Accepted Manuscript

A mobile application to report and detect 3D body emotional poses

Iván García-Magariño, Eva Cerezo, Inmaculada Plaza, Luca Chittaro

PII: S0957-4174(19)30021-1
DOI: https://doi.org/10.1016/j.eswa.2019.01.021
Reference: ESWA 12422

To appear in: *Expert Systems With Applications*

Received date: 20 January 2018
Revised date: 10 December 2018
Accepted date: 6 January 2019

Please cite this article as: Iván García-Magariño, Eva Cerezo, Inmaculada Plaza, Luca Chittaro, A mobile application to report and detect 3D body emotional poses, *Expert Systems With Applications* (2019), doi: https://doi.org/10.1016/j.eswa.2019.01.021

Highlights

- A novel app allows users to report 3D body emotional poses
- The app classifies poses into the six Ekmans emotional states and a neutral one
- The app used a set of emotional poses from the literature as prototypes
- We conducted a user study to assess the app
- Five out of the six emotional states were identified with accuracies above 70

A mobile application to report and detect 3D body emotional poses

Iván García-Magariño[a,c,*], Eva Cerezo[b], Inmaculada Plaza[d,c], Luca Chittaro[e]

[a]*Department of Software Engineering and Artificial Intelligence, Complutense University of Madrid, Spain*
[b]*Department of Computer Science and Systems Engineering, Instituto de Investigación en Ingeniería de Aragon (I3A), University of Zaragoza, Spain*
[c]*Instituto de Investigación Sanitaria Aragón. University of Zaragoza,Zaragoza, Spain*
[d]*Department of Electronics Engineering and Communications, University of Zaragoza, Spain*
[e]*Human-Computer Interaction Lab, Department of Mathematics, Computer Science, and Physics, University of Udine, Italy*

Abstract

Most research into automatic emotion recognition is focused on facial expressions or physiological signals, while the exploitation of body postures has scarcely been explored, although they can be useful for emotion detection. This paper first explores a mechanism for self-reporting body postures with a novel easy-to-use mobile application called EmoPose. The app detects emotional states from self-reported poses, classifying them into the six basic emotions proposed by Ekman and a neutral state. The poses identified by Schindler et al. have been used as a reference and the nearest neighbor algorithm used for the classification of poses. Finally, the accuracy in detecting emotions has been assessed by means of poses reported by a sample of users.

Keywords: affective computing, app, emotion detection, mobile application, pose detection, expert system

*Corresponding author, Telephone: +34 913947643, Fax: +34 913947547
Email addresses: `igarciam@ucm.es` (Iván García-Magariño), `ecerezo@unizar.es` (Eva Cerezo), `inmap@unizar.es` (Inmaculada Plaza), `luca.chittaro@uniud.it` (Luca Chittaro)

1. Introduction

Automatic emotion detection is becoming the cornerstone for natural man-machine interaction (Alonso-Martín et al., 2013) and is normally integrated into human-computer multimodal communication systems.

Many authors have studied the expression and detection of emotions (see (Zeng et al., 2009) for a review). Most visual emotion detection methods focus on facial expressions (Gunes and Hung, 2016). However, emotions can also be expressed and perceived through body language (De Gelder, 2006).

Emotional body language can be analyzed considering the motion of expressions or some static views of the corresponding body poses. Saneiro et al. (2014) developed a system for tagging body movements with emotions providing information which can be used with data mining techniques. In addition, Garber-Barron and Si (2012) found that body postures after changes were more representative for the automatic detection of emotions than still body poses. Some authors have proposed automatic techniques for classifying two-dimensional (2D) static images into a set of emotional states (Schindler et al., 2008; De Silva and Bianchi-Berthouze, 2004), starting a challenging research line in the field of affective computing.

Existing automatic emotion detection mechanisms from body poses mainly use computer vision techniques in which relevant information is extracted from images (see an example in (De Silva and Bianchi-Berthouze, 2004)) or videos (such as (Garber-Barron and Si, 2012)). This visual information is normally taken from either independent cameras (De Silva and Bianchi-Berthouze, 2004) or cameras integrated into consoles such as the Microsoft Xbox (Rázuri et al., 2015). Existing classification methods are then normally applied to detect the emotion from the information extracted. Examples of applied methods are the k-nearest neighbor (KNN) algorithm, the support vector machine (SVM), and neural networks (see (Rázuri et al., 2015) for a comparison of these classification methods).

Automatic emotion detection from body poses is harder than facial emotion detection, since the configuration of the human body has more degrees of freedom than the face (Schindler et al., 2008). There are many features that compose the human pose. For instance, Kollias and Karpouzis (2005) showed that hand gestures play a key role in the detection of emotions, based on an analysis of several works.

Recent work in affective computing increasingly relies on mobile technologies, as described in the survey of Politou et al. (2017).

Nevertheless, to the best of our knowledge, no reported work has used an avatar for allowing users to freely represent body poses by dragging their parts (i.e. without giving a restricted set of poses) for automatically detecting emotions from users.

In this context, we propose a novel mobile application called EmoPose as a mechanism for reporting three-dimensional (3D) body poses and detecting emotions from them. The main purpose of the app is to be useful for users who have difficulties in being aware of their emotions and how they are expressed by their poses. Recent research into Human-Computer Interaction has demonstrated the fundamental importance of the somatic relationship (in our case, supporting the user in establishing a connection between his/her lived experience of the body and the avatar that displays it, and then the associated emotion) in computer-based tools and techniques that aim at users' self-improvement in wellbeing (Loke and Schiphorst, 2018). More generally, a growing amount of work in the psychological and medical literature suggests that heightened body awareness can be beneficial in a number of conditions that affect people's mood such as, for example, post-traumatic stress disorder; see Mehling et al. (2009) for a review.

For these reasons, instead of adopting automatic systems that detect the user's body pose and automatically apply it to the avatar on behalf of the user, our app has been explicitly designed with a touch-screen interface that actively involves the user in reflecting about his/her body poses (which includes feeling them) and then reinforces such perceptions in detail by expressing the pose through a first-person representation process supported by the interface. From this perspective, an automatic system would have diminished the somatic approach, which is central to the purpose of the app.

