Although they have the knowledge that the anxiety was irrational, they seek to avoid social encounters whenever possible [1]. Therefore this research studies the extent to which the participant responds towards the virtual character with appropriate affect.

We have carried out an experimental study in an immersive (Cave) system, where shy and confident males interacted with a forward virtual woman, in a virtual bar [2]. We are interested in the extend to which the participants' responses are similar to real life. We have invited both shy and confident participants, which is determined by the SPAI (Social Phobia and Anxiety Inventory) questionnaire [3]. There were 24 male participants who took part in the experiment, 12 of whom are shy, 12 confident.

The scenario was a virtual bar which is populated by 5 virtual characters, 4 of whom are talking to one another, except for one lone female, 'Christina'. Once the participant enters the scenario Christina begins to stare at him for a few seconds, then makes her way towards him (Fig. 1(a)), and then initiates a conversation (Fig. 1(b)). It is important to note (as shown in Fig. 1(b)) that from the point of view of the participant Christina is life-sized and that the projection is active stereo. The participant stood approximately in the centre of the room and head tracking was used so that Christina could look at him in the eyes, and also the other characters in the room would occasionally look towards him.



(a) (b)

Figure 1. The virtual female first looks at the participant and then (a) approaches him and (b) a conversation ensues while the other characters continue their conversations.

Christina is modelled to be an attractive female, speaking English with a slight Portuguese accent (the accent of the real actress who recorded the script). She started the conversation with the participant by introducing herself and asking the participant questions. When he spoke, she appeared to listen carefully and showed her interest by nodding and smiling. She also showed her interest by leaning forward to him, smiling, looking at him, and also by breaking the norms of social distance (of the country in which the study was conducted) by approaching within 0.5m. If the participant asked where she was from, she told him that she is an air hostess and had just arrived in London. Finally she suggested they should meet up again.

Christina's behaviour consisted of a number of multi-modal utterances, containing recorded speech, body and facial animation, which were triggered from a control panel by an experimenter, who was listening to the experiment out of view of the participant. There were around 60 utterances prepared and pre-recorded; half of these are the core of the conversation and the rest of them are backups for unexpected situations. Each utterance is a synchronised combination of speech (audio file) and movements (animation). Our animation engine distinguishes between foreground behaviour, the multi-modal utterances, and background behaviour including gaze and proxemics. We use a model of proxemics that ensures that Christina mostly orients towards the participant and maintains an appropriate conversational distance, but breaking this as time goes on as she gets closer. The following is an extract from a typical conversation: (Christina - C; Participant - P)

C: This shirt looks great on you, how much was it?
P: Thank you! It is a gift.
C: Ah, I really want to find a pair of trousers something like those for my brother [glancing down at the man's trousers]. Where did you get those?
P: Haha, I cannot remember, but there are a lot of nice shops alone the Oxford Street if you are interested.

[1] University College London, London, UK
[2] Goldsmiths, University of London, UK
[3] University of Barcelona, Spain,

C: [Turns her face to the other virtual characters around] So, Do you know anyone here?
P: Well, not really anyone else, no.
C: I feel a bit shy about talking with the other people, do you mind if I talk with you for a bit longer?
P: Sure, no problem.

During the interaction, the participants' behaviour were recorded with a camera from behind. The video recording of the experiment (See Fig.2), together with a assessment form, were then passed on to a body movement expert who otherwise has no involvement with this research and had no knowledge of the purpose of our research. The Form consists of 3 questions each associated with particular events: first, when the virtual female asks about the participant's clothes; second, when the virtual female asks if the participant knows anyone there; third, some participants asked the virtual female to repeat what she has said. In real life these events would trigger the following reactions respectively: first, look down at their clothes (Fig.2(c)); second, look around at other people in the social environment(Fig.2(b)); third, lean forward in order to hear the question more clearly. Therefore the 3 questions we asked on the form were:

- When the virtual lady mentioned the participant's clothes (shirts and trousers), did he *look down* at his shirts/trousers?
- How many the participant ask the virtual lady to repeat; and when asking, how many times of these he *leaned forward* to the virtual lady?
- When the virtual lady asked if the participant knew anyone here, did he *look around* at the other virtual characters?

The answers to those questions would provide us with evidence if the participants have responded towards the virtual characters as if they were real.

(a) (b) (c)

Figure 2. (a), (b) and (c) shows a participant with normal head position, looking around, and looking down.

The body movement expert then filled in the assessment form for each participant by screening the video data from the experiment. The result is presented in Table 1, where for *looked down* and *looked around* we present the number of shy and confident participants who did the movement at the time over all participant observed; for *learned forward*, we present the number of times the participants has learned forward (when asking the virtual character to repeat) against the number of times they asked her to repeat. It can be seen that the majority of participants leaned forward when asking the avatar to repeat a question, and looked around when being asked about other "people". Note that the leaning forward made no sense from an objective point of view since the sound was not spatialised and therefore was not actually coming from the location of the virtual woman. These results suggest that the participants' bodily responses towards the female were similar to how they would respond to a real person.

Table 1. Event triggered body movement: 1, The number of participants who looked down when asked about their clothes / all participants observed. 2. The number of participants who looked around when asked about other "people" /all participants observed; 3. The times participants leaned forward when asking the virtual lady to repeat / times asked the virtual lady to repeat.

	1,Looked Down	**2,Looked Around**	**3,Leaned Forward**
Shy Participant	6/12	11/12	9/9
Confident Participant	7/12	10/12	13/15

The results support the notion that participants tend to act towards the virtual character with appropriate interpersonal behaviour. Bodily movement is a gross overall indicator of a person's state, and therefore could offer us an additional method to measure the responses of people in VEs, especially when the participant interacts with a virtual character. In our previous report [2] we have evaluated the reactions of participants and the results showed that the participants tended to respond to the situation at the subjective and physiological level as if it were real. In the previous study we have also evaluated their verbal responses to assess their behaviour, yet one can argue that verbal responses can be playfully delivered by the participants without being serious. This new evaluation of bodily responses, however, underlines the findings of our previous study, since it is unlikely that people deliberately and consciously choose their bodily responses - they are an automatic action.

REFERENCES

[1] 'Diagnostic and statistical manual of mental disorders', in *American Psychiatric Association*, volume Fourth Edition (DSM-IV).

[2] X. Pan and M. Slater, 'A preliminary study of shy males interacting with a virtual female', in *Presence: The 10th Annual International Workshop on Presence*, (2007).

[3] S. Turner, D. Beidel, C. Dancu, and M. Stanley, 'An empirically derived inventory to measure social fears and anxiety: The social phobia and anxiety inventory', *Psychological Assessment*, (1989).

Comparing Perception of Affective Body Movements displayed by Actors and Animated Characters

Ariel Beck[1,2], Dr. Brett Stevens[1] and Dr. Kim A. Bard[3]

Abstract. This paper reports on a comparative study, which investigated how emotional body language, from animated characters and real actors, are perceived. The results are discussed in relation to the uncanny valley [1], which is a drop in believability as animated characters become more realistic [2]. The results showed that, videos of the actor were found to be more emotional, more believable and more natural than the animated characters whilst displaying the same emotional body language (recorded simultaneously by Motion Capture technology). Moreover, there was a significant difference in the number of correctly interpreted negative emotions displayed. Although, not for positive emotions. This could be due to the physical appearance of the animated character or to the loss of micro-gestures inherent to Motion Capture technology.

1 INTRODUCTION

Virtual environments (VE) for training have been shown to be efficient for skill learning [3, 4] by immersing the trainee in simulated situations. VEs allow participants to interact in a risk free environment. For example, in a medical scenario, a real patient is not required, reducing ethical and practical issues. Moreover, once developed, a VE can provide continuity across assessment contexts, which are then not subject to unpredictable variations arising from human intervention, as could be the case for training using actors. VEs also have many advantages when compared to traditional methods, such as video training, as they are responsive and flexible, adapting to instructions from the trainee in real time, whereas video training is always limited to a pre-recorded set of answers. However, it has been suggested that the key advantage of VE training is that it induces a feeling of presence (i.e. the feeling of being there), providing trainees with experience of situations that are more comparable to those evident in real life [5].

However, research on Animated characters populating the VEs used for the training of medical students has mostly focused on mechanical/surgical tasks. Nevertheless, work is beginning on simulating a wider set of patient interactions as well. This requires a full range of participant behaviours and actions, including speech, lip synchronisation, realistic looking anatomy and appearance, and animated facial expressions, all of which have started to be researched from a functional perspective [6]. However, recent results show the importance of body language as a medium for expressing emotions [7]. Hence, the animated character also requires more subtle behaviours such as non-verbal cues, or body language which would also be essential for training medical personnel in assessing a patient's inner emotional state.

To date, it is not known how users would interact with highly realistic characters and whether they would be considered as social partners. However, it has been shown that the way humans interact with technology is 'mindlessly' social [8]. It was found that technology can trigger social scripts, which typically apply to human-human interaction but are inappropriate for human-computer interaction, as they ignore the essential nature of the technology [8]. Moreover, consistent with this early result, recent studies support the fact that animated characters are indeed perceived as social agents, which trigger natural and social protocols in human users, such as gaze [9] or different acceptance levels based on ethnicity [10]. Nevertheless, these experiments considered mindless social rules and it is not evident that it will hold true at a conscious level.

The aim of the study presented, was to explore if an animated character would be consciously perceived as a social partner. This is done by testing the implicit assumption that viewers would perceive and interpret humans and animated characters in a similar way. Moreover, a similar approach as [8] can be adapted by testing the ability to perceive (i.e. see) and interpret (i.e. attribute meaning to) emotional expressions when displayed through the body [11]. This emotional body language is an ideal start to an investigation as it is known that people can accurately distinguish among emotions when displayed through human body language [12].

2 THE EXPERIMENT

2.1 Design

The main aim of the experiment was to compare emotional interpretation (accurate labelling and perceived strength) of a recorded person and an animation, which were both displaying the same emotional body language (Actor Vs Animated). However, in order to avoid the believability paradox, or Uncanny Valley [1], which is a drop in believability as characters acquire greater similarity with humans [2], traditional animation, such as Disney [13], usually uses stylised display of emotion, which are adapted to the physical appearance of the character. This could effect the perception of the animated display (Participants on seeing an animated character, could expect stylised and exaggerated display of emotions). Therefore the experiment also investigated whether there is an 'uncanny effect' in the display, by comparing the perceived believability and naturalness, of Stylised and Natural displays of emotion

[1] School of Creative Technologies, Univ. of Portsmouth, PO1 2DJ, UK
Email: {ariel.beck, brett.stevens}@port.ac.uk
[2] ExPERT Centre, Uni. Of Portsmouth, PO1 2FR, UK
[3] Centre for the Study of Emotions, Univ. of Portsmouth PO1 2DY, UK
Email: kim.bard@port.ac.uk

(Stylised vs. Natural). Studies on emotional body language recognition [7, 14, 15], consider only small sets of prototypical emotions, which could have resulted in discrimination between emotions (a process of elimination) rather than true recognition. This point is supported by Kramer (2008) who argues that nonverbal behaviours rarely carry specific intrinsic meanings but are interpreted within context [16]. Therefore, this study also investigates the effect of body language in two different situations, one with contextual verbal support and one unsupported (Voice vs. No Voice) on the interpretation of the Actor and Animated display.

Due to the believability paradox, which models a drop in believability as characters become highly realistic, the study investigates whether there is a relationship between perceived physical realism of the characters presented and the believability of the display.

For this study, it was necessary to consider the participants' ability to accurately interpret emotional behaviour, displayed by the actor and the animated character, and how this may affect the results. One of the main components of Emotional Intelligence is the perception of emotions [17]. Therefore its effects, including whether it correlates with the viewer's ability to classify emotion in either the video or the animated display, were investigated. Moreover, previous experience in video games and animated characters could also affect the perception of the animated character. Therefore, the experiments investigated possible correlations between previous experience and the results, including the number of good interpretations, perceived believability and naturalness.

2.2 Participants

25 Participants were recruited, mostly members of staff of the University of Portsmouth (16 females and 9 males) ranging in age from 24 to 61. Participants were randomly allocated to one of the two groups (Voice vs. No Voice). The thirteen participants in the Voice condition ranged in age from 24 to 61 (M=39.30, SD=13.70). The twelve participants in the no voice condition ranged in age from 31 to 60 (M=44.90, SD=9.66). Participants were entered in a raffle in exchange for participation.

2.3 Apparatus

A

B

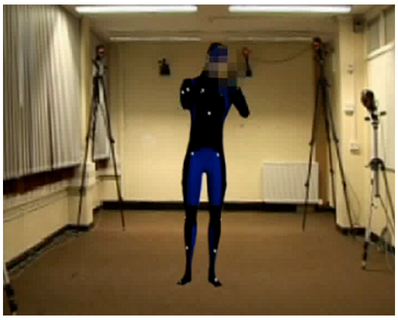


Figure 1. Screen shots of video condition (A) and animated condition (B).

A professional actor and a professional director were hired to build the material. The actor performed the following emotions: Anger, disgust, shame, fear, sadness, surprise, relief, happiness, pride and excitement. Each emotion was performed in two different ways, a natural version and an exaggerated stylised one closer to the one that can be seen in traditional animation. To ensure equivalency across conditions, the actor was video recorded (Figure 1 A) and motion captured simultaneously. The videos were recorded using a Sony PD170P. Motion capture data was recorded using a VICON motion capture system. The motion capture data was then used to animate a character (Figure 1 B) so that it displayed the same body language. The faces of the actor and the animated character were pixelated, so that this source of information was removed along with the possible uncanny effect that may come from facial animations. Moreover, to remove possible effects, such as differences in dress of the actor and animated character, both appeared in a motion capture suit (Figure 1), were physically similar (skin and face are not visible) and were put in the same context (Figure 1). Forty videos were created: 10 emotions x 2 versions x 2 types of display (see table 2).

These videos and animations were displayed on a 5m x 2.5m rear projection screen at life size. To record participants' answers, the material was embedded into custom made software, which was used for displaying the video clip as well as recording participants' answers.

2.4 Materials

The Emotional Intelligence (EI) was measured using the Emotional Intelligence Test-2nd Revision [18].
The physical realism of the actor and animated character were rated on a seven points Lickert scale using pictures showing them in the same 'neutral' position.

In order to record participants' interpretations of the emotional body language displayed, an existing questionnaire has been modified. It is based on the Geneva Emotional Wheel[1] that places twenty emotional labels on a two dimensional axis: valence and control. The Geneva Emotional Wheel is usually

[1] GEW; see [19], downloaded from http://www.unige.ch/fapse/emotion/ on the 01/07/2008

used for self-report (i.e. participants reporting their own emotional state). However, in this study, it has been used to report on participants' interpretations of the clips. The wheel's centre includes two additional options, "no emotion at all" and "none of the above", in case a different emotion is perceived by a participant. Each participant was also asked to indicate the strength for every emotional clip (five point Lickert scale radiating out from the centre).

For each emotional clip, participants were asked to rate the believability and naturalness on a seven point Lickert scale.
The Post-study questionnaire gathered personal information including gender, age, experience of video games and familiarity with animated characters. Finally participants were asked whether they noticed the presence of an animated character and if they found it disturbing.

2.5 Procedure

Participants were tested by the same experimenter in individual sessions. Each session began by obtaining consent, followed by an emotional intelligence questionnaire.

After completion of the questionnaire participants rated the realism of the actor and the animated character. It was necessary to rate realism before interpreting the video clips as if they watched the movements before, it would bias the rating. However, the fact that participants rated static pictures of the actor and animated character before interpreting the video clips may have affected the results. It can be argued that participants were made aware that it was a comparative study before watching the video clips and that this awareness could have biased their answers. However the experiment was advertised as a study on the recognition of emotional body language and the aim only explained during debrief. Moreover, several participants thought that they had been asked to interpret the exact same video clips more than once. This suggests that they were not aware of the comparative hypothesis during the test.

After rating the static pictures, participants started watching and assessing the video clips, which were played only one time. Once all video clips were interpreted, the post-study questionnaire automatically started. The whole procedure took one hour.

2.6 Results

Participants' mean score at the *Emotional Intelligence Questionnaire* was 111.92 out of 150 (SD=10.65). An average of 2.4 out of 7 (SD=1.47) was reported on *Experience in Video Games* and 3.8 out of 7 (SD=1.82) on *Familiarity with Animated Characters*. The mean rate of *physical realism* was 5.52 out of 7 (SD=1.83) for the actor and 2.76 out of 7 (SD=1.56) for the animated character.

Prior to any analysis, one-way ANOVAs have been computed to check if differences confounded the results between the two groups. There was no difference between groups in *EI* (F(1,23)=0.61, p=0.44), *Experience in Games* (F(1,18)=1.53, p=0.23) nor in *Familiarity with Animated Characters* (F(1,18)=1.53, p=0.47)

	Contextual Support	Presentation Condition	
		Acted Videos	**Animated Videos**
Believability		M=4.99 SD=1.01	M=4.38 SD=1.05
Naturalness		M=5.31 SD=0.99	M=4.51 SD=1.03
Perceived Strength		M=3.62 SD=0.63	M=3.33 SD=0.68
Number of Good Interpretations	**Voice** max=20	M=8.54 SD=1.94	M=8.08 SD=2.1
	No Voice max=20	M=5.83 SD=2.85	M=5.50 SD=2.15
Number of Good Interpretations (Negative)	**Voice** max=10	M=6.08 SD=1.32	M=5.08 SD=1.66
	No Voice max=10	M=3.25 SD=1.215	M=2.17 SD=1.115
Number of Good Interpretations (Positive)	**Voice** max=10	M=2.46 SD=1.39	M=3 SD=1
	No Voice max=10	M=2.58 SD=1.78	M=3.33 SD=1.67

Table 1. Means and Standart deviations of the results.

Repeated-measures ANOVA were carried out on the data. Assumptions of normality, homogeneity of variance and sphericity were met. Results showed that differences between conditions were unlikely to have arisen by sampling error. There was a significant difference in *Physical Realism* (F(1,23)=49.37, p<0.01), *Believability* (F(1,23)=12.4, p<0.01), *Naturalness* (F(1,23)=25.4, p<0.01) and *Strength* (F(1,23)=41.785, p<0.01). Repeated measures ANOVA did not show a significant difference in the *Number of Good Interpretations* (F(1,23)=0.9, p=0.4) for the acted and the animated condition, but there was a significant between subjects' effect for voice (F(1,23)=10.7, p<0.01). However, grouping the displays by valence revealed a significant difference for the negative ones (F(1,23)=14.5, p<0.01) for which participants were more accurate in interpreting the actor. There was no difference for the positive displays (F(1,23)=0.9, p=0.4).

To assess the correlation of *Emotional Intelligence* with the *Number of Good Interpretations*, Pearson correlations were computed. There was no correlation between *Emotional Intelligence* and the *Number of Good Interpretations* for the actor (r(23)=-0.3, p=0.09) nor for the animated character (r(23)=-0.2, p=0.1).

Similarly Pearson correlations were computed for *Emotional Intelligence* and *Believability*. There was no relation between *Emotional Intelligence* and the *Believability* of the actor (r(23)=0.27, p=0.10). However, *Emotional Intelligence* was positively correlated with the *Believability* of the animated character (r(23)=0.48, p<0.01). The results are similar with the perceived *Naturalness*, as there was no correlation between *Emotional Intelligence* and the *Naturalness* (r(23)=0.2, p=0.3) of the actor. However, *Emotional Intelligence* was positively correlated with the *Naturalness* of the animated character ($r(23)=0.5$, $p=0.01$).

Pearson's test showed a positive correlation between *Experience in Video Games* and the *Number of Good Interpretations* for the actor ($r(18)=0.37$, $p=0.05$) but not for the animated character ($r(18)=0.17$, $p=0.24$). However, *Familiarity with Animated Characters* was positively correlated with the *Number of Good Interpretations* for the actor ($r(18)=0.64$, $p<0.01$) as well as for the animation ($r(18)=0.38$, $p=0.05$).

There was a positive correlation between *Physical Realism of the Actor* with *Believability* of the actor ($r(23)=0.49$, $p<0.01$) and *Naturalness* of the actor ($r(23)=0.60$, $p<0.01$). However, there was no such correlation between *Physical Realism of the Animated Character* and *Believability* of the animated character ($r(23)=0.04$, $p=0.42$) nor with *naturalness* of the animated character ($r(23)=0.245$, $p=0.12$).

Finally, there was a positive correlation between *Experience in Games* and *Physical Realism of the Actor* ($r(23)=0.39$, $p=0.04$) as well as with *Physical Realism of the Animated Character* ($r(23)=0.75$, $p<0.01$). Similarly *Familiarity with Animated Characters* was positively correlated to *Physical Realism of the Actor* ($r(23)=045$, $p=0.02$) and to *Physical Realism of the Animated Character* ($r(23)=0.54$, $p<0.01$).

