What does touch tell us about emotions in touchscreen-based gameplay?

YUAN GAO, University College London  
NADIA BIANCHI-BERTHOUZE, University College London  
HONGYING MENG, Brunel University

The increasing number of people playing games on touch-screen mobile phones raises the question of whether touch behaviours reflect players' emotional states. This prospect would not only be a valuable evaluation indicator for game designers, but also for real-time personalization of the game experience. Psychology studies on acted touch behaviour show the existence of discriminative affective profiles. In this paper, finger-stroke features during gameplay on an iPod were extracted and their discriminative power analysed. Machine learning algorithms were used to build systems for automatically discriminating between four emotional states (Excited, Relaxed, Frustrated, Bored), two levels of arousal and two levels of valence. Accuracy reached between 69% and 77% for the four emotional states, and higher results (~89%) were obtained for discriminating between two levels of arousal and two levels of valence. We conclude by discussing the factors relevant to the generalization of the results to applications other than games.

Categories and Subject Descriptors: H5.2 [Information interfaces and presentation]: Miscellaneous; Input devices and strategies.


Additional Key Words and Phrases: Automatic emotion recognition, touch-based computer games, touch behaviour, affective touch.

I. INTRODUCTION

Mobile games are becoming very popular these days. According to NilsenWire’s blog [2011], games are the most popular mobile app category at 64%; ahead of weather apps, social networking apps, maps and search apps. Among all types of smart phones, the iPhone is the most successful gaming device. The statistics on the blog report that iPhone users spend almost twice as much time playing games on the phone than the average smart phone gamer, with 14.7 hours per month.

In parallel with this increasing use of touch-based devices to play games, there is also an increasing interest, within the game industry, in adding emotion recognition capabilities to games (e.g. see Microsoft XBOX 360 Milo project and Fable Journey).
To appear on TOCHI, Nov. 2012

To increase the engagement experience of the player. This capability could be used either to evaluate new games or to create games that make use of the player’s affective states to create a personalized, engaging game experience [Bianchi-Berthouze 2012; Yannakakis and Hallam 2007]. It could, for example, be used to adapt the difficulty level of the game to maintain the right balance between challenge and player’s skills [Csikszentmihalyi 1990], or to decide the evolution of the gameplay at run-time. An example of the latter is provided by [Plass-Oude Bos et al. 2010], where the player’s character shape and its powers depend on the player’s stress level. When the player’s stress level is high, the player’s character assumes the shape of an aggressive bear with combat capabilities, whereas it returns to being an elf with cast-spell skills when the stress level is low. The stress level is measured through a brain-computer interface that captures the player’s brain alpha waves.

Automatic emotion recognition is not only an important function for the game industry but for many other areas. In the entertainment industry as a whole, for example, we are seeing the emergence of technology that is able to retrieve media automatically (e.g. music, images) on the basis of the mood a person is in or wants to experience [Liu and Reiner 2008; Lu et al. 2006; Bianchi-Berthouze and Lisetti 2002; Thrasher et al. 2011]. In the communications industry (e.g. SMS, twitter), there is interest in automatically enriching the text message with the emotion of the sender. The work presented in [Sundström et al. 2007] allows the sender to add emotions to an SMS message through different shaking patterns of the mobile phone. The new smart phone promised by SAMSUNG4 should be able to detect the emotions of the sender by capturing the sender’s behaviour during tweeting (e.g. number of typing errors, speed of typing, use of special symbols, and amount of shaking of the device). Furthermore, with the growing understanding of the strong interaction between cognitive and affective processes [Damasio 1984; Clore and Palmer 2009], there is a need to create technology with affective capability that can better support its users. Palm-held electronic diaries are, for example, used by patients with chronic illnesses to monitor and reflect on their emotional states [Isaacs et al. 2012]. In the field of e-learning, affective capabilities are investigated to help the students to regulate their emotional states to facilitate cognitive processing and foster motivations [D’Mello and Graesser 2011]. As most of these applications are migrating from desktops to touch-based portable devices, it becomes critical to investigate how emotions can be automatically tracked in this context.

Biosensors, facial- and vocal-expression recognition (e.g. [Calvo and D’Mello 2010]), body movement/posture (e.g. [Kleinsmith and Bianchi-Berthouze in press]) and EEG (e.g. [Nijholt et al. 2009]) have been explored as modalities to build affective-aware interactive technology. However, these affective modalities can be intrusive and cumbersome in a mobile situation. Biosensors have the problem of being very sensitive to environmental conditions, whereas facial-expression recognition is computationally very costly and very dependent on illumination conditions. This raises the question of investigating touch behaviour as an affective communication modality. The monitoring of this modality would be less intrusive and at the same time less computationally expensive. Recent studies in psychology literature [Hertenstein et al. 2009] provide support for this line of research, as they have shown that touch behaviour may convey not only the valence of an emotion but also the type of emotion (e.g. happy vs. upset). Unfortunately, these studies have been carried out

---

mainly in a social context (person-person communication) and only through acted scenarios. This paper aims to explore the possibility of using this modality to capture a player’s emotional state in a naturalistic setting of touch-based computer games. The findings from this study could be extended to other application areas where touch-based devices are used.

This paper is organized as follows. First, we review the literature on touch behaviour as a modality to convey and recognize emotions; as well as the literature on related modalities, such as keystroke typing and body movement. Then, we present our protocol to collect and label the data for building and testing our emotion recognition system. We then analyse the extracted touch-behaviour features and discuss their relationship with the emotional state of the players. Automatic recognition models are then built and tested on a person-independent context. We conclude by discussing the results and the possibility and challenge of using automatic affective touch recognition capabilities in other areas such as entertainment, education and clinical contexts.

2. BACKGROUND

2.1 Emotion recognition by touch behaviour in social science

An underexplored modality for automatic emotion recognition is touch. This is possibly due to the fact that, even if of great social importance, this modality has received very little attention, even in the field of affective science [Hertenstein et al. 2009]. Initial studies on touch behaviour as an affective modality argued that the role of touch was mainly to communicate the valence of an emotion (i.e. positive vs. negative emotions) and its intensity [Jones and Yarbrough 1985; Knapp and Hall, 1997]. More recently, two consecutive studies by [Hertenstein et al. 2006; 2009] have instead shown that touch communicates much more about emotions in interpersonal communication. In their first study, [Hertenstein et al. 2006] showed that participants could discriminate between different emotions when they were touched by another person (a stranger) on their forearm without seeing the hand or the person. The recognition of anger, fear, disgust, love, gratitude and sympathy reached performances between 48% and 83%. The subsequent study [Hertenstein et al. 2009] was carried out in a more ecological context but still with acted touch expressions. The results confirmed the findings of the previous study and showed that happiness and sadness could also be decoded with recognition rates higher than chance level. The recognition performances reached with touch alone were indeed comparable with the ones obtained with other affective modalities [Elfenbein and Ambady 2002].

What then are the features that facilitate the discrimination of the different types of affective touch? In their recent study, Hertenstein et al. [2009] investigated 23 different types of tactile behaviour (e.g. stroking, squeezing, patting). They found that the tactile behaviour alone was not sufficient to discriminate between different affective messages, i.e. the same type of tactile behaviour was shared by different emotions - for example, stroking was observed when communicating sadness but also when communicating love or sympathy. From a more detailed analysis of touch behaviour, they found that two more tactile behaviour qualities were necessary for an accurate discrimination between emotions: duration of the tactile behaviour and intensity (i.e. “the amount of skin deformation created by the touch on the receiving person”).

Whereas these studies clearly promote touch behaviour as a fine-grained affective modality, they have been carried out in acted situations, and it is not clear how these findings would vary in a naturalistic context. Furthermore, these studies relate to
inter-personal communication and it is hence necessary to understand if and how they translate to a context of interaction with a touch-based device and in a gameplay situation.

2.2 Interpersonal touch as an affective modality in HCI

Studies in neuroscience have shown that our skin has special receptors dedicated to detecting pleasurable touch. In [Essick et al. 2010], the authors describe the mechanisms underlying the skin sensors in charge of detecting the pleasurable experience of touch. Their extensive experiments show that these sensors respond particularly to gentle dynamic touch. It has also been shown that hedonic touch can promote hormonal responses that can facilitate bonding [Uvanas-Moberg et al. 2005].

Given these properties, various studies have been carried out to explore if touch can also enhance bonding and engagement with or through technology. This modality has, for example, been explored in Human-Robot interaction to facilitate bonds. Robots have been integrated with touch sensors over their body to capture people’s touch behaviour and provide affective displays in response. Examples of such robots are Shibata’s seal Paro [Shibata et al. 2001] and the Huggable [Stiehl et al. 2005]. In these robots, the mapping between the touch qualities (e.g. intensity) are predefined (i.e. not automatically learnt) but the responses humans receive to their touch provides a sense of robot’s agency and desire of communication.

