Grounding Affect Recognition on a Low-Level Description of Body Posture

Andrea Kleinsmith

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London

Department of Computer Science

University College London

July 14, 2010

To my family

I, Andrea Kleinsmith, confirm that the work presented in thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The research presented in this thesis is centred in the rapidly growing field of affective computing and focuses on the automatic recognition of affect. Numerous diverse technologies have become part of working and social life, hence it is crucial to understand whether recognising the affective state of the user may be added to increase the technologies' effectiveness.

The contributions made are the investigation of a low-level description of body posture, the proposal of a method for creating benchmarks for evaluating affective posture recognition models, and providing an understanding of how posture is used to communicate affect. Using a low-level posture description approach, this research aims to create automatic recognition models that may be easily adapted to different application contexts. These recognition models would be able to map low-level descriptions of postural configurations into discrete affective states and levels of affective dimensions.

The feasibility of this approach is tested using an incremental procedure with three studies. The first study (*acted postures*), investigates the feasibility of recognising basic emotions and affective dimensions from *acted*, i.e., stereotypical, exaggerated expressions. The second study (*non-acted postures*), aims at recognising subtle affective states and affective dimensions from *non-acted* body postures in the context of a video game. In both studies, the results showed above chance level agreement and reliable consistency between human observers for the discrete affective states and valence and arousal dimensions. A feature analysis showed that specific low-level features are predictive of affect. The automatic recognition models achieved recognition rates similar to or better than the benchmarks computed. Extending the non-acted postures study, the third study focuses on how the affective posture recognition system performs when applied to *sequences* of non-acted static postures that have not been manually preselected, as if in a runtime situation. An automatic modelling technique was combined with a decision rule defined in this research. The results indicate that posture sequences can be recognised at well above chance level.

Acknowledgements

This research was funded by Marie Curie International Re-Integration Grant: AffectMe (MIRG-CT-2006-046434). This thesis would not have been possible without the support and encouragement of so many people.

I would especially like to thank my primary supervisor Nadia Bianchi-Berthouze. This work simply could not have been completed without her continuous advice, support and encouragement, for which I will always be grateful. I consider myself lucky to have benefited from your kindness, knowledge and friendship. I would also like to thank my secondary supervisor Anthony Steed for the advice and time given to help finish this research.

I would like to thank the former and current members of UCLIC, my good friends, who welcomed me and made the transition into a new life in a new country easier than I ever thought possible. We've had some very good times.

I would like to thank everyone who participated in my experiments, and who did not run when they saw me coming with the motion capture suit time and again.

I have been blessed with a wonderful and supportive family, to whom I am very grateful. Mom, Dad, Jeff and Katie, Linda, Juniper and Frances. Mom and Dad, I would never have thought this possible if it hadn't been for your love, commitment, and belief in me. Words can't really say thank you enough.

Heather Widup, my oldest and dearest friend. Even though we've spent the last eight years on almost opposite sides of the world, just knowing you were out there definitely made this journey tolerable. To my friends Iain Jackson, Georgia Davis, Charlotte Patten, Kerry Newton, thank you for always listening.

Ian Greatbatch, the amount of support you have given is immeasurable. You have been both a mentor and a best friend. I don't think I could have finished this journey without your encouragement, kindness and empathy.

Milo and Griffyn have been by my side and in my heart every step of the way.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Motivation

The research presented in this thesis is centred in the rapidly growing field of affective computing and focuses on the automatic recognition of affect. Coined by Rosalind Picard [Pic97], **affective computing** is a multidisciplinary field of research concerned with *"computing that relates to, arises from, or deliberately influences emotions".* In order to create systems aimed at the recognition of affect, an understanding must be gained of how people experience affect and emotion, both in conveyance and recognition. A challenge in creating these systems is that, as Picard [Pic98] points out, many factors affect the manner in which humans convey emotion or affective messages, such as age, gender, culture, context, etc.

Numerous affective computing studies have focused on affect that is expressed and perceived from facial expressions in particular whereas research on body posture has been much less emphasised. Reasons for this difference may be due to the lack of formal models for body posture as there are for the face (e.g., the Facial Action Coding System (FACS) [EF78]), as well as the complexity of the body. Only recently have bodily expressions of affect gained attention in affective computing research. As technologies encountered by the average person on a day-to-day basis become more and more ubiquitous [FT05], they afford a multimodal interaction in which body posture is becoming a focus of attention. The purpose of the research presented in this thesis is on testing the power of body posture in affect recognition.

Relevance

The relevance of this research and the benefits of developing applications into which affect perception can be integrated is evident in many areas of society, such as security, law enforcement, games and entertainment, education and health care. For instance, research on affective aspects of video games has become a hot topic in the last few years. Video games are meant to be fun and challenging as well as frustrating at times, and frequently, eliciting frustration in a player is considered important in order to intensify the gaming experience [Fre03]. However, if the player's level of frustration or boredom is too high for too long, s/he is likely to give up and quit [GD04]. To be more effective, technologies need to be created to account for such changes.

Educational applications face similar problems. Motivation plays a significant role in learning [AM94][Cov93][Sta96], and students lose motivation when high levels of affective states such as frustration, anxiety, fear of failure, etc. are experienced [JG93][KBP07]. Indeed, human teachers are taught how to read affective aspects of students' body language and how to react appropriately through their body language and actions [NC93] in an effort to help students maintain motivation. Educational applications that can act as human counterparts and recognise the learner's affective states from nonverbal communication modalities such as posture may be able to provide efficient guidance to the student.

In the health care industry, robots are now being introduced into traditionally human roles which also require affective awareness. For instance, robots are being implemented in patient rehabilitation situations [NR05][CPM^{+}07], taking the place of human physiologists. A patient's affective state and level of motivation plays a big role in whether or not the patient continues with the prescribed rehabilitation. Rehabilitation robots that can detect a patient's affective state may have a better chance of helping the patient regulate his/her emotions in an effort to maintain positive affect, motivation and interest level.

Terminology

Throughout this thesis, the terms **affect** and **emotion** are used repeatedly. There is much debate about these terms and how to define them and what triggers them. Generally speaking, the function of affect and emotion is a support to, and necessary for, cognitive processes such as memorisation, decision making, problem solving, intelligence and attention focusing [Dam94]. Damasio further asserts that rational thinking is not possible without emotion.

According to the Dictionary of Psychology [psy08], **affect** is defined as

> *"Behavior that expresses a subjectively experienced feeling state (emotion); affect is responsive to changing emotional states, whereas mood refers to a pervasive and sustained emotion. A subjective feeling or emotional tone often accompanied by bodily expressions noticeable to others".*

Many researchers consider the term affect as the umbrella term under which all emotional phenomena exists [You43][LR78] and the term emotion as one aspect underneath that umbrella term. This is the viewpoint taken in this thesis. Furthermore, there is a debate about whether affect is the result of a cognitive appraisal of a situation [Laz82], or if affect occurs prior to cognition [Zaj80]. The affective computing community tends to include cognitive states under the umbrella of affect [eKR04][AR09]. This is also the view taken in this thesis.

Conceptually, the term **emotion** and what it represents is almost universally understood, and yet almost impossible to define clearly [FR84][Rus03][PPB$^+$04]. No single definition seems to suffice within the various fields in which emotion is studied [Fox08][Iza07][Fri88] [KJK81]. Somewhat similar to the definition of affect, in her 2008 book Fox [Fox08] defines **emotion** as

> *"A relatively brief episode of coordinated brain, autonomic and behavioural changes that facilitate a response to an external or internal event of significance to the organism."*

In this thesis, the terms affect and emotion are used throughout as general terms and both discrete and dimensional approaches are used to describe affect.

1.1 Research Aims and Challenges

To have a more prominent place in today's society, new technologies may need to have affect recognition capabilities [FT05]. To develop systems with these capabilities, several issues have been identified which this thesis aims to address. These issues have been formulated into a set of hypotheses depicted in Figure 1.1 and explained below.

