HAL
open science

Identifying Emotions Expressed by Mobile Users through 2D Surface and 3D Motion Gestures

Céline Coutrix, Nadine Mandran

To cite this version:

Céline Coutrix, Nadine Mandran. Identifying Emotions Expressed by Mobile Users through 2D Surface and 3D Motion Gestures. UbiComp 2012 - 14th International Conference on Ubiquitous Computing, Sep 2012, Pittsburgh, PA, United States. pp.311-320, 10.1145/2370216.2370265 . hal-00752456

HAL Id: hal-00752456

https://hal.science/hal-00752456v1

Submitted on 20 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Identifying Emotions Expressed by Mobile Users through 2D Surface and 3D Motion Gestures

Céline Coutrix
CNRS, Laboratoire d'Informatique de Grenoble
Celine.Coutrix@imag.fr

Nadine Mandran
CNRS, Laboratoire d'Informatique de Grenoble
Nadine.Mandran@imag.fr

ABSTRACT
Only intrusive and expensive ways of precisely expressing emotions has been proposed, which are not likely to appear soon in everyday Ubicomp environments. In this paper, we study to which extent we can identify the emotion a user is explicitly expressing through 2D and 3D gestures. Indeed users already often manipulate mobile devices with touch screen and accelerometers. We conducted a field study where we asked participants to explicitly express their emotion through gestures and to report their affective states. We contribute by (1) showing a high number of significant correlations in 3D motion descriptors of gestures and in the arousal dimension; (2) defining a space of affective gestures. We identify (3) groups of descriptors that structure the space and are related to arousal. Finally, we provide with (4) a preliminary model of arousal and we identify (5) interesting patterns in particular classes of gestures. Such results are useful for Ubicomp application designers in order to envision the use of gestures as a cheap and non-intrusive affective modality.

Author Keywords
Mobile User Interfaces, Affective Computing

ACM Classification Keywords
H.5.2 User Interfaces: [Input devices and strategies]

General Terms
Experimentation, Human Factors

INTRODUCTION
Nowadays, the number of smartphones users is increasing. These devices are manipulated by their users everyday, everywhere, and for a lot of various daily purposes. As our subjects self-reported during first interviews, it is not rare to see users manipulating their device, not in a useful purpose, but as a support to explicitly express emotions, e.g., irritation by forcefully shaking the device or impatience by patting with fingertips on it. Such observations suggest that there might be a link between gestures performed by users and the affective states users wish to express. In this paper we address the problem of identifying to which extent such a link exists between *intentional expressive gestures* and *emotions users explicitly wish to express*. We neither address the identification of intimate emotion nor the implicit gestural activity of a user throughout the day.

The importance of including the affective aspect in computing systems in general have been already demonstrated by the pioneers of affective computing [16], for instance for adapting the interaction to increase performance or satisfaction or for communicating subtle emotions to distant others. Researchers then proposed techniques for identifying emotions based on technologies that are intrusive (e.g., electromyography of the face), expensive (e.g., thermal imaging) or not suitable for ubicomp (e.g., voice analysis subject to background noises). Yet, the benefits of affective interaction could also apply to the use of mobile systems [3]. For instance, users could benefit from precisely specifying an emotion in a mobile communication application, instead of typing ambiguous, too narrow and incomplete emoticons [3].

Motion [21] and surface [24] gestures are, respectively, performed in 3D with a mobile device and in 2D on the touchscreen of a mobile device, like a smartphone or a mediaplayer. Sensing these gestures is now widely done on many mobile devices. Thus, 3D motion and 2D surface gestures are ideal candidates for a cheap, discreet and mobile way of expressing affective parameters. However, little is known about the affective dimension of gestures: HCI researchers mainly focused on gestures mapping to commands and affective computing researchers mainly focused on non-ubicomp situations of use. The challenge to explore the area of affective 2D surface and 3D motion gestures is high as it has never been addressed before, but the possible benefits are promising, making this topic important to investigate.

After gathering and analyzing *in situ* and *in the moment* samples of the subjects' gestures and emotions, we found that 249 descriptors of each collected gesture, 104 of them were significantly correlated to at least one of the affective dimensions reported by subjects. More precisely, a high number of 3D motion gesture descriptors were correlated to affective dimensions and the arousal dimension of emotions was correlated to a high number of gestures descriptors. As correlations were rather low, we gathered descriptors along more

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
UbiComp '12, Sep 5-Sep 8, 2012, Pittsburgh, USA.
Copyright 2012 ACM 978-1-4503-1224-0/12/09...$10.00.

significant axes to define an affective gestures' space. We identified groups of descriptors that structure this space. Two of them are related to arousal : minima of z projections of acceleration (back-front of device) and its derivative on the one hand, and duration and spectral descriptors on x and y projections (left-right and bottom-top of device) of either raw or low-pass filtered acceleration on the other hand. Finally, we provide with a preliminary model of arousal and we showed that qualitative analysis of the data allowed us to identify interesting patterns in particular classes of gestures, leading to more targeted future studies.

The results of our study will be useful for ubicomp system designers and developers in order to knowingly and reliably use affective gestures as one of the affective modality in their applications in a cheap and non-intrusive way.

In the paper, we present our field study and its results. We then discuss the results and their implications, before concluding and proposing future work. First we start by presenting how others contributed in solving the problem of linking gestures and affective states in the next section.

RELATED WORK

Previous work contributed in solving the problem of sensing affective states through gesturing from the viewpoint of affective computing, gesture-based interaction and affective gestural interaction.

Alternative Sensing for Affective Computing

In the field of affective computing, modalities usually contributing to sense emotions are facial expressions, thermal imaging of faces, vocal intonations, language, galvanic skin response, electromyography of the face and heart rate. Unfortunately, these technologies are costly and intrusive, or not always usable in a ubicomp situation. In a business meeting for instance, it is not acceptable for an attendee to use speech as an input to his/her interface, as it would disturb the speaker. On the contrary, we aim to focus in this paper on modalities that are discreet, widely available, cheap and easy to use in ubicomp.

Researchers have introduced low-cost, non intrusive sensing of emotions. The keyboard through keystrokes dynamics has been shown as a promising way for sensing emotions [5]. Keyboards however are unfortunately not available while on the move.

