Posture, Gesture and Motion Quality: A Multilateral Approach to Affect Recognition from Human Body Motion

Daniel Bernhardt

Computer Laboratory, University of Cambridge, 15 JJ Thomson Avenue, Cambridge, CB3 0FD, UK

Abstract. This paper introduces a unified computational framework for the analysis of affective body motion. The approach is based on three main motion aspects: posture, gesture and motion quality. Building upon the idea of motion primitives, I suggest an approach to defining those aspects. The paper also briefly discusses the progress made by the author in implementing parts of the proposed model and considers which strands of research are likely to benefit from the results of this work.

1 Introduction and Motivation

When Charles Darwin published his observations on "The expression of emotion in man and animals" in 1872, he argued that emotional expressions of the face and body constitute innate, adaptive, physiological responses to events in our environment [1]. After over a century of research, the human body is today often regarded as a channel of interpersonal communication, conveying information relating to emotion and inter-personal attitudes. In fact, the structure and protocols of emotional expression have been found to be so sophisticated that one often encounters terms such as "non-verbal communication" or "body language" [2]. These descriptions allude to the fact that bodily expression of emotions is far from limited to archetypal body motions exhibiting basic emotions such as fear or anger [3].

Motivated by these earlier findings, my research focuses on finding computational methods to recover communicated emotions from body motions. In order to comprehensively analyse human body motion, I am developing a unified model which combines different aspects of body motion. My model distinguishes between three main aspects: posture, gesture and motion quality. These aspects have so far been studied very much independently from each other. However, considering the complex nature of human movement, combining these aspects in one model promises to make affect recognition from body motion more accurate and applicable to many different contexts. I will introduce the above aspects in Sect. 1.1 and subsequently present the developed model in Sect. 2. Based on this model, I have so far studied one of the aspects, motion quality, in detail. I will report the experimental results relating to this previous work in Sect. 3. In Sect. 4 I am going to outline the work which I am intending to complete in order to explore the proposed model as completely as possible within the scope of my PhD research.

1.1 Posture, Gesture and Motion Quality

We normally understand *posture* as a description of static body configuration. Examples belonging to this category are crossing one's arms, leaning forward, tilting one's head. Psychologists such as Bull [2] and more recently computer scientists such as Kleinsmith [4] have demonstrated that static body posture can communicate affect for contexts such as conversations or expressively acted emotions. Picard et al. demonstrated that affect can be correlated to *shifts* in posture in an e-learning scenario [5].

Most authors, however, admit that neglecting most of the dynamic information when using posture information is a big hindrance for affect recognition. Body movements such as raising one's arms in joy or the iconic look at the watch are much better captured as *gestures*, which describe a body's dynamic motion. For the purpose of my research I define a gesture to be a motion which, in itself, carries some affective information. Ekman and Friesen referred to these kinds of communicational gestures as Affect Displays [6]. Subsequently it has been the face for which many of those gestures have been studied in the form of smiles, eyebrow flashes etc. Some recent research on emotion recognition from the face further suggests that only basic emotions such as happiness or disgust are easily distinguished using static facial images. In order to distinguish more complex emotions such as interest or agreement we need to consider facial gestures [7]. I am hypothesising that emotion recognition from the body equally needs to consider dynamic gestures on top of static body posture information.

One complication arises when one tries to implement emotion recognition from body cues. In most natural scenarios people are using their bodies to achieve certain goals such as driving a car or cleaning the home. In those cases, it is not only the observed postures and gestures but above all the *quality* of the observed motions which gives us information about a person's mental state. Efforts towards analysing motion qualities such as velocity, acceleration and jerk are fairly recent, but tend to exhibit good results [8–10]. My recent work in this area has demonstrated the effect of emotion on velocity and acceleration in particular and related results are discussed in Sect. 3.

2 A Unified Model for Analysing Body Motion

Given that each motion aspect has been shown individually to hold some valuable affective information, this section proposes a unified model which is designed to integrate the three aspects. Such a unified model will give us the ability to investigate the interaction of the motion aspects and their relative contributions to the communication of emotional meaning.