2. Related work

2.1. Relation between body postures and emotions

Several works investigate the relation between body postures and emotions. For example, Rosario et al. (2010) examined this relationship considering emotions of sadness, happiness, concern and anxiety. A linear regression analysis showed significant correlations between raised shoulders and happiness. They also found a significant relation between shoulder extension and both sadness and concern.

Moreover, adopting certain body poses can influence or provoke certain emotions. For instance, Rossberg-Gempton and Poole (1993) highlighted the effects of open and closed postures on pleasant and unpleasant emotions, respectively. Hao et al. (2017) showed that the creativity of people was generally higher when their actual emotions were compatible with their body poses.

Mondloch (2012) performed an experiment analyzing whether participants properly distinguished emotions in both facial expressions and body postures. They presented several combinations, each with a facial expression and a body posture taken from two different pictures. Sometimes the facial expression emotion was different from the emotion expressed by body posture. Her work showed that human subjects (both adults and 8-year- old children) were able to properly distinguish an emotion from a body posture without seeing the face.

Regarding the bodily expressions of emotions in general, Gross et al. (2012) showed the relation of certain emotions with the kinematics of walking. In particular, the velocity of walking was relevant. For example, fast walking was usually related with joy and anger, while a slow speed was commonly associated with sadness.

In general, the relation between body poses and emotions motivates research into the detection of emotions from body poses in the field of affective computing.

2.2. Automatic detection of emotions from pose images

Several works propose automatic emotion detection mechanisms for the analysis of body poses from images. These works mainly apply computer vision techniques and classification methods for this purpose.

De Silva and Bianchi-Berthouze (2004) proposed some affective posture predictive models that were able to distinguish between four basic emotional states: anger, fear, happiness and sadness. Their approach was based on the affective gestures of 13 human actors. They used 8 cameras to obtain the most representative posture of each gesture. They then collected 109 gestures, and measured 32 markers in certain joints and body segments. The Discriminant Analysis statistical technique was used to properly distinguish the postures of different emotions.

Schindler et al. (2008) also analyzed body poses for detecting emotions using a neural network approach for this detection. The dataset contained static images associated with certain body poses. In particular, the dataset used six basic emotional states proposed by Ekman (1992b) (i.e. anger, disgust, fear, happiness, sadness and surprise) and the neutral state. Images were taken of 50 actors enacting the different emotions. In this study, facial expression was considered as a part of the body posture. The face was partially covered (e.g. with the hair or hands) in a large number of cases, and the number of pixels representing the face was actually low. It was argued that masking out the face would have had the risk of distorting the image pose. Principal Component Analysis (PCA) was used for obtaining some high-level features. A SVM approach was applied for classifying the high-level features into an emotional category. The Tilburg University stimulus set of emotional body poses was used in the experimental stage. This approach achieved an 82% recognition rate, compared to the 87% of human test subjects. Since the input was pictures, the authors argued that their emotion detection used 2D data.

Rázuri et al. (2015) used the sensors of a Microsoft Xbox 360 for extracting relevant information from user body postures in order to analyze emotional body language. Specifically, SVM classification was used for detecting emotions from body poses, which outperformed other classification methods such as multilayer perceptron (MLP) neural networks or the KNN algorithm.

Garber-Barron and Si (2012) analyzed body poses in non-acted scenarios of participants playing video games in which players felt triumph, frustration, defeat and concentration. Their analysis highlighted the fact changes in body posture were especially representative of participants emotions. These changes included limb rotation movements, posture movement groups and joint rotations, and were especially significant when the movements were sudden and rapid.

Kollias and Karpouzis (2005) reviewed several works addressing the key features in automatic emotion detection from body poses. Their analysis highlighted the importance of hand gestures in the detection of emotions.

Nevertheless, to the best of our knowledge, there is no reported work concerning the detection of emotions from 3D self-reported poses.

2.3. Body postures for expressing emotions with virtual humans

Virtual humans have played a key role in several analyses of emotional body language. This line of research can be classified into two main areas: (a) the integration of body language into virtual humans for multi-modal affective communication, and (b) the use of virtual agents as prototype models for establishing relations between body poses and emotions for advancing the automatic emotion detection field.

The first group of works belongs to the field of Human-Computer Interaction (HCI)in which virtual agents adopt different body poses for providing multi-modal communication from machine to human with affective content. Within this field can be found Maxine (Baldassarri et al., 2008), an animation engine for including embodied animated agents in the development of applications. This engine allows the creation of agents with multimodal emotional communication, including body postures among others (Baldassarri and Cerezo, 2012). The engine can communicate Ekmans six basic emotions and a neutral state. In addition, virtual human poses have represented emotions with different levels of intimacy (Sadowski and Lomanowska, 2018). In this study, physical contact among avatars was used for representing intimacy in human interactions. The corresponding experiments took place in an online virtual chat room platform where users could select animated poses.

The second group of works uses virtual humans in analyses of relations between body poses and emotions in different contexts. For instance, Kleinsmith et al. (2006) analyzed the relation between emotions and body poses in different cultures by means of embodied virtual agents. Their dataset comprised groups of static posture images of embodied virtual agents expressing anger, fear, happiness and sadness. Different 3D body posture models were obtained from experiments with subjects from three different cultures. In a broad study, Clavel et al. (2009) used virtual characters to analyze the perception of emotions considering both facial and body expressions and the relation between these. They concluded that human subjects had a clearer understanding of emotional states when the facial and body expressions were congruent. Furthermore, Marschner et al. (2015) studied the impact of the gaze of virtual humans on emotional communication as a specific feature of body language that affects body poses. Their analysis concluded that the direction of the body and head of virtual humans were also significant in the expression of emotions.

Nonetheless, in all these works the body poses were entered by the animator and not by the application users or participants. In fact, representing emotions through virtual characters is quite a challenging task. Professional animators spend appreciable amounts of time achieving this, sometimes resorting to motion capture techniques. Therefore, the idea of enabling users with no animation skills to accurately express emotions through an avatar's body is a challenging and quite unexplored field.