Others tried to sense emotions of mobile users. An idea has been to use wearable computing to sense emotions [18]. For this, the user has to wear special clothing. Sensors embedded in the clothing detect embrace, squeeze/press and stroke actions centered around the arm. Others enhance a mobile communication applications [23][8], using physiological sensors. However with these solutions, users would have to embed sensors in the sleeve of every piece of clothing - sometimes lacking sleeves - or wear a special equipment everyday. On the contrary, a smartphone or a media player is already carried everyday and everywhere by users.

Gesture-based interaction

Gesture-based interaction usually specifies commands, like e.g., in one of the earliest gestural interaction techniques on a surface [20] like a pinch gesture to pick up an object on the surface. Gestures sets are mapped to sets of commands, or less commonly, to parameters like in Control Menus [17]. Users need to implicitly or explicitly learn and train in order to be able to use such a gesture-based interface, whether with a device or on a surface.

Previous attempts tried to facilitate the learning of gestures by finding clues in spontaneous users gestures. In [24], the authors investigate spontaneous gestures proposed by users for a given command on a touch-sensitive tabletop UI. They find four dimensions for their taxonomy: form, nature, binding and flow. For instance, along the flow dimension a gesture can be discrete when the action occurs after completion of the gesture, or continuous if the action occurs during completion of the gesture. Even though these dimensions could apply to affective gestures, it has not been investigated yet that they actually are relevant to affective gestures. In this paper, we would like to take the same approach by finding clues in spontaneous gestures in order to find relevant dimensions for *affective* gestures, on the contrary to *command-targeted* gestures. Indeed, this work made clear the need to involve end-users in the design of gestures for interaction: Designer's gestures set was shown significantly different from user-defined gesture set.

Others tried to apply the same approach to find user-defined 3D motion gestures in mobile context [21]. From their taxonomy, physical characteristics can apply to affective gestures. Among these, the kinematic impulse is the range of jerk, e.g., change in acceleration; the dimensions of the gestures, from one to 6 degree of freedom; the complexity of a gesture, i.e. if it is composed of single, elementary gestures between spatial discontinuities like a pause, or if it is a single, elementary gesture. These three characteristics could apply to affective gestures, but such a link has not yet been explored.

Affective gestural interaction

In order to provide more expressivity, the combination of 2D surface and 3D motion gestures have been explored. For instance, it is possible to distinguish gently swiping a finger on the screen from "drags with a hard onset" [11]. However, the authors did not map them to affective states.

A promising link between the way one writes a letter with a pen and the stress due to task difficulty was shown [1]: Gesture duration, number of points, gesture length and pulse rate were related to difficulty. However subjects were not performing free gestures but had to write a specific letter and the results are linked to physiological measurements but not to any model of emotions.

Haptic gestures has been explored for emotional communication, through moving, squeezing, and stroking a custom device [19] and through spontaneous gestures [9]. However, the first experiment took place in a lab instead of in the wild

and participants were only provided a limited set of gestures to choose from - three different gestures only. Moreover, even though their scenario stimuli had a strong emotional component, no explicit link to an emotional model was proposed.

The expressivity of gestures have been explored through camera image processing [2]. Characteristics of expressivity such as overall activation, spatial extent, temporal, fluidity, power are computed. Another similar approach has been explored in [7], where quality of gestures based on dance characteristics has been detected through Kinect[1]. Movement quality is defined as "distinctly observable attributes or characteristics produced by dynamics independently of movement trajectory or shape". In their paper, the authors define three qualities: (1) breathing, an oriented increase of hand's velocity and a release of the hand with a slow velocity, (2) jumping: caress on the surface where the hand sweeps periodically from top to down at different position on the interface, and (3) expanding: a multi-directional travel of the hand on the surface. However, both of these approaches are not explicitly linked to any emotional model. Moreover, they need the user to stand in front of a camera and we would like to avoid this constraint in order to be usable in a ubicomp application.

The link between gestures and emotions was analyzed. Features of gross body movement and emotions categories were shown related [15]. However, they consider whole body movement, which is not applicable to gesturing with or on a mobile device. The link between hand gestures and emotions in theater plays was also examined [12]. Authors annotated handedness, hand shape, palm orientation and motion direction in two scenes of two version of the same play and found correlation between handedness and emotion. Unfortunately, there is no way to capture the handedness through a mobile device and it is not straightforward that users gesture the same way with free hands or while carrying a device.

Previous attempts to introduce affective gestures into mobile devices include eMoto [6], a mobile service for sending affective messages to others. For specifying a user's affective states, eMoto uses a language of gestures, combining two levels of (a) movement for pleasure and (b) pressure for arousal. These gestures have been arbitrarily designed. On the contrary, similarly to recent work in gesture-based interaction [24][21], we would like to investigate user-defined affective gestures. Towards this aim, we conducted the field study that we present now.

FIELD STUDY

In order to investigate the link between 2D surface and 3D motion gestures and expressed emotions, we gathered *in situ* and *in the moment* data of the subjects' gestures and emotions. In this section, we describe our method for collecting ecologically valid data for such an exploratory study (Experience Sampling Method (ESM) [10]), for ensuring a usable dataset and analyzing the data.

[1]http://www.xbox.com/en-US/kinect, retrieved June 2012

Data Collection

The study gathered 12 adult subjects, expert everyday users of mobile tactile devices like smartphones or media players. They were chosen to diversify as much as possible their profiles: 6 were males and 6 females, aged 25-47 (mean= 32.92, s.d.= 6.76). Their various areas of work were software, biology, didactics, social worker, executive assistant, ski instructor and mountain guide. Subjects did not get any compensation for participating in the study.

Subjects were asked to install an application on their personal, usual mobile device and use it to record gestures and emotions during 15 days at least. Eleven subjects used their Apple iPhones while a single subject used his iPod touch. As a consequence to using their usual personal device for ecological validity, we could not get gyroscopes data as these were not available on all devices.