Touch has also been explored through the use of haptic devices to facilitate distance communication [Noma and Miyasato 1997; Oakley et al. 2000; Strong and Gaver, 1996]. Most of these works have shown, through informal testing, that virtual touch received positive reactions from the participants (for a review see [Haans and IJsselsteijn 2006]). In [Chang et al. 2002], the authors designed a vibrotactile communication device where the two interlocutors could enhance their verbal message through vibrotactile signals. It was mainly tested to see if and how vibrotactile signals would facilitate turn-taking cues, emphasis of messages and the sense of co-presence. The preliminary results show that this information could enhance communication and could also provide information in case the verbal communication was limited. Following a similar line of investigation, the study presented in [Ho et al. 2000] explores the use of haptic feedback as a possible means to facilitate collaborative tasks in a shared virtual environment. The results show that participants not only produced a higher performance when touch feedback was provided but also reported an increased sense of togetherness. Another study is presented in [Bonanni et al. 2006], where a haptic scarf is used to send affectionate and nurturing touches to the wearer. Touch behaviour is captured and recorded by the scarf. The wearer can then play back the touches when apart from the person that recorded them. The recognition of the meaning of the touch behaviour is left to the receiver’s interpretation of it.

A study more in line with ours is reported in [Bailenson et al. 2007]. The authors investigated if a two-degree of freedom force feedback joystick could be used to communicate emotions. Participants were asked to communicate, through the use of the joystick, the emotion of anger, disgust, fear, interest, sadness and surprise to another person (called here receivers). For each joystick movement, measures of distance, acceleration, jerkiness and direction were computed and used to identify possible discriminative affective profiles. ANOVA test results showed that distance, speed and acceleration were much higher in joy and anger than for other emotions and showed the smallest values when expressing sadness. The direction measures were also quite informative, with fear showing smaller use of the major axis and
sadness showing smaller use of the minor axis. The measures were used to build seven classification models (one for each emotion) by using machine learning algorithms. The results were compared with the human receivers’ discrimination performances. The receivers were hence asked to classify the type of emotion that had been communicated through the haptic devices. The results show that the receivers recognized the correct emotion $33\%$ of the time (chance level $= 14\%$). The computational model built with a machine learning algorithm obtained similar recognition performances ($36\%$). The authors conclude that the emotions could be communicated well above chance level and that higher performances could be obtained with a more complex device.

These results are quite encouraging as they indicate the possibility of building a machine with the ability to recognize the affective content of touch. However, as in the previous section, these messages were acted and hence strongly stereotyped and forcefully communicated. As the authors [Bailenson et al. 2007] of this study acknowledge, as well as other researchers (e.g. [Wallbott and Scherer 1986; Nass and Brave 2005; Kleinsmith and Bianchi-Berthouze in press]) in the field of emotion communication, natural expressions differ from acted ones. Unfortunately, given the difficulties of collecting large naturalistic datasets, most of the work on automatic emotion recognition is still carried out on acted datasets [Sebe et al. 2004; Bänziger and Scherer 2007; Castellano et al. 2010; Kleinsmith and Bianchi-Berthouze in press]. Furthermore, whilst the studies on touch, discussed above, do show the importance of this communication modality, they still focus on interpersonal touch and cannot be directly translated into game context where touch may not always be used to purposely convey emotion to someone else. Before presenting our study, we first review works on automatic emotion recognition on two modalities that may share characteristics with touch behaviour: keystroke tapping and body movement.

### 2.3 Automatic emotion recognition by keystroke

Initial studies on the existence of finger-pressure profiles reflecting emotion were reported in [Clynes 1973]. In his work on sentography, Clyne investigated and showed the existence of vertical and horizontal finger-pressure profiles for different emotions. The data were collected in a controlled situation with no finger or arm movement (the finger is on a finger rest). Following this direction of investigation, a few studies in human-computer interaction explored the possibility of automatically detecting emotions from typing behaviour and mouse clicking.

In the study reported in [Khanna and Sasikumar 2010], the authors found that keyboard typing behaviour is affected by the user’s emotional state. People reported not only that their emotional states affected the frequency of selection of certain keys (e.g. backspace) but that in a positive mood their typing speed tended to increase. They carried out a further experiment to quantitatively characterize such behaviour. People were asked to type in either a positive or a negative mood. The results partially supported such self-reporting, showing an increase in speed in positive states and a decrease in speed in negative states. However, the authors state that the patterns were not as clear and this could be due to the fact that they have used broader emotional classes: positive vs. negative states. Each class includes a large variety of emotions that may indeed have different keystroke behaviour (e.g. negative states include anger as well as sadness). Given the large variability they identified, they developed a system that discriminates only between positive and neutral states or between negative and neutral states. They obtained a maximum performance of $87.7\%$. In a recent paper [Lv et al. 2008.], the authors further these investigations
using a novel pressure sensor keyboard. The keyboard produces a pressure sequence when the user enters characters. Six different emotions (neutral, anger, fear, happiness, sadness and surprise) were elicited in the participants by getting them to immerse themselves in a set of predefined stories. They were then asked to type a set of words not specified in the paper. Using three different types of features (global pressure sequences, dynamic time warping and keystroke dynamics) and using fusion techniques, they reached a 93.4% recognition performance over the six emotions. Unfortunately, no confusion matrices or discussion of emotion-keystroke pressure patterns were provided.

Matsuda et al. [2010] investigated the possibility of using finger behaviour to recognize automatically the emotional state of deaf people when communicating through a finger Braille device. In this study, the authors use duration and acceleration of finger dotting to discriminate between neutral, joy, sadness and anger states. The participants were asked to input two different sentences multiple times by expressing, each time, one of the four emotions. By using ANOVA, their results show that the duration of the dotting behaviour in the joy condition was significantly shorter than in the other conditions, whilst the dotting behaviour in the sadness condition was significantly longer. The finger load was significantly stronger in the anger condition. These results differ from the results on touch behaviour presented in [Hertenstein et al. 2009], where joy is longer than anger. This could indicate that different types of touch behaviour (e.g. tapping vs. stroking) may be characterized by different patterns. However, also in [Hertenstein et al. 2009], touch behaviour in sadness is longer than joy but of similar duration than in other positive states. Both studies show an increase in pressure for the anger state. Using Discriminant Analysis, the automatic discrimination model reached an average of 55% on absolute data and 82.5% with standardized data (taking into account daily variations in dotting strength). Even if the collected expressions were acted and hence prone to prototypical emotional behaviour, these results are quite promising.

These studies confirm the importance of duration and pressure qualities to discriminate between emotional states. They also identify speed as an important discriminating modality. However, they investigate one type of touching behaviour (tapping). In our study, we aim to investigate finger stroking touch behaviour (as discussed in section 3) as it is an important type of touch behaviour with touch-based game devices. To complete our review on automatic emotion recognition, we draw on the relationship between finger movement and body movement, an area much more explored than the one of touch. Touch and movement are also often coupled as part of the proprioceptive system.

2.4 Automatic emotion recognition by body expression

The literature shows that body movement is an important modality to decode the affective states of a person (for a review see [Kleinsmith and Berthouze in press]). Automatic recognition systems based on this modality have shown comparable performances to systems for emotion recognition in facial expressions. Most work has been done in acted situations by exploiting both body configuration and dynamic information. Gross level descriptions of body movement have been initially investigated by Camurri in dance context [Camurri et al. 2003a]. Inspired by Laban’s studies [von Laban 1971], they propose a set of gross features capturing the configuration of the body (shape) and its dynamics (i.e. shape and effort in Laban’s terminology). Effort is described in terms of directionality, flow (bound vs. free), time (quick vs. sustained) and weight of the movement. Their emotion recognition system
from dancers’ body movement [Camurri et al. 2003b; Camurri et al. 2004], aimed at
discriminating between four discrete emotion categories (anger, fear, grief and joy),
reached recognition performances (average of 40%) well above chance level. It should
be noted that the results obtained by human observers on the same data set are also
far from optimal (56%), even if higher than the ones obtained with the system. Their
analysis of body-movement features showed that joy movements were characterized
by high fluidity as opposed to grief movements that demonstrated more jerkiness.
Using a more complex architecture of the emotion recognition system, [Kapur et al.
2005] obtained much higher performances, with recognition rates ranging from 62% to 93%.

Another study presented in [Bernhardt and Robinson 2007] highlighted the
importance of idiosyncrasy in the modelling of emotion recognition from movement.
Their system was built to discriminate between three discrete emotional states (i.e.
angry, happy, sad) expressed by acted knocking arm movements. Using machine
learning classifiers, the correct recognition rate of affective states reached 50%. However, by subtracting individual idiosyncrasies from the description of the
movement, the performances increased to 81%. The recognition performances were
comparable to human observers’ performances (varying between 59% and 71%) for
the same set of stimuli, as discussed in [Pollick et al. 2001]. Similarly, other studies
(e.g. [Castellano et al. 2008]) explored further the discriminative power of dynamic
and shape body movement characteristics through the use of automatic analysis,
demonstrating its possible interest in human-machine interaction.