Fig. 1 shows a diagrammatic representation of the three aspects. The concepts are arranged in an pyramid for two reasons. Firstly, as we move upwards

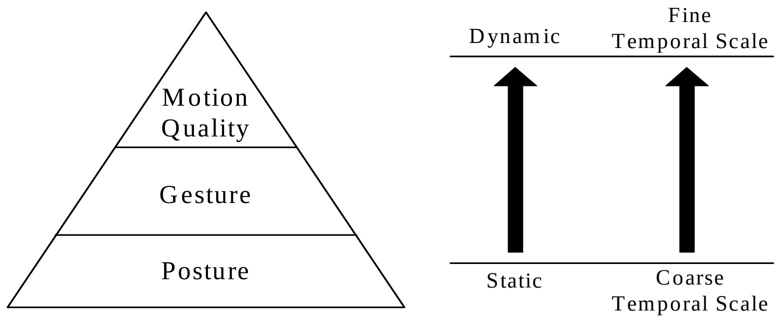

Fig. 1. A schematic representation of the three discussed aspects of body motion.

through the pyramid, the cues we are utilising are increasingly *dynamic*. While body posture denotes the long term static properties of the body, gestures describe how basic movements are composed together to create dynamic displays. Finally, motion qualities are best described in terms of concepts such as force and energy or by using a multi-dimensional space such as Laban's Effort and Shape dimensions [11]. These concepts capture the motion dynamics of the gestures and other body motions, such as speed and acceleration.

The second way in which the pyramid can be interpreted is in terms of *temporal scale*. Posture tends to change at a coarse scale. Depending on the context, the overall body posture might remain relatively constant for periods of between 10 seconds to many minutes. At a finer scale, we will often find gestures modulating the basic posture. The typical duration of a gesture might be up to several seconds. Finally, the motion dynamics can potentially be defined at every time frame and therefore represent the finest temporal scale.

2.1 Motion Primitives

Given the conceptual model in Fig. 1, we have to find a definition of posture, gesture and motion quality which lends itself to a computational implementation. I am proposing to define the meanings of these concepts in terms of the smallest isolable elements from which complex movements are constructed — atomic or primitive motions. This set of motion primitives, or kinemes, can be understood as analogous to a set of phonemes in linguistics. Just as phonemes are the smallest meaningful units of sound, kinemes are the smallest meaningful units of motion. Some authors have in the past suggested possible lists of such primitive motions. My argument is largely inspired by the anthropologist Ray Birdwhistell [12] who devised a set of kinemes in order to record the events which he believed significant for interpersonal communication. Much like in a natural language, syntactic rules are followed to combine these kinemes into more complex motion structures with social meanings. This idea leads directly to the definition of a gesture. Once we have defined a set of motion primitives, we

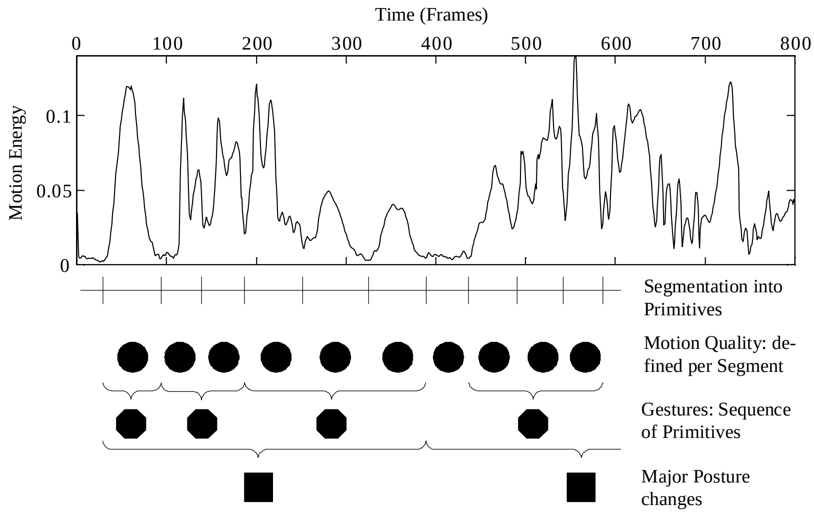

Fig. 2. The derivation of postures, gestures and motion qualities from a complex motion. The motion is represented by a measure of motion energy as a function of time.

can describe or detect particular gestures by considering the dynamic sequence of these primitives.

We can now turn to the definitions of posture and motion quality. I will think of posture as the long term average body configuration. Given that the formulation of posture should be robust with respect to any higher frequency gestures, the posture at any moment should be calculated as the average body configuration over the duration of many past motion primitives. Motion quality, on the other hand should be calculated over a much smaller time span. As the experiments discussed in Sect. 3 illustrate, considering motion dynamics calculated over the duration of individual motion primitives can lead to very discriminative features. The dependencies between the original motion signal, the motion primitives and the three affect-communicating motion aspects are summarised in Fig. 2.

2.2 Dealing with Different Contexts

Even if we manage to break down a complex movement sequence into information concerning posture, gesture and motion quality, we cannot readily make a decision on the expressed emotion unless we know the context in which this motion occurred. One of the major elements of context is the action being performed by a person or group of people. If we were only allowed to use body motions as a cue, we could try to represent different contexts by the different distributions of various motion primitives and postures over time. Different activities and therefore contexts might be discriminable by these distributions.