Emotion Displayed	**Condition**	**Good Answers in the Acted Condition**	**Good answers in the animated condition**
Anger	Voice	Natural: 13/13 Stylised: 10/13	Natural: 12/13 Stylised: 9/13
	No Voice	Natural: 3/12 Stylised: 10/12	Natural: 1/12 Stylised: 9/12
Disgust	Voice	Natural: 12/13 Stylised: 11/13	Natural: 13/13 Stylised: 9/13
	No Voice	Natural: 0 Stylised: 2/12	Natural: 2/12 Stylised: 1/12
Shame	Voice	Natural: 0 Stylised: 1/13	Natural: 1/13 Stylised: 3/13
	No Voice	Natural: 0 Stylised: 3/12	Natural: 0 Stylised: 0
Fear	Voice	Natural: 7/13 Stylised: 3/13	Natural: 6/13 Stylised: 2/13
	No Voice	Natural: 7/12 Stylised: 3/12	Natural: 4/12 Stylised: 1/12
Sadness	Voice	Natural: 11/13 Stylised: 9/13	Natural: 4/13 Stylised: 7/13
	No Voice	Natural: 4/12 Stylised: 7/12	Natural: 3/12 Stylised: 5/12
Surprise	Voice	Natural: 4/13 Stylised: 4/13	Natural: 3/13 Stylised: 0
	No Voice	Natural: 3/12 Stylised: 2/12	Natural: 2/12 Stylised: 1/12
Relief	Voice	Natural: 5/13 Stylised: 5/13	Natural: 6/13 Stylised: 9/13
	No Voice	Natural: 2/12 Stylised: 5/12	Natural: 5/12 Stylised: 6/12
Happiness	Voice	Natural: 1/13 Stylised: 2/13	Natural: 4/13 Stylised: 2/13
	No Voice	Natural: 1/12 Stylised: 4/12	Natural: 0/12 Stylised: 7/12
Pride	Voice	Natural: 0/13 Stylised: 3/13	Natural: 0/13 Stylised: 7/13
	No Voice	Natural: 0/12 Stylised: 4/12	Natural: 1/12 Stylised: 6/12
Excitement	Voice	Natural: 2/13 Stylised: 6/13	Natural: 0/13 Stylised: 8/13
	No Voice	Natural: 4/12 Stylised: 6/12	Natural: 5/12 Stylised: 7/12

Table 2. Proportion of good answers per emotions and conditions (13 participants in the Voice condition and 12 in the no voice condition). Chance level would be 1/22.

As shown in Table 2, recognition rates depended on emotion and condition, although most of the clips were recognised above chance level.

3 DISCUSSIONS

The results showed no difference in the number of video clips that were correctly interpreted between the actor and the animated character. However, it cannot be concluded that the interpretation of an actor and an animated character is similar, as the animated character has been perceived as less natural, less believable and less emotional. Moreover, grouping the displays by valence shows that for the negative emotions, participants were better at interpreting the actor. There was no difference for the positive emotions. It is important to emphasize, that these differences occurred with animated movements that were captured from the acted performance. However, four participants out of twenty-five reported not noticing the presence of animations, whilst only one participant reported being disturbed by the presence of animation.

The fact that no difference was found on the number of good interpretation may be due to a floor effect on the positive displays (i.e. Excitement, Happiness, Pride, Surprise, Relief): The recognition rate of these videos was very low in both conditions (See table 1). This may be due to the fact that it was impossible for participants to distinguish between so many positive emotional states watching emotional body language even with voice as well. However, this low recognition rate is consistent with Ekman's basic theory of emotion [20], which considers that happiness is at the core of all positive emotions. The basic theory of emotion suggests that the positive displays are very hard to accurately distinguish because of their similarities.

Motion Capture is not perfect; it fails to capture secondary animations such as skin deformation. This could explain why participants' interpreted more accurately the actor displaying a negative emotion than the animated character. This is shown by the significant difference for the video clips where a negative emotion (i.e. Anger, Fear, Disgust, Sadness, Shame) was expressed. Similar differences were found in the strength of the emotion, in believability and in naturalness where it was found that the actor was more emotional, more believable and more natural than the animated character. It can be suggested that these differences are due to the gestures that were not recreated by the Motion Capture Technology.

However, the differences in the physical appearance of the actor and the animated character could also explain the significant drop in believability and naturalness (Figure 1). The highly significant difference between the *realism of the actor* and the *realism of the animated character* (which were rated using pictures) revealed that the animated character failed to reach the right side of the Uncanny Valley. Nevertheless, it can be argued that the animated character used for this study is at least as realistic as the one that would be used for medical interview training. These characters have to be animated in real time and thus would also be susceptible to the same drop in believability and naturalness.

Participants who scored high on the Emotional Intelligence questionnaire were less affected by the missing micro-gestures and the lack of realism of the animated character. This was shown by the positive correlation between emotional intelligence, believability and naturalness of the animation. However, the lack of a correlation between Emotional Intelligence and the number of good interpretation could indicate that either the questionnaire failed to assess participants' ability to accurately interpret emotional body language or that the videos presented lacked the appropriate emotional cues. The second possibility can be dismissed by the fact that for almost all emotional clips the recognition rate is above chance level (Table 2).

The positive correlation between familiarity with animated character and the number of good answers in both conditions is very encouraging for medical interview training. It shows that participants who reported being familiar with animated characters were more accurate in interpreting the video clips including the acted ones. However, it is not possible to conclude that these participants learned from previous experience how to interpret emotional body language. The positive correlations between experience in video games and familiarity with animated character with the physical realism of the actor and animated character could simply imply that participants' familiar with animated character were less affected by the appearance of the actor and the animated character which was different from every day displays: they wore the motion capture suit and their faces were pixelated. This could also explain why participants familiar with animation were more accurate. Therefore, it will be necessary to determine whether it is linked to the appearance and whether it is related to specific types of experience with animated characters (different type of animations, interaction…).

Nevertheless, these differences were not unexpected and it can be argued that animated characters cannot be, in terms of believability and naturalness, as good as videos of a real person. However, the results show that participants were able to accurately interpret many emotions displayed by the animated characters.

4 CONCLUSION

The results clearly established that emotional body language animated with motion capture is significantly different when compared to the video motion. It was not possible to clearly establish if this loss was due to the micro-gestures that Motion Capture failed to recreate or if it was due to the physical appearance of the animated character, which was poorly rated when directly compared to the actor.

The next research step will investigate the effect of changing the level of physical realism as well as the effect of reintroducing the secondary animations that were missing. It is expected that increasing the level of realism will affect the difference in believability, naturalness and will explore the real shape of the Uncanny Valley. Whereas, secondary animations should minimize the difference in the number of good interpretations as they should give complementary cues necessary to accurately interpret the display.

This line of research is important because the feeling of presence is crucial for medical training to be efficient. Therefore, in order to be useable animated characters need to overcome the believability paradox.

ACKNOWLEDGEMENTS

This project is funded by the ExPERT Centre at the University of Portsmouth as part of a PhD programme. The authors would like to thank Dr. Guy Van De Walle for his advices as well as the director Matt Ghrist and the actor Teo Ghil for their excellent work. Kim A. Bard was partially supported by FEELIX GROWING, an European Commission FP6 project, IST-045169.

REFERENCES

[1] M. Mori, "Bukimi no tani [The un-canny valley] ." *Energy,* vol. 7, pp. 33-35, 1970.
[2] H. Brenton, M. Gillies, D. Ballin, and D. Chatting, "The Uncanny Valley: does it exist and is it related to presence ?," *Presence-connect,* vol. http://www.presence-research.org/, April 2005 2005.
[3] R. Aylett, A. Paiva, S. Woods, L. Hall, and C. Zoll, "Expressive Characters in Anti-Bullying Education," *publication13280,* 2005.
[4] W. Swartout, J. Gratch, R. W. Hill, E. Hovy, S. Marsella, J. Rickel, and D. Traum, "Toward virtual humans," *AI Magazine,* vol. 27, pp. 96-108, Sum 2006.
[5] M. Slater, A. Antley, A. Davison, D. Swapp, C. Guger, C. Barker, N. Pistrang, and M. V. Sanchez-Vives, "A virtual repirse of the Stanley Milgram obedience experiments," *PLoS ONE,* vol. 1, December 2006 2006.
[6] V. Vinayagamoorthy, M. Gillies, A. Steed, E. Tanguy, X. Pan, C. Loscos, and M. Slater, "Building Expression into Virtual Characters " in *Eurographics 2006*, Vienna, 2006.
[7] B. de Gelder, "Towards the neurobiology of emotional body language," *Nature reviews Neuro-Science,* vol. 7, pp. 242-249, 2006.
[8] C. Nass and Y. Moon, "Machines and mindlessness: Social responses to computers," *Journal of Social Issues,* vol. 56, pp. 81-103, 2000 2000.
[9] H. Prendinger, C. Ma, and M. Ishizuka, "Eyes movements as indices for the utility of life-like agents: a pilot study," *Interacting with Computers,* vol. 19, pp. 281-292, 2007 2007.
[10] J. Pratt A, K. Hauser, Z. Ugray, and O. Patterson, "Looking at human-computer interface design: Effects of ethnicity in computer agents," *Interacting with Computers,* vol. 19, pp. 512-523, 2007.

[11] S. Pichon, B. De Gelder, and J. Grèzes, "Emotional modulation of visual and motor areas by dynamic body expressions of anger," *Social neuroscience,* vol.?, p.?, ? 2007.
[12] J. V. den Stock, R. Righart, and B. de Gelder, "Body expressions influence recognition of emotions in the face and voice," *Emotion,* vol. 7, pp. 487-494, 2007.
[13] F. Thomas and O. Johnston, *The illusion of life*. New-York: Abbeville Press, 1995.
[14] J. V. den Stock, J. Grèzes, and B. de Gelder, "Human and animal sounds influence recognition of body language," *Brain Research,* vol. in press, 2008.
[15] A. P. Atkinson, W. H. Dittrich, A. J. Gemmell, and A. W. Young, "Emotion perception from dynamic and static body expressions in point-light and full-light displays," *Perception,* vol. 33, pp. 717-746, 2004.
[16] N. C. Kramer, "Nonverbal communication," in *Human behavior in military contexts*, J. J. Blascovitch and C. R. Hartel, Eds., 2007.
[17] J. W. Kalat and M. N. Shiota, *Emotion*. Belmont: Thomson Wadsworth, 2007.
[18] I. Jerabek, "Emotional Intelligence Test - 2nd Revision," PsychTests.com, 2001.
[19] K. R. Scherer, "What are emotions? And how can they me measured?," *Social Science Information,* vol. 44, pp. 693-727, 2005.
[20] P. Ekman, "An argument for basic emotions," *Cognition and Emotion,* vol. 6, pp. 169-200, 1992.

Computational model and the human perception of emotional body language (EBL)

Marina Fridin[1], Avi Barliya[1], Edna Schechtman[3], Beatrice de Gelder[2], Tamar Flash[1]

Abstract. Emotions are undeniably a central component of human existence. In recent years, the importance of developing systems which incorporate emotions into human-computer interaction (HCI) has been widely acknowledged. However, research on emotion recognition has been dominated by studies of facial expression of emotion. In comparison, the study of EBL has received relatively little attention.

Here we study the phenomena of EBL, specifically of *static body postures expressing emotions*, from two different perspectives. First, we have built a computational model for the recognition of four basic emotions which achieves a relatively high recognition rate (70 %). Secondly, to study perception of EBL, we examined what body parts attract the observer's attention during the perception of EBL. This is done by tracking eye movements of human subjects during the observation of static postures expressing emotions. Although invaluable information can be inferred from motion, this study will show that information about static body posture is rich enough for both automatic recognition and human perception.

The present study contributes in an applicative way both to the development of automatic recognition systems of EBL and provides insight into the nature of human recognition of EBL.

1 INTRODUCTION

Research on EBL is rapidly emerging as a new field in Human-Computer Interaction (HCI) and affective computing.

Darwin in his groundbreaking work on the evolutionary significance of emotion stressed the importance of bodily movement [4]. Nonetheless, the dominant theory of Ekman [8] denied that specific gestures, body movements, or postures could indicate emotions. Contrary to the above claims, Wallbott a pioneering researcher in the field study of EBL [16] , has demonstrated that specific body movements, postures and gestures, generally allow to differentiate among different emotions.

Studies of EBL are still rare but it has been already repeatedly shown that distinct expressions of at least the basic emotions are readily recognized even in the absence of facial and vocal cues, when emotions are portrayed by static body postures and by whole body movements, as was described in [1], [7] and [16] . In recent years, Pollick and colleagues [14] have investigated the perception of EBL through movement and have shown that the speed of movement plays a major role in the perception of EBL. In the study by Cammuri [4] of emotional dance performance, hand configurations and locations with respect to the body were also indicated as important features for emotion discrimination. The neural mechanisms underlying perception of EBL were examined by de Gelder et al. [7] , where in a fMRI study they have shown that the amygdala is a key structure in the processing of static, fearful body expressions, along with the fusiform cortex.

In this study we propose a novel computational approach towards the understanding of EBL. The strength of the proposed model is that it is highly modular, that is, it is independent of the database being examined and the feature extraction processes.

We have also studied of the perception of EBL by investigating gaze patterns of human subjects when perceiving EBL of other individuals. In this study, the first to describe such gaze patterns, we examined whether fixation behavior is emotion-specific. Fixation behavior was described in terms of which body parts receive most visual attention. We also examined whether fixation behaviour remains stable under different task conditions, such as observation vs. recognition, for images which were correctly vs. incorrectly recognized, and as a function of the familiarity with the stimuli, which may result in perceptual learning

As it is impossible to cover a large spectrum of human emotions, in this study we have strived to characterize and model the four basic emotions: ***JOY***, ***SADNESS***, ***ANGER***, and ***FEAR***.

The paper is organized as follows. First, we describe the database of EBL that was used for both parts of the research (Section 2). Next, in Section 3 we discuss how EBL is perceived and recognized. In Section 4, we present the computational model of EBL. Section 5 presents our conclusions and contrasts this paper with related works. Section 6 provides directions for future work.

2 EBL DATABASE

The input material in both studies consisted of still images derived from video films in which 80 semi-professional actors and 47 ordinary subjects freely portrayed body postures expressing four basic emotions. The subjects were of different genders, cultures, ages and socioeconomic status. The images from both groups were randomly mixed together. Therefore the entire database consists of 508 images (127x4). In the eye tracking study the faces on the photographs were blurred.

[1] Dept. of Computer Science and Applied Math., Weizmann Institute of Science, Rehovot, Israel. Email: {**marina.ousov-fridin**, **avi.barliy**, **tamar.flash**}**@weizmann.ac.il**.

[2] Lab. of Cognitive and Affective Neuroscience, Tilburg Univ., the Netherlands. Email: {b.degelder}@uvt.nl.

[3] Dept. of Industrial Engineering and Management, Ben Gurion Univ. Beer Sheva, Israel. Email: {ednas}@ bgu.ac.il.

3 GAZE AND THE PERCEPTION OF EBL, AN EYE TRACKING STUDY

In this study we examined eye fixation patterns during the perception and recognition of EBL. Participants were asked to perform two tasks while wearing an eye-tracking system. In the first task (observation task), subjects were asked just to look at photographs and in the second (recognition task) to tell verbally which emotion they have recognized.

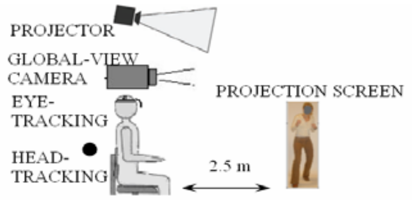


Figure 1. Schematic representation of the experimental set up.

An "ISCAN" eye-tracking apparatus was mounted on the participant's head (Figure 1) and was connected to the eye-tracking computer. To estimate global gaze direction, the eye-tracking apparatus was coupled with a Polhemus Liberty head-tracker. A separate computer projected the stimuli on the screen and captured the presentation data. A global view camera captured the scene observed by the subject and produced a video showing the scene and the estimated gaze directions

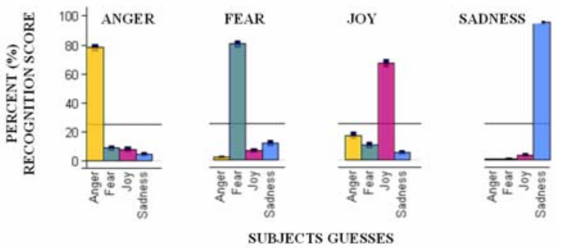


Figure 2. The recognition score (RS) for each emotion shown in static postures presented to the subjects. Subjects' errors in the recognition task are also shown. RS of all emotions were clearly well above the 25% chance level (black line).

We first evaluated subjects' recognition rate of EBL in the recognition task. Figure 2 shows the recognition accuracy for each emotion (when the emotion is actually shown) and the respective confusion rate. Choices were not made randomly: the emotions shown were very frequently recognized, with sadness being the most easily recognizable emotion, with a recognition rate of 94.83% $\pm$ 1.19%. The other emotions were also well recognized, when the most confused emotion being fear, which was confused with sadness (11.54% $\pm$ 1.77%), and joy, which was confused with anger (17.11% $\pm$ 2.13%).

Response times (RT) for emotions that were incorrectly recognized were significantly longer than the response times for correctly recognized emotions (means 38543 $\pm$ 29329 msec (n=134) and 27004 $\pm$ 22720 msec (n=586), respectively).

We then quantified and performed a statistical analysis of body regions on which the eyes fixated. There were significant differences among the eye fixation patterns observed for the different emotions (Figure 3). Most attention during the perception of images showing anger was paid to the hands (7.32 $\pm$ 1.64) and the arms (4.28 $\pm$ 0.83), with similar observations for fear. For images showing joy most of the time was spent gazing at the head (3.55 $\pm$ 0.89), while the hands and the arms also attracted significant attention. The fixation times for images showing sadness were distributed equally between the head, the arms and the hands. The trunk and the legs never attracted significant attention. There was no change in the fixation behavior in the observation versus the recognition task, and while viewing correctly vs. incorrectly recognized images.

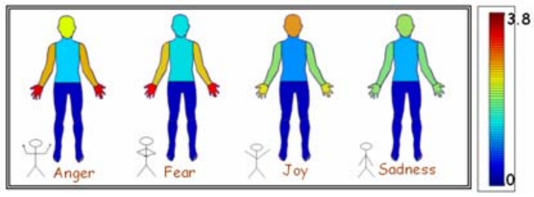


Figure 3. Distribution of fixation behavior on body segments for each emotion during the recognition task. Fixation behavior is represented by color: from blue (marked by 0) to red (3.8 and more).

These fixation patterns were not influenced by the task and were not correlated with recognition performance nor with RT. Our results suggest that fixation behavior is emotion-specific and is automatically triggered by the stimuli. The observed differences in recognition accuracies and RTs could not be explained by the observed fixation behavior. Finally, by analyzing changes in recognition accuracy, RT, and fixation patterns over task repetition we concluded that no perceptual learning could be detected.

4 A COMPUTATIONAL MODEL

We applied a heuristic approach for the modeling of EBL. To that end, we assumed that a set of well defined features with the aid of some combinatorial syntax would be sufficient to represent any EBL. We identified the major feature types, which we then extracted from the pictures as follows.

Figure 4. Extracted features: ***a***: Body Joint Angles; ***b***: Gesture; ***c***: Body Silhouette

The photographs were treated with different image processing algorithms, which included body segmentation from the background, labeling body parts, building a multi-joint human body model, estimating head position and hand gestures. Since some of these image processing steps are very complicated to be automatically performed, part of the analysis was performed manually. The feature types that were extracted are: head pose characterized by the direction of the normal vector to the plane of the head, using an idea suggested by Gee&Cipolla [10] , upper arms and trunk joint angles, (as illustrated on the Figure 4a) legs joint angles, right/left hand gestures (Figure 4b) body silhouette (Figure 4c), defined by a polygon and body space. On the whole, there were ***seven different feature types.*** These feature types are commonly used in the literature dealing with recognizing emotions [16] , imitation [11] and affective computing [4]

Then, an algorithm that computes the maximum mutual information between the different features and the various emotions (Equation 1) was developed [15] . Mutual Information (MI) between class C and feature f_i is defined as:

$$\mathrm{I}(\mathrm{C}; f_i(\theta_i)) = -\sum_{C}\mathrm{P(C)Log(P(C))} + \sum_{F} P(f_i)\sum_{C}\mathrm{P}(\mathrm{C}\,|\,f_i)\mathrm{Log}(\mathrm{P}(\mathrm{C}\,|\,f_i))$$

By measuring frequencies of detecting f_i for images within different classes C, we could evaluate the mutual information of the feature with respect to the class. Examples of the most informative features of type "Body Silhouette" for the class Joy are shown on Figure 5.

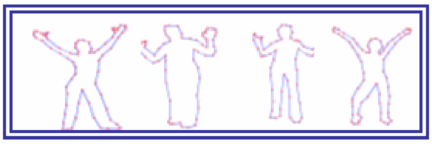


Figure 5. Examples of the most informative features of type "Body Silhouette" for the class of emotion Joy.