The overall main hypothesis is that **affect can be recognised from whole body postures using a low-level description of the body**. More specifically, it is hypothesised that levels of affective dimensions, acted basic emotions and non-acted, non-basic affective states can be recognised from posture. In this research *actors* are defined as participants who are explicitly asked to enact emotions through their body movements, whereas *non-actors* are not aware of the purpose of the study, thus their body movements are considered to be naturally expressive. Both acted and non-acted situations were chosen in order to implement an incremental approach to affect recognition. The use of an acted situation was chosen first because it is a typical starting point for affect recognition research [DCCS$^+$07][LNP02]. It is thought that if acceptable levels of recognition are not possible at this level, then recognition at a non-acted, more complex level will not be possible either.

To investigate the hypothesis more fully, two sub-hypotheses are made. One, that **human observers** (i.e., the participants who judge the affective bodily expressions as opposed to those participants whose bodily expressions were recorded) can recognise affect from body posture at **above chance level**. Chance level is the percentage that would be achieved with choosing, in this research, an affective state or affective dimension level at random. Two, that **automatic recognition models** can achieve accuracy rates similar to benchmarks based on the **human observers**.

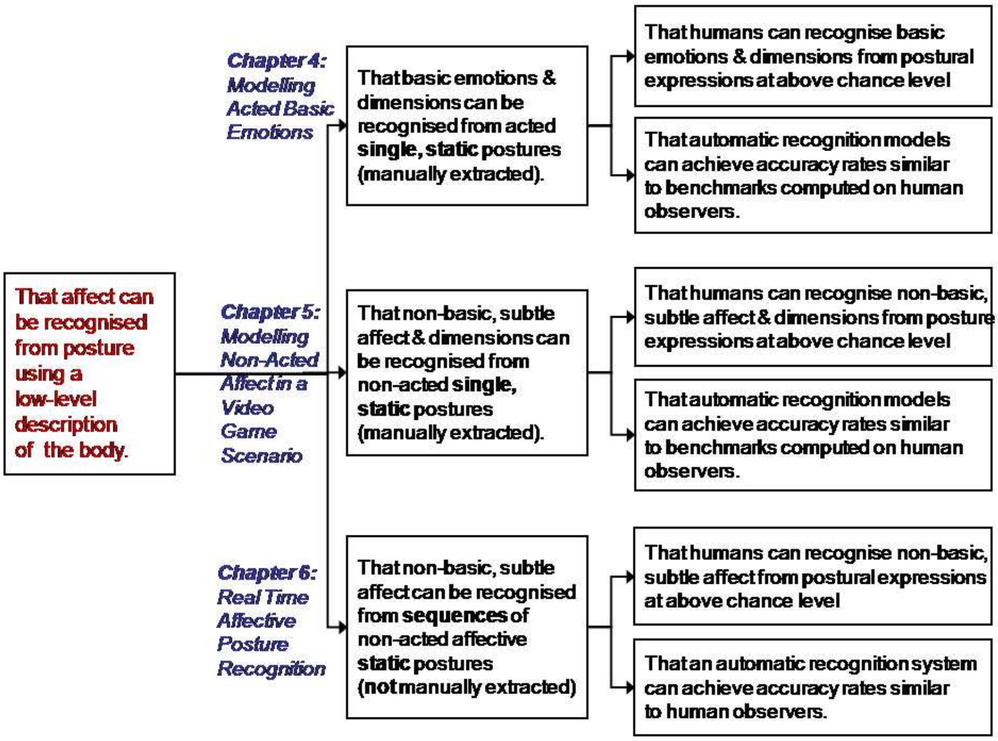

That affect can be recognised from posture using a low-level description of the body.
Chapter 4: Modelling Acted Basic Emotions
That basic emotions & dimensions can be recognised from acted single, static postures (manually extracted).
That humans can recognise basic emotions & dimensions from postural expressions at above chance level
That automatic recognition models can achieve accuracy rates similar to benchmarks computed on human observers.
Chapter 5: Modelling Non-Acted Affect in a Video Game Scenario
That non-basic, subtle affect & dimensions can be recognised from non-acted single, static postures (manually extracted).
That humans can recognise non-basic, subtle affect & dimensions from posture expressions at above chance level
That automatic recognition models can achieve accuracy rates similar to benchmarks computed on human observers.
Chapter 6: Real Time Affective Posture Recognition
That non-basic, subtle affect can be recognised from sequences of non-acted affective static postures (not manually extracted)
That humans can recognise non-basic, subtle affect from postural expressions at above chance level
That an automatic recognition system can achieve accuracy rates similar to human observers.

Figure 1.1: Hypothesis tree

Recognising affect from posture using a low-level description of the body

The study of body *posture* within the affective computing field is still novel while automatic affect recognition for facial expressions and voice have a stronger history. To date, most of the work has focused on stereotypical affective *movements* such as dance [CMR+04][KKVB+05], and with a limited or gross description of the bodily expression. Affect and emotion may be expressed in many situations and activities that involve the use of the entire body, and this makes the modelling process quite complex. That may be why most of the work has focused on body movement instead of posture, looking at dynamic aspects of body movement as suggested by dance (where body movement is purposely used to express emotion) and only with a gross description of the configuration of the body. Static information about the configuration of the body may provide sufficient information for discriminating between affective states. Can an approach be found that allows for the creation of more general postural configuration models that are grounded on cues that are largely independent of the context? This is a novel approach because, as previously mentioned, there are no widely agreed upon models for affective bodily expressions as there are for the face.

Building automatic recognition models based on human observers

In this thesis, human observers are chosen to assign ground truth labels to affective bodily expressions as opposed to using the expressers' labels. The reason for this is that at the moment, it is not possible to know reliably what the expresser's true affective state is. Asking the expresser herself is not sufficient and not always possible, as will be discussed in Chapter 2.

Judging the recognition accuracy of the automatic models

In this thesis, the recognition accuracies of the automatic models are validated against benchmarks that are created based on the agreement levels achieved by human observers. This method was chosen because benchmarks for evaluating the performance of affective posture recognition systems have yet to be defined in the affective computing field. Benchmarks for facial expression recognition systems exist but they are generally based on an expert coder's ability to classify Action Units (AUs)[1] [DBH$^+$99][BLF$^+$06]. Furthermore, the intention is to create software focused towards replacing a human interaction partner. The automatic recognition of affect from body posture will be created using a random repeated sub-sampling method to test the models' ability to generalise to new observers, and cross-validation to test the models' ability to generalise to new postures.

[1] An Action Unit is *"a visually distinguishable and anatomically based unit of facial muscle movement"* [TCDFBP02] as part of FACS [EF78].

1.2 Contributions

This thesis makes contributions to the affective computing field in the study of human and automatic recognition of affect from whole body posture. The main contributions are listed below.

The investigation of a low-level description of body posture

Testing whether a low-level description of body posture can encode enough information to support the automatic recognition and discrimination of bodily expressions of affect. Bodily expressions of affect are thus established as a powerful modality to mediate human-technology interaction. While it is acknowledged that to obtain more reliable automatic recognition of affect, systems should contain the integration of affective information from multiple modalities, the research carried out for this thesis provides recognition tools for whole body posture only.

Proposing a method for setting benchmarks for affective expressions

The benchmarks for automatic affect recognition are developed using human observer judgments as the ground truth labels as opposed to the expressers themselves. The recognition models are tested against human observer judgments of the affective postures. While some research has also considered observers' judgments as ground truth, the variability that occurs between observers is not addressed. Furthermore, the observers recruited typically represent a narrow band of the population. Therefore, to take these issues into account, the approach proposed in this research project is a repeated sub-sampling method to determine benchmarks against which the automatic recognition models are evaluated.

Understanding how affect can be communicated through body posture

An understanding of the affective information in body posture that can be recognised by human observers. This is examined through an analysis of human observers' judgments on acted and non-acted affective expressions. By studying factors such as culture and understanding cross-cultural differences that exist in the way people read affective body expressions, the information can be used to inform researchers in other fields such as behavioural psychology and emotion synthesis. For instance, the information about specific low-level features that contribute to the conveyance of a specific affective state can be used to create embodied avatar agents and robots (outside the scope of this thesis).