3. An app to report and classify emotional poses with a 3D avatar

The goal of the app is to explore new computer-aided and interactive mechanisms of reporting and detecting emotions from 3D poses. This app is based on a 3D avatar that a user with no animation skills can easily manage using touchable devices. In particular, the user interface is designed to allow users to change the aspects of poses that are normally related with emotions, such as hand gestures, the look-at direction influencing the neck and shoulder positions, and generally the relative positions of limbs. This app can be useful for users that have difficulties in being aware of their emotions such as people with autism or troubled teenagers.

Section 3.1 introduces the user-centered process for designing the app interface. Section 3.2 presents the resulting user interface for representing poses with the 3D avatar. Section 3.3 describes the novel mechanism for detecting emotions from the 3D emotional poses represented with the app.

3.1. Designing the app interface for representing poses

The challenge was to design an easy-to-use interface so that users with no previous training could report body poses through a 3D avatar. Therefore, we decided to apply a user-centered design methodology. Firstly, we developed an initial prototype and carried out a pilot study with three users who were not involved in the design of the app and had HCI expertise. We asked these users to represent the 21 emotional poses from the work of Schindler et al. (2008) with the app. We collected comments and suggestions for improvements from the three users, implemented them and then checked again with each user to check whether the changes were satisfactory.

During the design process, we considered three different ways of managing the pose of the avatar which were tested with the following prototypes:

- *Touchpads for controlling joints*: This prototype had six touchpads for controlling the muscles of the avatar attached to certain joints. Each touchpad controlled either one or two groups of muscles, and each of these was controlled along one axis (vertical or horizontal). The upper touchpad inclined the head up or down on the vertical axis, and rotated the head to the right or left on the horizontal axis. The upper corner touchpads controlled respectively the forward/backward movement of the shoulders. The lateral touchpads controlled the bending of the forearms on the horizontal axis and the up/down movement of the shoulders on the vertical axis. The lower touchpads opened and closed the hands. The touchpads were fixed to certain positions.

- *Draggable distant controls*: This prototype used movable controls attached to the body limbs. The user could drag these movable controls which moved the body limbs. In particular, the user could move each hand or foot, and the bone positions were calculated by inverse kinematics. Each control was separated from the controlled body part. This interface was inspired by the avatar control mechanism proposed in (Chittaro, 2016).

- *Dragging hands, feet and the "look-at" position*: In this prototype, the user can directly drag the hands, the feet and the look-at position. The draggable look-at control was placed in the head, although the look-at point could be moved to any position on the screen. All the muscles of the avatar were positioned by inverse kinematics considering the target points for the hands, feet and look-at position. The draggable controls were distinguished by a light-gray rectangular shadow. When the user pointed down to any control area, the corresponding rectangular shadow darkened with respect to the others so that the user had feedback of his/her action.

A consensus emerged that the third prototype was the most intuitive and easy to use. Figure 1 shows this initial interface of the app, which was later improved during the user-centered design process.

After this, we incorporated other details. One of the most relevant was the inclusion of colors for distinguishing and selecting different body parts in rotation when these overlapped. We added the possibility of rotating the avatar to obtain three different views (left, frontal, right). The dragging of the body parts was in the plan perpendicular to the camera. The transitions between views were made smoother to improve the user experience. By using the different views, the avatar could put the hands and feet on any possible 3D point as long as it was not too far from its body or inside it. We also added the possibility of dragging the joints (i.e. knees and elbows). Since this feature was not useful in all cases, we included a menu for enabling/disabling it. In the side views, the movement of the spin was made more natural when dragging the head. The feet were automatically horizontal when they were on the floor.

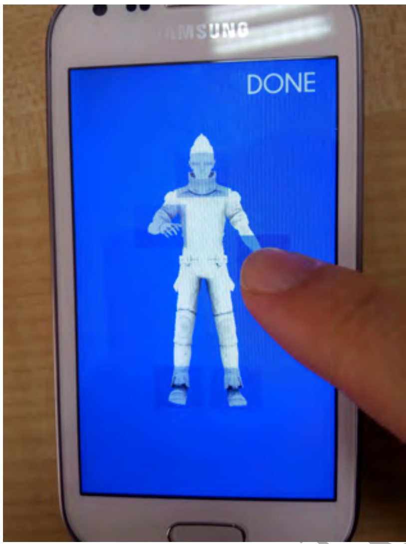


Figure 1: App interface in the initial stage of the user-centered design process

We added the possibility of rotating the hands and opening/closing them, which were considered essential for representing some poses such as those related to anger, for example clenched fists. Furthermore, we added some tutorials so that the user could understand how to use the app.

3.2. User interface

Figure 2 shows some screenshots of the user interface (UI) of the EmoPose app, displaying the representation of poses from the front view. The first time the user enters this function, the app shows the tutorial illustrated in Figure 2(a). In this tutorial, users can drag different body parts of the avatar. Basically, they can (a) drag hands and feet, (b) determine the position of the head by dragging it in the direction of view (e.g. up, down or to either side), (c) change from the front view to a side view, and (d) focus on the hands to adopt specific features.

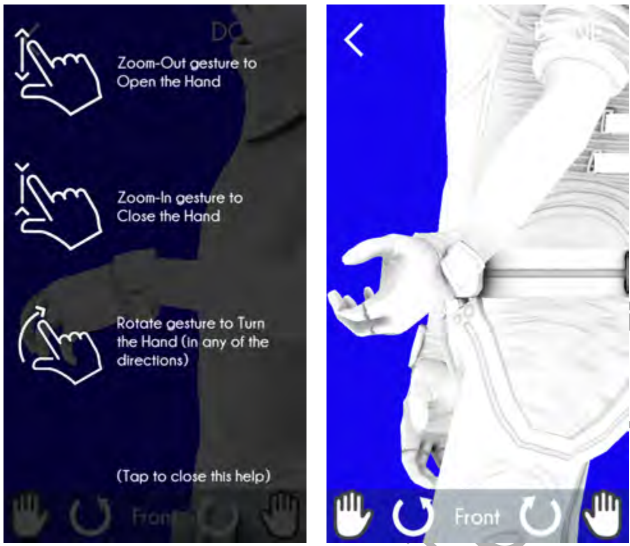
Figure 2(b) shows an example in which a user is dragging the left hand. A surrounding square highlights the body part that is being dragged. The control areas for dragging body parts are represented with light-colored semi-transparent squares. In some cases, two body parts can overlap either in the front view or the side view. In these cases, users can touch the body parts, and the selected one is visually shown by highlighting a square of the same

(a) Main tutorial for representing poses (b) Front view of EmoPose app for controlling poses

Figure 2: Representation of poses from the front view

color as the control square but opaque. In this way, users know which body part they are dragging. If users wanted to drag another of the overlapped body parts, they can just touch again, and then the app selects another body part in rotation. In this manner, users can represent poses in which several body parts act together. For example, in some poses expressing fear the avatar can hide its face with both hands.