Then we re-arranged the lists of most informative features so that each feature added to the list would supply the maximum possible additional MI concerning the emotional class (Equation 2). The selection process was initialized by selecting a feature that provided maximum MI: $f_1 = \arg\max I(C; f_i)$

Therefore, each of the following selected features was determined by:

- First, to make sure that information contributed by any candidate feature f_{k+1} is not already represented by any previously selected feature f_i , for each f_{k+1} feature we found the most similar feature f_i (min stage), by computing the conditional MI between the new feature f_{k+1} and the class given the most similar feature f_i : $I(C; f_{k+1} \,|\, f_i)$;
- Then, from all candidate features f_{k+1} the feature with the maximum contribution (i.e. maximum conditional MI) was selected (max stage) as follows:

$$f_{k+1} = \arg\max_{f}\min_{i} I(C; f_{k+1} \,|\, f_i)$$

Then, different feature types were combined into one list, based on calculating the additional MI. The result of the training process consisted of sets of lists which included all possible combinations of the feature types, for each class of emotion. In each list, features were arranged according to their additional *MI*. The results of this stage are illustrated in Figure 6: the six most informative features of type ***Body Joint Angles*** are shown for the class Joy (the first row) in a descending order, whereas the order is changed (Max-Min, second row) when considering the highest additional information score for each of the features.

To recognize a new image *I*, we first measured the function S_I, indicating whether the features from a currently examined list were found in the image or not. Then, to check to which class the image *I* belongs, we need to know the distribution of the functions S of images from the training set and compare it with the computed function S_I. Therefore, for each of four classes of emotions, and for each image in the training set the distribution of S given C_j class of emotion $p(S|C_j)$ was estimated in advance by using the EM algorithm. A mixture of Gaussian model was chosen as an appropriate model for representing the distribution of S. Finally, the image *I* was assigned to a specific class of emotion, j, according to the class for which $p(S_I|C_j)$ is maximal.

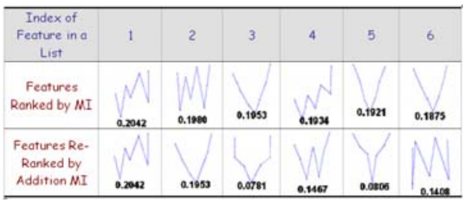


Figure 6. An example showing the first two steps of the algorithm, MI and Max-Min**.** The six most informative ***Body Joint Angles*** are shown for the joy emotion on the first row in a descending order of MI, whereas the sequence of the features is changed (Max-Min, second row) when considering the most additional information provided by each of the features.

We evaluated the recognition rate for the different lists consisting of different feature combinations, and for different sets of features, a few examples of which are presented in Table 1. For each list, we present the average recognition rate (RecRate), taken across all the different classes of emotion, and the required number of features in the combination that maximizes the recognition rate. Using only BODY SILHOUETTE achieves a recognition rate of 46.83% averaged over all emotions. However, using a combination of BODY SILHOUETTE, BODY SPACE, LEFT HAND, and HEAD the best recognition rates: a 70.25% averaged across all emotions was achieved. It should be emphasized that this is a relatively high recognition rate, considerably above chance levels, using only static images as input.

Feature Types Combinations	**Rec. Score (%)**	**Num. Features**
Body Silhouette (BS)	46.83	12
Left Hand	43.65	96
BS , Body Space, Left Hand, Head Norm	70.25	396
BS, Body Joint Angles, Left Hand, Head Norm	69.45	285
BS, Body Space, Body Joint Angles, Right Hand	68.25	196

Table 1: Examples of lists with mixed feature types. For each list, we present the average recognition score and the number of features in the list (over all possible number of features) that maximizes the recognition rate.

This work resulted in a model for the recognition of EBL in static images. For a given new image, even if some features cannot be extracted, the model defines the combination of feature sets that has the highest probability of predicting the expressed emotion in the picture.

5 CONCLUSIONS & RELATED WORK

The results of both parts of the presented study, e.g. the computational model and the analyse of fixation behaviour, are consistent with each other. First we have compared the recognition rates achieved by the computational model with the recognition rates of human observers using the same data set.

We have found, on the average that human observers can recognize emotions with an accuracy of 82%, while the recognition scores of the algorithm were around 70%. That is, human performance is not extremely better than the performance by the algorithm. . Similar recognition scores were reported by a number of EBL studies for both human perception [5] and computational models [4, 1]. Nevertheless, human observers may use more information than is available to our model and the ability to extend and improve the model will be discussed below.

Here for the first time eye tracking has been used to examine fixation behavior during perception of EBL When perceiving pictures associated with joy, people tend to fixate on the head, whereas for pictures associated with anger and fear, most attention is devoted to the hands and arms. The feature involving the hands in the eye tracking study encoded information regarding both hand position and configuration. For pictures associated with sadness, people fixate on all body parts: heads, arms, and hands. The legs almost never drew the observers' attention. Also, the effects of hands' and arms' locations were statistically analyzed and it was found that the fixation behavior is emotion rather than location dependent (e.g. the location within the picture: at the center vs. on the sides of the body). For instance, the location of the hands at the center is a feature that is common for both fear and sadness; however, the amount of time that subjects gazed at the hand was considerably longer for fear versus sadness. These results are consistent with features that were found important for the recognition of EBL, both by our computational model and in a number of related studies [1, 4]. It was shown that hand position and configuration, and the orientation of the head are important features for the recognition of EBL.

Finally, it is important to mention perceptuo-motor primitives [9] [11] and their relation to our study. Perceptuo-motor primitives can be defined as a basic set of body segments and postures that can be learned from perceptional behavior (including fixation behavior) and mapped onto the elements of a recognition system (both human and automatic). Therefore, the sets of emotion-specific body segments and postures, identified in our eye-tracking study, were also identified as important features by the computational model. Hence these body segments and the different features characterizing the location and configuration of the body parts which are associated with the different emotions could be defined as emotion-specific perceptuo-motor primitives of EBL.

6 POSSIBLE APPLICATIONS AND FUTURE WORK

The possible applications of the presented computational model to the automatic recognition of EBL are numerous, ranging from the development of robotic systems capable of mimicking and expressing emotions, through having more natural looking characters in computer games (virtual worlds) and animated movies.

In addition, our data suggest that many more experiments are needed in order to further clarify the mechanisms sub-serving the perception of EBL. For example, to unravel the processes involved in the recognition of EBL, the influence of face versus the body on the eye fixation and recognition behavior could be studied by using stimuli which combine incongruent face and body EBL stimuli. Employing this approach in a recent fMRI study, conducted by de Gelder's group [11], it was found that the ability to correctly recognize different facial expressions is strongly influenced by EBL. Another avenue for future research is to investigate the fixation and recognition behavior during perception of dynamic stimuli where information about the expressed emotion may be gathered by focusing on movement kinematics.

We also suggest that several of the widely used paradigms should be reevaluated with respect to EBL, e.g., the influence of affective priming on fixation and recognition behavior [13]; the influence of gender [2] and of the symmetry of different body parts [3] and their effects on the perception and fixation behavior during the observation of EBL.

Acknowledgements: This research was supported in part by the part by grants from the Frontier of Science Program HFSP-RGP0054/2004-C and FP6-2005-NEST-Path Imp 043403,. Tamar Flash is an incumbent of the Dr. Hymie Moross Professorial chair.

REFERENCES

[1] A. Atkinson, W. Dittrich, A. Gemmell, A. Young, Emotion perception from dynamic and static body expressions in point-light and full-light displays, *Perception*, 33: 717-746 (2004).

[2] L. Barrett, L. Robin, P. Pietromonaco K., Eyssell, Are women the "more emotional sex?" Evidence from emotional experiences in social context. *Cognition and Emotion*, 12: 555-578 (1998).

[3] W.Brown, L. Cronk, K. Grochow, A. Jacobson, C.K. Lu, Z. Popovic and R. Trivers. Dance reveals symmetry especially in young men. *Nature*, 438: 1148-1150 (2005).

[4] A. Camurri, B. Mazzarino, M. Ricchetti, R. Timmers, G. Volpe. Multimodal analysis of expressive gesture in music and dance performances. *"Gesture-based Communication in Human-Computer Interaction", LNAI 2915, Springer Verlag* (2004).

[5] C. Darwin, The Expression of the Emotions in Man and Animals ,Univ. Chicago Press (1965).

[6] B. deGelder, Towards the neurobiology of emotional body language, *Nature Reviews: Neuroscience*, 7: 242–249 (2006)

[7] B. de Gelder, N.Hadjikhani. Non-conscious recognition of emotional body language. *Neuroreport, 17*(6): 583-586 (2006).

[8] P. Ekman, Facial Expressions, *In T. Dalgleish and T. Power (Eds.) the Handbook of Cognition and Emotion., 301-320. Sussex, U.K.: John Wiley & Sons, Ltd.*, 1999.

[9] T. Flash, B. Hochner, Motor primitives in vertebrates and invertebrates, *Current Opinions in Neurobiology*, 15: 660-666(2005).

[10] A. Gee , R. Cipolla, Determining the gaze of faces in images. *Image and Vision Computing*, 12(10): 639-647 (1994)

[11] M. Mataric, Sensory-motor primitives as a basis for imitation: linking perception to action and biology to robotics. In: Nehaniv, C. and Dautenhahn, K., Editors, *Imitation in animals and artifacts*, MIT Press (2001).

[12] H. Meeren, C. van Heijnsbergen, B. deGelder, Rapid perceptual integration of facial expression and emotional body language. *Proc. Natl Acad. Sci. USA*, 102:16518–16523 (2005).

[13] S. Murphy, R. Zajonc, Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. Journal of Personality & Social Psychology, 64(5), 723-739 (1993).

[14] Paterson H. M., Pollick F. E., Sanford A. J., The role of velocity in affect discrimination, *Twenty-Third Annual Conference of the Cognitive Science Society*.(2001)

[15] S. Ullman, M. Vidal-Naquet, E. Sali, Visual features of intermediate complexity and their use in classification, *Nature Neuroscience*, 57: 682–687 (2002).

[16] H.Wallbott, Bodily expression of emotion. *European Journal of Social Psychology*, 28: 879-896 (1998).

Audio Description of Emotions in Films using Eye tracking

Paula Igareda and Alejandro Maiche

Abstract. Dealing with the Audio description (AD) and taking into account the many works and studies carried out in the field, the aim of this paper focuses on the moment when emotions and gestures must be described —avoiding any personal interpretation by the one who describes the audiovisual material. In this sense, we will carry out an experiment evaluating the eye movement from two groups of people: one watching a film excerpt with AD and another one watching the same film excerpt without AD. The aim of this paper is to present the on-going research on the eye-tracking technology applied to the media accessibility.

1 INTRODUCTION

AD is a recent field within Translation Studies, though we already have access to many guidelines for doing it properly. AD is defined as a feature that makes performing arts, audiovisual products, countries' natural, historic and cultural heritage, and all kinds of socio-cultural events accessible to everyone. AD consists of "a set of techniques and abilities, whose main objective is to compensate for the lack of perception of the visual component in any audiovisual message, by providing suitable sound information which translates or explains, in such a way that the visually impaired perceive the message as a harmonious work which is as similar as possible to that which is perceived by the sighted" [1].

Taking some existing studies and guidelines as ITC [2], Benecke [3], Remael [4], Ofcom [5], Orero [6], Puigdomènech, Matamala and Orero [7], Remael and Vercauteren [8], Vercauteren [9] into consideration, it seems that there are some recommendations and ideas on what, when and how to audio describe.

When analyzing a given piece of work we have to take into account both the non-verbal visual elements and the written visual elements. When we study the verbal visual elements we have to deal with the *where* (places and spaces, dimension of the places, location of the characters, lighting, relation between the characters and the places); *when* (the time/period when everything happens); *who* (the characters in general, describing who talks, their clothing, their physical appearance, their facial expressions, body language, age, attitudes); and *what* (action, colours, music and other noises). In relation to the written visual elements, important items are the closing credits, the possible subtitles appearing on the screen, the written elements belonging to the piece of work, etc. [7]. The above studies also focus on how to audio describe, dealing with the objectivity and subjectivity during the process of AD, the elements of motion and the filmic imagery (such as symbols, associative images, etc.).

The issue that is common to all countries and guidelines developed for AD is the lack of any scientific research for its practice; hence, there is not any agreement on how to produce good quality and meaningful AD. On one side, there is not enough research on this topic and that affects the quality of ADs. On the other side, we still find that one of the biggest challenges when doing any AD emerges when emotions and gestures must be described and the it would be desirable to avoid any personal interpretation by the one who describes the audiovisual material.

2 CULTURAL AND TRANSLATION STUDIES

For this study, we have considered Cultural Studies, especially in relation to Translation Studies and Studies in Psychology.

Many studies confirm that there is a series of elemental obstacles in order to have a successful intercultural perception and understanding in translation. According to Samovar and Porter [10], these factors are:

- *attitudes*, which include religion, politics, customs, food, dress, or art; the world view perspectives that may use different frames of reference; the absolute values regarding good and bad, ugly and beautiful, true or false, etc.; or the stereotypes and prejudices;
- *social organization*, as regards government, family, law, etc.
- *patterns of thought*, mainly between the Western and the Oriental worlds, but also between Anglo-Saxon and Latin ways of thinking about the family ties, moral values, dating, etc.; which cause so much confusion;
- *role expectations* obviously derived from patterns of thought;
- *conceptualization of time*;
- *handling space*;
- *language barrier*; which does not arise only with respect to correct or incorrect translation;
- and *non-verbal communication*, nonverbal stimuli in a communication setting that are generated by both the source [speaker] and his or her use of the environment and that have potential message value for the source or receiver [listener].

In this way, we can see that apart from verbal communication, physical appearance, and the use of things and the environment, among many other issues, it is possible to find other cultural signs with a communicative value. This communicative non-verbal activity within a conversation's verbal behaviour comprises a system of signs, gestures, manners, postures, ways of behaving, attitudes, etc. Some studies say that not every culture has the same way of expressing its feelings and emotions, and that not every human being uses the same gestures: Everyone experiences emotions in a particular way, depending on his/her previous experiences, learning, character, culture and situation [11, 12].

2. 1 Non-verbal Communication

Non-verbal communication can facilitate a large amount of information in a given situation or context. Cestero [13] classifies the non-linguistic elements into the following categories: the basic systems (paralinguistic and kinesics); the cultural systems (cronemic and proxemic); and the physical and cultural signs systems.

According to Poyatos [14], the **paralinguistic** system will be determined by the non-verbal communicative activities accompanying verbal behaviour during a conversation. This system is composed of non-verbal vocal signs that work towards communication and are part of the communicative interaction. On the other hand, the **kinesic** system refers to the study of the significant body movements, including facial expressions. This system embraces the analysis of gestures, manners and postures: *Gestures* are movements with communicative value which cover facial expressions (including glances) and body movements; *manners* are the ways in which movements are expressed and they show how people do certain gestures or postures according to their culture, age, sex, emotional state, socioeconomic level, etc.; and *postures* refer to the static body positions that are able to communicate in a conscious or unconscious way. These three categories are related to each other and yet they are independent.

Poyatos also distinguishes the following cultural signs with communicative value: physical appearance, the use of artifacts and the environment. The **cronemic and proxemic** systems are part of the cultural signs with the ability to modify or add significance to the rest of the non-verbal systems or to act alone. These systems have, as a main objective, to study human beings' use and perception from space and time, respectively [13, 11].

2.2 Emotions and Gestures

At the beginning, we mentioned that in a conversation between at least two persons the communicative non-verbal activity comprises a system of signs, gestures, manners, postures, ways of behaving, attitudes, etc. We base our conclusion on the assumption that non-verbal communication has the ability to express a wide range of feelings, emotions and thoughts; and that there are certain emotions associated with certain gestures. In this way, we will deal with the relations between emotions and gestures in this section.

Studying works on emotions we can observe that some of the physiological and behavioural reactions derived from emotions are innate and some of them can be acquired. It appears that every emotion has its functions for different reasons, and has its consequences on the nervous system and on body expression. According to the study of Hager [15], the component that can be the core of common sense approaches to emotion is the feeling component. But this author also mentions other descriptive components of emotions such as the set of behaviours that may be performed and observed in conjunction with an emotion like the gross behaviours of the body affected by the skeletal muscles and emotion expression; the set of internal bodily changes caused by the smooth muscles and glands; the ideation, imagery, and thoughts that occur along with emotion; the circumstances that give rise to emotions; and the neural processes that underlie much of the preceding activities.

Focusing on facial expressions, Hager, between many others, shares his opinion saying that these expressions are associated with particular human emotions and that they can provide accurate information about emotions. Hager and Ekman [16] work on how emotions and faces relate to each other, if they relate to each other at all, if the facial expressions are culturally bound or universal, etc.

Regarding the cultural relationship between emotions and facial expressions, we found works by many researchers showing the emotions on people's faces in a similar way around the world. Ekman, Sorenson and Friesen [17] and Izard [18] maintain the idea that "there are constants across cultures in the emotional meanings of certain facial expressions". On the other hand, others say that "the fact that universal connections between expressions and emotions could arise from learning which has a high probability of occurring in all cultures or from a functional role of the movements in the emotional situation" [16].

Although it is quite clear that there are universals in emotions and gestures, it is also important to bear in mind when dealing with AD that emotions and gestures have their own cultural associations and, therefore, we cannot stereotype; they are an integral part of individual and collective culture-transmitting values. If we want to fully understand the meaning of an expression, we have to be able to interpret facial expressions, gestures and body language.

In order to successfully interpreting this non-verbal communication, we focus on the possible problems we can find because of these cultural differences in body expression. According to Poyatos [14], we find out that:

- the way people make their gestures can vary between cultures;
- one gesture can be characteristic of one culture and it is possible that this gesture does not exist in the other culture; that is: it has no gesture equivalent;
- the gesture or facial expression is the same in two cultures but the meaning of this gesture is different;
- the gesture exists in two countries but it has different variants or its use is more extensive in one country than in the other one.

These are some of the problems we can find interpreting the gestures and facial expressions of people from different cultures or countries. Since we are aware that every gesture, manner or facial expression communicates an emotion, the other problem that we observe is the classification of these emotions and, even more difficult: the direct relation between one emotion and the gesture to express it.

There are studies classifying every possible human facial expression as the *Facial Action Coding System (FACS)*of Ekman and Friesen [19], that measures all visible facial movements and it estimates that there are several hundred thousand possible visibly distinguishable facial expressions, most of which are never seen on people's faces in everyday life. This system has been used to score pictures of faces that researchers have seen to express emotions and to score faces of people in emotionally arousing situations. Between their conclusions, we can find that many emotional expressions are synonyms or convey different connotations of particular emotions and that the number of expressions conveying emotional meanings is much greater than researchers thought, but it is smaller than the number of possible expressions.

Researchers continuously argue about classifications of emotions, the existence of primary and secondary emotions, that there are studies, which say there is a series of basic emotions from which the rest of the secondary emotions originate, etc. After many readings and secondary literature, we will take for our analysis the classification of the basic emotions and all the possible "secondary" emotions that Goleman [12] facilitates. This author divides emotions into eight groups or basic families:

- *Anger*: fury, outrage, resentment, wrath, exasperation, indignation, vexation, acrimony, annoyance, irritability, hostility, and, perhaps at the extreme, pathological hatred and violence.
- *Sadness*: grief, sorrow, cheerlessness, gloom, melancholy, self-pity, loneliness, dejection, despair, and, when pathological, severe depression.
- *Fear*: anxiety, apprehension, nervousness, concern, consternation, misgiving, wariness, qualm, edginess, dread, fright, terror; as a psychopathology, phobia and panic.
- *Enjoyment*: happiness, joy, relief, contentment, bliss, delight, amusement, pride, sensual pleasure, thrill, rapture, gratification, satisfaction, euphoria, whimsy, ecstasy, and at its farthest point, mania.
- *Love*: acceptance, friendliness, trust, kindness, affinity, devotion, adoration, infatuation, and agape.
- *Surprise*: shock, astonishment, amazement, and wonder.
- *Disgust*: contempt, distain, scorn, abhorrence, aversion, distaste, and revulsion.
- *Shame*: guilt, embarrassment, chagrin, remorse, humiliation, regret, mortification, and contrition.

The question now is **how to identify these emotions in a facial expression, in a gesture and which is the most efficient method to describe them to the target audience**. Hager and the extensive research mentioned in his work say, "The expression of a given face at a specific time is conveyed by a composite of signals from several sources of facial appearance. All these sources include the general shape, orientation (pose), and position of the head, the shapes and positions of facial features (e.g., eyes, mouth), coloration and condition of the skin, shapes of wrinkles, folds, lines, and so forth. Some of these sources are relatively fixed; others, more changeable. The most important source of change in facial expression is the set of muscular movements produced by facial muscles, which provide the most substantial changes in facial appearance over short time durations and contribute most to nonverbal communication by the face".