Studies on semi-spontaneous multi-modal communication have also shown that
body gestures strongly improve the automatic recognition of the emotional content of
a message and they are often more informative than facial expressions [Gune and
Piccardi 2009; Kapoor et al. 2007]. This has brought about a shift of focus from acted
expressions to non-acted contexts, where the expressions are less stereotypical and
possibly more subtle [Kapoor et al. 2007]. The shift has been dictated by the
emergence of full-body technology and the need to use such recognition systems in
real-life human-machine interaction. Of particular relevance to our paper, the
emergence of full body computer games has led to increased interest in developing
software that not only tracks the players’ body movement [Bleiweiss et al. 2010;
Roccetti et al. 2012] but also recognizes their affective state. The idea is to design
games that can provide a more engaging embodied experience [Pasch et al. 2009;
The results are very promising and similar to the ones obtained in acted context. The
following two research works are examples of affective body expression recognition
that has been developed in game context.

The work presented in [Kleinsmith et al. 2011] aimed to detect the emotional state
of the players when they were looking at the replay of the game just after having
scored or missed a point (i.e. between game points). The game used was a Wii-
Nintendo sport game. The results showed that, by exploiting low-level body
configuration features (i.e. angle between body segments), the systems reached, on
average, correct recognition rates of 60% for four affective states (concentrating,
defeated, frustrated and triumphant). Their results were comparable with the human
observers’ level of agreement (i.e. 67% recognition rate) reached for the same set of
stimuli. Their work also showed very high performances for the recognition of
affective dimensions’ levels (arousal=energy level, valence=level of positivity of the
emotional state).

The expressions investigated in this study were quite static (i.e. time between
gameplay) and hence body configuration features alone produced high results [Kleinsmith and Bianchi-Berthouze 2011]. A study more relevant to our work is presented by [Savva et al. 2012]. They proposed a system to recognize continuously the emotional states of players while playing Wii tennis games. Dynamic features describing the movement of the arm, the orientation of the body and the quantity of motion were used for this purpose. The best results were obtained by using angular velocity, angular speed and amount of movement. A recurrent neural network, that takes into account the history of the classification, was used to classify, frame by frame, the collected motion-capture data. Overall the system was able to classify correctly 64% of the high intensity negative expressions, 67% of the low intensity negative expressions, 58% of the happiness expressions and only 36% of the concentration expressions. An analysis of the results highlighted the large diversity of the players’ playing styles also pointed out by other studies [Pasch et al. 2009]. Some of the participants tended to play the game using only their hand/wrist in comparison with the other group that used their arm and shoulder as well.

Building on the promising results obtained on automatic recognition of naturalistic whole body affective expressions and of keystroke affective behaviour, in our study, we investigate the possibility of using finger-stroke behaviour when interacting with touch-based devices to detect the emotional state of the user. We investigate this question in the context of touch-based computer games because of their spread and also because they do elicit quite strong emotions in a short period of time. In the following sections, we present the data collection, modelling process and testing. We then conclude, discussing the results and the possibility of using this modality to create less invasive affective-aware technology that could be used on the move.

### 3. DATA COLLECTION

This section presents the material and the procedure used to capture and analyse players’ finger-stroke behaviour with respect to their emotional state. The most discriminative features are then used in section 4 to build person-independent emotion recognition models.

#### 3.1 Game

The iPhone game Fruit Ninja\(^6\) was selected for this study. Fruit Ninja is an action game that requires the player to squish, slash and splatter fruits. To play, players swipe their fingers across the screen to slash and splatter fruit, simulating a ninja warrior. Fruit Ninja has been a phenomenon in the app store with large sales, and has inspired the creation of many similar drawing-based games. As the Fruit Ninja is not an open source app, an open source variation of it (called Samurai Fruit), developed by Ansca Mobile, was used instead.\(^6\) It has been developed with the Corona SDK. In order to elicit different kinds of emotions from players and to be able to capture players’ touch behaviour, a few modifications to the Ansca’s game were implemented.

Our new version of the Samurai Fruit contains 20 levels in total. Each level lasts 30 seconds and has a target point to reach. The goal is to reach this target point by slashing enough fruits. Each fruit is worth 1 point. The players have to pay attention and avoid the bombs; since slashing a bomb makes the player lose 5 points. There are no constraints on how to slash the fruits. Players can slash the fruits one by one or

\(^5\)http://www.fruitninja.com/

wait for multiple fruits to be lined up and slash them at once. The degree of difficulty of the game increases as the level increases. The difficulty is based on the speed of the game (shooting speed and angular speed of the objects) and on the ratio between the number of fruits and the number of bombs. If the player reaches the target point within the time limit, they win the level, otherwise they lose it. After having finished all the 20 levels, a final score is given according to the number of levels they won. Figure 1 shows the screenshots of New Samurai Fruit.

Fig. 1. Screenshots of the adapted Samurai Fruit game: slashing of a single fruit (left) and of multiple fruits (right). @ANSCA Inc.

Fig. 2. Coordinates of the points for two different strokes. Blue and red dots represent the points captured along the two strokes. The point size represents the finger contact area (at that point) after being normalized by using the player’s finger pressure baseline. Larger dots correspond to stronger finger pressure. The resolution of the display used in this study is 480×320 pixels with 163 ppi. @ANSCA Inc.

The developed New Samurai Fruit game software captures and stores players’ finger-stroke behaviour during gameplay: the coordinates of each point of a stroke, the contact area of the finger at each point and the time duration of a stroke, i.e. the time occurring from the moment the finger touches the display to the moment the finger is lifted. This information is directly gathered by using tracking functions of the Corona SDK. The contact area is used here as a measure of the pressure exerted by the participants, as the device does not allow pressure to be measured directly. Hereafter, for simplicity, we call this feature pressure. In order to take into consideration physical differences between individuals, i.e. the dimension of the tip of a finger, a baseline pressure area is computed before the game starts. Each participant is asked to perform a stroke. The average of the contact area of each point of this stroke is used as a pressure baseline for that participant. The pressure values collected during the game are hence normalized by subtracting from them the participant’s baseline value. Figure 2 shows an example of two different strokes after
normalization with the player’s finger pressure baseline.

3.2 Data collection and labelling procedure

Our experiment set-up included the following items: a laptop with a camera to videotape the whole experiment process, an iPod touch (display: 480×320 pixels with 163 ppi) and a self-assessment questionnaire used to capture the player’s emotional state. The self-assessment questionnaire consisted of a list of emotion words. This list was developed through a pilot study with 7 participants. During this pilot study, participants were allowed to use any words to express their emotional state. The most common emotional words used were: Delighted, Excited, Annoyed, Frustrated, Satisfied, Relaxed, Concentration and Bored. The list was further refined. First of all, concentration was eliminated as it was considered a neutral state and also because, as shown in other studies [Kleinsmith et al. 2011] and from the result of our interview with the 7 participants, the term concentration was often used when participants were not sure of their emotional states. Secondly, the terms that belonged to the same quadrant of the valence-arousal (V-A) emotion space [Russell 1980] were grouped: Excited and Delighted, Annoyed and Frustrated, and Satisfied and Relaxed. The aim was to ensure that a large set of data was obtained for each emotion class necessary for the analysis and modelling process. The final set of words was: Excited, Relaxed, Frustrated and Bored.

The experiment procedure was as follows. We introduced and explained the experiment process, the iPod touch device and the game rules of New Samurai Fruit to participants and answered their questions. The participants were then asked to play the game and get familiar with it. After the training session, the participants were asked to play and try to complete 20 consecutive levels of the game. As a form of motivation, they were informed that the participant with the highest final score would have been rewarded with an iTunes gift card worth £25. The game was video recorded. In order to decide the ground truth (i.e. the emotional state associated with each level), participants were first asked to fill out a self-assessment questionnaire after every game level. To reduce the effect of the game results on the labelling, the Cued Recall Brief approach [Bentley et al. 2005] was used. At the end of the 20 levels, the player was shown a recorded video of the gameplay s/he had just carried out. While the video was playing s/he was encouraged to recall his/her thoughts and emotional states and if necessary relabel the levels.

![Emotion label frequency](image.png)

Fig. 3. Distribution of the emotion labels over game levels. Each bin of the histogram is divided into 4 sections. Each section of a bin represents five consecutive levels of increasing difficulty. The blue sector at the bottom of each bin represents the number of times an emotional state appeared in the first five levels of the game (levels 1-5); the second sector (red) from the bottom represents levels 6-10 and so on.
Fifteen participants (9 male and 6 female whose average age varied between 18 and 40 years old) were invited to take part in the experiments and play the 20-level game. One of the participants asked to play the 20-levels game a second time and these data were also recorded. This provided data samples for a total of 320 game-levels. During the relabelling phase, only 12% of the levels were relabelled. Only 300 game-levels data were finally used in the study to avoid biasing the model towards one participant. The data from the participant that played twice were selected in order to balance the data and avoid having one emotion under-represented. 8 Excited, 6 Relaxed, 3 Frustrated and 3 Bored were randomly selected from the 40 session levels of that participant. It should be noted that for this participant this resulted in 4 samples for the first 5 levels (L1-L5), 7 for the second five levels (L6-L10), 2 for the third five levels (L11-L15) and 7 for the final five levels (L16-L20). As the modelling presented in this study is level-independent, we thought that, given the availability of data, it was better to prioritize the number of emotions vs. the number of levels. This resulted in a dataset for the study formed by 90 Excited, 70 Relaxed, 70 Frustrated and 70 Bored samples. The number of excited labels is much higher than the rest as this emotion was the most frequent one. Figure 3 shows the distribution of the emotion labels over the 15 participants. Figure 4 shows the distribution of the emotion labels by players. A certain variability is present in the frequency of each emotion label. Most players have all 4 labels represented in the dataset. Only players 7, 8 and 9 miss one label in their dataset and player 13 misses two labels.