Figure 3 shows the screens of the app displaying the representation of hand postures. Users can enter this view by pressing any of the two buttons with hand icons on the bottom bar. Then, a smooth camera transition makes the camera focus on the corresponding hand. The first time users enter this view, the app shows the tutorial illustrated in Figure 3(a). Users can perform several gestures to manage the hand posture. They can either open the hand with a zoom-out gesture or close it with a zoom-in gesture. The app uses a multi-touch interaction mechanism. If users bring the two touching fingers closer, the avatar hand closes. Otherwise, if they move one touching finger away from the other, the avatar hand opens. Users can represent any intermediate position between an open hand and a closed one. Users can also rotate the hand in the two different directions by rotating two touching



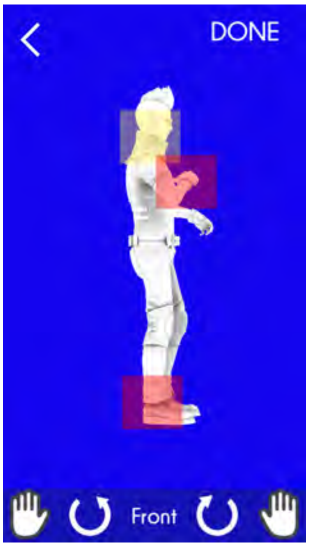
(a) Tutorial for determining the hand poses (b) Hand view for controlling the hand postures

Figure 3: Self-report of hand postures

fingers. Figure 3(b) shows an example in which the app is focusing on the posture of a hand of the avatar rotated by the user.

The EmoPose app can also control the avatar pose from any of the two side views. Figure 4(a) shows an example in which a user has entered a side view by touching one of the circled arrows on the bottom bar. The app allows users to see the avatar from either side view in order to better observe the actual pose in 3D. Even more importantly, the interaction with the side views allows users to truly represent the avatar pose in 3D since, for example, they can move the hands and feet in front of its body or behind. Only the front view cannot bring the hands closer to the front camera or away from it. However, this is possible by dragging the body parts from a side view.

The side view allows users to better represent poses by easily considering the third dimension in the direction of the front camera. Without the side view, the user could only control 3D points with a 2D screen, and consequently the depth of each point would be very hard to freely determine. However, the side view of the app allows users to turn the avatar in order to represent this dimension of depth. This is an extra view meant to make it easier for the user to understand the pose and also define it more conve-



(a) Side view for reporting poses

(b) Options menu

Figure 4: Options of EmoPose app

niently. In the feature extraction, the front view allows users to determine the x and y components of vectors, while the side view allows users to determine the y and z components of vectors. Thus, the side view is essential to represent the z component (i.e. the depth) of every vector, for determining respectively the positions of the hands, feet, and look-at point.

The app also includes the possibility of dragging more body parts than just the hands, feet and head. It allows users to optionally drag some joints. Figure 4(b) shows the options menu in which users can enable/disable the dragging of the knees and/or elbows. In the pilot study, no users' consensus emerged about what was the more intuitive option between the circular arrows for representing the rotation of the view in one direction or the other. The app thus includes an option called "natural rotation" to either enable one spin direction convention or the other.

3.3. Classification of the poses represented with the app

In order to classify poses into emotions, we selected the poses validated by Schindler et al. (2008). We chose this set of poses as it includes Ekmans six basic emotions and a neutral state, unlike other works that only consider smaller groups of emotions (De Silva and Bianchi-Berthouze, 2004; Kleinsmith et al., 2006). In addition, Schindler et al. (2008) introduced the poses by using images with higher resolution than other similar works (De Silva and Bianchi-Berthouze, 2004; Kleinsmith et al., 2006). The selected poses were represented with the app, and the corresponding relevant information of the poses was stored in it.

For the classification, the app extracts the relevant information from each user-defined pose. This information includes the positions of the hands, the feet and the head. It also contains the look-at direction as a normalized vector. This vector is obtained by subtracting the head position from the look-at position and then normalizing the resulting vector. It also contains a value for each hand in the [0, 1] interval that represents the opening state of the hands, where zero represents the completely open hand and one the completely closed hand. This information is compared with that from each of the aforementioned selected prototype poses.

The distance between two poses is considered by a weighted sum of the distances of each feature. Some distances are calculated from the vectors that respectively represent the position of each hand, each foot and the head, and the direction in which the avatar is looking. The distance of each feature is obtained by calculating the magnitude of the subtraction between each pair of vectors. In the case of the open/closed state of each hand, the distance was obtained with the absolute difference between the corresponding values.

In the calculation of the distance of poses, the current approach also considers symmetric poses along the vertical axis. This is supported by the previous literature concerning emotional detection from poses (Schindler et al., 2008), which assumed this kind of symmetry for categorizing poses into emotions and validating them. In the current approach, the final distance between two poses is calculated as the minimum of (a) the distance between the two poses and (b) the distance between one pose and the symmetric pose of the other.

In order to calculate the information of the symmetric pose, the positions of the hands are transformed into their symmetric vectors considering the vertical axis and are interchanged between each other. The same is done for the feet. The head and look-at direction are simply transformed into the symmetric vectors. The values that represent the opening of the hands are switched.

To classify a user-defined pose into an emotion, the app applies the nearest neighbor (NN) algorithm (Cover and Hart, 1967). The selected validated

Figure 5: Result of the pose detector app

poses are stored in a pose base associated with their corresponding emotions. When the user wants to detect the emotion from a pose, this is compared with all the poses in the pose base. The app selects the pose case with the shortest distance from the user-defined pose, and provides the associated emotion as the classification result to the user. The app provides the classifier output on a screen, such as the example shown in Figure 5.