Conversations, for example, should be characterised by a standard repertoire of common hand gesture primitives as well as a relatively constant upright posture. Walking, on the other hand would be well described by a balanced distribution of arm swing and locomotion primitives. This formulation of context is comparable to an approach by Barbic et al. who are segmenting sequences of actions into distinctive behaviours, based on the derivation of motion data distributions [13].

Once we have decided which contexts we are interested in for a particular application, we can train statistical models to describe and recognise each of these contexts. Based on the motion aspects we can then train emotion models and apply them to the different contexts.

3 Experimental Results

My work so far has concentrated on the investigation of motion dynamics [10]. It has examined a subset of the overall motion model described in Sect. 2. In particular, I have looked at the definition of motion primitives and the use of dynamic motion qualities such as speed and acceleration to discriminate between different affects. I therefore analysed a database of knocking motions performed in four different affective styles (neutral, happy, angry, sad) by 30 different individuals. This is an example where nearly all of the affective information has to be derived from motion quality. While the posture was very much prescribed by the performed task, subjects did not show any affective gestures.

First, I segmented the database of complex motions into atomic parts based on motion energy. I then defined motion primitives by clustering all of these segments. The four derived clusters and associated primitive motions were found to correspond to "raise arm", "lower arm" and the smaller scale "knocking" and "retracting" motions during the actual knocking action. Based on the motion dynamics of the segments, I then trained Support Vector Machines for each motion primitive to distinguish between the four different aforementioned affects. Individual primitive-based classifications were combined to produce an overall classification for a complex motion. This methodology achieved an overall recognition rate of 50%. Using a Wilcoxon Signed-Rank Test this performance was shown to be significantly better than chance at a level of $p < 0.0001$.

3.1 Dealing with Individual Motion Bias

It was found that this recognition rate could be boosted to 81% if we consider individuals' motion idiosyncrasies. I found that some people tend to move faster than others, a phenomenon I call motion bias. We can define a person's motion bias by the average motion dynamics exhibited across several affective states. My results suggested that deriving motion quality features as a measure *relative* to this motion bias leads to a much better discrimination between emotions. The improvements achieved with this insight are illustrated in Fig. 3. The figure shows the recognition rate for each subject before and after removing motion

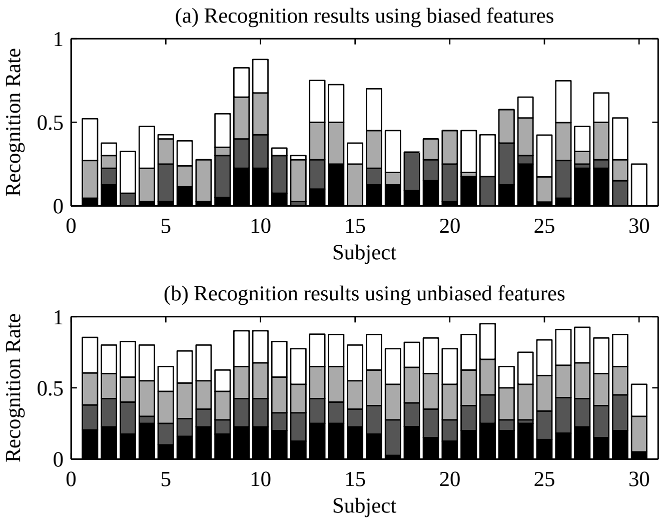

Fig. 3. Performance for emotion recognition from motion dynamics. The individual performances for the distinguished emotions neutral (black), happy (dark grey), angry (light grey), sad (white) are stacked to visualise the overall recognition rate.

bias. Removing the bias improves both the overall recognition rate and uniformity of the recognition performance. Again, the improvement was significant at a level of $p < 0.0001$.

4 Proposed Future Work

Having confirmed the discriminative power of motion dynamics for one particular context, my immediate work is going to focus on several related questions:

1. Does the same approach work for other, similar everyday motion contexts?
2. What is the connection between contexts? Do we need to train different emotion models for different contexts?
3. Is a person's movement bias context-dependent? Do we need to establish motion bias for each context individually or can we predict it from a small set of sample motions?

Answers to these questions will be found using the extended database which has been used for my previous work [10]. One of my primary goals, however, is the investigation of how the different motion aspects posture, gesture and quality interact in order to convey affective meaning. I am therefore planning to expand the used data set considerably. Preparations are underway at the time of writing to conduct a data acquisition effort which will exhibit a reasonably controlled set of body motions. The goal will be to analyse the data according to the three presented aspects and establish their relative importance for different contexts present in the data. Such contexts might be conversational or task-focused. While the latter context is hoped to confirm my findings on motion qualities, the former should provide information about the role of gestures in the communication of affect. Given the data, I will focus on a more complete implementation of the motion model described in Sect. 2.