<table>
<thead>
<tr>
<th>Emotion distribution by players</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Excited</td>
</tr>
<tr>
<td>Frustrated</td>
</tr>
<tr>
<td>Relaxed</td>
</tr>
<tr>
<td>Bored</td>
</tr>
</tbody>
</table>

Fig.4. Distribution of the emotion labels over the 15 participants.

4. FEATURE ANALYSIS

4.1 Visual Inspection

As discussed earlier, the coordinates and the pressure area of each point of each stroke, as well as the time duration, were recorded within each game-level. For each game-level, sixteen features were extracted as listed in Table I: average, median, maximum and minimum values of the length, speed, directionality index (DI) and pressure area of the strokes within that game-level. The features were based on the literature on touch and body movement reported in section 2. In particular, the set of features selected can be related to the effort characteristic of the movement proposed.
To appear on TOCHI, Nov. 2012

by Laban [von Laban 1971]. In Table I, \( N \) denotes the total number of strokes in a recorded game-level, \( b_{ij} \) denotes the \( j \)th point on the \( i \)th stroke and \( x_{ij} \) and \( y_{ij} \) denote respectively its x-coordinate and its y-coordinate. \( n_i \) denotes the total number of points on the \( i \)th stroke and \( t_i \) denotes the time duration of the \( i \)th stroke.

At first, a visual inspection of these features was carried out to identify possible emotion-feature patterns. Figure 5 shows the boxplots of the average and median values of each type of feature. The boxplots show that the positive states (i.e. Excited and Relaxed states) tend to have longer strokes with Relaxed having overall the longest ones. The strokes for Frustration present the largest variation. Boredom has the shorter strokes. The pressure features show a clearer separation between Frustration and the other emotions, with Frustration showing overall higher values. Speed seems, instead, to separate the states along the arousal dimension, with Frustration and Excitement showing higher values than the other two states. The directness index (DI) is another possibly useful feature, showing that Excitement and Frustration have less direct strokes than the other two states with Boredom showing the most direct ones (i.e. lower values).

Table I. Stroke Features.

<table>
<thead>
<tr>
<th>Feature Name</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of strokes in a level</td>
<td>( N )</td>
</tr>
<tr>
<td>Average stroke length</td>
<td>( \overline{l} = \frac{\sum_{i=1}^{N} l_i}{N} ), where ( l_i = \sum_{j=1}^{n_i} \sqrt{(x_{ij+1} - x_{ij})^2 + (y_{ij+1} - y_{ij})^2} ) and ( n_i ) is the number of points for the ( i )th stroke.</td>
</tr>
<tr>
<td>Median stroke length</td>
<td>( l_{med} = \text{median} (l_i) ), for all ( i = 1, 2, \ldots, N )</td>
</tr>
<tr>
<td>Max stroke length</td>
<td>( l_{max} = \text{max} (l_i) ), for all ( i = 1, 2, \ldots, N )</td>
</tr>
<tr>
<td>Min stroke length</td>
<td>( l_{min} = \text{min} (l_i) ), for all ( i = 1, 2, \ldots, N )</td>
</tr>
<tr>
<td>Average stroke speed</td>
<td>( v = \frac{\sum_{i=1}^{N} v_i}{N} ), where ( v_i = \frac{1}{t_i} )</td>
</tr>
<tr>
<td>Median stroke speed</td>
<td>( v_{med} = \text{median} (v_i) ), for all ( i = 1, 2, \ldots, N )</td>
</tr>
<tr>
<td>Max stroke speed</td>
<td>( v_{max} = \text{max} (v_i) ), for all ( i = 1, 2, \ldots, N )</td>
</tr>
<tr>
<td>Min stroke speed</td>
<td>( v_{min} = \text{min} (v_i) ), for all ( i = 1, 2, \ldots, N )</td>
</tr>
<tr>
<td>Average stroke Directness Index (DI)</td>
<td>( d = \frac{\sum_{i=1}^{N} d_i}{N} ), where ( d_i = \sqrt{(x_{ij} - x_{im})^2 + (y_{ij} - y_{im})^2} )</td>
</tr>
<tr>
<td>Median stroke DI</td>
<td>( d_{med} = \text{median} (d_i) ), for all ( i = 1, 2, \ldots, N )</td>
</tr>
<tr>
<td>Max stroke DI</td>
<td>( d_{max} = \text{max} (d_i) ), for all ( i = 1, 2, \ldots, N )</td>
</tr>
<tr>
<td>Min stroke DI</td>
<td>( d_{min} = \text{min} (d_i) ), for all ( i = 1, 2, \ldots, N )</td>
</tr>
<tr>
<td>Average Pressure (i.e., average contact area)</td>
<td>( p = \frac{\sum_{i=1}^{N} p_i}{N} ), where ( p_i = \frac{\sum_{j=1}^{n_i} p_{ij}}{n_i} ), ( p_{ij} ): the size of the touching area of the ( j )th point on the ( i )th stroke and ( n_i ) is the number of points in the ( i )th stroke. The contact area for each individual point is normalized with respect to a person-based baseline pressure value as discussed in the text.</td>
</tr>
<tr>
<td>Median Pressure</td>
<td>( p_{med} = \text{median} (p_i) ), for all ( i = 1, 2, \ldots, N )</td>
</tr>
<tr>
<td>Max Pressure</td>
<td>( p_{max} = \text{max} (p_i) ), for all ( i = 1, 2, \ldots, N )</td>
</tr>
<tr>
<td>Min pressure</td>
<td>( p_{min} = \text{min} (p_i) ), for all ( i = 1, 2, \ldots, N )</td>
</tr>
</tbody>
</table>

In order to verify that these patterns were independent from the difficulty of the
To appear on TOCHI, Nov. 2012

levels, we look at the marginal means of these features. Figure 6 shows the estimated marginal means graphs for average length, pressure, speed and directness index over different levels of difficulty of the game. Each 5-consecutive levels are represented by one line. We can see that, for all the four features, the lines show a similar trend independently of the difficulty of the level. Only the stroke length values for the first 5 levels (L1-L5, i.e. the easiest levels) show a different trend, with Frustration reaching the highest values. It is possible that normalizing these data with respect to the difficulty of the level may bring better classification results. However, only the first 5 levels show a slightly different pattern that could be due to outliers.

Fig 5. Boxplots of the average and median values of length, pressure, speed and DI features.

The comparison of the observed emotion-touch behaviour with the one reported in section 2 is not so straightforward. Our results seem to support the results on touch behaviour reported in [Hertenstein et al. 2009] on the difference between Excitement and Frustration (respectively joy and anger in [Hertenstein et al. 2009]). In fact, the duration of the stroke was longer in joy states compared to anger. Also, higher pressure values were observed for anger in [Hertenstein et al. 2009]. Different results were observed for the low energy negative states i.e. sad in [Hertenstein et al. 2009] and Boredom in our study. Sadness in [Hertenstein et al. 2009] has produced the longest stroke, whereas in our study Boredom produced the shortest one. The differences show that the stroke does not only inform us about the valence of the
To appear on TOCHI, Nov. 2012

arousal of an emotion but also its type. Whereas Boredom may indicate the desire to disengage from the game, this may not be the case for a feeling of sadness in the social context.

Figure 6. Estimated marginal means by level bins: average length, pressure, speed and directness index.

More difficult is the comparison with the characteristics of the typing and dotting behaviour reported in [Matsuda et al. 2010] and [Khanna et al. 2010]. In [Matsuda et al. 2010], differently from our study, anger and sadness produced longer dotting duration than the other two states. In [Khanna et al. 2010], speed appeared to discriminate along the valence dimensions rather than the arousal one, as seems to be the case in our study. However, in [Matsuda et al. 2010], frustration produced higher pressure values, as in our study. This could mean that, in the case of typing or dotting, the duration of the behaviour is more strongly affected by pressure actions, bringing the time spent on the key/dot to be longer and hence slower. Unfortunately, a more detailed comparison of stroking versus typing is not possible, as in [Khanna et al. 2010] the states are separated only along the valence dimensions.
To appear on TOCHI, Nov. 2012

Table II. Eigenvalues and Wilk’s Lambda for the first 3 discriminant functions.