Figure 6 shows an example of a user adopting a pose that represents sadness, to illustrate the use of the app with poses different from those used for training the app (i.e. those taken from Schindler et al. (2008)). The figure shows two pictures of this pose taken from two different angles (i.e. a front view and a side view). Using two pictures made it easy to represent the pose in the app with the two views.

4. User study and case report

EmoPose was assessed with a user study to determine its capability of representing emotional poses and properly detecting the emotions from them. This study also measures the time taken by users to represent emotional poses. Furthermore, a case report on emotion induction and the detection of emotions in new poses was carried out.

(a) Picture for the front view (b) Picture for the side view

Figure 6: A user adopting a pose of sadness

4.1. Participants

Two independent experimenters, not involved in the design or development of the app, recruited participants through convenience sampling. The inclusive criteria indicated that the participants should be familiar with using a smartphone.

The study was performed with 30 participants (15 males and 15 females). Their average age was 27.8 (SD = 9.6). Only 3 participants were computer science students or professionals.

4.2. Measures

To assess the capability of the app to represent the poses of Schindler et al. (2008) and recognize emotions from these, the experimenter asked each user to represent four of the poses explicitly included in that work with the app, following the order indicated in the next section. We measured the accuracy of the app in detecting the right emotions according to the associations proposed by Schindler. This accuracy was calculated as the percentage of correct classifications divided by all the classifications made by the app. We also measured the corresponding Cohen's Kappa coefficient for obtaining a measurement that was not affected by the null error rate. These same measurements were also calculated independently for each emotion.

4.3. Procedure

The experimenter wrote down the date and times of the beginning and ending of the test for each participant. In this manner, we were able to corroborate the data annotated by the experimenter with the information collected in our database. Notice that the app sends information to the database about the poses entered by the user (i.e. 3D positions and directions of draggable body parts), the result of the emotion classification, the feedback of the user determining whether the outputted emotion properly matched the pose, the distance to the nearest case in the database of poses, and the time the user spent in representing the pose. The session was videorecorded, and the camera was pointed in such a way not to include user's face for privacy reasons.

The evaluator asked each user to look at the postures proposed by Schindler et al. (2008), by saying "Please look at this pose, noticing the positions of hands, feet and head. Also notice whether the hands are open or closed, and how much they are rotated". The evaluator pointed to each photo. At the beginning, the app showed the tutorials to each user when entering each functionality for the first time. The evaluator asked the user to learn how to use the app from the tutorials. Each user represented four poses. The first user represented the first four poses. The second user represented the next four poses. Each user continued in this order representing four poses each. When the poses were finished, the evaluator started again from the beginning. For each pose, the evaluator followed the steps below:

- The evaluator asked the user to represent the corresponding pose, as closely as possible to the photo indicated by the evaluator.
- Users could use the options they preferred, for example the possibility of dragging knees and/or elbows.
- The evaluator asked users to press the "Done" button when they had finished the representation.
- The evaluator took a snapshot showing the pose and the estimated emotion.

- The evaluator checked whether the emotion recognized by the app matched the emotion associated with the photo from the article by Schindler et al. (2008). The evaluator pressed “Right” if the emotion matched, or Wrong otherwise.

- In case of pressing Wrong, the experimenter selected the correct emotion from a list of emotions provided by the app for the current experiment.

- Then, the evaluator took another snapshot showing the list of emotions with the correct one selected.

The experimenter told the participant that (s)he could make any comment about the app during the experiment. The experimenter collected all the comments.

Finally, the experimenter interviewed the participants about the ease of learning, the main difficulties, the enjoyment, and the utility of the app. The experimenter asked them to provide suggestions or comments for improving the app so that we would be able to detect new design opportunities.

4.4. Exploratory self-induced emotions case report

Given the risks involved, we ran an experiment of inducing emotions with one volunteer. This volunteer was fully informed about all the risks and she gave her consent, indicating that she was fully aware of the risks and she freely took them. She was 37 years old, had two children, and was educated to the level of vocational training within the Spanish educational system.

In this case report, the experimenter asked the participant to adopt the appropriate poses in the hypothetical situations in which: (1) one of her children hit her other child hard, provoking anger, (2) she suddenly found a dead animal in her office, showing disgust, (3) she saw an actual hungry lion, expressing fear, (4) she received an outstanding mark for one of her works done with a lot of effort, showing happiness (5) her husband left her, expressing sadness, and (6) she suddenly found herself at an unexpected surprise party for her, showing surprise. For each situation, we asked her which emotion she felt and took a picture of her pose. We then asked her to represent her pose with EmoPose app and recorded the emotion indicated by the app. These poses were completely new to the app, since the app had been trained only with the poses contained in Schindler et al. (2008). In case of any emotion not properly induced, the experimenter asked the participant

	Angry	Disgusted	Fearful	Happy	Sad	Surprised	Neutral	Total
Frequency	18	18	17	18	17	16	16	120
Accuracy(%)	44.44	72.22	70.59	88.89	100.00	81.25	18.75	68.33
Kappa	0.352	0.676	0.657	0.870	1.000	0.781	0.052	0.631

Table 1: Accuracies and the Cohen's Kappa coefficients

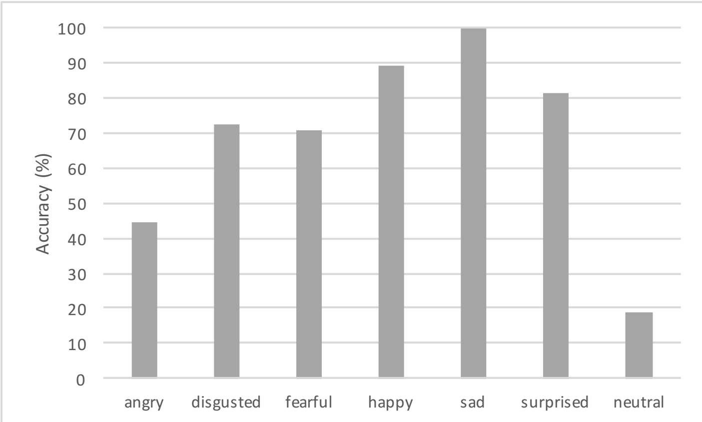


Figure 7: Accuracies in detecting the different poses

to make up a story that would make her feel the emotion, to confirm she felt the emotion in question, and to adopt and represent the corresponding pose.