1.3 Scope of Thesis

This thesis is focused on the mapping of body posture (i.e., configuration information as opposed to dynamic information) described by low-level configuration features into affective labels and dimensions, and investigating how humans and automatic recognition models use this low-level posture description to discriminate between the affective states and dimensions. Motion capture systems were chosen over vision based systems in order to more easily obtain precise numerical data that describes the relations between the joint positions of the person's body in a three-dimensional (3D) space.

To create models for recognising posture according to affective states and levels of affective dimensions, postures are described according to low-level features (e.g., distances between joints, etc., explained in detail in Chapters 3-5) and labelled according to observer judgments. Posture judgment surveys are employed to determine whether affective states and levels of affective dimensions can be recognised by human observers from static images created from original motion capture data of acted and non-acted postures mapped to simple 3D avatars. In contrast to most affective computing research, the observers' labels are used to determine the ground truth of each posture as a step toward creating automatic recognition models that may act as a substitute for a human partner. Thus, observer judgments were chosen over other methods such as the actors' labels, self-report and physiological measures because the aim is to achieve recognition rates similar to human observers. The recognition models are built using machine learning techniques. Due to a lack of existing benchmarks in the affective computing field, the models are evaluated by comparing their performance rates to benchmarks defined in this research. Acted postures are examined first. In the second phase, non-acted postures are obtained while users play a video game as a step toward examining the real-world applicability of this type of system. In the third phase, an affective posture recognition system is built to evaluate how the system may perform when applied to *sequences* of non-acted static postures in a runtime situation where the postures to be recognised have not been manually selected.

Using a video game context for the second and third studies, the research presented in this thesis aims to validate the use of posture as an input modality for automatic affective recognition systems. While the integration of multiple modalities into one recognition system is recognised as an important, overall goal of affective computing, it is outside the scope of this thesis. As will be shown in Chapter 2, the study and understanding of posture in affective computing is still new and emerging while the study of other modalities is more advanced. Thus, it is important first to determine the information that can be recognised from posture in order to assess how much emphasis to place on posture alone within a multimodal affective recognition system.

1.4 Thesis Structure

- *Chapter 2:* Background

  This chapter highlights the main issues in affect recognition and outlines the novelty of the research presented in this thesis. It provides a discussion on the state-of-the-art in both human and automatic recognition of affect from whole body postures. Affect can be recognised from a variety of modalities which are briefly discussed throughout the chapter. However, the main focus is on affect recognition via the nonverbal communication channel of body posture and the need for a new modelling and evaluation approach in affect recognition in general.

- *Chapter 3:* Research Hypotheses and Methods

  This chapter explains the hypotheses of this research and describes the method adopted to prove the hypotheses. This method involves creating a set of benchmarks using a random repeated sub-sampling method on the agreement levels between subsets of human observers to evaluate the automatic recognition of affective posture. Both basic and non-basic affective states and levels of affective dimensions are examined. Also discussed are the modelling techniques implemented for building and testing automatic affective posture recognition models. Three case studies are outlined which involve collecting affective posture data to be described by a set of low-level posture description features and training and testing automatic recognition models built with this data.

- *Chapter 4:* Case Study 1: Modelling Acted Basic Emotions

  This chapter describes the acted postures study which addresses basic emotions from acted, stereotypical affective postures. The motion capture participants are called actors because they were explicitly instructed to enact affective postures for four basic emotion categories, *angry, fearful, happy,* and *sad.* After obtaining ground truth labels for a set of static postures that have been manually extracted from the motion capture data, the modelling technique described in Chapter 3 is implemented for creating and validating the recognition models. The posture features important for classifying affective posture are analysed for both human observer and automatic recognition of emotion categories and affective dimensions. The results obtained are discussed and reflected upon at key points throughout the chapter.

- *Chapter 5:* Case Study 2: Modelling Non-Acted Affect in a Video Game Scenario

  This chapter describes the non-acted postures study which investigates affective posture in a non-acted, natural context instead of an acted situation. The participants have not been apprised of the purpose of the study as an attempt to collect unsolicited postural displays of affect, resulting in postural expressions that are typically more subtle than in the acted postures study. Instead of basic emotions, *concentrating,* *defeated, frustrated* and *triumphant* affective states are examined as well as affective dimensions. Static postures manually extracted from the motion capture data have been evaluated by a small group of observers across several evaluations per observer. As in Chapter 4, automatic recognition models are built and tested, and important features are identified and analysed for both the observers and the automatic models. The results obtained are discussed and reflected upon at key points throughout the chapter.

- *Chapter 6:* Case Study 3: Real Time Affective Posture Recognition

  Using the video game scenario described in Chapter 5, the aim of Chapter 6 is to evaluate how the affective posture recognition system performs in a runtime situation on a sequence of postures that have not been manually extracted. The system is built by combining the classification method used in Chapter 5 with a decision rule that has been developed in this research.

- *Chapter 7:* Conclusions

  This chapter summarises the conclusions drawn from the case studies presented. It contains a discussion on the overall findings and contributions of the research project. The limitations encountered due to the various methods and techniques employed are examined, and ideas on possible solutions are contemplated as directions for future work.

Chapter 2

Background

The role played by affect in human development and everyday functioning is well recognised [IASF02]. Indeed, its importance in intelligent and social behaviour has been accepted and researched within several fields including psychology, neuroscience, biology and affective computing. Within the affective computing field, emotion theories are applied to create technologies that are more aware of human emotions in order to make the technologies more able to react and respond appropriately.

Several issues arise in determining how to create effective models for the automatic recognition of affect. How is affect described? What role does bodily information play in affect recognition? Which communication modalities should be used to recognise affect in humans and how should they be modelled? Are there universal aspects to affect expression and recognition or are they affected by human factors? Is it possible to create automatic recognition systems capable of recognising affective states and dimensions as well as human observers can? How can the ground truth for affective expression data be reliably determined? The purpose of this chapter is to examine these issues and present the state-of-the-art in the field.

The chapter is organised as follows: Section 2.1 briefly outlines two main schools of thought for describing affect and emotion and how they may be organised. Section 2.2

discusses the motivation for understanding and studying affect expression and recognition from whole body posture. A discussion on how specific cues of bodily expressions may be mapped to specific affective states is provided in Section 2.3. Section 2.4 briefly explores the universality argument surrounding affect recognition and conveyance. Section 2.5 reports on the state-of-the-art of automatic affective recognition systems. Section 2.6 tackles some of the methodological issues discussed in the literature that exist for creating automatic affective expression recognition systems. Section 2.7 ends the chapter with a summary of the main issues encountered in the field of affective computing and how these issues have informed the work in this thesis, thus highlighting the novelty of this research project.

2.1 Theories of Affect: An Overview

> *"The mysticism of ineffability and freedom that surrounds emotions may be one reason why the psychology of emotion and feeling has advanced so slowly over the last 100 years"* [Fri88].

There is a wealth of research on emotion in psychology and the discussion of emotion and affect extends back hundreds of years. Given this understanding, the question to be asked in this section is: how should affective expressions be described? Research in the affective computing field involving classification of affective expressions typically focuses on two schools of thought, discrete emotion categories and affective dimensions [VK02][Laz91]. A complete discussion on emotion theories is beyond the scope of this thesis. The remainder of this section is focused on an overview of some of the main theorists and components of both schools of thought.

Discrete emotion theories

Discrete emotion theorists consider emotions as instances of unique and separate states, e.g. anger or happiness. Furthermore, many discrete emotion theorists also consider a number of emotions as *basic* or *primary* yet there is no consensus on either the number of categories or which emotions are considered basic [OT90]. According to Fox [Fox08], *“the ‘basic’ emotions might be those that are critical for the survival of: (a) the species - biological criterion, (b) the society - social criterion, or (c) the self - psychological criterion.”*

The core concept in the discrete emotions approach is that obvious behavioural and physiological responses occur after an emotion has been elicited [Fox08]. The activation of an emotion is likened to a computer program in that the triggered emotion causes a series of responses by various systems [Tom62][Tom63]. This model of emotion activation is depicted in a schematic model adapted from Levenson [Lev94] shown in Figure 2.1.