We used both *event-contingent sampling* and *signal-contingent sampling*, as subjects were asked to use the application either:

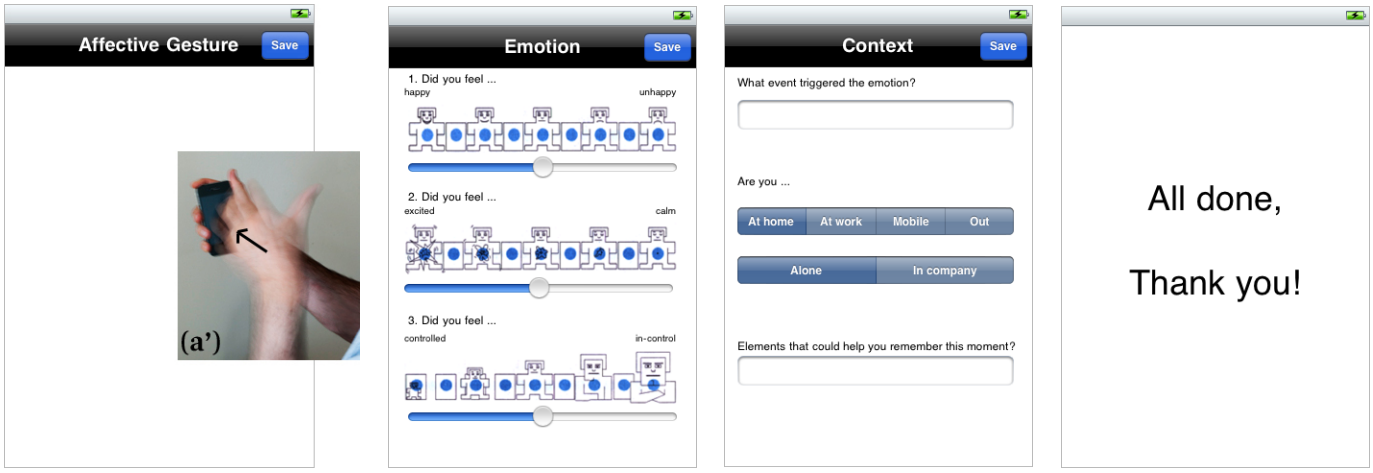
1. when they were experiencing a particular affective state. Doing so allowed us not to miss intense, seldomly occuring affective states.

2. when they received a text message asking them to do so. One text message was sent everyday at a random time of the awoken day to every participant. Doing so allowed us to capture subsequent data per subjects, but also to capture more neutral affective states that are not usually reported.

Figure 1 shows screenshots of the data collection application manipulated by the participants. Sequentially, subjects were prompted to:

1. Perform their gesture (Figure 1.a), either on the device's screen and/or with the device in the air. There were no other restrictions on their gesture, as subjects were encouraged to provide us with unrevised behavior without being influenced by capabilities of a system. While the subject was performing a gesture, the application recorded accelerometer and touches data, widely available on smartphones and media players. For these gestures, no feedback was provided through the interface, as we did not want subjects to tailor their gesture to obtain a particular feedback, e.g., drawing, but rather to express their emotion through the gesture.

2. Specify their affective state (Figure 1.b). For this, we used the PAD model [14] describing every emotions as points in a orthogonal, three-dimensional and continuous space of Pleasure (happy vs. unhappy, top slider of Figure 1.b), Arousal (excited vs. calm, middle slider of Figure 1.b) and Dominance (controlled vs. in-control, bottom slider of Figure 1.b). PAD scales were instrumented with the graphical Self-Assessment Manikin (Figure 1.b) [13] to collect subjective report of emotion as PAD values. Among available emotional models, we chose one that was not categorical in order to provide users with subtle way of reporting their emotions. Indeed, it has been



(a) (b) (c) (d)

Figure 1. Data Collection Application Screenshots: Collecting gesture (a), emotion (b), context (c) and Ending (d); Tapping Gesture (a').

shown that a categorical model leads users to very often use custom label for emotions as they do not always find a proper one in the list [3].

3. Give some context information (Figure 1.c): The event that triggered the emotion, whether they were at home, at work, on the move or out, whether they were alone or in company, and any other elements that could help them remember this moment. Figure 1.d shows the final screen when leaving the application.

During the 15 day study, we performed 3 interviews with each participant. All interviews were recorded and later transcribed. First, just before the study begins, we explained its purpose, collected background data, trained subjects with the SAM questionnaire with 3 contrasting scenarios (delighted, shocked and relaxed) and had them try the application. Then, half way during the study and at the end of the study, we solicited their feedback about the study, asked them to comment on their reported gestures and emotions of the past week and to demonstrate the reported gestures. We also asked them if they ever wanted to express emotions in a way that was not taken into account by the mobile phone (e.g., with pressure), and about the application they wish, or not, for identification of emotions through gestures, their benefits and drawbacks.

Dataset and Analysis

We collected 188 valid samples from all 12 subjects. Samples were not considered valid when unfinished or performed during interviews for demonstration or try. Individual subjects reported between 6 and 27 samples (one report every two days vs. two reports a day). We collected on average 15.7 samples ($median = 15.5$, $s.d. = 6.84$) per subjects - approximately one per day, as expected.

We prepared the data for analysis. From raw gesture logs like represented in Figure 2 and 3, we computed descriptors of gestures. Apart from general ones like duration, we computed the 3D motion and 2D surface gestures descriptors to be used in analysis.

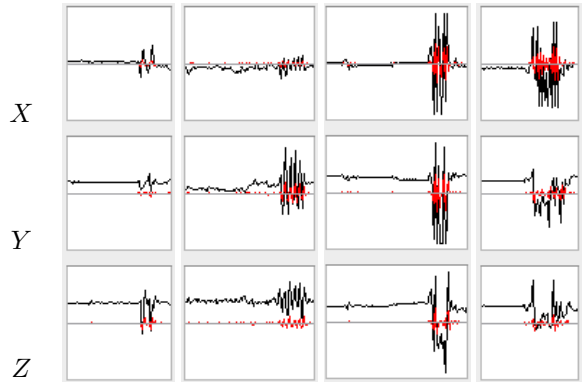


Figure 2. Four of the 3D motion logs characterized by a pause where the device is hold still and then an important movement in all three directions, expressing high arousal.

3D motion gesture descriptors

Among the 188 gestures, 36 are 3D only and 152 are combined 2D/3D gestures. For describing 3D motion, we considered (1) acceleration values and acceleration's derivative (Jerk) and (2) spectrum. First, for acceleration and jerk, we considered projections of these on x (left-right of device), y (bottom-top of device) and z (back-front of device) axes, and also 3D vector length. We computed minimum, maximum, amplitude, median, mean and variance of their signed and absolute values. These descriptors were computed based on both raw acceleration values and high-pass filtered acceleration values, i.e. with gravity filtered out.