Cortese and Butterfield [21] stated that there is a big problem in decoding facial-emotional expressions. Considering many studies, they carried out a code for every basic facial expression. As said by these authors, this code constitutes a system of psychological interpretation of the messages people show in their faces. This code is divided in three zones on the face: eyebrows-forehead zone, eyes-eyelids-nose zone and cheeks-mouth-jaw zone. This would be the association between facial expressions and emotions:

- *Happiness*: Eyebrows are raised up; mouth is opened and can express different kinds of smiles, the body moves, and the eyes move and have a fixed look.
- *Anger*: Staring at the offence, eyes and eyebrows have a peculiar, fixed, stiff incline, teeth may be clenched.
- *Sadness*: Lack of expression, look falls, corner of one's mouth falls, eyebrows move, and general depression of the features.
- *Surprise*: Increase in muscle tone, interrupted breathing, dilated pupils, eyes and mouth open wide, internal and external part of eyebrows and eyelids lift up, jaw goes down, knees bend and body lends forwards.
- *Fear*: Shudder, startle, immobility for a while, possibility of falling silent or screaming, eyes and mouth open wide, corner of one's mouth moves backward.
- *Disgust*: Eyebrows fall and come close together; cheeks, chin and upper lip go up, puckered nose, corner of one's mouth and lower lip fall, eyes become smaller.

2.3 Audio description of emotions and gestures

We have to refer to the studies carried out by Salway and Graham [20] *Extracting Information about Emotions in Films* where they developed a method to extract information about characters' emotions in films. These authors say that "the language of AD is rich in information concerning the characters and their external appearance, but information about their cognitive states, including their emotions, is not described directly. However some insight into a character's emotional state is given by AD when the emotion is being depicted visually in the film".

Other works associated can be the one carried out by Orero and Vercauteren [22] concerning the AD of facial expressions in animated movies and the difficulties derived from this work. They detected that facial expressions are complex especially for three reasons: the type of message that can express no emotion at all or just a mood; the fact that the face can combine more than one emotion which will make it even more difficult to come to the right conclusion; and the fact that emotions can be expressed in a more or less intense way.

Apart from these studies, we have to think about all the guidelines for AD we mentioned at the beginning of this paper: the difficulties in AD, the lack of time between dialogues, sounds, and many other elements for doing it properly. Where are the limits? At what point is it too much information when we audio describe emotions in films? When is it too obvious? How can we find a balance between these problems? How important is to describe the emotion one person has at a certain moment?

3 ANALYSIS

One of the biggest problems is whether AD has the ability to avoid incorporating personal interpretations or not. In this sense, a good and objective AD could be one that does not significantly alter the film-watcher's eye movements compared to a person watching a clip without AD.

In order to obtain some objective and usable results in relation to AD of emotions and gestures we are carrying out an experiment using the eye-tracking technology with 10 participants. We will follow a similar methodology as the one developed by Masuda *et al*. [23], but instead of working with static pictures, we will use film excerpts, that is, pictures in movement; which will bring to light new and innovative results.

The procedure is as follows: these participants watch an excerpt of a film without AD and the eye movements are recorded at the same time. The main objective of our analysis is to detect the points where the majority of the group, coincides focusing their attention in the gestures and emotions.

We consider these points the places where the person resorts by the film's own dynamic: without intervention of the AD.

After that, we will take a group of people with similar characteristics and they will watch the same film excerpt but this time audio described. This excerpt contains Audio descriptions of emotions such as "she is sad" or "he feels alone". Once again, we will extract the points with a high level of coincidence.

The first hypothesis would be that there will be differences between the eye movements of the people watching the film with AD and the people watching the same film without AD: The eye movement patron will be influenced by the AD, falling on the points that the person in charge of the AD previously selects.

In relation to this, the participants watching the film excerpt with AD will have a higher number of eye movements and saccades and the other participants will have longer eye gazes.

4 CONCLUSIONS & FUTURE WORK

Once the analysis is done, we will analyze the achieved results, considering the mentioned studies, in order to obtain some guidelines, and when possible, to objectively audio describe the gestures and emotions in these films. We also take as starting point the fact that the facial expressions, gestures and all the non-verbal communication are essential in the understanding of the film.

If we observe the film excerpts chosen we find many obstacles, apart from the obvious of the AD, which can make the work more difficult dealing with non-verbal communication and emotions, such as the lack of time and the confusion with other sounds, the simultaneity of gestures and their corresponding emotions, or the fast change of gestures and, therefore, emotions, between many others.

Other problem is the description of these gestures and emotions: How should audio describers use words in this case? In an attempt of not raising once again the subjectivity or objectivity problem in AD, how should be the vocabulary, which words describe better and easily the emotions and feelings? This issue has its importance since "by only describing what we see, we are depriving the Audio description of its meaning and not creating a holistic and comprehensive Audio description" (Orero and Vercauteren, 2008).

Besides the mentioned questions, many more are appearing during the research:

- Where is the limit of the objectivity dealing with emotions in AD?
- What about the importance of the silence? Where do we have to stop and let the film/the sound goes on?
- Would it be possible to translate directly the AD of a Spanish spoken language film into English, German, or any other language different from Spanish and from another culture?
- Do we all have the same way to audio describe facial expressions and emotions of people in films, and therefore, their basic emotions? The same interest?

After this first part of our study, we want to extend our study in the future with participants from other countries – including subtitles when necessary in the film excerpt – in order to prove if there are cultural differences in the perception of the emotions.

REFERENCES

[1] AENOR. Norma UNE: 153020. *Audiodescripción para personas con discapacidad visual. Requisitos para la audiodescripción y elaboración de audioguías*. Madrid, Spain, p.4. (2005).

[2] ITC. ITC Guidance on Standards on Audio Description, http://www.ofcom.org.uk/tv/ifi/guidance/tv_access_serv/archive/audio_description_stnds. (2000)

[3] B. Benecke. Audio-Description. In: *Meta* 49:1, pp. 78-80. (2004).

[4] A. Remael. Audio Description for Recorded TV, Cinema and DVD. An Experimental Stylesheet for Teaching Purposes, http://www.hivt.be (2005)

[5] Ofcom. Guidelines on the provision of television access services, http://www.ofcom.org.uk/tv/ifi/guidance/tv_access_serv/guidelines (2006)

[6] P. Orero. Algunas consideraciones sobre la audiodescripción comercial en España. In: *Sociedad, integración y televisión en España*. R. Pérez-Amat, A. Pérez-Ugena (Eds.). Laberinto, Madrid, Spain, pp. 277-292. (2006)

[7] L. Puigdomènech, A. Matamala and P. Orero. Audio Description of Films: State of the Art and a Protocol Proposal. In: *Intermedia 2007 Lodz Conference on Audiovisual Translation and Interpreting*. Lodz, Poland (2007).

[8] A. Remael and G. Vercauteren. Audio describing the exposition phase of films. Teaching students what to choose. In: *Trans* 11, pp. 73-94. (2007).

[9] G. Vercauteren. Towards a European Guideline for Audio Description. In: *Media for All. Accessibility in Audiovisual Translation*. J. Díaz-Cintas, P. Orero and A. Remael (Eds.). Rodopi, Amsterdam, Holland, pp. 215-229. (2007).

[10] L.A. Samover and R.E. Porter. *Intercultural Communication*. Wadsworth/Peter Lang, New York, U.S. (1997).

[11] M.L. Knapp. *La comunicación no verbal. El cuerpo y el entorno*. Paidós, Barcelona, Spain. (1982).

[12] D. Goleman. *Emotional Intelligence. Why it can matter more than IQ*. Bantam Books, New York, U.S. (1996).

[13] A.M. Cestero. *Comunicación no verbal y enseñanza de lenguas extranjeras*. Arcos/Libros, Madrid, Spain. (1999).

[14] F. Poyatos. *Nonverbal Communication in Translation: Theoretical and Methodological Perspectives*. John Benjamins, Amsterdam-Philadelphia. (1997).

[15] J.C. Hager. Emotion and Facial Expression, http://face-and-emotion.com/dataface/emotion/expression.jsp. (2003).

[16] J.C. Hager and P. Ekman. The Inner and Outer Meanings of Facial Expressions. In: *Social Psychophysiology: A Sourcebook*, J.T. Cacioppo and R.E. Petty (Eds.). The Guilford Press, New York, U.S. (1983).

[17] P. Ekman, E.R. Sorenson and W.V. Friesen. Pan-cultural elements in facial emotion. In: *Science* 164, pp. 86-88. (1969).

[18] C.E. Izard. *The face of emotion*. Appleton-Century-Crofts, New York, U.S. (1971).

[19] P. Ekman and W.V. Friesen. *The Facial Action Coding System*. Consulting Psychologists Press, Palo Alto, U.S. (1978).

[20] A. Salway and M. Extracting Information about Emotions in Films. In: *Proceedings of ACM Multimedia*, Berkeley, U.S., pp. 299-302. (2003).

[21] A. Cortese and E. G. Inteligencia Emocional. http://www.inteligencia-emocional.org/articulos/lasexpresionesfaciales.htm (2008).

[22] P. Orero and G. Vercauteren. The Audio Description of Facial Expressions. In: *Audiovisual Translation: Multidisciplinary Approaches, Conference at the University of Montpellier 3*. (2008)

[23] T. Masuda, B. Mesquita, S. Tanida, P.C. Ellsworth, J. Leu and E. Van de Veerdonk. Placing the Face in Context: Cultural Differences in the Perception of Facial Emotion. In: *Journal of Personality and Social Psychology*, 94:3, pp. 365-381. (2008).

Towards a System-Theoretic Model for Transition of Affect

Michelle Karg and **Kolja Kühnlenz** and **Martin Buss**[1]

Abstract. This study investigates the dynamic properties of affect from a system-theoretic point of view. A psychological study was conducted, in which affect was induced sequentially using pictures of the International Affective Picture System. Based on this experimental data, a piecewise linear model is formulated which describes average affective reactions of a human. System-theoretic analysis of this model reveals that the system shows complex dynamic characteristics, which can not only be explained by an additive influence of the stimulus on the current affective state. It suggests that there exist internal fluctuations. Furthermore, the joyous region contains a stable attractor. A second model concentrates on individual differences of affective reactions between humans. A Markov Chain estimates the probability that a person feels a specific affect depending on an external stimulus and the previous affective state. Using this model estimates on how a sequence of different affective stimuli influences the affective state can be calculated.
The study proves that a system-theoretic approach is suitable for modeling emotions, in particular affective states, and can give additional insights on the dynamics of emotions.

1 INTRODUCTION

Emotions influence cognitive processes, like decision making, learning, recognition and motivation. Studies analyze the source of emotions, classification of emotions and how emotions relate to cognitive processes[23, 19, 7]. Recent research investigates the integration of emotions in Human-Computer Interaction (HCI). It is presumed that recognition of human emotions, modeling emotional reactions as well as interactions, and also expressing synthesized emotions on a virtual avatar or robotic head enhances HCI [21, 20]. This requires emotion recognition, expression and dynamic modelling of emotions. A dynamic emotion model offers the possibility to predict emotional behavior, to enhance recognition systems by previous knowledge about temporal development of emotions and gives insights in complex emotional reactions. Furthermore, implementation of such a model on an emotion-expressive virtual avatar or robotic head has the benefit of compact programming compared to state machines for showing complex and versatile behavior.
This study focuses on the temporal characteristics of affect. From a system-theoretic point of view, it analyzes how the previous affective state and an affective stimulus, corresponding to external events, effect the current affective state. A time-discrete model is developed based on experimental data. As there usually exists a large variation of responses to emotional stimuli between humans, one model considers the average affective reaction of a human and analyzes the average behavior with system-theoretic tools. A second model gives estimates of how an individual would react to a specific stimulus in respect to the current affective state.
This paper is structured as follows. Section 2 gives a brief overview of related work. Then discrete affective states are defined for our purpose in section 3. Section 4 describes the psychological experiment conducted to estimate the parameters for each model. Section 5 introduces a model based on a piecewise linear system (PL), which models average affective reactions. The parameters of the model are adapted to the experimental data, followed by a system-theoretic analysis. It is concluded that internal fluctuations in affective reactions exist. Section 6 uses a Markov Chain to estimate the next affective state of a person depending on the previous affective state and an external stimulus. The paper concludes with plans for future work.

[1] Technische Universität München, Germany, email: {mkarg, koku, mb}@tum.de

2 RELATED WORK

Probabilistic models are currently investigated to model HCI for prediction the next emotional state of a human, analysis of the interaction and design of the affective reaction of virtual avatars or robots. Gockley at al. propose a model which is capable to calcluate synthesized social reactions based on mood, emotion and attitude for a roboceptionist [9]. Their main goal is to model long-term effects to enhance long-term relationships beteen human and robots.
Cattinelli et al. [4] consider an interaction model based on a probabilistic finite state automaton. The automaton contains a finite set of states for the robot and the user. They conclude that the interaction is stable if both agents have similar personalities. If personalities differ no quick convergence of emotional state is exprected.
Hiddden Markov Models (HMM) can be used to connect actual felt emotion with emotion expression. In [11], a HMM is proposed as a emotional core for a robot. Blewitt et al. present a computational model for defining the emotional state of an agent based on a fuzzy logic system [3, 2]. Further approaches regarding modeling emotional behavior of virtual agents or robots can be found in [1, 24, 14, 8, 5]
In this study, we focus on a model for transition of affect itself and analyze it with system-theoretic tools.

3 ASPECTS OF MODELING AFFECT

3.1 Affective States

Various theories exist to categorize emotions and affect. According to Lazarus, an emotion itself contains physiological disturbance, action tendencies that are not necessarily acted out, and affect, whereas affect is the subjective experience of a person during an emotion [13].

Most common theories are the dimensional PAD model [16], the basic emotions [6] and the primary emotions [22]. The PAD model spans a 3-dimensional space with the independent and bipolar axes valence, arousal and dominance. The current emotion of each individual can be described as a state within this space. The basic emotions are anger, disgust, fear, joy, sadness, and surprise. They are derived from facial expressions common among all cultures. In contrast to the basic emotions, the primary emotions can be blent together, so that each emotion consists to a certain part of the primaries which are acceptance, anger, anticipation, disgust, joy, fear, sadness and surprise.

In this work, we combine the dimensions arousal and valence with the discrete emotions sad, joyous, neutral and quiet, which are derived from the basic emotions, see Fig. 1 . Mikels et al. collected descriptive emotional categories for IAPS to identify images which elicit one discrete emotion more than others. Based on their work, the discrete state sad represents the area in the arousal-valence space with which participants most associated the state sad in Mikels et al.'s study.

Descriptive emotional categories as awe, content, and amusement are mixed for pictures in the upper plane of the arousal-valence space [17]. We choose the area where arousal values range from 4 to 6 and valence values range from 6 to 8 for the state joy.

Morris scored advertisements and emotion adjectives using SAM in his study. The number of pictures available in the IAPS data set shows a V-structure with its peak at medium pleasure and low arousal. In Morris studies, adjectives like quiet and solemn correspond to this area. We defined this area as the discrete state quiet.

The discrete state neutral represents the area between sad and joyous. For equality of the states, each region spans a two times two square in the arousal-valence state space. Furthermore, the states sad, neutral and joyous have the same range of arousal, so that state transitions depending only on pleasure can be investigated.

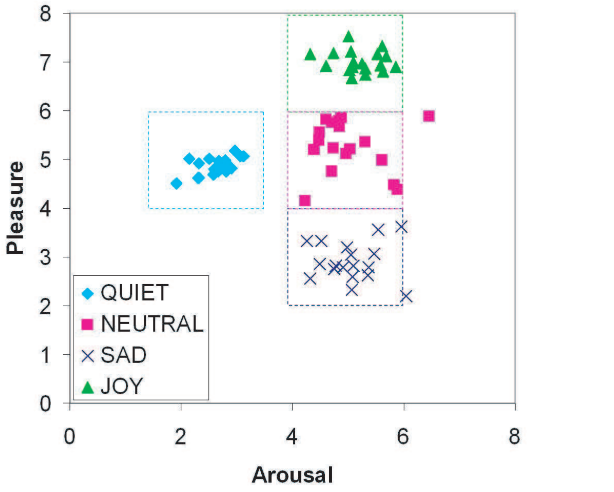


Figure 1. The four discrete emotional states quiet, sad, joyous and neutral are arranged in the two-dimensional arousal-valence subspace of the PAD model. The markers display the affective states of the IAPS pictures used for induction of a specific affect.

3.2 System-theoretic Model

There exists a large variety of system-theoretic models. In general, they can be divided in stochastic and deterministic models. This study uses a Markov Chain as a representative for a stochastic approach and a Piecewise Linear System for deterministic modeling. A data-driven concept is chosen to determine the properties for each model, as there is no physical law applicable to emotions in order to derive state equations. A psychological experiment is conducted for this purpose. The experiment is repeated over a number of participants, so that a model can be formulated which describes the mean affective behavior. This model specifies the average affective reaction of a human. The analysis of this model concentrates on the effect of the previous affective state on the current affect in comparison to the affective stimulus. Furthermore, it investigates if in general more stable or instable affective regions exist.

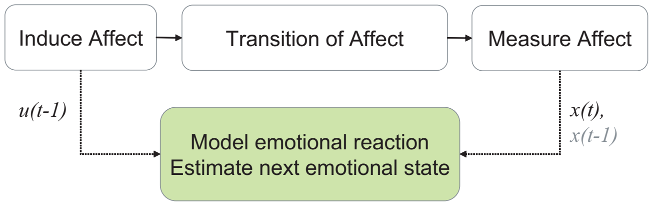


Figure 2. A stochastic as well as a deterministic model are designed based on knowledge about the affective stimulus $u(t-1)$ and the actual affective state $x(t)$ of a person. The previous affective state $x(t-1)$ is estimated based on ratings in the IAPS manual.

However, emotional reactions vary within subjects and in-between subjects. This leads to a Markov Chain which estimates the probability that an individual person feels a specific affect. This model gives estimates for each affective state considering the previous emotional state $x(t-1)$ and the stimulus $u(t-1)$.

4 PSYCHOLOGICAL EXPERIMENT

A psychological experiment is contucted in order to identify parameters of the Piecewise Linear System and the Markov Chain. It is essential to set the scene for interpretation and system-theoretic analysis of each model.

4.1 Preparation of Emotion Induction

IAPS was choosen for emotion induction because it is continously updated and evaluated. The IAPS pictures are sorted by their valence, arousal and dominance values. 18 pictures for each emotional state were selected from the IAPS database. Each picture was shown only once. Criterion for selection was the standard deviation of arousal, valence and dominance, which ranges between 1 and 3 in the data base. In the three subsets for sad, quiet and neutral, pictures were selected with a standard deviation lower than 2 for arousal and valence. Pictures which induce joyous vary more in their standard deviation so that the maximum standard deviation for arousal and valence is 2.14.

The pictures were aligned in a presentation so that each transition between the four states is repeated three times. It was achieved to keep arousal almost constant for transition between the states sad, neutral and joyous. This allows independent investigation of transitions along the pleasure axis.

Due to different reactions of humans to erotic pictures, they were excluded.

4.2 Test Procedure

The presentation which includes the IAPS pictures was shown with a video projector. The slide transitions were determined by a timer. Each picture was shown 5 seconds. The participants had 15 seconds time to fill out the Self Assessment Mannequin (SAM) questionaire. After that a start slide was shown with the number of the next transition for 5 seconds.
A maximum of 8 persons completed the experiment at the same time. Overall 50 persons participated at the experiment.
The handout for the participants contained a how-to-do manual for SAM, test questionaires, the SAM questionaires for the transitions and an EPI personality test at the end.
The general procedure of the experiment was as follows

- Initial hellos
- Supervisor explains the experiment procedure and reads out the manual for the SAM
- Time for questions
- 1 test cycle (IAPS picture + SAM evaluation)
- 1 IAPS pictures of each state to identify a participant's peculiar response to emotion induction
- 3*3*4 test sequences (3 transitions for each state, 3 repetitions, 4 states)
- EPI personality questionaire.

One procedure took about 30 minutes.

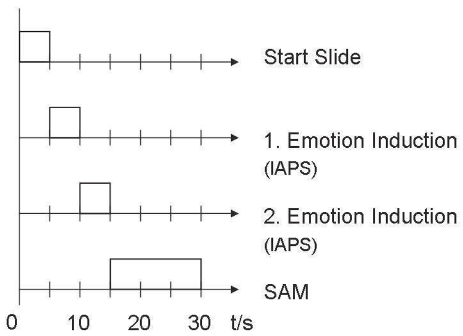


Figure 3. This graph shows a single cylce to investigate the transition from one emotional state to another. First the start slide with the number of the cycle was shown. After 5 seconds the first IAPS picture was presented. The second emotion induction followed 5 seconds later and the picture was also shown for 5 seconds. Then the SAM questionaire was filled out. This cycle was repeated 36 times.