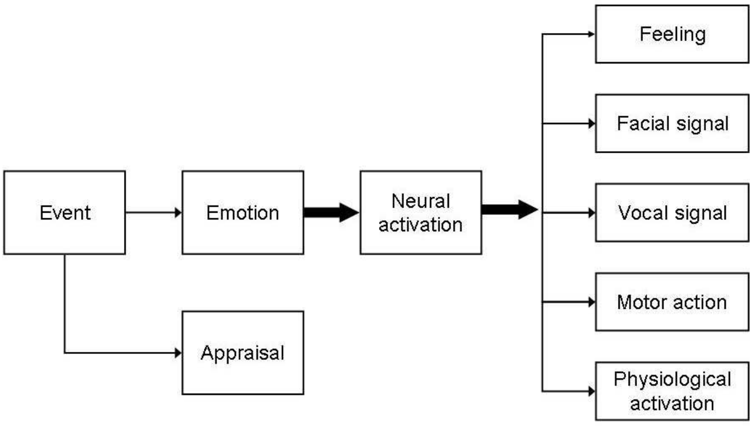


Figure 2.1: A schematic model of emotion depicting the key concept of the discrete emotions approach. Taken from Fox [Fox08] and adapted from [Lev94]

Table 2.1 lists some of the discrete, basic emotions theorists. As can be seen in the Table and noted by Fox [Fox08] and Ortony and Turner [OT90], many of the discrete emotions theorists have accepted anger, fear, happiness and sadness in their sets of basic emotions. Ortony and Turner go on to point out that an emotion term in one theorist’s list, such as anger or happiness, may be termed differently in other theorists’ lists. For example, anger may be labelled as anxiety or rage, and happiness may be referred to as joy or elation. When taking this view, the lists of basic emotions are not as different as they first appear.

Ortony and Turner go on to question, why basic emotions? They state that a common

Table 2.1: Several theorists' lists of "basic" emotions. Adapted from [OT90].

Reference	Fundamental emotions
[Arn60]	Anger, aversion, courage, dejection, desire, despair, fear, hate, hope, love, sadness
[EFE82]	Anger, disgust, fear, joy, sadness, surprise
[Gra82]	Rage and terror, anxiety, joy
[Iza71]	Anger, contempt, disgust, distress, fear, guilt, interest, joy, shame, surprise
[Jam84]	Fear, grief, love, rage
[McD26]	Anger, disgust, elation, fear, subjection, tender-emotion, wonder
[OJL87]	Anger, disgust, anxiety, happiness, sadness
[Pan82]	Expectancy, fear, rage, panic
[Plu80]	Acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise
[Tom94]	Anger, interest, contempt, disgust, distress, fear, joy, shame, surprise
[Wat30]	Fear, love, rage

reason is to be able to discuss emotion observations. The ability to discuss emotions in terms of labels can be important when evaluating recognition performances of humans. For instance, observations such as *"the fact that some emotions appear to exist in all cultures [...] [and] that some emotions appear to be universally associated with and recognizable by characteristic facial expressions"* [OT90]. The universality of emotions debate is discussed in more detail in Section 2.4.

Dimensional theories

The dimensional approach to affect is focused on how the world is experienced. Many dimensional theorists conform to the view that emotion labels are learned [Fox08]. *"An underlying assumption is that emotions are defined to a large extent by the verbal labels we use to describe them. [...] A genuine problem for science, however, is that these everyday labels often describe very broad and 'fuzzy' semantic categories that may not necessarily represent facts of nature [RF99]."* [Fox08].

As a response to this problem, dimensional theorists consider affective states as existing in a continuous, multidimensional space with the dimensions being bipolar and independent. Psychological research over the last century has reported that there are three dimensions of affect: valence (levels ranging from pleasure to displeasure), arousal (levels of alertness ranging from calm to excited) and dominance/potency (levels of control over an event) which cover the majority of affect variability [Meh68][OST57][Wun73][Dav64]. However, in a recent article, Fontaine and colleagues [FSRE07] asserted that understanding which affective dimensions are represented in the emotional domain remains an open issue.

A lot of the psychological research to date has focused mainly on the valence and arousal dimensions [FSRE07]. Indeed, the valence and arousal dimensions have been shown to be present in all cultures [Wie95]. Several bipolar dimensions were found in all the languages studied in [OMM75] as reported in [Fox08]. These bipolar concepts seem to reflect a combination of discrete emotion categories and dimensions, e.g., 'I was very happy/very sad'. As discussed by Feldman Barrett [Fel04], in reporting subjective experiences, some individuals are shown to use broad, global terms while others tend to use discrete emotion labels. Feldman Barrett [Fel06] further describes results from eight different studies conducted in her lab over a span of 10 years, stating that while *all* the participants of these studies described their emotional responses in terms of pleasure and displeasure, there were wide ranging differences for individuals in their use of discrete emotion labels.

The two-dimensional model of Watson and Tellegen [WCT88] proposes two dimensions of experience, a horizontal dimension ranging from low to high positive affect, and a vertical dimension ranging from low to high negative affect, as reflected in Figure 2.2. Their view is that these two dimensions (both are dimensions of valence) are completely separate and unrelated, i.e., orthogonal. In yet another view, represented in Figure 2.3, Thayer's model also comprises two independent dimensions, yet instead of valence, both dimensions represent levels of arousal: tension along the vertical dimension and energy along the horizontal dimension [Tha89].

Differing from Watson and Tellegen's and Thayer's views, Russell's view [Rus80] is that all emotions can be represented as consisting of variations of pleasantness, i.e., valence as

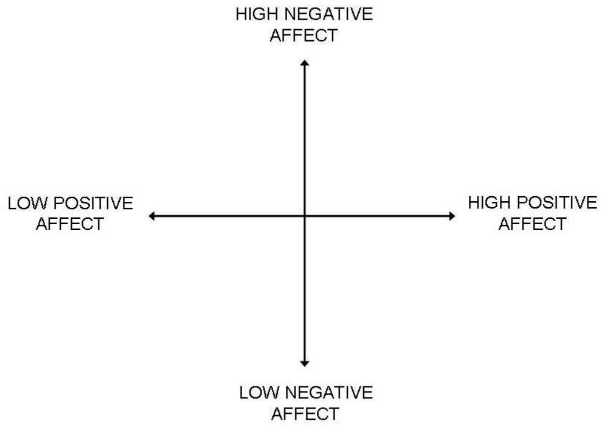


Figure 2.2: Watson and Tellegens's two-dimensional model of affect. From [WCT88]

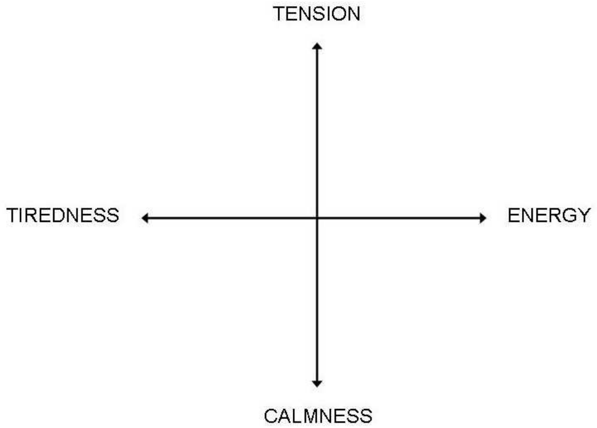


Figure 2.3: Thayer's two-dimensional model of affect. From [Tha89]

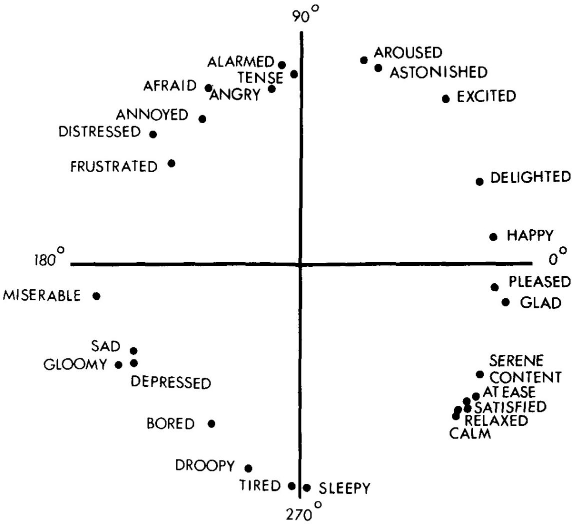


Figure 2.4: Russell's circumplex model of affect. Activation (also known as arousal) is presented as the vertical dimension and valence is shown on the horizontal dimension and discrete emotion labels are mapped to the dimensional space. From [Rus80]

the horizontal dimension and variations of activation, i.e., arousal as the vertical dimension. Russell does not consider the discrete and dimensional spaces as independent. Instead, they are linked and discrete categories can be mapped onto the dimensional space and create a 'circumplex model' (i.e., a circular structure) as shown in Figure 2.4. This mapping can be used to aid discussions and comparisons between modalities.