Second, for spectrum analysis, we computed the gap (G) that maximizes the difference between most important and less important frequencies in the spectrum of the acceleration signal in order to reveal the periodicity of a gesture. We also computed the number (N) of important frequencies in the spectrum, as well as the most important frequency (M).

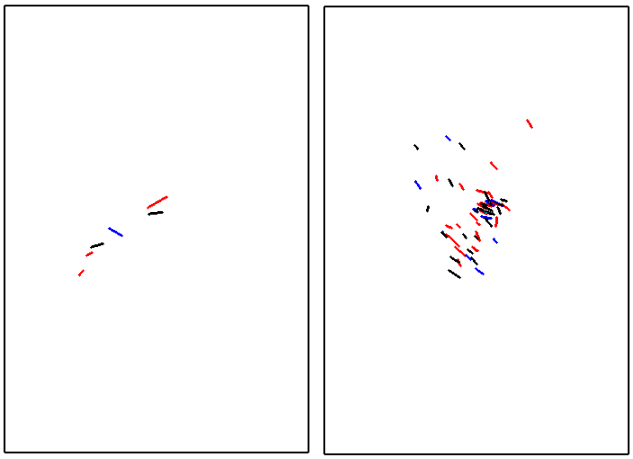


Figure 3. Two of the 2D surface tapped gestures, where pleasure *or* arousal are related to the number of strokes.

These three descriptors were computed for x, y and z projections of raw acceleration values, and for high-pass and low-pass filtered acceleration values.

2D surface gesture descriptors

For describing 2D surface gestures, we considered (1) touches and (2) strokes levels. For instance, the mean of a gesture's location can be (2) the mean of the mean for each strokes or (1) the mean of all touches, regardless of the stroke they belong to. We consider both as either one could be relevant to our purpose.

First, we considered touches' locations, touches' slopes (angle between two consecutive elementary drawn segments), touches' speeds (instantaneous speed between two consecutive elementary drawn segments). Second, we considered the number of strokes, the strokes' total length and strokes' lengths, strokes' locations, strokes' speeds and strokes' slopes. When applicable, we computed these descriptors for projections on x and y axis and also for vectors' length. Similarly to 3D motion gestures descriptors, when possible, we computed minimum, maximum, amplitude, median, mean and variance of their signed and absolute values.

In total, 249 descriptors were computed for single gestures. For both affective reports and gestures descriptors, we used raw values on the one hand, and standard Z-Scores for each subject on the other hand. Raw values for pleasure, arousal and dominance were reported between 1 and 9, 1 being respectively pleasant, excited and controlled and 9 respectively being unpleasant, calm and in control (Figure 1.b).

RESULTS

In order to be able to unveil the dependance between gestures and affective states, we first examined which of the computed descriptors is relevant to affective states and which are not. We then focused on identifying a gestures' space relevant to affective interaction. We also provide a preliminary model of arousal and qualitatively identified interesting patterns in particular classes of gestures.

Relevant and Unrelevant Descriptors

We performed a study of Pearson correlation coefficient r in order to evaluate the relationship between affective variables and descriptors of gestures. The coefficient reflects the noisiness and direction of a linear relationship, but not the slope of that relationship, nor many aspects of nonlinear relationships. A coefficient of 1 means that both variables have the same direction and no variance, -1 means that variables have opposite direction and no variance, while 0 means no relationships between them, i.e. much variance. Statistical significance of this relationship is evaluated with a Student test. We report here only the significant relationships ($|t| > 1.96$, i.e. $\alpha = 5\%$).

Pleasure

Pleasure is slightly positively correlated to:

- the median of the projection on the y axis of the acceleration ($r = 0.17$, $t = 2.42$),
- the mean of the projection on the z axis of the jerk of filtered acceleration ($r = 0.16$, $t = 2.22$).

This means that the more unpleasant the feeling is, the higher these descriptors are.

If we consider Z-score of pleasure, then the number of important frequencies found in the spectrum of the high-pass filtered value of the projection on the z axis of the acceleration is found positively correlated to ZP ($r = 0.17$, $t = 2.41$).

Arousal

We found 97 3D motion descriptors that are correlated to the arousal dimension of emotions. This very high number can be explained by the fact that arousal is defined as the strength of the emotion: the stronger the emotion is, the more expression we can expect through gestures. Out of these 97 correlations, the strongest ones are with:

- the mean of the absolute value of the x projection of the high-pass filtered acceleration ($r = -0.28$, $t = -3.97$),
- the maximum of the signed value of the x projection of the high-pass filtered acceleration ($r = -0.28$, $t = -3.90$)
- the maximum of the x projection of the derivative of the raw acceleration ($r = -0.30$, $t = -4.29$)
- the maximum of the length of the derivative of raw acceleration ($r = -0.28$, $t = -4.01$)
- the amplitude of the length of the derivative of raw acceleration ($r = -0.28$, $t = -4.02$)
- the maximum of the x projection of the derivative of the high-pass filtered acceleration ($r = -0.28$, $t = -3.93$)
- the maximum of the length of the derivative of high-pass filtered acceleration ($r = -0.31$, $t = -4.46$)
- the amplitude of the length of the derivative of high-pass filtered acceleration ($r = -0.31$, $t = -4.47$). Extreme

examples of amplitude and arousal are shown in Figure 4a.

We can add that duration of the gesture is significantly correlated to arousal ($r = 0.18$, $t = 2.56$ for A and $r = 0.20$, $t = 2.77$ for ZA). Interestingly, correlations with 2D surface gestures descriptors were also found with arousal: the projection on the x axis of the maximum of the absolute speed of the strokes is negatively correlated to arousal ($r = -0.17$, $t = -2.09$ for A and $r = -0.17$, $t = -2.11$ for ZA), and the amplitude of the slopes of strokes is correlated to ZA ($r = -0.16$, $t = -2.04$).

Dominance

Dominance is slightly positively correlated to the number of important frequencies found in the spectrum of the high-pass filtered value of the projection on the x axis of the acceleration ($r = 0.21$, $t = 2.76$). Dominance is also positively correlated to the mean of the absolute value of y projection of touches' speed ($r = 0.17$, $t = 2.11$). This means that the more the person feels in control, the higher these descriptors are.