5. Results

Table 1 shows the accuracy of the app globally and for each specific emotion. This table also includes the Cohen's Kappa coefficients. Figure 7 shows the accuracy results for each pose graphically with a bar chart. One can observe that the emotions of disgusted, fearful, happy, sad and surprised were all detected with accuracies above 70% and Kappa coefficients of 0.65. By contrast, angry was only properly classified with 44%, and neutral was the worst classified with a significant difference from the others.

In order to better show the errors, table 2 shows the confusion matrix. For example, one can observe that angry poses were frequently wrongly classified as sad (i.e. 44%), and neutral poses were frequently mismatched with the sad emotion (i.e. 50 %). All the other possible combination errors were less frequent (i.e. below 19%). In addition, 25 out of the 30 possible combinations had an error of 6.3% or below.

		Actual emotions						
		Angry	Disgusted	Fearful	Happy	Sad	Surprised	Neutral
Predicted emotions	Angry	44.4	0.0	0.0	0.0	0.0	0.0	12.5
	Disgusted	0.0	72.2	17.6	5.6	0.0	0.0	0.0
	Fearful	5.6	5.6	70.6	5.6	0.0	6.3	0.0
	Happy	0.0	0.0	5.9	88.9	0.0	6.3	0.0
	Sad	44.4	16.7	5.9	0.0	100.0	0.0	50.0
	Surprised	0.0	0.0	0.0	0.0	0.0	81.3	18.8
	Neutral	5.6	5.6	0.0	0.0	0.0	6.3	18.8

Table 2: Confusion matrix

	Angry	Disgusted	Fearful	Happy	Sad	Surprised	Neutral	Total
Time (s)	134.00	139.56	120.18	149.06	84.47	163.56	71.00	123.66
Distance (%)	36.83	25.12	25.14	31.51	25.17	29.84	35.26	30.59

Table 3: Average times for defining poses and average distances to the nearest prototype with the correct emotion

Table 3 shows the average times users spent in determining each pose for each emotion and the average distances to the most similar pose of the case base with the correct emotion. It displays the total values of both measures. One can observe that it took around 2 min on average to represent each pose.

Figure 8 graphically shows the average times taken to represent the postures associated with each emotion. One can observe that participants represented neutral with the shortest time on average (i.e. 71 s), but this emotion also obtained very low accuracy results (19%). On the other hand, the participants also took a short time to represent sad poses (the second shortest time with 84 s on average), and it obtained very high accuracy results (i.e. 100%).

The most relevant and frequent suggestion of users was to include the possibility of rotating feet.

The participant in the emotion induction experiment stated that she felt the following for each situation previously indicated in section 4.3 (1) sadness, (2) disgust, (3) fear, (4) happiness, (5) sadness, and (f) surprise. The app estimated respectively the emotions (1) neutral, (2) fear, (3) fear, (4) happiness, (5)sadness and (f) surprise. Since the participant did not feel anger, we asked her to make up a story that would make her feel this emotion. She adopted a pose, and then represented her emotion from a picture of this pose. We asked her whether she actually had felt anger, and she replied yes. The app provided the anger emotion. Considering all these experiments, the app obtained 5 successes out of 7 trials which represented an accuracy of 71.4% and a Kappa coefficient of 65.7%.

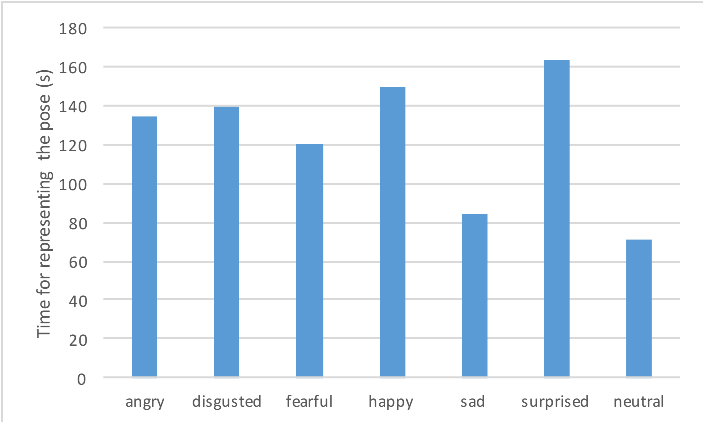


Figure 8: Times taken to represent the different poses

6. Discussion

The emotion represented the fastest by the users was neutral, but the resulting poses were the ones that the app found most difficult to classify. The reason may be that users perceived it as an easy pose and did not consider all aspects of the body such as the rotation and pose of the hands. Indeed, the neutral state may be the least characteristic as it does not actually represent any specific emotion but rather the absence of emotions.

It is worth highlighting that some poses representing different emotions in the experiments were quite similar when ignoring facial expressions. For example, the first angry pose, the third fearful pose and the second happy pose shared most features. In all these poses (a) the hands were clenched in fists, (b) the fists were relatively close to the head, (c) the legs were straight, and (d) the body was twisted with a similar look-at direction (i.e. looking up-front instead of down or up). A human could easily distinguish these images because of the facial expressions. However, without the facial-expression information, even humans could have difficulty in distinguishing the emotions expressed by these poses.

It is worth noting that the feeling of emotions differs from person to person. In addition, the number of poses used also has an influence. As pointed out in Schindler et al. (2008), each emotion could be expressed in three completely different poses. For example, anger could be represented by (a) looking down with the arms crossed, (b) with fists held at middle height as though looking for a fight, or (3) by twisting the whole body with one fist much higher than the other. This explains why anger had lower accuracy than the other emotions. However, all the other emotions also have three different pose patterns. In total, we used 21 different pose patterns, and the accuracy may be considered not so low given the amount of prototype poses. Our previous work (García-Magariño et al., 2018) shows that even well-known approaches used in the context of emotion detection usually obtain relatively low accuracy values if these are compared to results in other less subjective domains.