The dimensional approach has focused on experienced affect through subjective, self-reported affect (either through verbalising experiences or filling out standardised questionnaires) or physiological responses in order to identify the dimensions. However, recent neuroscience evidence indicates that both experience and perception may occur when viewing another person's actions [DG99][RFGF96] meaning that these dimensions can be implemented to gain an understanding of how human observers perceive and interpret affect.

Research exists for mapping facial expressions to affective dimensions [Rus97][Bre03]. Breazeal [Bre03] has mapped a series of facial expression photos onto Russell's [Rus80] arousal-valence dimensions. However, in the realm of bodily expressions, while Paterson and colleagues [PPS01] have mapped arm movements into affective dimensions (discussed in more detail in Section 2.3), research mapping whole body posture into affective dimensions has not been found. This is important in affective computing for comparing systems to understand if body posture can be discussed in terms of affective dimensions as can be done for facial expressions.

Integrating discrete and dimensional approaches

While the debate between the use of dimensions and discrete emotion categories continues, emotion science research appears to make use more and more of a hybrid approach to the description of affective experience. *"It is difficult at this point to determine whether discrete emotions perspectives or dimensional perspectives provide a more accurate view of emotions and moods and how they are structured. A key challenge is to examine these different research traditions and determine whether the empirical evidence from both approaches can be integrated in a sensible way to provide a comprehensive understanding of affect"* [Fox08].

Identifying bodily expressions as combinations of discrete labels and levels of affective dimensions may provide a more complete and accurate description of the affective state exhibited. A single label may not always be enough to reflect the complexity of the affective state conveyed. Indeed, in the realm of affective computing, research is now focusing on an integration of the discrete emotions and affective dimensions approach, as evidenced by recent studies examining the effectiveness of recognising combined facial expressions and bodily expressions [GP06][CPM$^+$09].

2.2 The Importance of the Body in Affect Recognition

> *"Considering the emotional value of bodily expressions, it is somewhat surprising that the study of perception of whole-body expressions lags so far behind that of facial expressions."* [VdSRdG07]

Armed with a better understanding of how affect and emotion are viewed, attention can now be focused on how affect and emotion are expressed and recognised in humans. Affect expression occurs through combinations of verbal and nonverbal communication channels such as vocal prosody, eye gaze, facial expression, and body posture, among others [Pic98]. Yet given this wide range of modalities, the majority of research on nonverbal affect recognition has concentrated on recognising facial expressions in particular [Wal98][Ekm94][EMA$^+$02][Rus94] [ADGY07]. Thus, a fair amount is known and accepted about affective facial expressions, such as some of the ways in which they are conveyed and recognised, their neurobiological bases [dG06], and an understanding towards how to code them [EF78]. As explained in Chapter 1, there is a well established, well known coding system for facial expressions, (FACS), that took nearly a decade to develop [EF82] in the 1970's by Ekman and Friesen [EF78]. The examination of facial expression perception has been the basis for learning how humans process affect neurologically [Ado02].

The same cannot be said for affective bodily expressions. Only recently has affective computing research focused on body movement and posture. Indeed, in a 2009 article, de Gelder [dG09] states that 95% of the studies on emotion in humans has been conducted using facial expression stimuli. Research using information from human voice, music and environmental sounds make up the majority of the remaining 5%, with research on whole-body expressions comprising the smallest number of studies.

What role does bodily information play in affect recognition? Bodily expressions have been recognised as more important for nonverbal communication than was previously thought [MF69][Arg88]. According to Mehrabian and Friar [MF69] and Wallbott [Wal98], changes in a person's affective state are also reflected by changes in body posture. Mehrabian and Friar found that bodily configuration and orientation are significantly affected by a communicator's attitude toward her/his interaction partner. Ekman and Friesen [EF67][EF69b] conjecture that postural changes due to affective state aid a person's ability to cope with the experienced affective state. In fact, as seen in the behavioural studies presented in the remainder of this section, some affective expressions may be better communicated by the body than by the face [Arg88][Bul87].

Darwin [Dar72] surmised that people are able to control bodily movements during felt emotions. Ekman and Friesen [EF69a] refer to this as the *"face>body leakage hypothesis"*. However, they conclude that there is a lack of evidence to support Darwin's claim by stating that *"most people do not bother to censor their body movements"*. Specifically, the authors hypothesised that in Western cultures more emphasis is placed on an individual's facial activity during conversation than on bodily activity. The individual may therefore make a conscious attempt to control her facial expressions. Thus, it may be advantageous instead to note a person's body posture when in situations where deception may be a concern.

Ekman and Friesen [EF74] carried out a study to evaluate how people control facial and bodily expressions during deception. Participants consisted of twenty-one female nursing students at the start of their studies. The participants were video recorded while viewing a stress film and a pleasant film separately. During the stress film the participants were instructed to try to convince an interviewer that they were viewing a pleasant film. The same instructions were given when the participants really were viewing the pleasant film. In questioning the participants after the deception task, the majority (17 out of 21) reported that they focused on trying to control facial expressions more often than bodily expressions.

Furthermore, in the same study, Ekman and Friesen hypothesised that when behaviour is appraised by observers as either deceptive or honest, observer ratings for the body will be more accurate than ratings for the face. Using the videos collected in the first part of the study, judgment tasks were implemented in which observers were asked to rate the recordings of the nonverbal behaviour as either deceptive or honest. The recordings consisted of face only and body only samples. 103 university students participated. The results showed significantly higher recognition rates for the body (63.5%) than for the face (47.69%) when the judgments indicated deception. These results attest to the need for recognising bodily information as well as facial information. As stated by the authors, the necessity now is to measure and identify the specific body information exhibited during each of the tasks.

Studies have also examined the role played by the body in communicating emotions when human observers are presented with affective displays containing a combination of facial

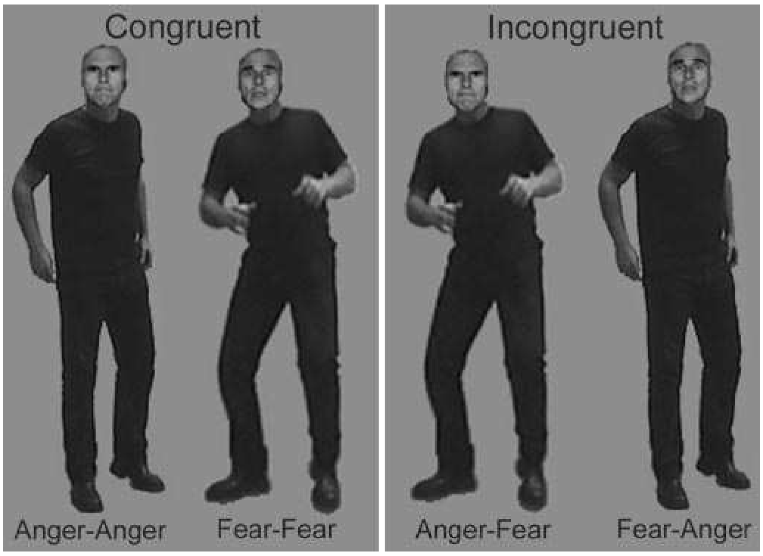


Figure 2.5: Examples of the congruent and incongruent face and body posture images evaluated in [MvHdG05].