<table>
<thead>
<tr>
<th>Function</th>
<th>Eigenvalue</th>
<th>% of Variance</th>
<th>Cumulative %</th>
<th>Canonical Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.065</td>
<td>57.6</td>
<td>57.6</td>
<td>.823</td>
</tr>
<tr>
<td>2</td>
<td>1.605</td>
<td>34.2</td>
<td>91.8</td>
<td>.745</td>
</tr>
<tr>
<td>3</td>
<td>.290</td>
<td>8.2</td>
<td>100.0</td>
<td>.479</td>
</tr>
</tbody>
</table>

Test of Function(s) Wilk’s Lambda Chi-square df Sig.
1 through 3 .111 259.351 36 .000
2 through 3 .344 126.069 22 .000
3 .771 30.687 10 .001

4.2 Feature Extraction

After the visual inspection, aimed at investigating the relevance of the computed features, a further evaluation of their discriminative relevance was carried out. IBM SPSS version 20 was used for this purpose. Using the Bernoulli function of SPSS with a parameter set to 0.4, the 300 instances were randomly split into two subsets. 173 instances were assigned to the training set and 127 to the testing set. Given the similarity and high correlation between the median and average values for length, speed and pressure features (see Figure 5), their median values were not used in this analysis.

A Discriminant Analysis (DA) was carried out to identify the most relevant features for the discrimination between the four emotions categories. The results showed that only the Average-Stroke-DI feature was not significant (p-value = .589). This feature was hence removed from further analysis. All the other features obtained p-values <.000 but for Min-Stroke-DI which was still significant with p-values < .010).

Table III. Structure matrix for discriminant analysis for 4 emotions.

<table>
<thead>
<tr>
<th>Features</th>
<th>Function 1</th>
<th>Function 2</th>
<th>Function 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max Pressure</td>
<td>.856</td>
<td>-.221</td>
<td>.321</td>
</tr>
<tr>
<td>Average Pressure</td>
<td>.774</td>
<td>-.328</td>
<td>.322</td>
</tr>
<tr>
<td>Min Pressure</td>
<td>.658</td>
<td>-.275</td>
<td>.309</td>
</tr>
<tr>
<td>Max DI</td>
<td>.360</td>
<td>.575</td>
<td>.469</td>
</tr>
<tr>
<td>Max Speed</td>
<td>.505</td>
<td>.537</td>
<td>.361</td>
</tr>
<tr>
<td>Average Speed</td>
<td>.463</td>
<td>.529</td>
<td>.348</td>
</tr>
<tr>
<td>Min Speed</td>
<td>.364</td>
<td>.498</td>
<td>.319</td>
</tr>
<tr>
<td>Min DI</td>
<td>-.050</td>
<td>-.213</td>
<td>.026</td>
</tr>
<tr>
<td>Average Length</td>
<td>-.095</td>
<td>.047</td>
<td>.808</td>
</tr>
<tr>
<td>Min Length</td>
<td>-.244</td>
<td>-.073</td>
<td>.714</td>
</tr>
<tr>
<td>Median DI</td>
<td>.281</td>
<td>.522</td>
<td>.543</td>
</tr>
<tr>
<td>Max Length</td>
<td>.077</td>
<td>.144</td>
<td>.472</td>
</tr>
</tbody>
</table>

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions. Variables ordered by absolute size of correlation within function. **Bold**. Largest absolute correlation value between a variable and the discriminant functions.

Using the remaining features, the DA model was built. The results are reported in Tables II-IV. The results identified three discriminant functions whose Eigenvalues are reported in the top part of Table III. The first discriminant function explained 57.6% of the variance (canonical R²=.823) and the second explained 34.2% (canonical R²=.745). The third explained only 8.2% of the variance (canonical R²=.479). The
bottom part of Table II shows that, in combination, these three functions significantly differentiate between the four emotion groups ($\Lambda=.111$, $\chi^2(48) = 259.351$, $p=.000$).

Table III shows the correlations between the stroke features and the three discriminant functions. The table reveals that pressure features loaded highly on the first discriminant function. Speed and DI features loaded highly on the second discriminant function. Finally, the length features loaded highly on the third discriminant function. The discriminant function plot in Figure 7 shows that the first function discriminates between the strokes performed in a Frustrated state from the strokes performed under the other affective states. The second function, loading on speed and DI features, discriminates Excited states from low arousal states (Relaxed and Bored) and from Frustration. Finally, length features discriminate between Bored and Relaxed states.

Table IV shows the confusion matrix for the testing set. The results show that 76.4% of the testing set was correctly classified. The best results are obtained for the classification of Frustrated and Excited strokes. 86% of Frustrated strokes are correctly classified and the performances for Excited are just below 80%. The lowest performances (63.6%) are obtained for Relaxed samples. Most misclassifications occur between Relaxed and Bored (37.9%) and between Relaxed and Excited (23.9%).

![Fig. 7. Results for the Discriminant Analysis for 4 emotions. Each stroke is represented by a symbol in the 3-dimensional space. The type of symbol indicates the predicted class for that stroke. The centroid of each emotion group is represented by a square. The centroids illustrate how each 3 discriminant function (and hence the features that load on them) support the discrimination of the four emotion classes.](image)

A discriminant analysis was also performed to identify the most relevant features for arousal (low vs. high arousal state) and for valence (negative vs. positive valence states). The results are shown in Tables V and VI. Table V shows that the first discriminant function explains 100% of the variance for both arousal (canonical $R^2=.793$) and valence dimensions (canonical $R^2=.717$). It also shows that the
discriminant functions significantly differentiate between, respectively, two levels of arousal ($\Lambda=.372, \chi^2 (13) = 162.773, p=.000$) and two levels of valence ($\Lambda=.486, \chi^2 (13) = 118.846, p=.000$).

Table IV. Confusion matrix for the discriminant analysis.

<table>
<thead>
<tr>
<th>Testing Set (127 Sample data)</th>
<th>Count</th>
<th>Excited</th>
<th>Relaxed</th>
<th>Frustr.</th>
<th>Bored</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arousal</td>
<td>Count</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excited</td>
<td>27</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>Relaxed</td>
<td>4</td>
<td>21</td>
<td>1</td>
<td>7</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>Frustrated</td>
<td>2</td>
<td>0</td>
<td>26</td>
<td>2</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Bored</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>23</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Excited</td>
<td>79.4</td>
<td>11.8</td>
<td>5.9</td>
<td>2.9</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>Relaxed</td>
<td>12.1</td>
<td>63.6</td>
<td>3.0</td>
<td>21.2</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>Frustrated</td>
<td>6.7</td>
<td>0</td>
<td>86.7</td>
<td>8.7</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>Bored</td>
<td>0</td>
<td>16.7</td>
<td>6.7</td>
<td>76.7</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

76.4% of testing set cases were correctly classified.

Table V. Eigenvalues and Wilk’s Lambda for the first 3 discriminant analysis functions.

<table>
<thead>
<tr>
<th>Function</th>
<th>Eigenvalue</th>
<th>% of Variance</th>
<th>Cumulative %</th>
<th>Canonical Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arousal</td>
<td>1.653</td>
<td>100.0</td>
<td>100.0</td>
<td>.789</td>
</tr>
<tr>
<td>Valence</td>
<td>1.040</td>
<td>100.0</td>
<td>100.0</td>
<td>.715</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test of Function</th>
<th>Wilks' Lambda</th>
<th>Chi-square</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arousal</td>
<td>.377</td>
<td>160.963</td>
<td>12</td>
<td>.000</td>
</tr>
<tr>
<td>Valence</td>
<td>.489</td>
<td>117.926</td>
<td>12</td>
<td>.000</td>
</tr>
</tbody>
</table>

Table VI shows the importance of speed and DI features to separate between low and high arousal states. Pressure appears to have less discriminative power than speed and DI for discriminating along the arousal dimensions. Length does not show any relevant weight on this affective dimension. Differently from arousal, pressure and length are mainly used to separate strokes along the valence dimension. The Discriminant Model built for Arousal obtained 89% of correct classification on the testing set. The results for the Discriminant Model for Valence were slightly lower, with 83% of correct classification on the testing set.

5. EMOTION RECOGNITION SYSTEM

5.1 Modelling

The results of the Discriminant Analysis showed that finger-stroke behaviour allows for the discrimination of the four affective states we investigated, as well as for the discrimination between two levels of arousal and between two levels of valence. The next step in our analysis is to investigate if person-independent emotion recognition models could reach similar results. Creating a person-independent model is generally harder, as the data of the person on which the model is tested are not used to build the model. This testing is very important as it shows the generalization capability of the stroke features.

To this purpose, three modelling algorithms were selected: Discriminant Analysis (DA), Artificial Neural Network (ANN) with Back Propagation and Support Vector Machine (SVM) classifiers. The last two learning algorithms were selected for their
To appear on TOCHI, Nov. 2012

popularity and their great ability for generalization.

Table VI. Structure matrix for discriminant analysis for each affective dimension. The features that load higher on the discriminant functions are in bold.