4.3 Discussion of Experimental Data

The averaged ratings of our participants over all pictures differ from the ratings listed in the IAPS manual. The mean arousal value is 1.8 lower in average to IAPS. The mean valence value is 0.8 lower in average to IAPS. This results in a shift of the discrete regions for quiet, joyous, neutral and sad in the valence-arousal space (see Fig. 6) . The

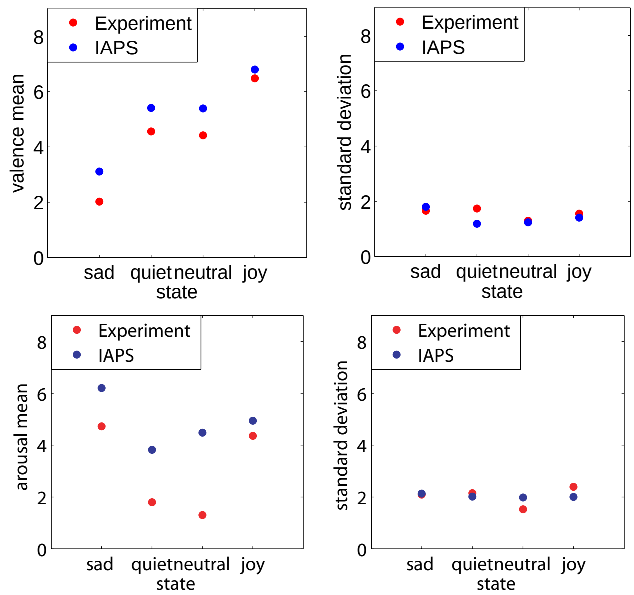


Figure 4. The ratings for valence and arousal are lower in our experiment comparing to the lists in the IAPS manual. The standard deviation is almost equal.

mean valence values over all participants are plotted separately for each affective state in Fig. 5.The degree of valence is usually higher for the second affective state if the previous state was quiet or neutral. If the first affective state was sad, participants rated in average their next emotional state with less valence. In average, the discrete induced affect for the second state was always reached, nevertheless what the previous affective state was. The averaged arousal values were also within the expected discrete states. It is aimed that the following state model is capable to model slight differences in emotional reaction to a stimulus depending on the previous state.

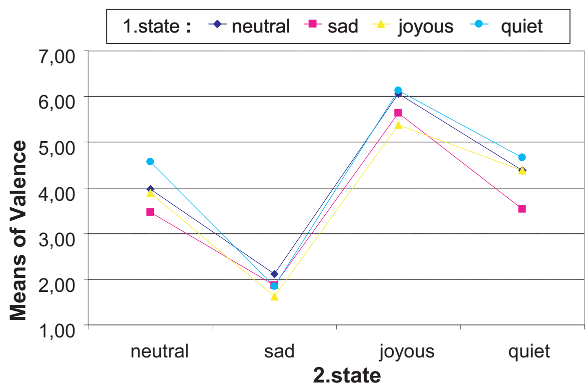


Figure 5. This plot shows the mean ratings of valence depending on the previous affective state. In average, participants rated valence higher if the previous affective state was neutral or quiet.

5 MODELING THE MEAN TRANSITION OF AFFECT: STATE MODEL

A time-discrete, time-invariant, Piecewise Linear System (PL) with state-depented switching is proposed to model transition from one affective state to another [15, 10]. This model covers the average behavior over all participants. As it is supposed that affective reactions are complex, dynamical and nonlinear, a set of linear dynamic equations are chosen for approximation.

The state vector X_k consists of two dimensions arousal and valence. It is measured with the SAM rating during the experiment. As the SAM test ranges between 1 and 9, the components of X_k also range continuously between 1 and 9.

$$X_k = \{valence, arousal\}^T \tag{1}$$

The influence of emotion induction is modeled by the input vector u_k. It reflects the difference of the expected PAD values of 2 consecutive IAPS pictures. Furthermore, the influence of the emotion induction u_k is delayed and affects the state X_{k-1}.

$$u_k = \Delta u = \{\Delta valence, \Delta arousal\}^T \tag{2}$$

The output vector Y_k is a unit vector e_q of $S_Y = \{e_{joy}, e_{neutral}, e_{sad}, e_{quiet}\}$. Each unit vector e_q represents one discrete affective state sad, neutral, quiet or joyous.

The state space for arousal and valence is divided in four regions according to Fig. 6. The behavior in each region is approximated by the following linear dynamic equations:

$$X_{k+1} = A_q X_k + B_q u_k + X_q \tag{3}$$
$$Y_k = C X_k + Y_0 + w_k \tag{4}$$
$$q \in \{sad, joyous, neutral, quiet\}$$

The probability matrix A_q describes internal fluctuations in human emotions. Even if no external stimuli is present, the emotion will drift. The impact of the external stimuli u_k is defined by the matrix B_q.

The mapping between the continuous state X_k and the discrete affective output Y_k is realized by the matrix C. This mapping can be derived from our definition of the discrete outputs. The martingale w_k is required so that Y_K reaches the most probable output calculated by CX_k (4). The value of the martingale increment is adapted for each time step k. The sample rate is 0.2 Hz.

If the present state X_k would not influence the next emotional state and if the input underlies no damping or amplification, the state equation would look as follows:

$$X_{k+1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} X_k + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} u_k + \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{5}$$

The matrices A_q and B_q are identity matrices for all q. If mean emotional behavior is completely covered by Equ. (5), affective reactions are barely a reaction to affective stimuli and the previous affective state does not influence the current affective state.

If this hypothesis does not hold, the conclusion can be drawn that complex dynamics underlie emotional reactions.

5.1 Estimation of the Output

The output describes the mapping between the continuous states valence and arousal to the four discrete states joyous, neutral, sad and quiet. The matrix C and the offset Y_0 need to be estimated.

$$Y_k = C X_k + Y_0 + w_k \tag{6}$$

The switching boundaries between the states are defined in Fig.6. Their mathematic description is as follows:

$$\begin{aligned} f_1 &= (1\,0)X_k - 3,2 \\ f_2 &= (1\,0)X_k - 5,2 \\ f_3 &= (0\,1)K_k - 2,2 \end{aligned} \tag{7}$$

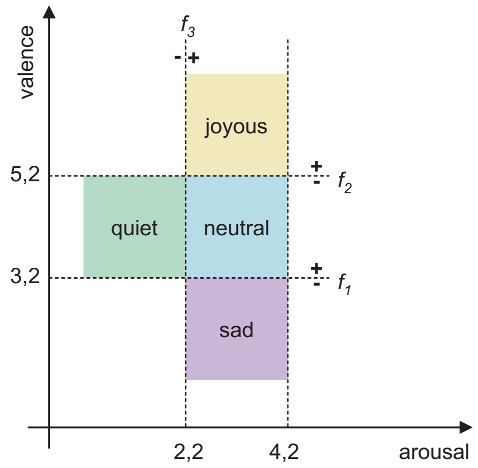


Figure 6. This graph shows the switching boundaries between the regions. The most probable discrete state Y_k is estimated by superposition of the functions f_1, f_2 and f_3 .

The signs of f_1, f_2 and f_3 differ depending on the discrete states, e.g. for a specific state in the region for neutral, f_1 and f_3 are positive and f_2 is negative.

The first element of the vector calculated by $CX_k + Y_0$ is asso-

[htbp]

Table 1. Values of functions in discrete regions

region	f_1	f_2	f_3
joy	+	+	+
neutral	+	-	+
sad	-	-	+
quiet	+	-	-

ciated with the discrete state joy, the second with neutral, the third with sad and the last with quiet. So that the first row represents the superposition of the functions for the state joy in that way, that the resulting function is maximum if the state is in the region of joy otherwise less (according to table 1). This procedure is analog for the following three rows of Y_k. The following equation combines the superposition of the decision functions for the four discrete states.

$$Y_k = \begin{pmatrix} 2 & 1 \\ 0 & 1 \\ -2 & 1 \\ 0 & -1 \end{pmatrix} X_k + \begin{pmatrix} -10,6 \\ -0,2 \\ 6,2 \\ 4,2 \end{pmatrix} + w_k \tag{8}$$

Usually, probabilities range between 0 and 1. The above equations are shift to positive values for valence ranging between 0 and 9, and arousal ranging between 0 and 9, so that the result is positive for all possible SAM ratings. Also, they are normalized to guarantee a

maximum value of 1 and a minimum value of 0.
The martingale w_k is required so that Y_k reaches the most probable state and can be represented by unit vectors.

$$Y_k = \frac{1}{28,2}\begin{pmatrix} 2 & 1 \\ 0 & 1 \\ -2 & 1 \\ 0 & -1 \end{pmatrix} X_k + \begin{pmatrix} 0.0426 \\ 0.4113 \\ 0.6383 \\ 0.5674 \end{pmatrix} + w_k \quad (9)$$

5.2 Estimation of State Equation

Actually, one state equation was considered. However, it is not possible to reflect complex behavior within one linear equation. So the state equation were extended to a PL system. A_q, B_q, and X_q are adapted separately for each state.
The average arousal and valence values over all participants were calculated for each experimental cycle. They are the values for X_{k+1} and represent the emotional state of a participant after the second stimulus. The emotional state after the first stimuli was not measured and is derived from the values in the IAPS manual. Besides X_{k+1} and X_k, u_k is known. It is the difference in arousal and valence between the first and the second stimulus.
Least Square was used for estimation of the parameters. Exemplarily, the calculation of $a_{11}, a_{12}, b_{11}, b_{12}$ and $X_{0,v}$ is shown for the discrete state joyous. These parameters are required to estimate the valence component of X_{k+1}.

$$\begin{aligned} X_{k+1,v} &= a_{11}X_{k,v} + a_{12}X_{k,a} \\ &+b_{11}u_{k,v} + b_{12}u_{k,a} + +X_{0,v} \end{aligned} \quad (10)$$

The unknown parameters are combined in the vector $x = [a_{11}, a_{12}, b_{11}, b_{12}, X_{0,v}]^T$. Each transition was repeated three times. Also, if the first emotional state is joyous, the next state can be sad, neutral or quiet. In total, we have nine times measured state variables and input variables for the above equation. So that a data matrix D is introduced, whose rows j contain the input and state values for each of the nine transitions.

$$D_j = [X_{k,v,j}\ X_{k,a,j}\ u_{k,v,j}\ u_{k,a,j}\ 1] \quad (11)$$

$$\text{with } j \in 1...9$$

$$d = [X_{k+1,v,1}...X_{k+1,v,9}]^T \quad (12)$$

The next states $X_{k+1,v,j}$ are combined in the vector d. Least Square is used to solve the overdetermined system

$$d = Cx. \quad (13)$$

This procedure was repeated for arousal, which is the second component of X and for each region. This results in the following state equations.
State equation for the region joy:

$$X_{k+1} = \begin{pmatrix} 0.4 & -0.6 \\ 0.7 & -0.4 \end{pmatrix} X_k + $$
$$+ \begin{pmatrix} 0.9 & -0.4 \\ -0.3 & 0.8 \end{pmatrix} u_k + \begin{pmatrix} 5.2 \\ 0.2 \end{pmatrix} \quad (14)$$

State equation for the region neutral :

$$X_{k+1} = \begin{pmatrix} 1.0 & -1.2 \\ 0 & 1.0 \end{pmatrix} X_k + $$
$$+ \begin{pmatrix} 1.3 & -0.3 \\ -0.3 & 1.0 \end{pmatrix} u_k + \begin{pmatrix} 3.6 \\ 0.4 \end{pmatrix} \quad (15)$$

State equation for the region sad:

$$X_{k+1} = \begin{pmatrix} 1.6 & -0.8 \\ -0.6 & 1.3 \end{pmatrix} X_k + $$
$$+ \begin{pmatrix} 1.0 & -0.3 \\ -0.1 & 0.8 \end{pmatrix} u_k + \begin{pmatrix} 0.5 \\ 0.7 \end{pmatrix} \quad (16)$$

State equation for the region quiet:

$$X_{k+1} = \begin{pmatrix} 1.3 & 1.1 \\ -0.8 & 0.8 \end{pmatrix} X_k + $$
$$+ \begin{pmatrix} 1.1 & 0.0 \\ -0.2 & 0.6 \end{pmatrix} u_k + \begin{pmatrix} -2.4 \\ 4.8 \end{pmatrix} \quad (17)$$

There are large variations in the parameters of A_q for each region. This proofs that an individual model for each region is required. The elements of B_q are similar. As expected, the elements on the diagonal are larger than the elements on the secondary diagonal so that the component arousal of the stimulus influences arousal more than valence of the state and vice versa.
So far, this model covers only the average emotional behavior of humans.

5.3 Evaluation

5.3.1 Evaluation of A_i, B_i

The root mean-squared error e_j is calculated between the measured values $\overline{X}_{j,l}$ over all participants for each transition l and the simulated value X. The errore_j is calculated for each discrete state j separately.
It is averaged over the three repetitions of each transition and over all transition from one discrete state.

$$\begin{aligned} e_j &= \sqrt{\sum_{l=1}^{n}(\overline{X}_{j,l} - X_{j,l})^2/(n-1)} \\ &= \sqrt{\sum_{l=1}^{3}(\Delta X_{j,l})^2 \cdot 0.5} \end{aligned} \quad (18)$$

The following table shows the mean quadratic error for valence (v) and arousal (a) for each transition:

Table 2. Root mean-squared error of simulation

1.state	v/a	sad	2.state joy	neutral	quiet	overall
sad	v	\	0.9	0.1	0.3	0.5
	a	\	0.5	0.4	0.3	0.4
joy	v	0.5	\	0.2	0.2	0.3
	a	0.5	\	0.5	0.6	0.4
neutral	v	0.1	0.3	\	0.2	0.2
	a	0.3	0.1	\	0.2	0.2
quiet	v	0.4	0.4	0.4	\	0.3
	a	0.3	0.4	0.6	\	0.4

The error alternates between 0.2 and 0.5. However, emotion induction has a large variation (standard deviation between 1 and 2), so

that the simulation errors are minor comparing to variations caused by individual emotional responses or reactions.
The total root mean-squared error is 0.32 for valence and 0.35 for arousal.

5.3.2 *Evaluation of State Order*

Fig. 5 shows that participants felt more valence, if the previous state was neutral or quiet. The same plot was generated by the PL system. Comparison of both plots showed that the model predicts the same behavior for joyous, neutral and sad as the second state. If the second state is quiet, the mean values for a neutral and joyous previous state lie close to each other. In the simulation, their order is vice versa. Same results are gathered for analyzing the mean values of arousal.

5.3.3 *Reachability of the Discrete States*

In average, all participants reached the induced emotional states for all transitions. The same holds for simulation of all transitions.

5.4 Systemtheoretic Analysis

Methods from control and system theory are applied to the obtained state equations, so that conclusions can be drawn about stability, observabilty and detectability for each affective region. This analysis interprets the mean affective behavior over all participants. The external stimulus u_k is set to zero. This means that the external stimulus stays constant, e.g. in case of our experiment showing the same joyous or sad IAPS picture for a longer time period and try to induce the same affective state for a longer time period. The reaction differs depending on the affective state at the beginning.

5.4.1 *Stability*

First, the fixed point for each affective region is calculated using the corresponding linear dynamic equation. The fixed points of the dynamic equations for joyous and sad lie within their regions, the others outside. Table 3 lists the coordinates for each fixed point.

Table 3. Fixed points

region q	valence	arousal
sad	2.6	2.7
joy	5.4	2.9
neutral	12.3	2.9
quiet	5.8	0.5

As the system is time-discrete, the absolute value of the eigenvalues must be lower than one for a stable fixed point. Applying Lyapunov's Indirect Method to each fixed point, yields that the system has one stable fixed point, one instable fixed point and two saddle points (see Table 4).
The phase diagram of the PL system is plotted in Fig.7. However, this diagram is plotted for $u_k = 0$. This case, repeating the same stimulus, was not included in the experiment, so that the state equations are only approximated by data for alternating stimuli. Interpreting the phase plot leads to assumptions which need to be checked in a second experiment which includes repeating the same stimulus. For this model, the following interpretations are drawn. If a joyous picture

Table 4. Eigenvalues of the state matrix A_q

region q	1.Eigenvalue	2.Eigenvalue	
sad	2.1	0.8	saddle point
joy	$0.5i$	$-0.5i$	stable fixpoint
neutral	0.8	1.2	saddle point
quiet	$1.0 + 0.9i$	$1.0 - 0.9i$	instable fixpoint

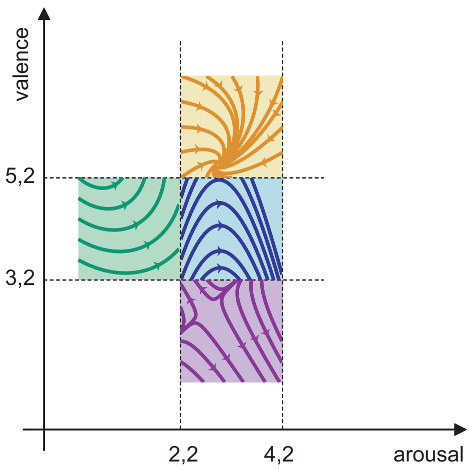


Figure 7. This graph shows the phase diagramm for each region for $u_k = 0$

is shown repetitively, the mean reaction over all participants would be, that a stable joyous affect is reached. However, a sad stimulus with less pleasure and medium arousal would lead to more sadness and eventually anger. Starting with a quiet or neutral affect, arousal increases over time.

5.4.2 *Controllability*

Kalman method is used to analyze controllability of the PL system for each region. For all regions, the rank of the controllability matrix $Q_{B,q} = [B_q, A_q B_q]$ was equal the number of states, so that each region is controllable.

5.4.3 *Observability*

In analogy, observability is studied for each region. The observability matrix $Q_{O,q} = [C^T, A_q^T C^T]$ is calculated. As the rank for all observability matrices $Q_{O,q}$ is equal the number of states, all regions are observable.

6 MODEL OF AFFECTIVE TRANSITIONS FOR ONE INDIVIDUAL: MARKOV CHAIN

A Finite Markov Chain is considered to model individual fluctuations from mean affective behavior [12]. This model calculates the probability that a specific person feels a specific affect influenced by an external stimulus. The affective state and the stimulus can take four discrete values joyous, neutral, quiet and sad. The probability vector p_k at a specific time step k contains the probability for each affective state.

$$p_k = (p_{joy}, p_{neutral}, p_{sad}, p_{quiet}) \quad (19)$$

The parameters of the probability matrix $\Omega_{stimulus}$ depend on the external stimulus.

$$\Omega_{stimulus} \in \{\Omega_{joyous}, \Omega_{neutral}, \Omega_{sad}, \Omega_{quiet}\} \quad (20)$$

The probability that a specific person feels a specific affect depends only on the external stimulus and on the previous affective state.

$$p_{k+1} = p_k \Omega_{stimulus} \quad (21)$$

The probability matrices differ for each type of stimulus. Table 5 to 8 lists the matrices for joyous, neutral, sad and quiet stimuli separately. As the same stimulus was not repeated in this experiment, the according row is filled with equally distributed probabilities. These probabilities need to be estimated in a continuative experiment. The interpretation of the tables is as follows. If the previous state is quiet and the external stimulus is joyous, see Table 5, the probability that a person feels joyous is 74%. However if the previous state is sad, the probability is only 36% that a person reacts to a joyous stimulus with joy. This model can be used to simulate the effect of a sequence of stimuli. Alternating a sad and a joyous stimulus leads to a higher probability that a person feels sad. So it can be concluded that the sad stimulus has a stronger effect.

Table 5. Probability matrix for joyous stimulus

	predicted state			
previous state	joyous	neutral	sad	quiet
joyous	0.25	0.25	0.25	0.25
neutral	0.68	0.12	0	0.2
sad	0.36	0.16	0.14	0.34
quiet	0.74	0.02	0.04	0.2

Table 6. Probability matrix for neutral stimulus

	predicted state			
previous state	joyous	neutral	sad	quiet
joyous	0.2	0.14	0.44	0.14
neutral	0.25	0.25	0.25	0.25
sad	0.06	0.16	0.52	0.26
quiet	0.5	0.08	0.14	0.28

Table 7. Probability matrix for sad stimulus

	predicted state			
previous state	joyous	neutral	sad	quiet
joyous	0.02	0.04	0.88	0.06
neutral	0.1	0.06	0.64	0.2
sad	0.25	0.25	0.25	0.25
quiet	0	0.02	0.94	0.04

Table 8. Probability matrix for quiet stimulus

	predicted state			
previous state	joyous	neutral	sad	quiet
joyous	0.3	0.06	0.12	0.52
neutral	0.12	0.1	0.08	0.7
sad	0.04	0.12	0.44	0.4
quiet	0.25	0.25	0.25	0.25

7 DISCUSSION

The first model based on a piecewise linear system analyzes average affective transitions of a human. The linear approximations in each region joyous, neutral, quiet and sad differ from equation 5. This concludes that superposition of external stimulus and affective state is not sufficient to explain affective transitions. Furthermore, the states valence and arousal are controllable and observable in each region of the state space. As properties of nonlinear systems are only locally valid, in this case in each region, a reachability analysis of the complete PL system is considered in future works. Within the joyous region lies a stable fixpoint. It is assumed that repetition of joyous stimuli maintains joy. However, reactions to a repetition of sad stimuli depends on the initial value of arousal and valence. It can drift to anger or neutral. As the underlying experimental data did not contain repetition of the same stimulus, these are assumptions. A continuaitive experiment is ongoing.
A second model predicts the probability for the next affective state depending on the current and an external stimulus. A Markov Chain ist proposed with a separate probability matrix for each stimulus joyous, neutral, quiet and sad. Mostly, the probability for the next state is highest for affective state intended by the external stimulus. For neutral and quiet stimuli, the previous state influences the next stronger comparing to sad or joyous stimuli. This model can be used to calculate the probabilities for each affective state for a sequence of stimuli. For HRI interaction scenarios, this model offers the benefit, that combinations of different stimuli can be simulated with this model beforehand.
The Markov property holds for both models. For simplification, we assume that the future affective state depends only on the present emotional state and is independent of its history. Long-term effects are neglected within this study.