expressions and posture or movement. In agreement with a study by Ekman and Friesen [EF69a], McClenney and Neiss [MN89] hypothesised that the body may be a less controllable channel of communication and thus is susceptible to spontaneous affect leaks. McClenney and Neiss's study examined recognition accuracies of *happiness, sadness* and *anger*. The goal was to ascertain the recognition rates of these emotions when the stimuli contained either face only or body only information. The participants, 36 female undergraduate students, were asked to relive either a happy, sad or angry experience from their lives after which an interview session was videotaped. A group of 37 undergraduate students were recruited to view the video recordings and rate them on a five point Likert scale for the emotions. The results on the comparison between the communication channel viewed by the observers showed that sadness (F:(5, 175) = 5.69, $p < .01$) and anger (F:(5, 175) = 8.41, $p < .01$) were more accurately recognised from body information than from facial information with no differences seen for happiness. The authors attributed the recognisability of anger to the symbolic gesture of clenched fists.

According to studies by de Gelder and colleagues, body posture may also provide more information than the face in the case of fear and anger [MvHdG05] and fear and happiness [VdSRdG07]. In an article by de Gelder [dG06] the author postulates that for fear specifically, by evaluating body posture, it is possible to discern not only the cause of a threat but also the action to be carried out (i.e., the action tendency), while the face communicates only that there is a threat. In a neuroscience study considering both fear and anger, de Gelder and colleagues [MvHdG05] examined the significance of emotional body posture on facial expression recognition of incongruent displays. The authors hypothesised that facial expression recognition is directly affected by affective bodily expressions that are presented at the same time. Sets of corresponding and opposing, realistic-looking fear and angry face-body images (e.g., Figure 2.5) were created using photographs. The body stimuli were created from the researchers' own, previously used data of individuals displaying a wide variety of arm positions. The face stimuli were taken from Ekman and Friesen's database [EF76]. The same face and body stimuli were prepared as individual, separate stimuli to create a control condition. Twelve observers participated in the study which comprised two experimental conditions. The compound face-body stimuli were used in one condition and the individual, isolated stimuli (e.g., face or body only) were used in the control condition. The observers were instructed to evaluate the emotion displayed by the face for the compound stimuli (and the individual stimulus in the single stimulus condition). The behavioural findings indicate that when the affective information displayed by the two channels is incongruent, body posture is the influencing factor over the recognised emotion. There was a significant decrease in facial expression recognition when face and body information were incongruent (67% with a reaction time of 840 ms) than when they were congruent (81% with a reaction time of 774 ms). The results were replicated by de Gelder and colleagues in a more recent study [VdSRdG07] aimed at extending the set of emotions by investigating fear and happy congruent and incongruent face-body images. Using a newly developed set of images and a similar method as described above, compound stimuli were created as a five step continuum from fear as one extreme morphing into happy as the other extreme. Again, observers were asked to evaluate the emotion expressed by the face only. The results showed that the affective information displayed by the body has a significant affect on the affective state recognised from the face.

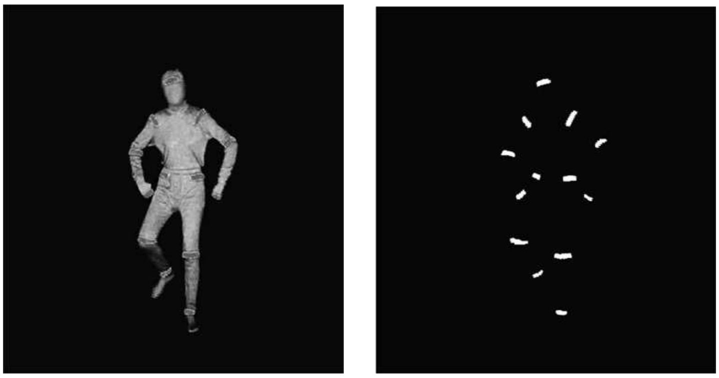
Preliminary studies by Pollick and colleagues [PPJ02][PPM04] examined high, low and neutral saliency facial expressions (saliency was determined according to recognition accuracy) combined with motion captured arm movements representing knocking motions for *angry, happy, sad* and *neutral*. The facial expression information was computer generated, and was obtained from a pre-existing database. In one condition, affective face information was paired with a neutral body and vice versa. In the second condition, face and body information were congruent. The results showed that when the modalities were viewed separately, the movement information for angry was more heavily weighted than facial information [PPJ02] and that angry knocking motions were perceived as more intense and with higher recognition rates than low saliency angry facial expressions [PPM04].

As discussed in this section, the body in general appears to be a sufficient nonverbal communication channel from which affective information can be accurately recognised. In the following section, as suggested by Ekman and Friesen [EF74], the discussion turns to studies that have examined individual elements of bodily information to which the recognition of particular affective states may be attributed.

2.3 Mapping Bodily Posture and Movement into Affect

What type of information about the body is necessary for recognising the affective state displayed? According to a neuroscience study by Giese and Poggio [GP03], there are two separate pathways in the brain for recognising biological information, one for form information (i.e., the description of the configuration of a stance) and one for motion information. Findings of neuropsychological and neurophysiological research indicate that form information can be instrumental in the recognition of biological motion through point-light displays [HH06][PWD06].

A recent study by Atkinson and colleagues [ADGY07] determined that both form and



(a) (b)

Figure 2.6: (a) Full-light and (b) patch-light examples. Taken from [ATD04]

motion signals are assessed for affect perception from the body. Participants viewed short clips of people acting out specific emotions as well as displaying affect through everyday actions such as bending and walking, in either full-light (Figure 2.6(a)) or patch-light (Figure 2.6(b)).[1] The clips were shown upright and upside-down, forward-moving and reversed. Results showed that for all conditions, affect could be recognised at above chance levels. However, recognition rates were significantly lower for the upside-down, reversed, patch-light displays, indicating a difficulty in recognising a human form when the information is presented in a non-human-like configuration. The authors conclude that these results indicate the importance of form-related, configurational cues of posture for recognising emotion.

Many psychological studies over the decades have examined bodily configurations to evaluate if specific features of the body can be identified that may be attributed to the recognition of specific affective states. These studies have sought to understand some of these features according to two main levels of bodily detail. Several studies have focused on gross body configurations while fewer studies have focused on the role played by fine-grain, individual postural features. Overall details of some of these studies are listed in Table 2.2. Table 2.3 lists the specific affective states and features examined in each study. The study by Atkinson and colleagues [ADGY07], described in the previous paragraph is listed in the

[1] In **full-light** conditions, the entire stimulus is visible. In **patch-light** conditions, the stimulus information is only partially preserved [ADGY07].