<table>
<thead>
<tr>
<th>Affective dimension</th>
<th>Features</th>
<th>Function</th>
<th>Affective dimension</th>
<th>Features</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arousal (2 levels)</td>
<td>Max Speed</td>
<td>.766</td>
<td></td>
<td>Max Pressure</td>
<td>.568</td>
</tr>
<tr>
<td></td>
<td>Average Speed</td>
<td>.733</td>
<td></td>
<td>Average Pressure</td>
<td>.545</td>
</tr>
<tr>
<td></td>
<td>Max DI</td>
<td>.693</td>
<td></td>
<td>Min Pressure</td>
<td>.532</td>
</tr>
<tr>
<td></td>
<td>Min Speed</td>
<td>.639</td>
<td></td>
<td>Min Length</td>
<td>-.335</td>
</tr>
<tr>
<td></td>
<td>Median DI</td>
<td>.594</td>
<td></td>
<td>Average Length</td>
<td>-.328</td>
</tr>
<tr>
<td></td>
<td>Max Pressure</td>
<td>.316</td>
<td>Valence (2 levels)</td>
<td>Median DI</td>
<td>-.221</td>
</tr>
<tr>
<td></td>
<td>Average Pressure</td>
<td>.285</td>
<td></td>
<td>Max DI</td>
<td>-.173</td>
</tr>
<tr>
<td></td>
<td>Min Pressure</td>
<td>.211</td>
<td></td>
<td>Max Length</td>
<td>-.163</td>
</tr>
<tr>
<td></td>
<td>Min DI</td>
<td>.188</td>
<td></td>
<td>Min Speed</td>
<td>-.101</td>
</tr>
<tr>
<td></td>
<td>Max Length</td>
<td>.180</td>
<td></td>
<td>Min DI</td>
<td>.096</td>
</tr>
<tr>
<td></td>
<td>Min Length</td>
<td>.175</td>
<td></td>
<td>Average Speed</td>
<td>.056</td>
</tr>
<tr>
<td></td>
<td>Average Length</td>
<td>.006</td>
<td></td>
<td>Max Speed</td>
<td>.036</td>
</tr>
</tbody>
</table>

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant function. Variables ordered by absolute size of correlation within function. Only the higher correlations are reported.

A 3-layered feed-forward backpropagation ANN was used. The parameters used for the ANN are shown in Table VII. The selection of these parameters was based on [Wu et al. 2008]. The setting of the parameters for the ANN was the same for all the experiments. The inputs of ANN were the stroke features listed in Table III. Three models were built: one for arousal with two output nodes corresponding to low and high level states; one for valence with two output nodes corresponding to positive and negative states; and one for the discrimination of the 4 emotional states with hence 4 output nodes. The ANN classifiers were implemented in MATLAB using the Neural Network toolbox.

Table VII. ANN Parameters.

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of hidden layers</td>
<td>3</td>
</tr>
<tr>
<td>Number of input layer nodes</td>
<td>Equal to the number of features</td>
</tr>
<tr>
<td>Number of hidden layer nodes</td>
<td>Same as input layer</td>
</tr>
<tr>
<td>Number of output layer nodes</td>
<td>Equal to the number of emotion classes (i.e. either 4 or 2)</td>
</tr>
<tr>
<td>Transfer Function (TF) for input layer to 1st hidden layer</td>
<td>Tansig</td>
</tr>
<tr>
<td>TF for 1st to 2nd hidden layer</td>
<td>Tansig</td>
</tr>
<tr>
<td>TF for 2nd to 3rd hidden layer</td>
<td>Tansig</td>
</tr>
<tr>
<td>TF for 3rd hidden layer to output layer</td>
<td>Purlin</td>
</tr>
<tr>
<td>Back-propagation transfer function (BTF)</td>
<td>Trainlm (Levenberg-Marquardt backpropagation)</td>
</tr>
<tr>
<td>Learning Rate</td>
<td>0.03</td>
</tr>
<tr>
<td>Training Epoch</td>
<td>1000</td>
</tr>
<tr>
<td>Momentum</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Linear SVM is a very powerful and popular classifier and has achieved excellent performance for pattern recognition in many applications. It was used with both linear (default configuration) and non-linear kernel. The non-linear kernel function Radial Basis Function (RBF) [WU, et al. 2008] was used:
To appear on TOCHI, Nov. 2012

\[ K(x_i, x_j) = e^{-\gamma|x_i - x_j|^2} \]  

where \( \gamma \) is the parameter of the RBF kernel function. LIBSVM toolbox in MATLAB was used in the implementation. The parameters of the kernel SVM were estimated from the first fold of the 15-fold cross-validation process. The tuned parameters were then kept unchanged for all the other folds. Since SVM is a binary classifier, 4 SVM classifiers were trained, one for each emotion category: Excited, Relaxed, Frustrated, Bored. A winner-take-all approach was used for the multi-class classification, i.e. deciding the winner between the four emotions. Two binary SVM classifiers were also built for arousal and valence dimensions.

Table VIII. 15-fold cross-validation results.

<table>
<thead>
<tr>
<th>Classes</th>
<th>Features</th>
<th>Classifiers</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 Emotion labels</td>
<td>12 features listed in Table III</td>
<td>DA</td>
<td>75.3%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANN</td>
<td>72.7%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Linear SVM</td>
<td>77.0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kernel SVM</td>
<td>69%</td>
</tr>
<tr>
<td>2 levels of arousal</td>
<td>12 features listed in Table III</td>
<td>DA</td>
<td>89.3%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANN</td>
<td>88.3%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Linear SVM</td>
<td>89.7%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kernel SVM</td>
<td>86.7%</td>
</tr>
<tr>
<td>2 levels of valence</td>
<td>Pressure, length and speed features</td>
<td>DA</td>
<td>83.0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANN</td>
<td>84.7%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Linear SVM</td>
<td>82.3%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kernel SVM</td>
<td>86%</td>
</tr>
</tbody>
</table>

Prior to the training of the models, all feature values were normalised to the interval [0,1] by using maximum and minimum values observed in the database. The 12 selected features listed in Table III were used as input features to the classifiers. For 4-emotion classification and Arousal classification, the best results were obtained by using all the 12 features. For Valence classification, the best results were obtained by using only stroke length, pressure and speed features. The leaving-one-person-out cross-validation method was used to build and test person-independent models, i.e. at each trial of the cross-validation process, a model was built over 14 participants and then tested on the participant left out. The process was repeated for all 15 participants.

5.2 Results

Table VIII shows the summary results over the 15 cross-validation trials for the indicated sets of input features. The models for Arousal produced the best correct recognition results ranging from 86.7% (kernel SVM) to 88.7% (linear SVM). In the case of Valence, the performances ranged from 82% (linear SVM) to 86% (kernel SVM). The results for the discrimination of the 4 emotion labels were also well above chance level; ranging from 69% (kernel SVM) to 77% (linear SVM).

For economy of space we report only the confusion matrix obtained for the 4-emotion linear SVM classifier (see Table IX). We observe the same pattern observed in Table IV for person-dependent models. The best results are obtained for the classification of both Frustrated and Excited strokes. In each case, more than 80% of
the stroke samples were correctly classified. The lowest performances (67%) are obtained for Relaxed samples. Most misclassifications occur between Relaxed and Bored (37.1%) and between Relaxed and Excited (20%).

Table IX. Confusion matrix for 4-Label Classification (Excited, Relaxed, Frustrated, Bored) by linear SVM classifier.

<table>
<thead>
<tr>
<th>Actual Label (count)</th>
<th>Predicted Label</th>
<th>Valence: Length feat.</th>
<th>Arousal: All features</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Excit, Relax, Frus, Bor.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excited</td>
<td>74</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>Relaxed</td>
<td>7</td>
<td>47</td>
<td>1</td>
</tr>
<tr>
<td>Frustrat.</td>
<td>5</td>
<td>0</td>
<td>60</td>
</tr>
<tr>
<td>Bored</td>
<td>2</td>
<td>11</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Actual Label (%)</th>
<th>Predicted Label</th>
<th>Valence: Length feat.</th>
<th>Arousal: All features</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Excit, Relax, Frus, Bor.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excited</td>
<td>82.2</td>
<td>10.0</td>
<td>6.7</td>
</tr>
<tr>
<td>Relaxed</td>
<td>10.0</td>
<td>67.1</td>
<td>1.4</td>
</tr>
<tr>
<td>Frustrat.</td>
<td>7.1</td>
<td>0</td>
<td>85.7</td>
</tr>
<tr>
<td>Bored</td>
<td>2.9</td>
<td>15.7</td>
<td>10.0</td>
</tr>
</tbody>
</table>

As shown in Table VIII, the results for the person-independent model built with DA (75%) are comparable to the results obtained with DA for the person-dependent model (76.4%) shown in Table IV. This indicates that the selected features have high generalization capability independently of people’s differences.

5.3 Future improvements

The results shown in the previous section are comparable to the results on acted affective movements (e.g. [Camurri et al. 2003, 2004; Kapur et al. 2005; Bernhardt et al. 2007; Kleinsmith and Bianchi-Berthouze in press]), on acted key-pressing and dotting Braille devices [Matsuda, et al. 2010; Khanna et al. 2010] and also to results on acted emotion recognition in interpersonal touch-based communication [Bailenson et al. 2007]. They are also comparable to and even higher than the performance obtained for naturalistic affective body movements [Kleinsmith et al. 2011; Savva et al. 2012].