Dominance is also negatively correlated to the gap that maximizes the difference between most important and less important frequencies in the spectrum of the projection on the x axis of the low-pass filtered acceleration signal ($r = -0.18$, $t = -2.43$) and also to the gap that maximizes the difference between most important and less important frequencies in the spectrum of the projection on the x axis of the acceleration signal ($r = -0.18$, $t = -2.46$). This means that the less the person feels in control, the higher these descriptors are.

If we consider Z-scores of dominance, then three descriptors adds up to the significant ones : The amplitude of the projection on the y axis of the strokes' locations ($r = -0.16$, $t = -2.01$), the minimum of the projection on the y axis of the strokes' locations ($r = 0.16$, $t = 1.98$), the median of the length of the acceleration vector ($r = 0.16$, $t = 2.19$).

Summary of relevant and unrelevant descriptors

In this section we presented descriptors that are significantly correlated to affective dimensions P, A and D and their standardized value per subject. Designers should consider the descriptors we reported in order to include gesture as an affective input of an application, but also avoid the descriptors that were not found correlated to any of the affective dimensions. Yet, developers should be aware that the lower $|r|$ is, the larger the predictable interval of the affective dimension: we showed that Ubicomp system designers could predict a *range* of affective values. We found a much larger number of correlations, much higher correlation coefficients and much higher significance for the arousal dimension compared to pleasure and dominance. This indicates that this dimension is the most reliable for ubicomp designer to use in their application.

As correlation coefficients are all below 0.40, little variance of affective variables are explained by single descriptors. This shows that there is no unique descriptor perfectly linked to affective states. In order to provide a usable model of affective gestures, we now explain how we gathered descriptors in new axes defining an *affective gestures' space* that explain as much variance as possible.

From Single Descriptors to a Gestures' Space

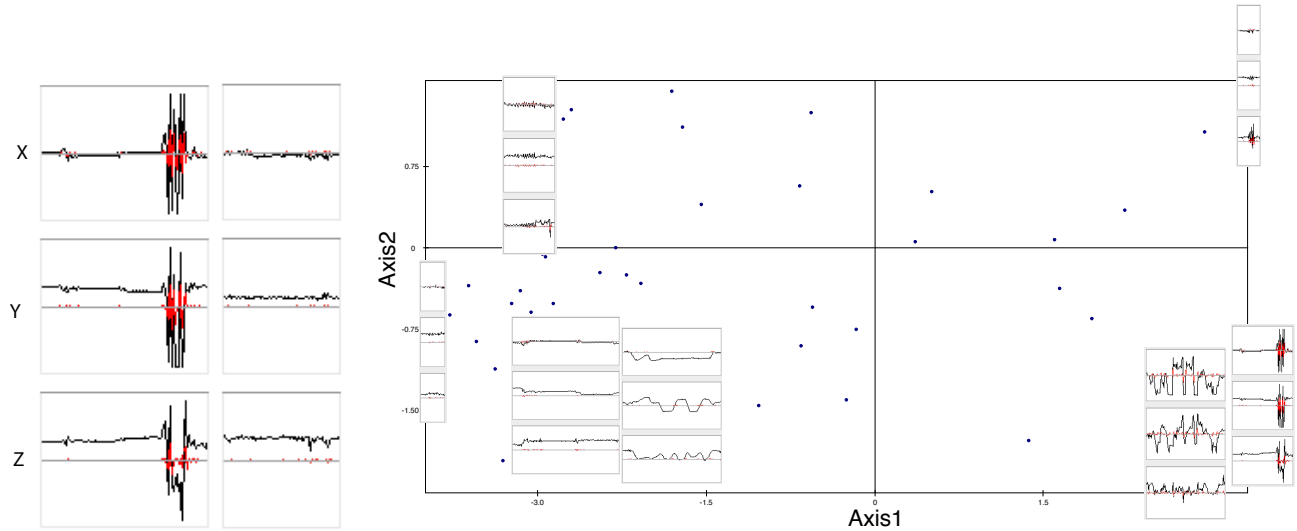
For a developer to identify an expressed emotion, considering all 249 descriptors is not practical. To reduce this number, a developer could consider only the most correlated descriptors. But, due to descriptors' correlation to each other, it could be a waste of time and/or one could miss relevant information. To find the best subset of descriptors, we contribute with an *affective gestures' space* that best describe affective gestures.

We performed a principal component analysis (PCA) in order to extract axes from the list of single descriptors. Figure 5 show the descriptors and the affective variables in the space of the new significant axes. To take part in the axes, we only considered descriptors that have the most significant correlation coefficients ($|t| > 2.50$, i.e. $\alpha = 1\%$). For each of descriptor, its 188 values are taken into account. Gestures' variables were used as active variables in PCA and PAD variables as illustrative variables (black arrows in Figure 5). We only considered the three most significant axes, i.e. with eigen values > 3. Figure 5 shows the projection of descriptors on the subspaces defined by new axes 1 and 2, and by new axes 2 and 3. The closer to unit circle a descriptor is, the more it is important to structure the 3-axes space. As one can see, the first axis (left horizontal axis of Figure 5) includes most of the information (64%), but the second axis (left vertical axis and right horizontal axis of Figure 5) includes 6% of the information and the third axis 5% (right vertical axis of Figure 5).

Figure 4b shows some of the typical 3D logs of gestures located in the subspace defined by new axes one and two, i.e. whose coordinates on these axes are higher than others gestures. Their positions are determined by a linear combination of severals descriptors. We can see on Figure 4b that similar gestures are close to each other and that different gestures are far from each other. For instance, the gestures at the bottom right corner show more 3D movement than the gestures at the top left corner. However, it is difficult to interpret in a straightforward way how such a position in the space of affective gestures can be obtained when *executing* a gesture. We now explain how developers can locate gestures in the space.

A benefit of these new axes for the identification of expressed emotions is that it gives the information on which descriptors to consider in order to locate a gesture in the space of affective gestures. Now a developer can, for instance, consider the three most important descriptors per direction of each axis, in order to better represent the space of affective gestures. Doing so, she/he would consider 18 descriptors:

- the minima of x, y and z projections of high-pass filtered acceleration, respectively contributing to -0.91, -0.90,



(a) (b)

Figure 4. (a) Extreme examples of the arousal correlated amplitude of the length of the derivative of high-pass filtered acceleration. Left: User is excited and moves the device in a very inconstant way. Right: User is calm and does not move the device. (b) 3D logs located in the subspace defined by axes one and two.