Table 2.2: Research mapping bodily features into affect. (Obs. = observers; + = attitude, not affectve states; comp. gen. = computer generated)

Reference	**Affective states**	**Posture or movement**	**Acted or non-acted**	**Perception study obs.**	**Stimuli Type**	**Num of Samples**	**Ground truth**
[AWH92]	2	Both	Acted (56)	6	Video	not reported	Actor
[Jam32]	not defined	Posture	Acted (1)	3	Photos	347	Observer
[Wal98]	14	Posture	Acted (12)	14	AV	224	Actor
[Cou04]	6	Posture	Comp. gen.	61	images	528	Observer
[dM89]	9 & 3^+	Movement	Acted (3)	85	Video	96	Observer
[MF69]	$^+$	Movement	Acted (48)	3	one-way mirror	384	Observer
[ADGY07]	5	Movement	Acted (36)	32	patch- & full-light	60 & 60	Actor
[PPS01]	10	Movement	Acted (2)	14	point-light	118	Actor

Table 2.3: Details of the affective states and body features studied for the research outlined in Table 2.2

Reference	Affective states	Features	Feature details
[AWH92]	Warm, threatening	Diagonal poses	-
		Arabesques	-
		Arms:	% round, % straight, % angular
		Movement:	% round, % straight, % angular
[Jam32]	Not pre-defined	Head, trunk, feet, knees, arms	"Each factor except the knees was varied separately in every one of its possible ways, and then in combination with variations of the other factors."
[Wal98]	Elated joy, happiness, sadness, despair, fear, terror, cold anger, hot anger, disgust, contempt, shame, guilt, pride, boredom	Upper body:	Away, collapsed
		Shoulders:	Up, backward, forward
		Head:	Downward, backward, turned sideways, bent sideways
		Arms:	Lateralized hand/arm movements, stretched out frontal, stretched out sideways, crossed in front of chest, crossed in front of belly, before belly, steemed to hips
		Hands:	Fists, opening/closing, back of hands sideways, emblem, self-manipulator, illustrator, pointing
[Cou04]	Anger, disgust, fear, happiness, sadness	Abdomen twist, chest bend, head bend, shoulder swing, shoulder adduct/abduct, elbow bend, weight transfer	A low-level approach using joint angles. Two different degrees for each feature except weight transfer. Degrees used are dependent on emotion label.
[dM89]	Joy, grief, anger, fear, surprise, disgust interest, shame, contempt, sympathy, antipathy, admiration	Trunk movement:	Stretching - bowing
		Arm movement:	Opening - closing
		Vertical direction:	Upward - downward
		Sagittal direction:	Forward - backward
		Force:	Strong - light
		Velocity:	Fast - slow
		Directness:	Direct - indirect
[MF69]	Attitude	Eye contact	-
		Distance from addressee:	Straight ahead, lateral
		Orientation to addressee:	Head, shoulders, legs
		Openness:	Arms, legs
		Relaxation:	Hand, foot, trunk (all measured by backward & sideways angles of lean)

Table 2.4: The discriminating features for the two affective states studied in [AWH92]

Aronoff et al [AWH92]

Affective states	**Discriminating features**
Warm	Roundedness (and lessened diagonality and angularity) of both arms and movement, more static and moving arabesques
Threatening	Diagonality and angularity of both arms and movement, more diagonal poses, more straight arms

second to last row of Table 2.2.

Using acted ballet movements and postures Aronoff and colleagues [AWH92] (first row of both Tables) concluded from their study that angular and diagonal postural configurations can be adopted to signify threatening behaviour, while rounded postures are intended to demonstrate warmth (results are summarised in Table 2.4). Other studies have acknowledged the important role that leaning direction plays in the perception of a particular affective state [HR83][Meh68][Jam32]. In James' behavioural study [Jam32] (second row of both Tables), he discovered the importance of more specific whole body features of posture (summarised in Table 2.5), such as leaning direction and openness of the body and head position, such as up, down, and tilted for discriminating between a variety of affective states.

Listed in the third row of Tables 2.2 and 2.3, Wallbott carried out a study to examine which postural cues of body configurations afford humans the ability to distinguish between specific emotions. Wallbott videotaped and audio recorded 12 professional actors displaying 14 emotions (*elated joy, happiness, sadness, despair, fear, terror, cold anger, hot anger, disgust, contempt, shame, guilt, pride* and *boredom*) from scenarios aimed at eliciting the emotions. A group of observers viewed the 224 recordings and coded all bodily activity displayed. Two additional coders were recruited to construct a set of postural cues that defined the configuration of the body, referred to by Wallbott as 'a category system', from the previous coders' descriptions. The system consisted of body movements, postures and

Table 2.5: The discriminating features in [Jam32]. Adapted from James [Jam32]. *a)* The weight of the body is thrown on the forward foot. Weight on the back foot is both unnatural and incongruent. *b)* Arms backward is unnatural and when in that position are either ignored by the observer or else their meaning is the same as 'arms at the side.' Body weight is on the rear foot or is equally distributed on both feet with wide base. *c)* Bent knees change the withdrawal to a contraction which is further specified as weakness, helplessness. *d)* The feet and arms point in the direction of, and the trunk and head away from, the observer

	James [Jam32]	
	I. *Trunk, head, and arms forward, knees straight*[a]	
a. Approach	with palms up	acceptance, offering, coaxing, supplication, beseeching - all with humbleness
	with palm outward	active repulsion, avoidance, holding off, opposition, command, disapproval
	with palms down	soothing, calming, blessing; groping, balancing in movement; with slightly bent knees, reaching for support
b. Contraction	palms up	servitude, surrender (the figure offers itself)
	palms outward	dejection, grief, anguish, shame, defeat (with a refusal of aid or sympathy)
	II. *Trunk and head backward, arms forward, knees straight*[b]	
a. *Withdrawal*[c]	palms up	prayer, proud offer or acceptance
	palms outward	extreme negation, exaggerated refusal, repulsion, disgust; with bent knees sudden recoil, horror, withdrawal from a dangerous position, arrogant refusal, pride
	palms down	blessing, proud dismissal, refusal
b. Expansion	palms up	joyful offering or receptiveness, welcome with great pride; with knees bent prayer
	III. *Trunk, head turned, arms forward*[d]	
a. Withdrawal	palms forward	emphatic refusal, utter disgust or scorn or disdain, the object of observer is cast aside

Table 2.6: The discriminating features reported for 11 of the affective states studied in [Wal98]. (x = no significant results reported)

Wallbott [Wal98]	
Affective states	**Discriminating features**
Cold anger	Lateralized hand/arm movements, arms stretched out frontal
Hot anger	Lifting shoulders, lateralized hand/arm movements, arms stretched out frontal, hands opening/closing
Boredom	Collapsed upper body, head bent backwards
Contempt	x
Despair	Shoulders forward, hands opening/closing
Disgust	Shoulders forward, head downward, arms crossed in front
Fear	Shoulders forward, hands opening/closing
Guilt	x
Happiness	x
Elated joy	Lifting shoulders, head bent backwards, arms stretched out frontal, hands opening/closing
Pride	Head bent backwards, arms crossed in front
Sadness	Collapsed upper body
Shame	Collapsed upper body
Terror	Arms stretched sideways

movement quality. Inter-observer reliability was determined by computing the percentage of agreement between the two coders. The categories for which >75% agreement was obtained were considered in the analysis of the category system. One-way ANOVAs were implemented to detect if differences exist in how people evaluate posture in order to distinguish between emotions and the relevance played by the features in discriminating between emotions. The results showed significant differences for 17 of the 26 categories. The results are summarised in Table 2.6 and confirm the importance of and information carried by specific postural cues to discriminate between emotions. However, Wallbott himself stated that this was an initial study and asserted that additional studies need to be carried out. In particular, Wallbott stated the need for studies examining spontaneous (e.g., non-acted) expressions and cross-cultural studies.

In another study, Coulson [Cou04] (shown in the fourth row of Tables 2.2 and 2.3) attempted to ground basic emotions into low-level static features that describe the configuration of posture. Computer generated avatars expressing Ekman's [EF69b] six basic emotions (*angry, fear, happy, sad, surprise,* and *disgust*) were employed to examine postural elements necessary for attributing a specific affective state to body posture. His proposed body description, summarised in Table 2.3, comprises six joint rotations (head bend, chest bend, abdomen twist, shoulder forward/backward, shoulder swing, and elbow bend). Judgment survey results showed that concordance rates between observers reached 80% for some postures in associating angry, happy, and sad labels. Coulson then used a statistical method to determine the role each joint rotation played in determining which emotion label was associated to each posture, and found that specific bodily features could be used to differentiate between the emotions studied. Refer to Table 2.7 for a complete list of which features were predictive for which emotions. While the overall findings were above chance level (16.7%) and all of the postures were kinematically plausible, according to Coulson himself, *"the complexity of the stimuli meant that some postures looked rather unusual"* [Cou04].