Whilst these results clearly show the possibility of using this modality in the context of affective interaction, a number of improvements could be made. First, from the perspective of the modelling process, improvement could be considered at the level of feature computation and labelling process. Specifically, the normalization of some of the features was based on a very simplistic process. A more refined computation of the pressure baseline, that takes into consideration the nominal contact area of the player’s finger, could be considered. Also, the use of devices that measure pressure directly, rather than through inference from other features, could provide more discriminative power. Finally, a better tuning of the learning parameters could further improve performance. In our study, tuning of the machine learning algorithms was performed using only the data from the first fold in the 15-fold cross-validation process. A larger dataset would therefore improve this tuning by allowing to set aside a dedicated set of data for this phase of the modelling process.

With respect to the labelling process, a triangulation method could be used to validate the self-report of the participants. For example, bio-sensing data as well as facial expressions data could have been collected to provide a further measure of the affective state of the player. However, interpretation of such data does also carry a certain level of uncertainty and subjectivity. Understanding how to combine them in building the ground truth is a research question in itself.
Another important improvement could be obtained by considering the variability between players and the difficulty of the game. Finger dexterity, as well as touch behaviour idiosyncrasies, may affect people’s stroke patterns independently of their emotional state. The study reported in [Bernhardt et al. 2007] and discussed in section 2 showed that, by removing idiosyncratic movements from knocking patterns, automatic emotion recognition rates improved significantly (from 50% to 80%). In [Khanna et al. 2010], it was shown that, by considering daily variability in people’s strength, the recognition performance could be improved. We could hence expect that personalized models may reach higher performance. However, differently from [Khanna et al. 2010; Bernhardt et al. 2007], our recognition performances are already quite high without taking into account such differences. The person-independent model did, in fact, reach performances similar to the person-dependent model.

Looking at Figure 8, we can also see that the variability is not only between participants but also and mainly within participants (long boxplots). This variability could be explained by the use of different strategies at different levels of difficulty of the game. In our game, the player could, for example, decide to slash many fruits at once, generating a less direct stroke; or, by using shorter and faster strokes, could decide to slash one fruit at a time. These strategies could be person-dependent but also depend on the level of difficulty of the game. Figure 3 showed that the four emotional states were triggered at any level of the game. However, the figure showed that certain states are more frequent in different levels of difficulty of the game. For
example, the number of Boredom states was higher towards the end of the game, whereas Relaxed states increased after the first levels and decreased again as the game got faster.

To investigate a possible effect of game difficulty on stroke misclassification, we further analysed the results of the DA person-dependent model obtained in section 4. The classification results are provided in Figure 9. The results are shown in the confusion matrix at the bottom of the figure. Each graph shows the results for the set of strokes belonging to the class indicated in the graph title. For each graph, the labelled lines correspond to the cumulative percentage (vertical axis) of strokes that were misclassified with that label given the level number (horizontal axis). The unlabelled line denotes the cumulative percentage (vertical axis) of strokes that were correctly classified.

The top two graphs show that the misclassifications between Excited and Relaxed strokes reach 80% within the first 10 game levels. The graphs for Relaxed and Bored strokes show, instead, that the misclassification between these two classes occurs in the second part of the 20-level game. Given that Figure 6 shows that length and speed, in particular, decrease as the game becomes faster, the normalization of these features, according to the speed of the game, may remove some of these errors. Similar patterns can be observed between Frustrated and Excited states and between Frustrated and Bored states, indicating that similar normalization for pressure and DI could also improve the results.

6. DISCUSSION AND CONCLUSIONS
The paper presented a non-obstructive approach to the detection of players' emotional
states during gameplay with touch-based devices. The emotion recognition is based on the touch behaviour (stroke) of the player. Length, pressure, direction and speed of finger strokes are used for this purpose. A visual analysis of the stroke features, as well as the results of Discriminant Analysis, showed that the length and the pressure of the stroke are particularly informative in discriminating between two levels of valence (positive vs. negative states), whereas the speed and the direction of the stroke strongly inform the discrimination of levels of arousal. Pressure also supported such discrimination. The analysis further showed that pressure strongly discriminates Frustration from the other three states and length was particularly informative for the identification of the Relaxed state.

The informative power of the extracted finger-stroke features was also demonstrated through a computational model. Both person-dependent and person-independent models were created. They showed similar correct recognition performance and similar misclassification patterns. In the case of the person-independent model (most challenging one), arousal classification reached a classification performance between 86% and 89% using all selected features. In the case of valence, performance ranged between 82% and 86%, slightly lower than that for arousal. This is consistent with studies of automatic emotion recognition from body movement and posture [Kleinsmith et al., 2011]. For the four-emotion recognition case, performance ranged between 69% and 77%.

These experiments have shown that touch behaviour is indeed a powerful modality, able to provide information on the affective state of the players, even in touch-based computer games. In fact, the results are comparable to those obtained on acted and naturalistic body movements (e.g. [Camurri et al. 2003, 2004; Kapur et al. 2005; Bernhardt et al. 2007; Kleinsmith et al. 2011; Savva et al. 2012]), in acted key-pressing and dotting Braille devices and also to those on emotion recognition in interpersonal touch-based communication. Unlike studies in body movement, the discrimination between Frustration and Excitement appears to be strongly facilitated by the use of the pressure feature. This is also supported by studies on typing and dotting [Khanna et al. 2010; LV et al. 2008; Matsuda et al. 2010]. It would be of interest to explore the pressure feature in a body movement context by using a force plate device.

In the following, we discuss the use of our findings in the game context and in other application areas. We also discuss generalization issues to pertaining to using automatic emotion recognition models based on touch.

6.1 The use of affective touch recognition in games

Overall, these results support and extend previous studies by showing that touch behaviour is indeed an informative modality for automatic emotion recognition not just in an acted setting but also in a naturalistic setting when touch-based game devices are used. This shows that this modality has the potential to be used both for user-experience evaluation studies in a game context and for on-line adaptation of the gameplay. The level of the game, as well as the gameplay and the functions available to the player, could be adjusted accordingly to maintain the engagement of the player. In fact, an emerging trend in computer games is the creation of algorithms that generate the content of the game at run time, by taking into account the affective and cognitive experience that the player is going through [Yannakakis and Togelius, 2011; Liapis, Yannakakis, Togelius, 2012], with the goal of optimizing it. Most of these generative algorithms infer the players’ affective and cognitive states on the basis of choices the players make and their mistakes. More recently, there has been an interest in facilitating the modelling of the players by also monitoring their physiological signals [Yannakakis et al. 2007; Yannakakis et al. 2010; Nijholt, et al. 2009] and other non-verbal modalities, or from explicit player reports when computational power is an issue. The use of touch as an affective communicative channel would be an interesting modality for applying such generative algorithms to a touch-display game device, when facial expression and body-movement recognition,
or bio-signal detection, may not be feasible. As for any other affective channel, touch
can also be used to send fake signals of affect. Rather than this being seen as a
limitation, it should be considered a possibility to recreate a gameplay situation
comparable to real life gameplay, where acted expressions can be used to confuse the
opponent [Nijholt 2007]. As the player learns that the game reacts to his/her affective
touch behaviour, this becomes a new functionality or tool to use to bring a new level
of complexity to the game. Similarly to the variability in the ability of humans to
recognize affect in others, the game software (or the characters in the games) could be
modelled with such a range of emotional intelligent skills too.

6.2 The use of affective touch recognition in other applications
As the use of touch-based devices (e.g. iPad, tablet, tabletop) is spreading to different
contexts, these results also show the potential for using this non-intrusive affective
measure in other emotion-sensitive contexts. Not only could the entertainment
industry, in general, benefit from such an approach (e.g. mood-aware music player),
other areas where emotion and mood play a critical role could also profit, by offering
support to emotion enhancement, awareness and regulation.

A touch behaviour similar to the one studied in this paper and commonly used on
touch-based devices is scrolling. This touch behaviour is used to look for information
in various contexts. For example, the detection of the emotional state of a person
while s/he is scrolling through the available songs may be used to propose songs or
personalize the search [Liu et al. 2008; Thrasher et al. 2011]. Being able to detect the
emotional state of a person, while scrolling to look for a person to contact, may be
instead used to enrich the messages with emotional cues [Sundström et al. 2007], or
to make the senders aware of their state before sending the message or making the
call. A stroke-like touch behaviour (e.g., dragging) is also used to organize or group
together stored data over displays (e.g., photos in a digital album or material for
holiday planning [Marshall et al. 2011]). The emotion detected through the stroke
behaviour could be used to tag the stored material and to identify clusters of stored
information (music and photos) that reflect similar emotions [Su et al. 2011]. Such
meta-tagging could also be used to facilitate crowd-sourcing of media information and
to build better recommendation systems [Brew et al. 2010]. Zooming is another touch
behaviour that requires stroke-like patterns (although two fingers are needed). Being
able to recognize the emotional states of a person navigating a map could be used in
various ways. For example, the detection of anxiety, which affects how information is
processed, could be used to adapt how information is presented or to decide when to
display advertisements. Personalized advertisement is very important to favour
adoption and to reduce users’ frustration [Hristova and O’Hare 2004].