8 CONCLUSION AND FUTURE WORK

This study shows that system-theoretic approach is applicable to model transitions of affect. Two models are proposed. One models the average affective reaction of a human and gives insights on the dynamics of affect. Simple superposition of external stimulus and current affective state is not sufficient to predict the next affective state. It is concluded that affect underlies more complex dynamics. These dynamics are approximated by a piecewise linear system. System-theoretic analysis revealed that a stable fixed point exists within the joyous region. This suggests that joyous is a stable condition over time. However, a continuative study is required which considers repeating the same stimulus in the experiment. A second model based on a Markov Chain estimates the probability how a person reacts to an external affective stimulus. This model is capable to predict the affective reaction to a sequence of affective stimuli.
Up to this point, only transitions of affect have been considered. The

models need to be extended to cover also repetition of the same external stimulus. Furthermore, personal traits, like extrovert/introvert, degree of anxiety, gender or mood, can have an influence on the affective reaction. For both reasons, a similar psychological experiment is planned. The long-term target is to test its validity in a real-world HRI setting.

ACKNOWLEDGEMENTS

This work is supported within the DFG excellence initiative research cluster ”cognition for technical systems - CoTeSys”, see also www.cotesys.org.

REFERENCES

[1] T. Archevapanich, B. Purahong, M. Klingajay, and P. Sooraksa., ‘Facial visualization for robotic indicator by using fuzzy emotional system’, in *SICE-ICASE*, (2006).
[2] W. Blewitt, A. Ayesh, R. I. John, and S. Coupland, ‘A millenson-based approach to emotion modelling’, in *Conference on Human System Interaction*, (2008).
[3] W.F. Blewitt and A. Ayesh, ‘Modeling the emotional state of an agent through fuzzy logic with reference to the geneva emotion wheel’, in *European Simulation and Modelling Conference (ESM)*, pp. 279–283, (2008).
[4] I. Cattinelli and N. A. Borghese, ‘A simple model for human-robot emotional interaction’, in *KES*, (2007).
[5] C.Park, J. Ryu, J. Sohn, and H. Cho, ‘An emotion expression system for the emotional robot’, in *IEEE International Sysmposium on Consumer Electronics*, (2007).
[6] P. Ekman and W. V. Friesen, ‘A new pan-cultural facial expression of emotion’, *Motivation and Emotion*, **10**, 159–168, (1986).
[7] Tommy Garling Ellen Peters, Daniel Vastfjall and Paul Slovic, ‘Affect and decision making: A hot topic’, *Journal of Behavioural Decision Making*, **19(2)**, 79–85, (2006).
[8] D.-S. Kwon et al., ‘Emotion interaction system for a service robot’, in *IEEE International Conference on Robot & Human Interactive Communication*, (2007).
[9] R. Gockley, R. Simmons, and J. Forlizzi, ‘Modeling affect in socially interactive robots’, in *IEEE International Symposium on Robot and Human Interactive Communication*, (2006).
[10] Mikael Johansson, *Piecewise Linear Control Systems*, Springer, 2003.
[11] K. Kühnlenz and M. Buss, ‘Towards an emotion core based on a hidden markov model’, in *ROMAN*, (2004).
[12] G.F. Lawler, *Introduction to Stochastic Processes*, Chapman & Hall, 2006.
[13] R.S. Lazarus, *Emotion & Adaption*, Oxford University Press, 1991.
[14] S.-I. Lee, G.-Y. Park, and J.-B. Kim, ‘Natural emotion expression of a robot based on reinforcer intensity and contingency’, in *IEEE International Conference on Robotics and Automation*, (2007).
[15] D. Liberzon, *Switching in Systems and Control*, Birkhuser, 2003.
[16] A. Mehrabian and J. Russel, ‘Evidence for a three-factor theory of emotions’, *Journal of Research in Personality*, **11**, 273–94, (1977).
[17] J.A. Mikels, B.L. Fredrickson, G.R. Larkin, C.M. Lindberg, S. Magold, and P.A. Reuter-Lorenz, ‘Emotional category data on images from the iternational affective picture system’, *Behavior Research*, **37(4)**, 626–630, (2005).
[18] Morris, ‘Observations: Sam : The self-assessment manikin’, *Journal of Advertising Research*, 63–68, (1995, Nov/Dec).
[19] N. Naqvi, B. Shiv, and A. Bechara, ‘The role of emotion in decision making: A cognitive neuroscience perspective’, *Current Directions in Psychological Science*, **15(5)**, 260264, (2006).
[20] Christian Peter and Russel Beale, *Affect and Emotion in Human-Computer Interaction*, Springer, 2008.
[21] R. Picard, *Affective Computing*, Cambridge: MIT Press, 2002.
[22] R. Plutchik, ‘A general psychoevolutionary theory of emotion’, *Emotion Theory, Research, and Experience*, **1**, (1980).
[23] E. Rolls, *Emotion Explained*, Oxford University Press, 2007.
[24] J.D. Velasquez, ‘An emotion-based approach to robotics’, in *Intelligent Robots and Systems (IROS)*, (1999).

Ubiquitous Computing and Pervasive Adaptation of Social Norms in Workplace Design

Mónica Sara Santos and **Jeremy Pitt**[1]

Abstract. The design of workplaces, namely the physical arrangement of people and machines, has an active influence on work-related issues, such as productivity and efficiency.

However, even an ideal physical arrangement of machines is subject to the social and emotional intelligence of the people. It is possible to formalise the former using rules of social order, and to capture the latter using the ideas from Affective Computing.

Our intention, then, is to demonstrate that by integrating affective states and participatory/pervasive adaptation of social norms in workplace design, the quality of experience in a working environment can be improved.

1 INTRODUCTION

The design of workplaces, namely the physical arrangement of people and machines, has an active influence on work-related issues, such as productivity and efficiency. This arrangement can be an inhibitor or a facilitator to more social interactions [19]. It is a goal of today's organisations to find ways to effectively use office work environments as means to improve worker performance. Moreover, group performance has become a main question for businesses that rely on collaborative work to achieve the organisational goals [6, 14]. Thus, it is essential that office environments are designed with the purpose of providing dynamic, user-friendly space [22].

It has been argued that social and emotional intelligence are the parts of human intelligence that most influence aspects of success in life, especially in social interactions, learning, and adapting to what is important [20]. Emotions play a critical role in cognitive processes in humans, such as focus and attention [10], organisation of memory and perception [3], motivation and performance [7], planning [17], learning [11], goals generation, evaluation and decision-making [9], and communication [2, 12]. Therefore to study an office environment we should leverage the emotions of each of the individuals that are part of the workplace with technology, to improve the interaction.

Besides the psychological and neurological documentation suggesting the influence of emotions in human interactions, there is evidence to justify that more specifically, emotions influence human interaction with computers. It has been shown that providing positive affective interventions to people who are having difficulties solving a problem with a computer, increases their performance [21]. This conclusion was gathered in an experiment that studied the psycho-physiological effects of positive and negative affective interventions in human-computer interactions. Subjects were exposed to pre-programmed mouse delays, while trying to solve an interactive puzzle. After that, positive or negative interventions were provided via a speech synthesiser and the subjects' responses were recorded and analysed. Another experiment was conducted by Kapoor et al. with the aim of assessing user frustration. A set of children were asked to solve a computer version of the Towers of Hanoi puzzle. The subjects' non-verbal multimodal data was analysed, with the objective of trying to predict when the children were feeling frustrated [16]. Branco et al. observed the spontaneous facial expressions subjects portray while trying to format a document in Microsoft Word, with the objective of identifying adverse event occurrences in the user interface [4]. It has also been proven that while working with computers, people display emotions that are caused by the interaction with the computer. [26]. To assess this, an experiment where some people were continuously recorded while working in their every tasks with the computer, was conducted. Sequences where people were showing emotions that were caused by the interaction with the computer were extracted. Some of these video sequences were presented to 75 people on an online questionnaire, who clearly agreed in the labelling of affective states such as frustration, fatigue and concentration.

In any experimental study it is necessary to consider cultural variations. Cultural personal characteristics influence the way people interact and view each other [5]. For instance, certain behaviours can be seen as natural in a culture and as offensive in another. Also, people from different cultural backgrounds display emotions in different ways. Cultural differences have always influenced the way users interact with computers [23]. Even though cultural issues should not be ignored, "there seems to be a gap between notions of technology and culture, and a lack of appropriate and valid approaches to their synchronisation [27]". A reason for this lack of synchronisation can be explained by the conclusions Kamppuri et al. [15] gathered after examining literature from the main HCI-related journals and conference proceedings published from 1990 to 2005. They concluded that from the 3286 published papers only 28 referred cultural issues in HCI, which represents about 0.85% of all the publications. There are still only a few studies of the ways cultural diversity may influence the users' interaction with computers, so it "has become a new challenge for HCI [15]".

2 RESEARCH QUESTIONS

Ubiquitous Computing (ubicomp) aims at creating digital environments that are sensitive to human needs, and adapt and respond accordingly [25]. In these environments, pervasive applications become ideally invisible, which is made possible by their degree of integration and need for minimal human input. For all this to be achievable, systems and devices that are part of the digital environment need to be context-aware and use this context-awareness smartly. Ubicomp

[1] Department of Electrical and Electronic Engineering, Imperial College London, UK, SW7 2BT, email: {monica.santos, j.pitt}@imperial.ac.uk

requires systems and devices that perceive context in an accurate way, followed by intelligent control or action between machines and humans [25].

We believe group interactions in workplaces could be improved by using ubicomp techniques to detect affective states, followed by a participatory definition and pervasive adaptation of social norms.

We intend to study how social norms in workplaces can be created and adapted taking into account the emotional states of each person that is part of that workplace. After that we want to study in what way all elements of the workplace comply with the created social norms - whether or not there are elements that break the norms, the motivations for breaking the norms and procedures to rehabilitate trust. We also want to know in what way cultural factors affect the behaviour and display of emotions in the workplace.

The first step of the study is to analyse the workplace as a whole and also each of its human elements individually. Visual cues are expected to provide valuable information about the way people react to situations that might be distracting, break concentration, or in any case upset the work stream. Therefore video cameras and web-cameras should be placed around the workplace so that both the interaction between people and each person's reactions are filmed.

The web-cameras should be used to individually record each person, storing personal reactions and displays of emotions that happen when any of the previously referred situations happens. Since cultural differences influence the way people from distinct cultural backgrounds behave, the web-cameras might not be sufficient to provide full information about emotions being felt at a certain moment. Therefore, as an additional experiment we might use brain sensors to analyse different cultural behaviours and displays of emotions.

The video cameras should be strategically placed in the workplace so they capture the interactions and general actions that might be the cause of some group reaction.

We only need to keep the recordings of the minutes immediately before, during and after the moment when some situation affects several elements of the workplace. For those moments the recordings need to be synchronised so they can be analysed as a global event where all reactions and displays of emotions might have been caused by the same situation.

Evidence from several areas of research, such as, for instance, medical studies and social support groups, suggest that people reveal more socially undesirable information about themselves in Computer Mediated Communications (CMC) than when compared to the equivalent face-to-face interaction [13]. It is therefore expected that when people break the social norms of the workplace, their colleagues won't confront them. If this situation persists it might lead to a bad work environment, since it represents a break in the general trust that everyone will follow the norms. By not confronting the breaker of the norms, who might not even be aware s/he has broken them, the remaining elements of the workplace don't give her/him a chance to apologise and/or explain her/himself. And as Vasalou et al. argue, when an offender has the opportunity to repair the action (in this case, whatever broke a social norm), the victim's trust can be restored [29].

Since we want to understand the whole social norms' interaction process, i.e. the way people comply with the norms and how they react when something breaks them, it is important that people express their opinions freely without the social barriers a confrontational situation might impose. So, to help achieve real results, some mechanisms to facilitate the resolution of these conflicts will be developed, in a CMC way.

Some initial formulation of research questions has been made. Though there are already some tentative answers to some of the questions, only the research conducted throughout this study will provide full understanding of the problem.

2.1 What data will be used?

To define the social norms we will need input from everyone that is part of the workplace. This input is composed of descriptions and requests from people, and also of conclusions gathered from observations of people's behaviours. The latter is perhaps the most important input. Often people don't know what they need, especially when it comes to future interactions, which they're not used to dealing with [18]. As stated before, this observation will be made indirectly, by placing video cameras and web-cameras in the workplace, and eventually by measuring additional inputs, such as brain signals, for instance.

2.2 What are the characteristics of the system to be created?

The system to be created is one that provides a computer-mediated interaction between everyone in the workplace, especially between people that are affected by someone else's behaviour (the victims), and the person or persons that behave in a way that upsets other elements of the workplace (the offenders).

The set of social norms will be mapped in the system through a policy-based language. The system will also comprehend decision-making tools, which control all feedback provided, such as, for instance, emails warning the offender of his/her inadequate behaviour. These decision-making tools will receive as inputs the social norms policies and historical data about previous offences and everyone's reactions to those offences. Inputs from brain sensors will eventually be added at a later stage. The brain sensors should provide real-time information about affective states and facial expressions of each person of the workplace.

The envisioned system needs to provide a solution for victims who are too introvert to let the offenders know they're upsetting them, to indirectly do so. This interaction might be anonymous so that people who might feel inhibited by direct confrontations, do not feel constrained in signaling offences.

All violations of norms will be reported by individuals, usually by the elements of the workplace. In order to have have an automatic detection of these offences, there would have to be automatic video analyses, which are not part of the scope of this study.

The system should also collect and relate data, such as:

- Who are the offenders (who has ever broken a social norm),
- How often each offender breaks the social norms,
- How many of the victims signal each of the offences (the breaking of the norms),
- How the offenders react to negative feedback from the system (whether or not they try to change their behaviour, whether or not they provide an explanation for it, ...),
- How the victims react to the redeeming behaviour from the offender.

Additionally, the system should keep every element of the workplace informed about the social norms and their personal level of compliance with them. For instance, an element that has never broken the social norms has a higher level of compliance than someone who has been an offender before.

Given the system's automated reasoning and decision-making on what is signalled as a violation of the norms, classification errors might occur occasionally. Since the offences are always flagged by the human elements of the workplace, and the system then decides whether or not to signal flagged events as violations, if classification errors happen, they are expected to happen because the system was too conservative in considering something as an offence, rather than classifying an innocuous action as a violation.

2.3 What techniques will be used?

The techniques to be used comprehend, among others:

- Analyses of inputs from the web-cameras and video cameras placed around the workplace.
- Interviews with workplace elements.
- Analyses of data generated by the usage of the system.

2.4 What interfaces will be created?

The full set of interfaces to be created in the system will be defined in a later stage of this study, as some of the interfaces will be a reflection of the needs of the workplace.

The system will be installed in each computer of the workplace, so everyone has access to the same tools and information. There is information about the social norms and the self- and other people's behaviour towards them.

This information might be presented in a map of the workplace. Here each of its human elements might be represented by an avatar that is colour-coded, according to the person's level of compliance with the social norms. This colour codification is based on the reasoning made by the system about each element's compliance with the norms and should vary throughout time.

When someone breaks the norms, every element of the workplace that feels upset by that behaviour can (and should) click on the offenders' avatars and send feedback about their behaviour. This feedback should also indicate what kind of action caused the person to feel upset, either the breaking of an established social norm, or a different action that is not (yet) contemplated by the norms. Also, the offenders should know how many people are disturbed by their behaviours and exactly what action caused the disturbance.

An offender can redeem her/himself by sending feedback to the people who complained, explaining why s/he broke the social norms. After this interaction the victim(s) should decide whether or not to change the rating they initially gave to the offender.

2.5 How will the social norms be expressed and adapted?

The social norms will be expressed in the system through a policy-based language, probably ALP [1]. Policies are used to dynamically control the behaviour of system components without the need to change the code. They provide the flexibility of implementing variations in the system that reflect externally imposed constraints or environmental conditions [28].

The social norms will be initially defined after observation of the work group and the individuals, and by interviewing everyone who is part of the workplace. At least the majority of the workplace elements have to agree on a rule for it to be established as a social norm for that workplace.

The policies will then be set to define the rules of behaviour in the workplace. They will be one of the inputs for the system's decision making module.

After the system is implemented, the data collected from all the interactions will be periodically analysed. This new information might point towards the need to create new social norms, based on indications from victims of behaviours considered to be disturbing to the workplace harmony. Once again, whenever a new social norm or a change in an already existing social norm is proposed, there has to be a high level of agreement between the elements of the workplace, for it to be implemented.

2.6 How are emotions going to be related and mapped with the social norms?

Although there are various descriptions for affect, we are going to use the one proposed by Russell [24], as it is simple and capable of capturing a wide range of emotions and shades of emotions [8]. Russell defines a multidimensional emotional space where a horizontal and a vertical axis define positive and negative values of valence and arousal, respectively. Each emotion or affective state falls within one of the quadrants defined by these two axes.

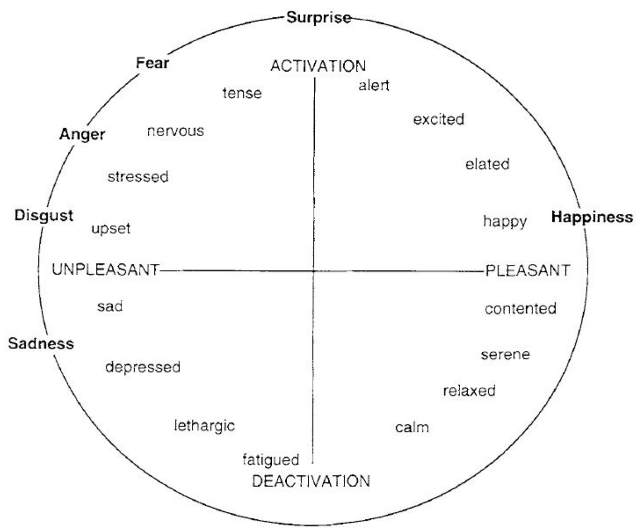


Figure 1. Russell's circumplex model of affect

The four distinct quadrants represent combinations of activation/deactivation and positive/negative affective states. Valence ranges from unpleasant to pleasant affective states. Arousal defines a level of emotional activation that ranges from low (i.e. fatigue) to high (i.e. alertness).

As we want to analyse what happens when some event upsets the regular work pace, the most relevant quadrant is the upper left one, as it represents unpleasant and active emotions. Hence, the main attention in the analysis of individual emotions should be given to cases where people are feeling upset, frustrated, distracted and annoyed, amongst others, when these affective states are caused by the actions of someone else in the workplace. If certain actions cause some kind of emotional reactions (with special focus on the ones from the up/left quadrant) in a continuous way, those actions might be proposed as a "what not to do" in the workplace, and originate new social norms.

2.7 How is the system going to be evaluated?

The overall objective of the system is to facilitate communication, so that group interaction in the workplace is improved. The evaluation of the system has to be done in two ways: assessing in which manner the group interaction is affected by the system; and evaluating the system's performance.

Immediate indicators to assess in what way the workplace environment has been improved, are the amount of offences reported in the system and the individual ranking of each element of the workplace. For the system to be successful these rankings should be improved and the total number of reports should decrease. Interviewing the workplace elements should also provide information about the way the workplace environment has changed by the introduction of the system.

When it comes to the system's performance, there should be additional filming that captures people's reactions to triggered breaking of the norms. These reactions should then be compared to the system's assumptions for those same situations, in establishing which of those actions were breaking the norms, and the consequent provided feedback.

3 CONCLUSIONS

One of the goals of today's organisations is to find ways of effectively using office work environments as a means to improve worker performance [19]. We believe one way of doing so is by improving the quality of experience in a working environment, by integrating affective states and participatory/pervasive adaptation of social norms in the workplace design.

We will start this study by analysing the workplace structure both collectively and individually. For this purpose, video cameras and web cameras will be placed in the workplace to be studied. The video cameras will capture global actions in the workplace, including the actions that might upset the work pace. The web cameras are expected to show individual displays of emotions by each of the human elements of the workplace, and will, therefore, be placed in front of each of those elements. With the outputs of these recordings and information gathered in interviews with the people of the wokplace, social norms will be defined.

There is evidence that suggests people reveal more socially undesirable information about themselves in CMC than when compared to equivalend face-to-face interaction [13]. We propose to build a system that will include the previously defined social norms, expressed by a policy language. This system will work as a mediator between the people in the workplace, and will help them to provide feedback about their satisfaction with the workplace environment. This feedback includes especially the cases where people feel harmed by the breaking of the norms by another element of the workplace. It has been shown that often when an offender has the opportunity to apologise and repair the damage s/he has caused, the victim's trust can be restored [29]. Hence it is another objective of the system to serve as a means of enabling trust recovery actions.

REFERENCES

[1] A. Bhusate, L. Kamara, and J. Pitt, 'Enhancing the Quality of Experience in Cultural Heritage Settings', in *Proc. 1st European Workshop Intelligent Technologies for Cultural Heritage Exploitation, 17th European Conf. Artificial Intelligence*, pp. 1–13, (2006).