-0.89 to the first axis; and the mean of the length of the high-pass filtered acceleration vector, the mean of the length of the acceleration derivative and the amplitude of the length of the high-pass filtered acceleration, respectively contributing to 0.96, 0.96, 0.95 to the first axis;

- the minima of the z projection of the high-pass filtered and raw acceleration's derivative and the maximum of the y projection of the high-pass filtered acceleration, respectively contributing to the second axis to -0.37, -0.33, -0.32; and the variances of the z projections of the high-pass filtered acceleration, raw acceleration's derivative and high-pass filtered acceleration's derivative, respectively contributing to the second axis to 0.41, 0.40, 0.39;
- G descriptors of y projection of raw acceleration and low-pass filtered acceleration and duration, respectively contributing to -0.75, -0.75, -0.63 to the third axis; and the median of the y projection of the raw acceleration, the mean of the z projection of the absolute value of the raw acceleration and the N descriptor of the x projection of the high-pass filtered acceleration, respectively contributing to 0.46, 0.42, 0.24 to the third axis.

By doing so, a developer would be ensured to have chosen the 18 most structuring descriptors of the affective gestures' space. Another way to benefit from this space would be to consider groups of descriptors that are the most related to an affective dimensions. We find 6 subgroups of descriptors structuring the first and second axes (Figure 5, left).

- In the top left quarter, the blue $\boxtimes$ group of descriptors, mostly gathering minimum of x and y projections of acceleration and jerk;
- In the bottom left quarter, the purple $\triangledown$ group, mostly gathering minimum of z projections of acceleration and jerk;
- In the top right corner, the green $\triangle$ group, mostly gathering maximum, amplitude and variance of z projection of acceleration and jerk; the brown $+$ group, mostly gathering median of length of acceleration and jerk; the yellow $\times$ group, mostly gathering maximum, amplitude, mean and variance of length of acceleration and jerk;
- In the bottom right quarter, the red $\circ$ group, mostly gathering maximum, amplitude and variance of x and y projection of acceleration and jerk.

We find a seventh interesting subgroup structuring axes 2 and 3 (Figure 5, right): the cyan $\diamond$ group, mostly gathering duration and G descriptors on x and y projections, of either raw or low-pass filtered acceleration.

Interestingly, on Figure 5 (left) the direction of standard arousal ZA is exactly the direction of the purple $\triangledown$ group and on Figure 5 (right) exactly the direction of the cyan $\diamond$ group. This means that a significant part of the variance of ZA can be explained by these subgroups of descriptors. As a consequence, for a developer to evaluate the arousal expressed by a gesture, she/he can choose to compute these subgroups of descriptors.

Towards Identification of Arousal

We performed a linear regression in order to investigate the prediction of the emotion expressed by the user. We applied a backward stepwise method on the most significant descriptors that defined the affective gestures' space. With this method, only the combination of descriptors that maximizes the explained variance of ZA is left in the model. We report here the model for the arousal dimension, as it has the strongest relationship with gestures. The model involves, respectively, the minimum of the x projection of the high-pass

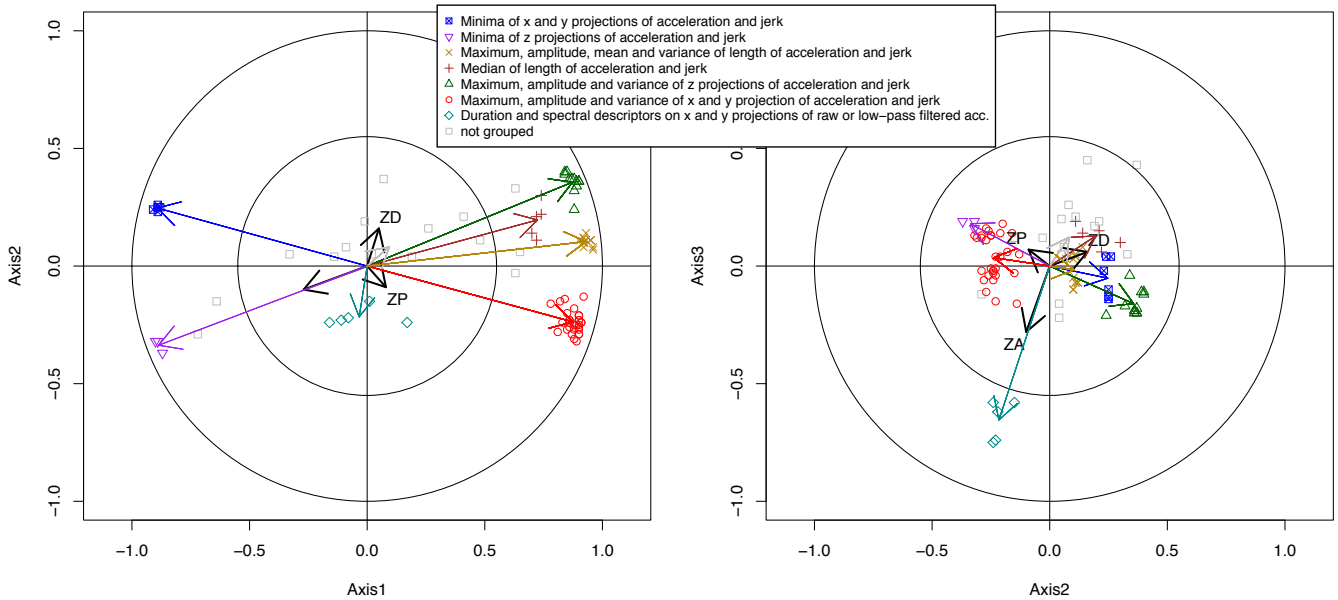


Figure 5. Projection of descriptors and affective variables on the subspaces defined by axes one and two (left) and two and three (right).

filtered acceleration, the minimum of the z projection of the high-pass filtered acceleration, the minimum of the z projection of the high-pass filtered acceleration's derivative, the duration of the gesture, the mean of the z projection of the absolute value of the acceleration and the N descriptor of the x projection of the high-pass filtered acceleration.

$$\begin{aligned} ZA = 0.33(\pm 0.214) \times FAccXMin \\ -0.62(\pm 0.476) \times FAccZMin \\ +0.56(\pm 0.428) \times FJerkZMin \\ +0.18(\pm 0.140) \times Duration \\ -0.21(\pm 0.137) \times AbsAccZMean \\ -0.14(\pm 0.140) \times NXHPF \end{aligned}$$

This model is significant (Fisher test with $F_{6,175} = 7.31$, $\alpha < 0.001$). Confidence intervals ($\alpha = 5\%$) of coefficients are indicated in the formula between parentheses. Not surprisingly, we find here a combination of the descriptors that belong both to (1) the 18 most structuring descriptors of the three axes of the affective gestures' space and (2) the groups that most relates to arousal, i.e. purple $\triangledown$ and cyan $\diamond$.