Participants only spent about two minutes on average to represent each pose in the app. This is an indication of ease of use, considering that it was the first time that they had used the app. Each participant represented four poses only, so they had just started to learn the app and were therefore inexperienced users.

The large distances from the different poses may be due to the high number of dimensions in which the poses are represented. This is supported by the statements of Schindler et al. (2008), who mentioned this challenge for emotion recognition from poses.

Regarding the set of emotions selected, there is some controversy about whether emotions can be represented as combinations of some basic ones. For instance, Ekman (1992a) highlighted the utility of determining emotions as combinations of other basic emotions, while Ortony and Turner (1990) indicated that there was no coherent and simple way of representing emotions as combinations of some basic ones. If we decide to adhere to the former approach, then the app could be adapted to be able to output more than one emotion considering, for example, a certain threshold of similarity from the prototypes.

The purpose of the current approach is to make users self-aware of their emotions and how these are expressed in their poses, so they can improve their wellbeing by understanding this somatic relationship (Loke and Schiphorst, 2018). In this context, final users can use two methods for detecting their emotions through their poses. In the first method, they can see themselves reflected in a mirror and memorize their poses for later representing them with the avatar and using the app for detecting emotions. The second method is similar to the first, but takes pictures of the final users so they can represent their own poses based on these pictures.

The app can represent any natural pose since the user can drag the head, the feet and hands to any natural position, and the avatar assumes a natural pose considering inverse kinematics and these positions. In addition, the user can represent more poses by also changing the positions of the main limb joints (i.e. elbows and knees) for configuring the orientation of the limbs. Furthermore, the user can rotate the wrists and select different degrees of closing/opening the fingers (i.e. between an open hand and a fist). By dragging the head, the user can establish the look-at direction distinguishing between looking up, ahead, down, or side to side, and with any angle between these. Most participants and the authors believe that the app can represent most natural poses, focusing on the relevant features concerning emotions. The app can classify all these features, since it searches the most similar pose among its set of prototype poses, and returns the emotion associated with this prototype pose. Hence, the app will always return a result. Whether the app properly outputs the right emotion is rather challenging, and the accuracy may vary depending on the set of poses used. We plan to conduct a further study in which users represent their own emotions. We had considered this previously, but realized we would need to overcome certain barriers. The first decision would be whether to try to induce emotions or not. Inducing emotions poses ethical issues: inducing emotions strong enough to cause a significant change in pose can have psychologically harmful effects, e.g. triggering an anxiety attack whose traumatic effect might extend well beyond the time of the experiment. In addition, another test should be carried out to confirm that each participant is actually feeling the intended emotion. Deciding not to induce emotions would probably imply that the set of emotions would be unbalanced (i.e. some emotions may be experienced frequently while others may hardly ever be experienced), as occurred in another of our studies regarding the detection of emotions from bodily sensation maps without inducing emotions (García-Magariño et al., 2018).

We are aware that the results of the emotion induction case report and the assessment of poses may not be representative due to the low sample size (7 poses from the same user), but they at least represent a promising outcome in relation to the real applicability of this tool for detecting emotions from the poses of any individual. It would be the first time that the app analyzed each pose. In the future, we plan to contact a specialized psychological team to induce emotions safely, in order to conduct the same study with a much larger sample of participants.

7. Conclusions and future work

This work presents an app that can detect emotions from self-reported poses with a 3D avatar. The app tackles two quite unexplored fields: the self-expression of emotions through 3D avatars by users with no 3D modeling or animation skills, and the automatic detection of emotions from the avatars' 3D poses. To design it, a user-centered methodology has been applied. The evaluation conducted on users revealed that the app is accurate enough to represent poses associated with certain poses extracted from the literature, with accuracies above 70% and Kappa coefficients above 0.65 for five out of the six non-neutral basic emotions. In addition, users were able to represent each pose in about two minutes on average in their first four pose representations with the app. The authors hope that the work will contribute to raising users' body awareness which is being recognized as crucial to promote personal well-being.

As future work, we will include the possibility of rotating the feet in the app as suggested by several participants. In addition, we plan to conduct more user studies to assess the app. We will use the app to collect free poses from users and ask them to classify the emotion expressed in such a way that we can create our own dataset of real poses. In this way, we can contribute to the literature on general emotion patterns as well as further test the app. Moreover, we will test other machine learning techniques such as neural networks to determine if they can improve the accuracy of the tool. Furthermore, we would like to carry out a study of the app with different collectives of people. This could be especially helpful to recognize the emotions of people who do not communicate their emotions easily, such as some of those in jail or suffering from autism.

Acknowledgments

This work has been supported by the program "Estancias de movilidad en el extranjero José Castillejo para jóvenes doctores" funded by the Spanish Ministry of Education, Culture and Sport with references CAS16/00075 and CAS17/00005. We also acknowledge support from "Universidad de Zaragoza", "Fundación Bancaria Ibercaja" and "Fundación CAI" in the "Programa Ibercaja-CAI de Estancias de Investigación" with references IT6/15 and IT24/16. This work has been partly financed by the Spanish Government and the European Union through the contract TIN2015-67149-C3-1R (MINECO/FEDER) and by the Aragonese Government and the UE through the FEDER 2014-2020 "Construyendo Europa desde Aragón" action (Group T25_17D). This work also acknowledges the research project "Construcción de un framework para agilizar el desarrollo de aplicaciones móviles en el ámbito de la salud" funded by University of Zaragoza and Foundation Ibercaja with grant reference JIUZ-2017-TEC-03. We also received support from project "Sensores vestibles y tecnología móvil como apoyo en la formación y práctica de mindfulness: prototipo previo aplicado a bienestar" funded by University of Zaragoza with grant number UZ2017-TEC-02.

References

Alonso-Martín, F., Malfaz, M., Sequeira, J., Gorostiza, J. F., Salichs, M. A., 2013. A multimodal emotion detection system during human–robot interaction. Sensors 13 (11), 15549–15581.

Baldassarri, S., Cerezo, E., 2012. Maxine: Embodied conversational agents for multimodal emotional communication. In: Computer Graphics. ISBN: 978-953-51-0455-1, InTech, pp. 195–211.