[2] R.L. Birdwhistell, *Kinesics and Context: Essays on Body Motion Communication*, University of Pennsylvania Press, 1970.

[3] GH Bower, 'Mood and memory.', *Am Psychol*, **36**(2), 129–48, (1981).

[4] P. Branco, P. Firth, L.M. Encarnação, and P. Bonato, 'Faces of Emotion in Human-Computer Interaction', in *Conference on Human Factors in Computing Systems*, pp. 1236–1239. ACM New York, NY, USA, (2005).

[5] R.W. Brislin and E.S. Kim, 'Cultural Diversity in People's Understanding and Uses of Time.', *Applied Psychology*, **52**(3), 363, (2003).

[6] SG Cohen, 'New Approaches to Teams and Teamwork', *Organization for the Future: The New Logic for Managing Complex Organizations, Jossey-Bass, San Francisco, CA*, 194–226, (1993).

[7] JA Colquitt, JA LePine, and RA Noe, 'Toward an Integrative Theory of Training Motivation: A Meta-Analytic Path Analysis of Twenty Years of Research', *Journal of Applied Psychology*, **85**(5), 678–707, (2000).

[8] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and JG Taylor, 'Emotion Recognition in Human-Computer Interaction', *Signal Processing Magazine, IEEE*, **18**(1), 32–80, (2001).

[9] A.R. Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*, Quill New York:, 2000.

[10] D. Derryberry and DM Tucker, 'Neural Mechanisms of Emotion.', *Journal of Consulting and Clinical Psychology*, **60**(3), 329–38, (1992).

[11] RJ Dolan, 'Emotion, Cognition, and Behavior', *Science*, **298**(5596), 1191, (2002).

[12] P. Ekman and W.V. Friesen, *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*, Prentice-Hall, 1975.

[13] A.N. Joinson, 'Self-Esteem, Interpersonal Risk, and Preference for E-Mail to Face-To-Face Communication', *CyberPsychology & Behavior*, **7**(4), 472–478, (2004).

[14] S.D. Jones, M. Buerkle, A. Hall, L. Rupp, and G. Maut, 'Work Group Performance Measurement and Feedback: An Integrated Comprehensive System for a Manufacturing Department', *Group & Organization Management*, **18**(3), 269, (1993).

[15] M. Kamppuri, R. Bednarik, and M. Tukiainen, 'The Expanding Focus of HCI: Case Culture', in *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles*, pp. 405–408. ACM New York, NY, USA, (2006).

[16] A. Kapoor, W. Burleson, and R.W. Picard, 'Automatic Prediction of Frustration', *International Journal of Human-Computer Studies*, **65**(8), 724–736, (2007).

[17] JE LeDoux, 'Brain Mechanisms of Emotion and Emotional Learning.', *Current Opinion in Neurobiology*, **2**(2), 191–7, (1992).

[18] J. Nielsen, *Usability Engineering*, Morgan Kaufmann, 1993.

[19] D.J. Oborne, *Ergonomics at Work: Human Factors in Design and Development*, Wiley, 1995.

[20] M. Pantic, N. Sebe, J.F. Cohn, and T. Huang, 'Affective Multimodal Human-Computer Interaction', in *Proceedings of the 13th annual ACM international conference on Multimedia*, pp. 669–676. ACM New York, NY, USA, (2005).

[21] T. Partala and V. Surakka, 'The Effects of Affective Interventions in Human-Computer Interaction', *Interacting with Computers*, **16**(2), 295–309, (2004).

[22] M.M. Robertson and Y.H. Huang, 'Effect of a Workplace Design and Training Intervention on Individual Performance, Group Effectiveness and Collaboration: The Role of Environmental Control', *Work*, **27**(1), 3–12, (2006).

[23] Y. Rogers, H. Sharp, and J. Preece, *Interaction Design: Beyond Human-Computer Interaction*, Jon Wiley & Sons, Inc, 2002.

[24] J.A. Russell, 'A Circumplex Model of Affect', *Journal of Personality and Social Psychology*, **39**(6), 1161–1178, (1980).

[25] D. Saha and A. Mukherjee, 'Pervasive Computing: A Paradigm for the 21st Century', *COMPUTER*, **36**, 25–31, (2003).

[26] M.S. Santos and M. Pantic, 'Do People Emote While Engaged in HCI Office Scenarios?', in *Proceedings of IADIS International Conference IHCI (Interfaces and Human Computer Interaction) 2008*, pp. 101–108. IADIS, (25-27 July 2008).

[27] S.T. Shen, M. Woolley, and S. Prior, 'Towards Culture-Centred Design', *Interacting with Computers*, **18**(4), 820–852, (2006).

[28] G. Tonti, J.M. Bradshaw, R. Jeffers, R. Montanari, N. Suri, and A. Uszok, 'Semantic Web Languages for Policy Representation and Reasoning: a Comparison of KAoS, Rei, and Ponder', in *International Semantic Web Conference*, pp. 419–437. Springer-Verlag; 1999, (2003).

[29] A. Vasalou, A. Hopfensitz, and J.V. Pitt, 'In praise of Forgiveness: Ways for Repairing Trust Breakdowns in One-Off Online Interactions', *International Journal of Human-Computer Studies*, **66**, 466–480, (2008).

Haptic Space and Bodily Expressions: A Bi-directional Relation of Affect

Myrto Karanika[1]

Abstract. Extensive research on bodily and emotional expression has followed the increased interest in virtual reality as well as the recent developments of motion tracking technologies. However, most of these technologies are vision-based, consequently lacking the physicality of bodily expression itself. Moreover, such technologies tend to isolate the expressive body from its surroundings, thus interfering in the relationship between the body's expressions and the environment that engenders it. This position paper presents an attempt to explore bodily expressions in a tactile manner through the tangible properties of physical space itself.

1 INTRODUCTION

Investigating the bi-directional relation that we share with our surroundings, my work is narrowing down the focus on the relationship between spatial experience and bodily expression. Historically, spatiality has been addressed as a matter of measures and distances, with little room left for its tangible, affective dimension. As a result, the variable array of bodily senses has been greatly disregarded in an attempt to emphasise on a distant, idealized visuality. However, spatial experience is always embodied and multisensory, equally dependant on vision, hearing, smell and touch.

In this paper, I will be briefly discussing the fundamental relation of the sensuous body with spatial experience, and I will be presenting my current work, which is an attempt to create a responsive haptic environment that shares a bi-directional relationship of affect with the body. I am proposing such an environment to be entirely constructed of a multi-textured fabric interface that not only evokes bodily expressions but also captures them in a tactile manner without use of sensors or vision-based tracking systems. Designed as a dense conductive grid, this textile spatial element can accurately translate bodily gestures into arrays of coordinates which are in turn fed into MAX/MSP to be translated into sound. Therefore, user engagement with the interface not only depends on their bodily gestures but also requires a close interrelation of their senses of vision, touch and hearing.

The following section will start with a short introduction to basic concepts of haptic space and its relation with embodied experience and emotional response. From there, I will continue with an overview of my work and how it is placed within the fore mentioned theoretical platform. The last section will be concerned with the technical details of the textile haptic interface I have designed and the gesture tracking method it employs. For the purposes of the AISB 2009 Symposium on Mental States, Emotions and their Embodiment, I am proposing a live demonstration of gesture tracking, using a sample of the fabric prototype.

2 HAPTIC SPACE: A CONTINUUM OF BODILY AND EMOTIONAL RESPONSE

Spatial experience is a synthesis of all of our senses; within this synthesis all senses are interrelated and co-dependent and that constitutes their distinctness or separation purposeless when it comes to spatial perception [1]. In their famous A Thousand Plateaus, Deleuze and Guattari [2] argue that haptic space 'may be as much visual or auditory as tactile', acknowledging that haptic embraces the sensory interrelation of the eye, the ear and the limbs. From this point of view, haptic is extended to address the essence of our embodied spatial perception; a perception that is simultaneously orchestrated by our vision, hearing and touch, and that therefore reflects our bodily experience of space's textural qualities: weight, mass, density, pressure, humidity, temperature, presences, and resonances.

However, haptic can also be extended to involve emotional connotations and to reflect affective response. Translating the words haptic, sense and emotion in Greek, my mother language, the interconnection of the three concepts becomes obvious at once. Haptic originates in the Greek word απτό, which means something that can be touched or grasp-ed. Sense, translated as aesthesi / αίσθηση, in Greek involves notions of feeling, grasp-ing and understanding. Consequently, the concept of 'grasp', in other words perceive, is core in both sense and haptic. Emotion on the other hand, translated in Greek as αίσθημα / aesthema, shares the shame root with aesthesi, as both derive from the word αισθάνομαι / aesthanome, whose ambiguous meaning can be equally translated as 'I sense' or 'I feel'. Among these three words -haptic, sense, and emotion- there is an underlying relation that, if examined closely, reveals the very nature of haptic as a sense that is ultimately bounded with emotional grasping.

The idea of 'haptic' embodying notions of emotional experience / attachment has been repeatedly used by theoreticians like Merleau-Ponty [3], Kant [4] and Paterson [5]. Berenson [6] notes that our bodily response to the 'tactile properties' of our surroundings –and space- highly depends on our understanding of their ability to affect and 'touch' us, while Fisher [7] addresses haptic as the merging of the bodily senses and the affective aspect of what creates them.

Drawing on the above, my study on the relation of bodily and emotional response with the space that encompasses them starts with the design of a responsive haptic environment that addresses all sensory data as an inseparable narrative pathway upon which our spatial experience is unfolded. That is an environment whose qualities can trigger our senses, affect our bodily expressions and can be affected by them. Such an environment should be able to

[1] Dept. of Computing, Goldsmiths Univ. of London, SE14 6NW, UK.
Email: ma701mk@gold.ac.uk.

not only evoke bodily expressions but also to capture them and 'feed' them back to its 'organism'. Of course, similar approaches have repeatedly taken place since the advance of computational systems that can provide interactive modes of communication between a space and its users. In most of the cases though, communication is established through distant modes of interaction such as sensors and vision-based tracking systems.

It is my intention to engender a bi-directional relation of affect between the body and its surrounding environment that is entirely based on the two agents of the interaction: the space and the body, without having to embed 'external' systems into their channel of communication. This mode of interaction springs, like Palasmaa [8] puts it, from the tactile sensibility of 'enhanced materiality, nearness and intimacy'. To model such a form of intimate, tangible interaction, my focus has been on the design of a spatial interface that is capable of 'perceiving' bodily expressions itself, and which also presents a range of textural qualities that challenge bodily responses. My approach is greatly influenced by the work of Finnish architect Juhani Pallasmaa [9] who notes that space should be re-sensualised 'through a strengthened sense of materiality, hapticity, [and] texture'; also by the work of Bloomer and Moore [9] which propose textural change as a generator of sensations that link the haptic materiality of a space with the bodies that inhabit it.

3 AN AUDIO-HAPTIC INTERFACE

To meet these goals, I have designed a custom-made fabric to be used as an enveloping interface for an installation space. This fabric prototype is knitted with non-conductive thread (PA, diameter of 0.20mm), and has conductive wire (tin copper, diameter of 0.10 mm) embedded on both its outer sides, horizontally on the one and vertically on the other, thus forming a conductive grid.

Figure 1. Example of Vertically Embedded Conductive Bands

The conductive bands are wired to a complex of keypad encoders, which is in turn connected to an Arduino microcontroller. That allows for the physical textile nodes to be perceived within the Arduino programming environment as elements of a matrix whose rows and columns are accordingly equivalent to the parallel and vertical conductive bands of the fabric. Eventually, that enables the prototype to simulate a tactile, numerical interface whose resolution depends on the density of the conductive grid. The conductive elements do not make contact within the same plane unless they are compressed by touch. When the fabric is being touched, the encoders detect which conductive elements make a connection.

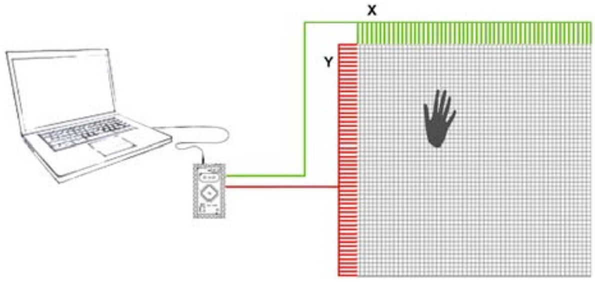


Figure 1. Interaction Design System

This way, the gestures of the users upon the interface are captured as arrays of compressed grid nodes, and are 'transduced' into arrays of integers that respond to the matrix elements. These integers are then passed to MAX/MSP to generate sound accordingly to the users bodily gestures.

Before, explaining in more detail how sound is produced from the gestural movements of the users upon the fabric prototype, it is important to refer to the physical qualities of the interface when exhibited in space as well as to the reasons for which I have decided to relate the interface with sound generation. Both sides of the prototype are layered with a translucent tulle surface upon which I am embroidering a variety of different stitches using yarns that vary in colour and weight. Apart from embroidery, I am also using a number of different techniques to process the tulle such as printmaking and collage. These processes result into a highly textured surface that acts as the skin of the prototype interface.

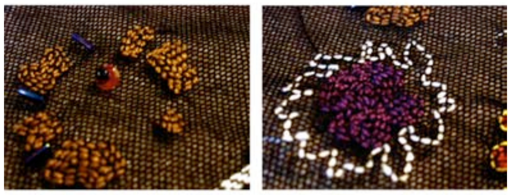


Figure 3. Details of the Embroidered Surface

With the conductive grid acting as the 'nerves' of the interface and the processed tulle acting as its skin, a quite abstract representation of the textile spatial element as a living organism evolves; a representation that sets the ground for a bi-directional relation of affect between the interface-enveloped space and the bodies it encloses. The textured surface of the envelope attempts to intrigue the users senses of vision and touch, aiming to evoke bodily engagement. As soon as the users engage with the interface through the medium of touch, their gestures are translated into sound. That enables a straightforward relation between the visual / haptic qualities of the interface and the generated sound, allowing for gestural patterns to be 'choreographed' and perceived both by the haptic qualities that engender them and by the audio output they generate.

A number of different audio samples map the different textural / chromatic qualities of the processed prototype skin, with 'warmer' sounds mapping the interface areas that are dominated by warm colours and/or smooth materials and vice versa. Within each textural area, a central grid node is assigned a given sound, and acts as the 'command centre' for its peripheral nodes. That means that within a certain radius –defined by the size of each distinct textural area- the sound of all neighbouring nodes is interpolating with respect to their distance from the central node.

Figure 4. Example of Audio Interpolation Mapping

When more than one person is engaging with the interface the sound is being produced as the merged outcome of their embodied engagement with the interface and with each other. The envelope can be approached from both its inner and outer side; as its weaving allows a certain level of translucency, the users' figures become part of the interface patterns. Thus, apart from an auditory-oriented collaboration of the users' gestures, a visual level of interaction among them holds also an important role in the orchestration of their bodily expressions.

4 CONCLUSION

In this paper I have presented my attempt to design a responsive haptic environment that explores bi-directional relations of affect between space and its users by addressing the close collaboration of the senses of vision, hearing and touch as a medium for a fully embodied spatial experience. Within this relation both space and body are considered as living organisms that can equally affect and be affected by each other. The mode of affection between the two agents is immanent in their interaction without the need for 'external' systems, such as sensors or camera tracking methods, into their channel of communication.

Such an environment consists of a space that is being enveloped by a highly-textured conductive fabric prototype, which can 'perceive' the users gestures as arrays of matrix elements. These elements are then being translated into sound, thus merging vision and touch (input) with hearing (output / and input) into a sensuous loop that 'orchestrates' the users bodily expressions and changes the space's audio qualities.

The work presented in this paper is still in a very early stage of development. The description I have provided so far is strictly based on small scale sample testing I have practiced myself. I am expecting improvements considering the accuracy of gestural tracking and sound generation as soon as I have user testings in larger scale pieces of the prototype. I therefore consider the AISB 2009 Symposium on Mental States, Emotions and their Embodiment to be an exceptional opportunity to present and perform the application live to a wider audience, and I am looking forward to their feedback.

AKNOWLEDGEMENTS

The work presented in this paper is being developed as part of my MFA Computational Studio Arts degree. I would like to thank my tutors Janis Jefferies, Jane Prophet and Andrew Shoben as well as AHRC for supporting my studies. Also Olly Farshi and Jeremy Keenan for their contribution to the sound design.

REFERENCES

[1] M. Paterson, The Senses of Touch: Haptics, Affects and Technologies. Oxford: Berg (2007).
[2] G. Deleuze and F. Guattari. A Thousand Plateaus: Capitalism and Schizophrenia. London: Athlone (1988).
[3] M. Mereau-Ponty. The Primacy of Perception and other Essays on Phenomenological Psychology, the Philosophy of Art, History and Politics. Evanston IL: Northwestern University Press (1964).
[4] I. Kant. Critique of Pure Reason. London: Macmillan (1990).
[5] M. Paterson. The Senses of Touch: Haptics, Affects and Technologies. Oxford: Berg (2007).
[6] B. Berenson. The Florentine Painters of The Renaissance. London: G.P. Putnam's Sons (1906).
[7] J. Fisher. Relational Sense: Towards a Haptic Aesthetics. *Parachute 87*, 1:4-11 (1997).
[8] J. Pallasmaa. Hapticity and Time. Notes on fragile architecture: *Architectural Review*, 207:78-84 (2000).
[9] J. Pallasmaa. The eyes of the skin: Architecture and the senses. London: Academy Editions (2005).
[10] K. Bloomer and C. Moore. Body, Memory and Architecture. New Haven, CT, Yale University Press (1978).

Affect in Autonomous Artificial Systems: Interfacing Technology and Philosophy

Chryssa Sdrolia[1]

Abstract. Inspired by the latest developments in affective artificial systems, this paper is concerned with investigating the insertion of emotion in artificial intelligent systems and mapping its implications on a philosophical, political, and cultural level. This goal is primarily pursued by drawing multiple parallels between the philosophies of representation and expression that have problematized the concept of 'emotion' as well as between the technological and philosophical uses of the loaded concepts of 'emotion' and 'affect.' What is argued is that the affective turn in artificial systems necessitates and at the same time creates an interface between technology and philosophy, which is vital for the articulation of a viable ethics within the larger framework of late post-industrial capitalism.

1. INTRODUCTION: THE CONTEXT

With networked systems having already infiltrated the patterns of everyday life and technology manifesting an ever-increasing ability to overcome its technical constraints, scientists are focusing their attention on the value of emotive expression as the condition and strategy *par excellence* for the success of human-machine interaction (Duffy 177). Efforts for the seamless insertion of machines into human ecologies are thus drawing upon an understanding of the body as an affective entity. Within the field of affective systems, this interest is especially manifest in evolutionary robotics for which the coupling of control software with the (machinic) body as a physical event of expression is vital for the creation of fully operational autonomous agents. It is known that since the late 1990s, at least, research has been revolving around the creation of emotionally grounded robotic architectures also capable of displaying real-time/real-world perception, mobile and multi-modal interaction competence (Breazeal 2003, Cañamero 2005, Picard 2003). Posing a number of practical and theoretical challenges, the affective turn in robotics and artificial intelligence, in general, is as challenging as it is vexed. The technical difficulty in synthesising emotion and emotional expression as well as the conceptual indeterminacy of 'emotion' vis-a-vis 'affect' has created the urgent and rather felicitous need for trans-disciplinarity. Cognitive psychology, neuroscience, ethology, and philosophy have been marshaled to aid the development of a computational framework that captures emotional processing and integrates it with other models of perception, behaviour, and motor control. Taking as a point of departure this productive tension, my interest in the area is thus primarily informed by the problematics posed by philosophy and cultural theory.

While scientists face their own problems in attempting to integrate emotional expression into intelligent robotic and computational models, in humanities the debates concerning the validity and implications of the concepts 'emotion' and 'affect' respectively are part of a long tradition and still raging. Yet before sketching out the philosophical theoretical approaches to the subject, it makes sense to mention in passing the historically troublesome relationship between the two fields. With the exception of a few schools to which I will shortly return, thinking alongside scientific developments and knowledge is a relatively recent trend in the humanities and one that is complicated enough in itself. As Rosi Braidotti notes, it 'runs against a well-established tradition of criticism, if not of actual rejection of the 'hard' sciences in social theory' (Braidotti, 2002 230). The mistrust has been primarily directed against the supposed 'objectivity' of modern human and hard sciences (physical, biological, psychological, and social). The main argument has been that these purport to offer universal scientific truths about human nature that are, in fact, often mere expressions of ethical and political commitments of a particular society. Notable currents of critical philosophy, including poststructuralism, feminism, social constructionism, and postcolonialism have undermined such claims by exhibiting how they are more often than not the outcome of contingent historical forces and not scientifically grounded truths. In Foucault's theorizations, for instance, as systems of representation the various sciences actually produce the rules and practices they need thus engendering meaningful statements and regulating discourses in different historical periods (Foucault 196-215). Whereas I would not like to ignore or invalidate the important implications mentioned above, I would nevertheless like to turn to the fruitful convergence of current technologies with the interests of contemporary philosophy, also keeping in mind the equally dubious part many a strand of philosophical thought has played in the shaping of political reality. After all, the mutually perceived gap between the sciences and philosophy and their respective responsibilities has not nearly been that clear-cut and is rather 'a symptom of the anxiety of contamination' (Huyssen ix) of each other's presumed purity.