As a final result I am hoping to gain insights into the following two questions:

1. How do the aspects of posture, gesture and motion quality interact in the bodily communication of affect?
2. To what extent do we need to take context into consideration when performing affect classification based on body motion?

5 Wider Perspective

Of course, body motion is only one of many channels that communicate affect. Although historically ignored, researchers from various backgrounds have recently been stressing the potential of analysing body motions for the recognition of emotions. Psychologists such as Bull and Pollick have argued that, contrary to prior beliefs, various aspects of body motion communicate affect independently from other cues such as facial expressions or verbal communication. Computational approaches have confirmed these experimental results.

Advances in affect recognition from individual channels have also led researchers to analyse the relative significance of bodily, facial, voice and physiological cues. The body in particular is promising to play an important part in these multimodal approaches. For one, body motion is the most easily accessible channel in a real-world application as the face is easily obstructed and hard to record due to the relatively small size. Similar observations hold when the body is compared to voice and physiological cues. Also, body expressions are generally regarded as harder to mask or fake than facial expressions [14]. A multimodal approach is promising to ease detection of masked emotions by exhibiting mismatching emotional displays from different channels.

The successful development of affect-sensing technology, in particular from the body, promises a wide range of interesting applications. The work described in this paper is part of the "Empathic Avatars" project which is investigating how affect recognition of the kind described in this paper could be used in the synthesis of affective expressions [15]. We are proposing the use of multimodal affect recognition from a real environment to drive the animation of emotions in a virtual environment. In a connected effort, I have been looking at the phenomenon of presence in Virtual Reality. One of the factors which gives people the feeling of being present in a virtual space is the appropriate actions and reactions of other living entities in the shared environment. In order to create this feeling of social presence, virtual agents need to understand the communicated affective information in order to react to them appropriately.

Another strand of research which increasingly seeks this kind of emotional intelligence is Human-Robot Interaction. Many research groups are investigating the problems of interacting with and, in particular, teaching robots. One strong hypothesis is that with the increasing use of robots in our domestic environment, it should be the robots who are getting increasingly adapt to our natural emotional feedback rather than the users adapting to artificial and unnatural computer interfaces — a premise which drives much of the Affective Computing research today.

Acknowledgments

I would like to thank my supervisor Prof. Peter Robinson for his continued support. This work is generously funded by EPSRC under the grant EP/D003180/1.

References

1. Darwin, C.: The Expression of the Emotions in Man and Animals. University of Chicago Press, Chicago (1872)
2. Bull, P.E.: Posture and Gesture. Volume 16. Pergamon Press (1987)
3. Cowie, R., Douglas-Cowie, E., Cox, C.: Beyond emotion archetypes: Databases for emotion modelling using neural networks. Neural Networks **18**(4) (2005) 371–388
4. Kleinsmith, A., de Silva, P.R., Bianchi-Berthouze, N.: Grounding affective dimensions into posture features. In: Affective Computing and Intelligent Interaction. (2005) 263–270
5. Mota, S., Picard, R.: Automated posture analysis for detecting learner's interest level. Technical report, MIT Media Laboratory (2003)
6. Ekman, P., Friesen, W.V.: The repertoire of nonverbal behavior: Categories, origins, usage, and coding. Semiotica (1969) 49–98
7. El-Kaliouby, R., Robinson, P.: Generalization of a vision-based computational model of mind-reading. In: Affective Computing and Intelligent Interaction. (2005) 582–589
8. Pollick, F.E., Paterson, H.M., Bruderlin, A., Sanford, A.J.: Perceiving affect from arm movement. Cognition **82** (2001) 51–61
9. Kapur, A., Kapur, A., Virji-Babul, N., Tzanetakis, G., Driessen, P.F.: Gesture-based affective computing on motion capture data. In: Affective Computing and Intelligent Interaction. (2005) 1–7
10. Bernhardt, D., Robinson, P.: Detecting affect from non-stylised body motions. In: Affective Computing and Intelligent Interaction. (under submission)
11. Chi, D., Costa, M., Zhao, L., Badler, N.: The EMOTE model for effort and shape. In: SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques. (2000) 173–182
12. Birdwhistell, R.L.: Kinesics and Context: Essays on Body Motion Communication. University of Pennsylvania Press (1970)
13. Barbic, J., Safonova, A., Pan, J.Y., Faloutsos, C., Hodgins, J.K., Pollard, N.S.: Segmenting motion capture data into distinct behaviors. In: Proceedings of Graphics Interface (GI 2004). (2004)
14. Argyle, M.: Bodily Communication. International Universities Press, Inc. (1975)
15. Empathic Avatars: input, processing and output of emotional state. (2005) http://www.cs.ucl.ac.uk/research/vr/EmpathicAvatars/.