Baldassarri, S., Cerezo, E., Seron, F. J., 2008. Maxine: A platform for embodied animated agents. Computers & Graphics 32 (4), 430–437.

Chittaro, L., 2016. Designing serious games for safety education:"learn to brace" versus traditional pictorials for aircraft passengers. IEEE transactions on visualization and computer graphics 22 (5), 1527–1539.

Clavel, C., Plessier, J., Martin, J.-C., Ach, L., Morel, B., 2009. Combining facial and postural expressions of emotions in a virtual character. In: Intelligent Virtual Agents. Vol. 5773 of Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 287–300.

Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. IEEE transactions on information theory 13 (1), 21–27.

De Gelder, B., 2006. Towards the neurobiology of emotional body language. Nature Reviews Neuroscience 7 (3), 242–249.

De Silva, P. R., Bianchi-Berthouze, N., 2004. Modeling human affective postures: an information theoretic characterization of posture features. Computer Animation and Virtual Worlds 15 (3-4), 269–276.

Ekman, P., 1992a. Are there basic emotions? Psychological Review 99, 550–553.

Ekman, P., 1992b. An argument for basic emotions. Cognition & emotion 6 (3-4), 169–200.

Garber-Barron, M., Si, M., 2012. Using body movement and posture for emotion detection in non-acted scenarios. In: Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on. IEEE, pp. 1–8.

García-Magariño, I., Chittaro, L., Plaza, I., 2018. Bodily sensation maps: exploring a new direction for detecting emotions from user self-reported data. International Journal of Human-Computer Studies 113, 32–47.

Gross, M. M., Crane, E. A., Fredrickson, B. L., 2012. Effort-shape and kinematic assessment of bodily expression of emotion during gait. Human movement science 31 (1), 202–221.

Gunes, H., Hung, H., 2016. Is automatic facial expression recognition of emotions coming to a dead end? the rise of the new kids on the block. Image and Vision Computing 55, 6–8.

Hao, N., Xue, H., Yuan, H., Wang, Q., Runco, M. A., 2017. Enhancing creativity: Proper body posture meets proper emotion. Acta Psychologica 173, 32–40.

Kleinsmith, A., De Silva, P. R., Bianchi-Berthouze, N., 2006. Cross-cultural differences in recognizing affect from body posture. Interacting with Computers 18 (6), 1371–1389.

Kollias, S., Karpouzis, K., 2005. Multimodal emotion recognition and expressivity analysis. In: Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on. IEEE, pp. 779–783.

Loke, L., Schiphorst, T., 2018. The somatic turn in human-computer interaction. interactions 25 (5), 54–5863.

Marschner, L., Pannasch, S., Schulz, J., Graupner, S.-T., 2015. Social communication with virtual agents: The effects of body and gaze direction on attention and emotional responding in human observers. International Journal of Psychophysiology 97 (2), 85–92.

Mehling, W. E., Gopisetty, V., Daubenmier, J., Price, C. J., Hecht, F. M., Stewart, A., 2009. Body awareness: construct and self-report measures. PloS one 4 (5), e5614.

Mondloch, C. J., 2012. Sad or fearful? the influence of body posture on adults and childrens perception of facial displays of emotion. Journal of experimental child psychology 111 (2), 180–196.

Ortony, A., Turner, T. J., 1990. What's basic about basic emotions? Psychological review 97 (3), 315.

Politou, E., Alepis, E., Patsakis, C., 2017. A survey on mobile affective computing. Computer Science Review 25, 79–100.

Rázuri, J. G., Larsson, A., Rahmani, R., Sundgren, D., Bonet, I., Moran, A., 2015. Recognition of emotions by the emotional feedback through behavioral human poses. International Journal of Computer Science Issues (IJCSI) 12 (1), 7.

Rosario, J., Diogenes, M., Leite, M., Persoli, R., Leite, J., 2010. Are there relations between human posture and emotions? International Journal of Psychophysiology 77 (3), 276.

Rossberg-Gempton, I., Poole, G. D., 1993. The effect of open and closed postures on pleasant and unpleasant emotions. The Arts in psychotherapy 20 (1), 75–82.

Sadowski, A. S., Lomanowska, A. M., 2018. Virtual intimacy: Propensity for physical contact between avatars in an online virtual environment. Computers in Human Behavior 78, 1–9.

Saneiro, M., Santos, O. C., Salmeron-Majadas, S., Boticario, J. G., 2014. Towards emotion detection in educational scenarios from facial expressions and body movements through multimodal approaches. The Scientific World Journal 2014.

Schindler, K., Van Gool, L., de Gelder, B., 2008. Recognizing emotions expressed by body pose: A biologically inspired neural model. Neural networks 21 (9), 1238–1246.

Zeng, Z., Pantic, M., Roisman, G. I., Huang, T. S., 2009. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE transactions on pattern analysis and machine intelligence 31 (1), 39–58.

Author Contributions

Iván García-Magariño contributed in the conceptualization of the research. He developed all the software of the presented mobile application. He actively participated in establishing the research methodology, the data acquisition and the interpretation of the data. He wrote the original first draft, and then improved it following the suggestions of other authors. He was the principal investigator of the two research projects that funded his two research stays in the Human-Computer Interaction (HCI) Lab at University of Udine (Italy) that made this international collaboration possible. He was also the principal investigator of another research project that funded this research.

Eva Cerezo proposed the original conceptualization of the problem and the original solution. She contributed in the design of the research methodology, and she supervised the whole research methodology. She supervised the user-centered design, and participated in it. She actively participated in reviewing and editing the article. She was the principal investigator of one of the projects that supported this work.

Inmaculada Plaza collaborated in the conceptualization of this research and the definition of the research methodology. She actively participated in the user-centered design. She reviewed the article. She actively helped the first author in acquiring funding for one of his research stays that made this international collaboration possible. She was the principal investigator of one of the projects that supported this work.

Luca Chittaro participated in the conceptualization of this research and the establishment of the research methodology. He organized and supervised the user study. He used his resources for finding the right experimenters for conducting this user study. He supervised the data acquisition of this user study. He actively participated in interpreting the data and writing the article. He hosted the first author in his HCI Lab at the University of Udine, to make this research possible.