2. EMOTION AND AFFECT IN THE PHILOSOPHIES OF REPRESENTATION AND EXPRESSION

Being pro-contamination, I thus endorse Gilles Deleuze's opinion that '[a]rt, science, and philosophy…are caught up in mobile relations in which each is obliged to respond to each other, but by its own means' (Deleuze xiv). My endeavour is not to criticise engineering techniques or scientific knowledge as unscientific but to pick up from that 'beyond the scope' point that many scientific papers naturally leave open when faced with the ethical and socio-cultural implications of their technological artifacts. The affective paradigm shift in intelligent systems and the sheer complexity of 'affect' can thus been seen as clearing up a promising plane of interaction, an interface between technology and philosophy. To resume from where I left it, the import of affective systems seems to be paralleled by a quite old interest within the humanities in the conceptual grounding of emotion. A first observation that could be made at this point is the different use both disciplines make of the concepts 'emotion' and 'affect' themselves. In scientific writings these two terms usually figure interchangeably or are subcategorised with emotion overriding affect. In a number of papers, the second is presented as an aspect that connotes 'moods' or 'drives' and hence is not considered to be a fully-fledged emotion which is more properly connected to distinctive self-conscious cognitive states (Khulood & Raed 696). What is important here is that this clarity-based categorisation is contested in philosophy which locates a fundamental disparity between the two. Different schools have drawn attention to the sociopolitical uses of 'emotion' as a representational construct. The argument is that as a concept and discursive practice it actually presupposes and propagates dominant notions of subjectivity and selfhood in the interests of hegemonic ideology and power structures. Deconstruction and transcendental empiricism have, each from a different perspective, countered the classical folding of emotion into unitary accounts of the mind. Yet while deconstruction retains the term 'emotion,' transcendental empiricism with which my own sympathies lie drops it in favour of 'affect.' The conflict between emotion and affect is thus caught between philosophies of representation and philosophies of immanence/ expression, a fact which could be fruitfully connected to the debates within the scientific community as to the efficiency of representation in the construction of artificial systems and their future with or without it (Brooks 1991b, Műller 2007).

As emotion is entangled in the mysteries of consciousness, its history has been locked inside the classical histories of mind and will, and hence, to subjectivity conceived through representation. Emotion is thus seen as the territory that remains to subjects after the realisation that even not strictly 'proper' subjects have affects. Through the exposure of the supposed transparency, universality and presence of representation as a 'white mythology,' deconstructionists have pointed to the fact that '[t]he classical picture of emotion already contraindicates the idea of the subject' (Terada 7). Approaching a theory of emotion without however stating it as such, Jacques Derrida shows how emotion is embedded within and calls forth the textual structures that belong to what is known within the humanities as the 'death of the subject' – a pattern that emotion actually surpasses. He does this by challenging Plato, Rousseau and mainly Husserl's conception of 'auto-affection' which conceives subjectivity on the basis of representational translucence (Derrida 1973 48-70). According to the Husserlian phenomenological schema, the mode of intentional, conscious self-reflexivity guides the route from affects, namely mere corporeal sensations, to meaningfully interpreted emotions that can be ascribed to cognizant subjects only. In this sense, and as an outcome of higher cognition, emotion is what happens when the subject practically represents itself to itself. Against the classical metaphysics of presence, Derrida thus undertakes the task of revealing the non-coincidence of represented meaning with a supposed fixed content.[1] The persistent *differánce* he traces within representation actually paves the way for a textual schema of emotional yet non-subjective experience. This entails the idea that because mental representations are never quite properly faithful *re*-presentations of a supposed unified subjectivity, they are also never quite equal to themselves. As such, subjectivity remains a fictional crossroads that is always deferred and never really crossed. Contrary to what Husserl and classical phenomenology would perceive as the complete effacement of emotion, this productive impossibility of the subject thus preserves emotional experience minus a central self-conscious subjectivity. To cut a long story short, emotion in poststructuralist theory 'does not demand a subject, unless an infinite abysm of transpersonal perspectives is your idea of a subject' (Terada 46).

The deconstructionist debunking of subjectivity in favour of emotive experience beyond intentionality is not very far from Daniel Dennett's theoretical formulations that non-subjective interpretation alone explains experience. Not unlike Derrida, his suggestions are concerned with the issue of experience that underlies most theories of emotion. In his formulation of 'heterophenomenology' and debate about 'qualia,' the notion of the self-generating, self-conscious subject is vividly undercut as the 'persuasive imagery of the Cartesian Theatre [that] keeps coming back to haunt us – laypeople and scientists alike – even after its ghostly dualism has been denounced and exorcised' (Dennett 1991 107). With this move Dennett challenges the content approach to emotion as a representation that requires a subject-overseer. Revealing the subject to be the personified correlative of qualia, he actually problematises the deeply rooted common intuition of a 'central Witness['s]' intentional stance. In his own words:

> These raw materials, whether they are called 'sense data' or 'sensations' or 'raw feels' or 'phenomenal properties of experience,' are props without which a witness makes no sense. These props, held in place by various illusions, surround the ideas of a central Witness with a nearly impenetrable barrier of intuitions (Dennett 1991 322).

Dennett offers an account in which qualitative states become hostages of the human subject but which nevertheless manage to live on (Terada 110). Besides the serious debates this model has generated among engineers and computer scientists as to whether it is possible or not to construct feeling machines, his 'sub-personal' (Dennett 1981 228-229) materialist philosophy most importantly points to the fact that 'self-differential selves are dead only as [unified] subjects' (Terada 156).

The abovementioned idea is stretched even further by the school of transcendental empiricism which is closer to that materialist framework than deconstruction. Notwithstanding the valuable conceptual and political implications of deconstructionist theory, the criticism against Derrida is that he never really breaks from the endless circle of representational iterability and remains more confined in the boundaries of philosophy 'proper.' As such, emotion is caught in the state of causing only a momentary setback to the supremacy of the representing subject since it cannot but constantly repeat it. By

[1] Deconstruction's central concern is a radical critique of the Enlightenment project and of metaphysics. It identifies in the Western philosophical tradition a 'logocentrism' or 'metaphysics of presence' (sometimes known as *phallogocentrism*) which holds that speech-thought (the *logos*) is a privileged, ideal, and self-present entity, through which all discourse and meaning are derived (Derrida 1976 141-143).

contrast, and resting on Spinoza on whom he heavily draws, Deleuze offers a way out of the impasse by mobilising the concept of 'affect' and making of his philosophy one of expression or immanence. His theory begins with a strong critique of representation as being part and parcel of the same 'state philosophy' that has haunted Western metaphysics since classical times (Deleuze 164-213). Representational thinking, Deleuze argues, constructs monstrous discursive monuments by forcing an analog of 'symmetrically structured domains' that stifle difference in favour of a slavish adherence to the ideality of transcendental Forms. As Massumi sums it up, '[t]he subject, its concepts, and also the objects in the world to which the concepts are applied have a shared, internal essence: the self-resemblance at the basis of identity' (Deleuze and Guattari 2004a xi). Contrary to the policing intentions of representational thought, the philosophy of immanence argues for affective expression by emphasising the importance of material bodily experience. In this framework, expression is not confused with the traditional subjective expressive hypothesis which is just another symptom of representation but taken to enact the material actualization of virtuality.

Bearing the flavour of theories of emergence by which it is influenced, Deleuzean theory is based on the dynamic bond between form and matter, thus tying expression to the body and engaging the two in a process of 'becoming' rather than 'being.' The force of this formulation lies precisely in the fact that it interrupts the ontological transcendental with the empirical. As Braidotti puts it, '[the] affective stratum makes it possible for Deleuze to speak of a pre-discursive moment of thinking' (Braidotti 2002 74). With a single stroke, he short-circuits interpretive subjective thought as much as he surpasses emotion. This, nevertheless, does not mean that the subject is completely done away with as common criticism against Deleuze has it; rather than discarding the subject, the focus is on its functioning as a virtual threshold through which transversal flows are constantly combined and dismantled not unlike the turbulent involutions and foldings of form and matter. The Deleuzian self is not bogged down to self-assuring ideality but is 'fascinated, always stretching to its breaking point, to the continuation of another multiplicity that works it and strains it from the inside' (Deleuze and Guattari 2004b 275). The state that becomes such a self is therefore not an emotional but an affective one. In this sense, it bears a remarkable affinity to Francisco Varela's description of auto-affection as the chiasmus of organic forces, the result of which is 'a nonsubstantial self that acts as if it were present, like a *virtual interface*.' 'The more we see the selfless nature of ourselves in various regions of the organism" Varela continues, "the more we become suspicious of our feelings of "I" as a true center" (Varela 61, original emphasis). In a similar vein, Deleuzian affect attempts to account for the multivalence and ubiquity of affective states in connection to the phenomenality of experience in a way that is surprisingly close to current developments in evolutionary affective systems: feeling-affects as the responses of affected bodies occupy the interval between affection and action. As he says, '[t]he pure self of the 'I think'…appears to be a beginning [of philosophy and thought] only because it has referred all its presuppositions back to the sensible, concrete empirical being' (Deleuze 164). Freeing the empirical being from the pretentions of the intentional subject, Deleuze thus restores thought from the status of mere reflection to an intensive process, the self as a threshold of flows, and the body as a plane of intensive affective states.

3. FROM AFFECT TO AFFECTIVE ARTIFICIAL SYSTEMS

For a number of reasons, including its openness to scientific knowledge, the Deleuzian framework proves to be particularly efficient in conversing with the late progress in affective computing and robotic autonomous systems. Recasting thought and the subject in an intensive loop with materiality and emphasising the body as a highly expressive medium, 'affect' serves to interface philosophy and technology. On a methodological level, it puts to the test the philosopher's universal self-indulging stance of assuming a bird's eye view of the developments of a supposedly mundane world. As such, it rids the interaction of the two fields of supremacist fantasies. As Deleuze puts it:

> Philosophy obviously cannot claim the least superiority, but….creates and expounds its own concepts in relation to what it can grasp of scientific function and artistic constructions. A philosophical concept can never be confused with a scientific function or an artistic construction, but finds itself in affinity with these in this or that domain of science or style of art. Philosophy cannot be taken independently of science or art (Deleuze xiv).

On a practical level, the regime of affect takes us full circle to the developments that are specific to current scientific debates over the '[synthesis] of emotions as the primary means to create believable autonomous synthetic agents' (Velásquez 70). Outside philosophy, engineers and computer scientists are equally troubled by the efficacy and uses of representation in the construction of artificial models, including the legitimacy of the concepts 'agent,' 'emotion' and 'affect' in themselves. In line with modern cognitive science that suggests we should dispose of the image of intelligent agents as central representation processors, roboticist Rodney Brooks for instance argues for artificial cognition without representation and without agents. In one of his papers, he claims that 'in the very simple level intelligence…explicit representations and models of the world simply get in the way' (Brooks 1991b 139). Stating that central representation is the 'wrong unit of abstraction,' he is in favour of intelligent systems that interface directly to the world through perception and action. Apart from their apparent practicality, these questions become increasingly philosophical when he criticises the von Neumann model of computation and the categories of thought and reason that have dominated artificial intelligence as a field (Brooks 1991a 569). Dreyfus moves to a similar direction when he questions representational models of intelligence using Merleau-Ponty's *Phenomenology of Perception* (Dreyfus 2002).

Needless to say, the implications of such claims for the construction of affective intelligent machines are not nearly resolved, as the voices in favour of the representational approach are equally strong. Some scientists support the view that physical embodiment is neither necessary nor sufficient as a basis for affective or any AI research. Instead, they propose the synthesis of emotion through the traditional approach, utilising symbolic rule-based agent systems in software environments (Etzioni 1993). As expected in any productive domain of research, these are countered by scientists, especially from the field of robotics, who argue in favour of a joint approach, evoking the implementation of emotional-based control architectures with physically embodied agents. In her paper on emotion understanding in autonomous robotic research, Cañamero outlines a number of different approaches regarding models, applications and theory (Cañamero 448-9. The latter include a wide range of emotion-based systems, from rule-based ones to 'emergent emotion' approaches influenced by evolutionary artificial life models and biologically-inspired emotion architectures (Braitenberg qtd. in Cañamero 448). Acknowledging a complex fit between an agent and its environment interactionist AI and robotics attempt to circumvent the Cartesian split precisely by assuming that material embeddedness is crucial for all let alone emotive/ affective intelligence (Maes 136). Obviously, what is at issue is the tortuous relationship between emotion, affect, mind, and embodiment. Having reached the 'beyond the scope' point and limitations of *this* paper, however, I would like to return to the productive problematic raised by the convergence between these areas of technological and philosophical knowledge.

4. CONCLUSION: PROBLEMATICS AND QUESTIONS

Often said to be 'last and impassable frontier of computationalist theories of mind,' emotion, affect and the necessity or not of their embodiment complicate the picture both in the humanities and hard sciences (de Sousa 70). Despite the common interest in the expressive body, the apparent concern that arises here is the difference in the use of conceptual tools. Thus, whereas for a certain number of scientists emotion is a means to enhance human-like robotic or artificial subjectivity and personality, for philosophers affect seems to dismantle subjectivity be it human or robotic. The paradox of the situation becomes even more complicated if we take into account the endo-disciplinary scientific and philosophical differentiations and the fact that the very concept of 'human/ agent' itself is under serious questioning by philosophers. At a time that robots tend to become humanoid and humans mechanoid through prostheses and their splicing with software environments, ethics have become more tentative than ever. In her account of the passage from the classical conception of human to the cybernetics-informed posthuman, Hayles draws attention to such issues; against a reading of the rising technologies as symptoms and harbingers of nihilism, she argues that the emergent discourses of distributed cognition may actually prove very helpful in shaking the solid ontological and teleological foundation of the liberal humanist subject (Hayles 281-7). Thus, from a cultural theorist point of view, the question would be whether it is possible to avoid the cliché, dangerously humanist discourses of anthropomorphism that affective systems could possibly bring along with them. If the discourse of 'emotion,' and 'autonomy' propagates the good old grand narratives, will the affective turn in robotics and artificial intelligence help us reconsider our sense of aliveness and corporeality in ways that will avoid falling back to social constructionism, essentialism, and neo-liberal relativism?

Given the complexity of the situation and the multiply-informed interaction among so many disciplines, such vexed questions should be approached with caution for catastrophist morality is even more dangerous than the fantasised or real harms of technology. After all, '[t]echnology is not just the expression of the desire for mastery, but also the object of desire, curiosity and affective involvement (Braidotti 2002 215). Yet I would also like to suggest that they cannot be considered apart from the centrality of affective systems in the late capitalist culture either. In many respects, and far from being just another humanist fantasy projected on metal, the recent advances in evolutionary robotics and computer science are partly a response to the demands of an aggressive capitalism that profits on the exchange of emotions/affects at a deeper level. Among other theorists, in his analysis of the political economy of post-industrial advanced capitalism, Massumi points to the commodification and management of anxieties, affective states and ambient fears (Massumi 187). After the transition from the personal computer era to the ubiquitous computing era, the 'social robot' seems to be the latest taste in the global market for the so-called 'personal robot.' This trend is a curiously apt example of the schizophrenic double-bind of capitalism that Deleuze and Guattari identify within contemporary culture (Deleuze & Guattari 2004a 242-260). Within this 'seamless' consumerist context that the paradox of 'user-controlled' robotic 'autonomy' becomes even more obvious: to meet the perceived desires of the market, robots must be autonomous/self-organising and emotionally limited/user-controlled at the same time. As such, they are part of a post-industrial complex that reasserts individualism and personalised commodities as the unquestionable standard while feeding off the breaching of individual sanctity by pushing commercial profit-making to the innermost boundaries of subjectivity (Braidotti 2006 11). In the end, it is not only a matter of whether computers and robots will be capable of bodily affective expression but of how users invest them emotionally, as well. If the possibility of constructing an emotionally intelligent (or affective?) agent that will throw up metaphysical punchlines is an elusive goal for many years to come, these areas are of crucial importance and certainly preclude any attempts to think technology outside the cultural and philosophical terrain and vice versa. As Bishop puts it, '[i]t is clear that the purpose of…*computations* is contingent on their *social use*. In Heideggerian terms, computing machinery doesn't exist in the world until it is put to some use' (Bishop 7). Reconsidering the symbiotic relation between the human and the technological is therefore vital and an ethics of mutual interrelation can be best achieved through the interfacing of technology and philosophy.

REFERENCES

Bishop, Mark. 'Why Computers Can't Feel Pain.' *International Conference on Computers & Philosophy*. Laval, France, 2006. 1-8.

Braidotti, Rosi. *Metamorphoses. Towards a Materialist Theory of Becoming*. Cambridge: Polity, 2002.

________*Transpositions: On Nomadic Ethics*. Cambridge: Polity, 2006.

Breazeal, Cynthia. 'Emotion and Sociable Humanoid Robots.' *Human-Computer Studies* 59 (2003): 119–155.

Brooks, Rodney. 'Intelligence without Reason.' *Proceedings of 12th International Joint Conference on Artificial Intelligence (IJCAI), Computers and Thought*. Ed. John Myopoulos Sidney: Morgan Kaufmann, 1991a. 569-95.

________'Intelligence without Representation.' *Artificial Intelligence* 47 (1991b): 139-59.

Cañamero, Lola. 'Emotion Understanding from the Perspective of Autonomous Robots Research.' *Neural Networks* 18.4 (2005): 445-455.

Deleuze, Gilles. *Difference and Repetition*. Trans. Paul Patton. London & New York: Continuum, 2004.

Deleuze, Gilles & Guattari, Felix. *Anti-Oedipus: Capitalism and Schizophrenia* Trans. Robert Hurley, Mark Seem & Helen R. Lane. London & New York: Continuum, 2004a.

________*A Thousand Plateaus: Capitalism and Schizophrenia*. Trans. Brian Massumi. London & New York: Continuum, 2004b.

Dennett, Daniel. 'Why You Can't Make a Computer That Feels Pain.' *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge: MIT Press, 1981. 190-229.

________*Consciousness Explained*. Boston: Little, Brown, 1991.

Derrida, Jacques. *Speech and Phenomena and Other Essays on Husserl's Theory of Signs*. Trans. David Allison. Evanston: Northwestern University Press, 1973.

________*Of Grammatology*. Trans. Gayatri Chakravorty Spivak. Baltimore: Johns Hopkins University Press, 1976.

de Sousa, Ronald. *The Rationality of Emotion*. Cambridge: MIT Press, 1987.

Dreyfus, Hubert L. 'Intelligence without Representation. Merleau-Ponty's Critique of Mental Representation: The Relevance of Phenomenology to Scientific Explanation.' *Phenomenology and the Cognitive Sciences* 1.4 (2002): 367-83.

Duffy, Brian. 'Anthropomorphism and the Social Robot.' *Robotics and Autonomous Systems* 42 (2003): 177–190.

Etzioni, Oren. ' Intelligence without Robots: A Reply to Brooks.' *AI Magazine* 14.4 (1993): 7-13.

Foucault, Michel. *The Archaeology of Knowledge*. London and New York: Routledge, 1989.

Hayles, N. Katherine. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago: UCP, 1999.

Huyssen, Andreas. 'The Vamp and the Machine: Fritz Lang's Metropolis.' After the Great Divide. *Modernism, Mass Culture and Postmodernism*. Bloomington and Indianapolis: Indiana Univeristy Press, 1986. 65-81.

Khulood, Abu Maria & Raed, Abu Zitar. 'Emotional Agents: A Modeling and an Application.' *Information and Software Technology* 49.7 (2007): 695-716.

Maes, Pattie. 'Modeling Adaptive Autonomous Agents.' *Artificial Life* 1.1-2 (1994): 135 – 162.

Massumi, Brian. *A User's Guide to Capitalism and Schizophrenia*. Boston: MIT Press, 1992.

Műller, Vincent. 'Is There a Future for AI without Representation?' Minds and Machines 17.1 (2007): 101-115.

Picard, Rosalind. 'What Does It Mean for a Computer to Have Emotions?' *Emotions in Humans and Artifacts*. Eds. Robert Trappl, Paolo Petta and Sabine Payr. Cambridge, Massachusetts: The MIT Press, 2003. 213-235.

Terada, Rei. *Feeiling in Theory. Emotion after the 'Death of the Subject.'* Chambridge, Massachusetts, and London: Harvard University Press, 2001.

Varela, Francisco. *Ethical Know-How: Action, Wisdom, and Cognition*. Stanford: Stanford University Press, 1999.

Velásquez, Juan. 'When Robots Weep: Emotional Memories and Decision-Making.' *Proceedings of the Fifteenth National/ Tenth Conference on Artificial Intelligence/ Innovative Applications of Artificial Intelligence*, 1998. 70-75.