This model is a first step towards identification of arousal through gestures, the closest to reality we can present after this first study. In order to open future direction of this work, we now present specific patterns that we found and that will lead to further studies.

Promising Findings

We manually classified gestures, in order to find interesting patterns that could lead to further studies. We report here two of these findings. First, Figure 2 shows four 3D motion gestures characterized by (1) a pause where the device is hold still and then (2) more important movement *in all three* x,y,z directions. Users holds their phones still before forcefully shaking the device in all directions. In our study, these gestures can be said to express a high arousal with a 95% confidence.

The second interesting pattern was found in tapped gestures, like those of Figures 1.a' and 3. In our study, for $2/3$ of them (triangles of Figures 6), the pleasure tends to be exponentially related to the number of strokes, until the number of strokes reaches 20 where ZP reaches a ceiling. This means that the more the user taps on the screen, the less happy she/he tends to feel. For the other third of the tapped gestures that do not follow the exponential relationship (circles in Figures 6), then the arousal tends to have a negative relationship to the number of strokes: the more the user taps on the screen, the more aroused she/he tends to feel.

Finding like these ones will lead to future studies in order to specifically investigate particular types of gestures and their relationship to affective dimensions.

DISCUSSION

Limitations of our dataset

First, as seen in the second screen of Figure 1, the default values of sliders for SAM forms were 5, i.e. the middle of the scale. As a consequence, we collected very few values between 4 and 5, and between 5 and 6. We explain this by subjects not moving sliders for such a small increment. Unfortunately, it leads to a gap around 5 values in our dataset. We believe this has little impact on the results, but a more appropriate way of collecting data would have been to present scales with no default values, forcing the participant to specify the closest values of PAD.

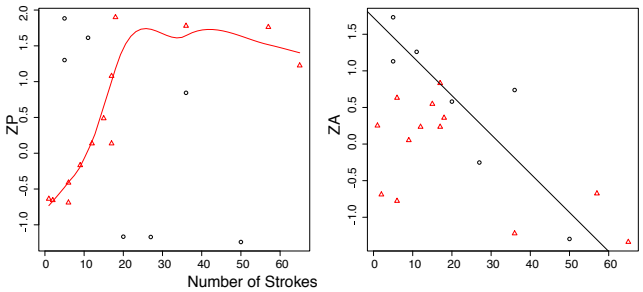


Figure 6. (Left) Relationship between the number of strokes of the tapped gestures and ZP: For triangles, the pleasure tends to be exponentially related to the numbed of strokes, until the number of strokes reaches 20 where ZP reaches a ceiling. (Right) Relationship between the number of strokes of the tapped gestures and ZA: For circles, the arousal tends to have a negative relationship to the number of strokes.

Second, several subjects reported that having ideas of gestures was sometimes difficult. To overcome this, we would now collect subjects' confidence in their gesture to express their emotion. This could allow using a confidence factor for the gesture and better balance the influence of gestures in our analysis.

The last limitation comes from the amount of samples. Ideally, the more samples, the more reliable the results. But as we had to propose an acceptable time for subjects to participate in the study, we balanced the need for a large amount of samples and for a short, lightweight study. Moreover, as this is the first study of its kind, its aim was to open areas for more targeted future work and the reasonable number of samples we collected allowed us to identify promising areas, e.g., for arousal identification or for particular types of gestures.

Implications and Applications

The obtained results are preliminary but have a number of perspectives to be *improved*, *refined* and *used*. Our results could be *improved* by conducting two studies. First, we could deploy a version of the data collection software that conducts modeling in the background thanks to the preliminary model provided by this paper and prompts the user to refine the system's proposal or to perform again when it detects a less clear affective gesture. Second, a study could be run in order to collect more samples to improve results and use methods, like classification, that demand a larger number of samples. Towards this aim, we could either (a) run a study in a lab, eliciting emotions from users and asking them to perform a gesture after each stimulus, or preferably for ecological validity, (b) run a longer term study in the field.

Our results could be *refined* in four directions: First, we could focus on particular types of gestures like the tapped ones of Figure 3 and investigate how different ways of performing a single type of gesture can relate to emotions. Second, we could focus on single subjects and investigate if the identification of emotion can be improved *for a single subject*. The limited agreement between subject can be an explanation of the low correlations between gestures and reported emotions. For instance, subject 9 and 12 show both strong, but opposite correlations between arousal and the amplitude of the length of the derivative of high-pass filtered acceleration. For similar descriptors, we could have better results when considering subjects independently. To do so, we would need more samples of specific subjects, as an average of 15 samples per subject is not enough.

Third, we could focus on introducing further low-cost, not intrusive modalities for the identification of emotions explicitly expressed by mobile users. A subject of our study proposed pressure as another modality he would like to use while mobile to express emotions. Gyroscopes could also be investigated, as they are now becoming more prevalent in mobile device and that palm orientation has already been found correlated to emotion categories for one actor in one play and motion direction was correlated to pleasure for two actors playing the same role [12]. Last, we could investigate if the gestural modality for expressing emotions is more efficient, expressive or pleasurable for users than, e.g., typing emoticons.

Our results could be *used* as an inspiration for future design of a gesture-based affective UI based on real-world data, as in [24]. The results of our study can lead to, e.g., a set of 4 gestures for each corner of the PA space. Such a set could then be evaluated against the existing, arbitrary one [6]. Another perspective for the use of our results is to integrate the gesture modality in an application in order to evaluate its benefits for mobile applications. During the study, participants envisioned a lot of possible applications for such an affective, gestural and mobile input. For instance, subjects envisioned self-monitoring of emotional states in order to control oneself. Gestures could be used to precisely specify an emotion in a communication application, instead of typing ambiguous, too narrow and incomplete emoticons. Earlier examples of using PAD values in applications can be found, e.g. with discretized and textual feedback [4] or continuous and graphical feedback [22].