The psychology literature (e.g. [Clore and Palmer, 2009]) has shown that emotional
states strongly affect our cognitive processes (e.g. by enhancing or inhibiting memory
processes). Being able to detect the emotional state of users, could help the
personalization of the task (e.g., presentation or content) to better support the users.
Other than the map navigation example discussed above, of particular interest is the
area of education, where the emotional states of students and their capacity to
regulate their states are strongly related to the students’ motivations, learning
capabilities and performances, as well as health [Schutz and Pekrun 2007]. By
detecting the emotional states of the students during self-directed study, the device
could provide some form of support, generally provided by a teacher; for example, by
tailoring the teaching material and the pace to the students’ needs [Conati and
Maclaren 2009], or offering empathic and encouraging behaviour to help an
emotional shift in the student. Studies in this area have shown that the ability of the
teaching application to provide such support can indeed show positive results
[Maldonado et al. Y., 2005]. However, most of these studies are based on facial-
expression detection or self-report and hence have their limitations. Facial expression
recognition requires high computational power to provide reliable results and it is
quite intrusive, as discussed in the introduction. It is also strongly affected by
illumination conditions and by how visible the face of the student is to the camera. Whereas, in the case of desktop conditions, this may be easier to address, the situation is quite different with touch-based devices, where the environment changes constantly and computational power is limited. On the other hand, having to self-report is often seen as an intrusive and frustrating activity that people will avoid.

Automatic recognition of emotional states through touch could also offer a non-obtrusive channel to facilitate awareness and hence reflection on emotional states in rehabilitation [Kapoor et al. 2007]. In clinical practice (e.g. mental health rehabilitation) and in clinical research contexts (e.g. pharmaceutical trials), patients are often asked to maintain diaries of their physical and psychological states in order to facilitate reflection on behavioural patterns but also to monitor the effect of drugs [Palmblad and Tiplady 2004; Keogh et al. 2010]. Generally based on touch screen and stylus of smart-phone or palm-held devices, these e-diaries appear to be easier to use than their paper counterpart. This is because they facilitate the repeated recording of automatically time-stamped data and hence allow for easy later retrieval and further elaboration. The frequent recording provides data that are less biased by late recall [Burton et al. 2007]. The possibility to capture emotion cues, automatically and continuously [Gune and Pantic 2010], from touch interaction with the device, would provide a non-obstructive and continuous mechanism to gather mood-relevant patterns stamped with the time and context of occurrence. Such cues could be provided to the user for reflection when the patients use the e-diary application. As proposed in [Lindstrom et al. 2006], rather than being used for a strict categorization of the emotional state of the patients, the collected touch patterns could instead be used as cues for the patients to reflect upon them without feeling that they have been categorized. It should be noted that, in the last two examples of application (educational and clinical contexts), the type of touch behaviour (stroke) studied in this paper may not be the only or the most frequently used one. As discussed in section 6.3, further studies may be necessary to understand how our findings generalized to other types of touch behaviour.

The application of this work to real-world scenarios raises the question of how to handle the fact that recognition performance is good but not perfect. According to the results shown in Section 5, recognition may be inaccurate, on average, between 20% and 30% of the time. Given the intrinsic subjectivity of reading someone’s emotions, this level of reliability may be acceptable in certain applications (e.g. game, entertainment) [Kleinsmith et al 2011]. For example, the study presented in [Iacobini et al. 2010] shows that, in the case of interactive-affective-aware art technology, users have the tendency to assign meaning to the reactive behaviour of technology that goes well beyond its real emotion recognition capabilities. Gaver et al. [2006] state, in fact, that error-induced randomness - which, in itself, is meaningless - appears to encourage a richer behaviour in people as they assign meaning to it.

However, this is not always the case and the misinterpretation may result in user frustration or may reduce the effectiveness of the support provided by the technology. Two possible solutions may be used to address the problem. From a modelling perspective, a probabilistic approach could be used. A confidence value on the reliability of the recognition of a particular emotional cue could be used to decide the type of personalization action to be taken and/or whether the action should be taken at all. This is a typical problem in any personalization task. The detected emotion should be used as one of the information available to decide what and if the personalization should occur. It would also be of interest to investigate the possibility of integrating touch-behaviour with other modalities when the use of multiple modalities is a possibility.

A different approach is used in [Lindstrom et al. 2006; Sundström et al. 2007]. Here, the mapping between emotional cues (e.g. physiological states or arm movements) and emotional states is used only to provide a canvas for reflection. Not only do the users keep complete control over the changes proposed by the technology (e.g. change of the colour and texture of the display or the objects in it), they are also
in charge of assigning meaning to the changes (the visual representations in these two studies). The reason for using a fuzzy interpretation of affective cues is not only that the recognition performance may not be sufficiently accurate, but also that emotions are very complex, subjective phenomena and knowing what the right response should be is a research question in itself.

6.3 Generalization considerations

Whilst our results show the potential of using touch behaviour as a source of information for personalization, the modelling of the relationship between touch features and emotional states may be affected by various factors, such as the type of tactile behaviour, the context of use and the activity performed as well as person-dependent factors.

It is very possible that the features we have identified as discriminative in the case of stroking behaviour may not be that discriminative for other types of tactile behaviour. It would be of interest to investigate if touching behaviour could be organized according to a set of characteristics that may be the factors that define the discriminative features. For example, tactile behaviour, such as stroking or caressing, are characterised by the fact that the behaviour is performed on a length of the touched surface. These two tactile behaviours may share the same, or a subset of the same, discriminative features (e.g. the duration). In contrast, tapping is a discrete localized behaviour and its frequency may have more discriminative importance than the duration of each unit of behaviour. As discussed in sections 2 and 4, the studies on key stroking and mouse clicking have indeed shown some differences with our results on stroking behaviour and with results in interpersonal touch [Hertenstein et al. 2009].

The context in which behaviours are performed may also play an important role, not only in determining the features, but also the baselines to be used for the normalization process. For example, activities such as gaming and organizing information may have different baseline levels because of their different speed constraints and because of the different cognitive processes involved. Furthermore, it is possible that the user emotions and cognitive states may bias the type of tactile behaviour used to perform an action, when multiple tactile behaviours are available to perform it. Indeed, the study by [Hertenstein et al. 2009] showed that, in the context of social interaction, each emotion has preferred touch behaviours associated with it.

It should also be questioned whether touch behaviour is a direct representation of the state of the user or if their correlation should instead be explained by a more complex process. For example, there is an emerging body of literature suggesting that behaviour may elicit emotional states. Studies show how proprioceptive feedback can bias the affective state of a person even when the movement or posture is imposed on them [see Niedenthal et al. 2005; Chandler and Schwarz 2009]. Studies in game technology have further investigated this relationship to inform the design of whole-body game technology [Niedenthal 2009; Isbister et al. 2011; Bianchi-Berthouze 2012; Savva et al. 2012]. Given the strong relationship between touch and proprioceptive feedback, it is very possible that this biasing phenomenon also occurs in the case of touch behaviour. This may have various implications for the design of touch-based applications, as touch behaviour could be used by designers not only to recognize how users feel but also to bias their affective states. As the user's experience evolves over time, Savva et al. [2012] proposed to continuously track the user's affective expressions to provide a better measure and understanding of such experience. In line with this work, Nijhar et al. [2011] propose to use players' body movement strategies to automatically detect their motivation to play the game.

Another direction to be considered is the possibility that touch behaviour could provide further information about other affective phenomena, such as the user's personality. Various studies have shown a relationship between personality and amount of tactile behaviour (e.g. [Deethardt and Hines 1983]). Furthermore, there is
increasing evidence that the response of a person's somatosensory cortices (e.g., response threshold to touch) is affected by personality traits [Schaefer et al. 2012]. It would be interesting to investigate if a relationship exists between the patterns of tactile behaviour and personality. This would help to improve the overall model investigated in this paper by providing a deeper understanding of the relationship between touch behaviour and affective processes.

In conclusion, this paper has investigated the possibility of using stroke behaviour to discriminate between a set of affective states in the context of touch-based game devices. The results are very promising and pave the way for further explorations with respect to not only different contexts but also different types of tactile behaviour. Further studies are needed in a variety of contexts to establish a better understanding of this relationship and identify if, and how, these models could be generalized over different types of tactile behaviour, activity, context and personality traits.

REFERENCES


LIAPIS, A., YANNAKAKIS, G.N., and TOGELIUS, J. 2012. Adapting Models of Visual Aesthetics for Personalized Content Creation. IEEE Transactions on Computational Intelligence and AI in Games, Special Issue on Computational Aesthetics in Games.


NIEDENTHAL, S. 2009. What we talk about when we talk about game aesthetics. Int. Conf. on Digital Game Research Association.


VON LABAN, R. 1971. The mastery of movement. MacDonald & Evans Ltd.

Received March 2012; revised June 2012; accepted September 2012