CONCLUSION

In this paper we presented the results of a field study aiming at enabling the identification of explicitly expressed emotions of mobile users through 2D surface and 3D motion gestures. Among 249 descriptors of each collected gesture, we found that 104 of them were significantly correlated to at least one of the affective dimensions reported by subjects. More precisely, a high number of 3D motion gesture descriptors were correlated to affective dimensions and the arousal dimension of emotions was correlated to a high number of descriptors of gestures. As correlations were lower than 0.40, we defined an *affective gestures' space*, gathering descriptors along more significant axes. We identified groups of descriptors that structure this space. Two of them are related to arousal : duration and G descriptors on x and y projections of either raw or low-pass filtered acceleration on the one hand, and minima of z projections of acceleration and jerk on the other hand. Finally, we provided with a preliminary model of arousal and we showed that qualitative analysis of the data allowed us to identify interesting

patterns in particular classes of gestures.

This work presents the first naturally-gathered affective gestures dataset and the first step towards unobtrusive and inexpensive identification of the link between 2D surface/3D motion gestures and affective states of a mobile user. This work is a broad, exploratory research in a new field. Before this study, as we show in the first section, no other approach tackled the feasibility of identifying emotion through 3D motion and 2D surface gestures. This work is important as it moves us closer to creating emotionally-aware ubiquitous computing that can be widely deployed.

ACKNOWLEDGMENTS

We would like to thank the participants of the study and Thomas Milcent, Kai Kuikkaniemi and the LIG EHCI group for their helpful comments.

REFERENCES

1. L. Anthony, P. Carrington, P. Chu, C. Kidd, J. Lai, and A. Sears. Gesture dynamics: Features sensitive to task difficulty and correlated with physiological sensors. In *MMCogEmS Workshop, ICMI'11*, 2011.

2. G. Caridakis and K. Karpouzis. Full body expressivity analysis in 3d natural interaction: a comparative study. In *AFFINE Workshop, ICMI'11*, 2011.

3. K. Church, E. Hoggan, and N. Oliver. A study of mobile mood awareness and communication through mobimood. In *Proc. of NordiCHI'10*, pages 128–137. ACM, 2010.

4. C. Coutrix, G. Jacucci, I. Advouevski, V. Vervondel, M. Cavazza, S. W. Gilroy, and L. Parisi. Supporting multi-user participation with affective multimodal fusion. In *Proc. of C5'11*. IEEE, 2011.

5. C. Epp, M. Lippold, and R. L. Mandryk. Identifying emotional states using keystroke dynamics. In *Proc. of CHI'11*, pages 715–724. ACM, 2011.

6. P. Fagerberg, A. Ståhl, and K. Höök. emoto: emotionally engaging interaction. *Personal and Ubiquitous Computing*, 8:377–381, 2004.

7. S. Fdili Alaoui, B. Caramiaux, and M. Serrano. From dance to touch: movement qualities for interaction design. In *Proc. CHI EA'11*, pages 1465–1470. ACM, 2011.

8. A. Gluhak, M. Presser, L. Zhu, S. Esfandiyari, and S. Kupschick. Towards mood based mobile services and applications. In *Proc. EuroSSC'07*, pages 159–174. Springer-Verlag, 2007.

9. J. Heikkinen, J. Rantala, T. Olsson, R. Raisamo, J. Lylykangas, J. Raisamo, V. Surakka, and T. Ahmaniemi. Enhancing personal communication with spatial haptics: Two scenario-based experiments on gestural interaction. *Journal of Visual Languages and Computing*, 20(5):287 – 304, 2009.

10. J. Hektner, J. Schmidt, and M. Csikszentmihalyi. *Experience Sampling Method: Measuring the Quality of Everyday Life*. Sage Publications, 2007.

11. K. Hinckley and H. Song. Sensor synaesthesia: touch in motion, and motion in touch. In *Proc. of CHI'11*, pages 801–810. ACM, 2011.

12. M. Kipp and J.-C. Martin. Gesture and emotion: Can basic gestural form features discriminate emotions? In *ACII'09*, pages 1 –8, 2009.

13. P. Lang. Behavioral treatment and bio-behavioral assessment: computer applications. *Technology in mental health care delivery systems*, pages 119–137, 1980.

14. A. Mehrabian. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14:261–292, 1996.

15. M. Meijer. The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior*, 13:247–268, 1989.

16. R. W. Picard. *Affective Computing*. MIT Press, 2000.

17. S. Pook, E. Lecolinet, G. Vaysseix, and E. Barillot. Control menus: excecution and control in a single interactor. In *CHI EA '00*, pages 263–264. ACM, 2000.

18. C. Randell, I. Anderson, H. Muller, A. Moore, P. Brock, and S. Baurley. The sensor sleeve: Sensing affective gestures. In *On-Body Sensing Workshop, ISWC'05*, 2005.

19. J. Rantala, R. Raisamo, J. Lylykangas, T. Ahmaniemi, J. Raisamo, J. Rantala, K. Makela, K. Salminen, and V. Surakka. The role of gesture types and spatial feedback in haptic communication. *IEEE Transactions on Haptics*, 4(4):295 –306, 2011.

20. J. Rekimoto. Smartskin: an infrastructure for freehand manipulation on interactive surfaces. In *Proc. of CHI'02*, pages 113–120. ACM, 2002.

21. J. Ruiz, Y. Li, and E. Lank. User-defined motion gestures for mobile interaction. In *Proc. of CHI'11*, pages 197–206. ACM, 2011.

22. P. Sundström, A. Ståhl, and K. Höök. In situ informants exploring an emotional mobile messaging system in their everyday practice. *Int. J. Hum.-Comput. Stud.*, 65:388–403, 2007.

23. H. Wang, H. Prendinger, and T. Igarashi. Communicating emotions in online chat using physiological sensors and animated text. In *CHI EA '04*, pages 1171–1174. ACM, 2004.

24. J. O. Wobbrock, M. R. Morris, and A. D. Wilson. User-defined gestures for surface computing. In *Proc. of CHI'09*, pages 1083–1092. ACM, 2009.