DeMeijer [dM89], listed in the fifth row of Tables 2.2 and 2.3 carried out a study to examine two questions: i) if specific gross body movements were indicative of specific emotions

Table 2.7: The discriminating features reported for the affective states studied in [Cou04]

Coulson [Cou04]	
Affective states	**Discriminating features**
Anger	Backward head bend, absence of backward chest bend, no abdominal twist, arms raised forward and upward.
Fear	Backward head bend, no abdominal twist, forearms raised, weight transfer either backward or forward.
Happiness	Backward head bend, no forward movement of the chest, arms raised above shoulder level and straight at elbow.
Sadness	Forward head bend, forward chest bend, no abdominal twist, arms at side of the trunk.
Surprise	Backward head and chest bends, any degree of abdominal twisting, arms raised with forearms straight.

and ii) which movement features accounted for these attributions. To this aim, seven movement dimensions, listed in Table 2.3, were utilised. The movements of three dancers were videotaped. The dancers did not explicitly enact the emotions, instead, they were instructed to perform specific movements derived from variations of the movement characteristics. As the face was not blurred for the judgment task, the dancers were instructed to maintain a neutral facial expression. A separate group of observers viewed the movements and judged each according to 12 four-point scales. The instructions were to rate each movement according to its compatibility with each emotion (i.e., nine emotions: *anger, contempt, disgust, fear, grief, interest, joy, shame, surprise*; three emotional attitudes: *admiration, antipathy, sympathy*). Answering the first research question of *if* specific movements denoted specific emotions, the results showed that *"all emotion categories, except disgust were attributed to certain movements"* with admiration attributed to the most (15) and interest attributed to the least (3) [dM89]. The results for the second question, summarised in Table 2.8, did indeed indicate that specific features could be attributed to specific movements. Trunk

Table 2.8: The discriminating features reported for the affective states studied in [dM89].

de Meijer [dM89]

Affective states	**Discriminating features**
Anger	Bowing, slow, strong, downward, backward, opening
Contempt	Bowing, backward
Disgust	Bowing
Fear	Bowing, downward, backward, strong, fast
Grief	Bowing, slow, downward, closing
Interest	Stretching, forward, light, slow, opening
Joy	Stretching, upward, strong, fast, forward, open
Shame	Bowing, downward, light, slow, backward
Surprise	Stretching, backward, fast
Admiration	Stretching, upward, forward, open
Antipathy	Bowing, backward
Sympathy	Stretching, forward, open, light, slow

movement, stretching or bowing, was the most predictive for all emotions except anger. In fact, trunk movement was found to distinguish between positive and negative emotions.

Other studies have focused on classifying affective body movements according to affective dimensions. With a focus on recognising emotion from biological motion, a study by Paterson and colleagues [PPS01] (the last row of Table 2.2) aimed to map part of the body, head and arm movements, to an affective space. Their aim was to examine not only how well affect may be recognised but also the structure of the representation of affect. To begin, two actors were asked to read stories aimed to elicit 10 affective states (*afraid, angry, excited, happy, neutral, relaxed, sad, strong, tired* and *weak*), after which the actors were motion captured while performing drinking and knocking motions. Each action was repeated three times by each actor for each of the affective states. They first obtained human observer judgments of the entire motion corpus using a forced choice experimental design. The motions were viewed as point light displays from a sagittal perspective. The results showed that the overall recognition rate across the 10 emotions was a mere 30% but still well above chance level (10%). The low performance of the participants was attributed to some motions being misjudged as a similar movement, e.g., weak movements were often judged as either weak, sad or tired. To construct the affective space, individual difference scaling (INDSCAL) was applied to a set of dissimilarity measures obtained from the observer judgments. A 2D affective space was obtained. Approximately 87% of the variance was accounted for by the two dimensions (70% for dimension one and 17% for dimension two). The mapping was shown to reflect a circumplex model of affect with levels of arousal depicted on the first dimension and levels of valence depicted on the second dimension. These results show that similar to research on the structural representation of *experienced* affect, valence and arousal dimensions are also used by human observers when describing affective posture expressions. The high percentage of variance covered by the arousal dimension may indicate that arousal is considered more important by the observers.

Overall, the results of these studies show that particular features of bodily expressions can indeed be reliably attributed to a variety of different affective states. However, there are still some important issues to tackle. In particular, as shown in the acted/non-acted column of Table 2.2, all of the studies except Coulson's evaluated bodily information that had been acted. Coulson used computer generated postures, which may be considered contrived in some way instead of natural. Another issue to tackle is how to determine the ground truth labels of the affective expressions. A more detailed discussion on both of these issues is provided in Section 2.6.

The discussion in this section has focused on research which has attempted to *identify* cues used to map body posture and movement into affect. From a different perspective, research in dance for example, has focused on *defining* a set of posture and motion cues (e.g., an affective language) that are strongly affective. One model in particular that was created specifically as a method for mapping affect into human movement is Laban Movement Analysis (LMA)[2] created by Rudolph Laban [vL71]. LMA considers general features of human posture and movement, and comprises four key components for describing human

[2] LMA is Laban's entire work. Within LMA he defined a system for annotating bodily expressions, called Kinetography. From Kinetography, Labanotation was developed by Ann Hutchinson [Hut87]

movement, *Body, Effort, Shape*, and *Space*. The *Body* component describes the body itself in terms of which parts are moving. The volume of the movement is specified in the *Space* component, while the general form of the body is indicated through the *Shape* component. The *Effort* component identifies the dynamic aspects of the movement, such as force, speed, etc. Rozensky and Feldman-Honor [RFH82] cite a strength of the system being that it *"accurately depicts quantitative and qualitative aspects of body movement such as magnitude of movement, direction of movement [and] some spatial arrangements"*. As weaknesses, the authors state that the system *"does not provide for the precise quantification of small-scale NVBs [3]"* and that it is most suitable for dance movements. Although the use of LMA is a positive start for creating computational models of affective body expressions, it is not enough. Research is still required to validate the mapping of affect into more subtle, non-dance movements in order to be accessible to a wide variety of situations.

2.4 Are There Universal Aspects of Affect Expression and Recognition?

> *"Perhaps no issue has loomed larger or permeated the study of bodily communication than the extent to which such expressions are universal, which implies that they have a common genetic or neurological basis that reflects an evolutionary heritage shared by all humans, or relative, which implies that their form, usage, and interpretation are tied to individual cultures and contexts"* [BJM$^+$05].

As Picard [Pic98] points out, the manner in which humans convey emotion or affective messages in general is affected by many factors, such as age, gender, posture, culture, and context. One factor that is being given significant attention by the affective computing community is culture (defined as *"a shared system of socially transmitted behaviour that describes, defines, and guides people's ways of life"* [Mat05]). Indeed, the need for understanding how different people and cultures recognise and express affective body language

[3] Nonverbal behaviours

has become more and more important in a number of real-life affective computing situations. For example, embodied museum agents are gaining much attention [KGKW05][LAJ05]. Due to the diversity of people visiting, museums are a particularly appropriate arena in which to have an agent capable of recognising differences due to personality, culture, etc. E-Learning systems may also benefit by taking into account various human factors. Research in the UK found high dropout rates for eLearning due to 'culturally insensitive content' [DM06]. As systems replace humans, it is important that how they express and perceive non-verbal behaviours in a multi-cultural community is as natural as possible so that the user is not made uncomfortable.

There is evidence that the way in which affective states are expressed and controlled [MF69], as well as the interpretation of affect [KEGB03] is shaped by culture. Many researchers have used cross-cultural emotion recognition studies to validate evidence in favor of emotion universality [EMA+02]. For some emotions, there is cross-cultural support for the universality of many modes of non-verbal behaviour, including face, voice and body expressions, as well as changes in a person's physiology [Mes03]. However, the majority of the research on emotion universality has concentrated on the recognition of facial expressions using still photographs [Ekm94][EMA+02][Rus94]. An issue with this method is that the culture of the person in a photograph may bias the observer's judgments. The relationship between the expresser's culture and the perceiver's culture affects the perceiver's judgments [EMA+02] as people may have preconceived ideas about other cultures and how they behave.

Facial expressions

In the realm of facial expressions, much research has been conducted comparing emotion differences between various cultures. Friesen [Fri72] reported cross-cultural differences in comparing the facial expressions of Japanese and American participants while viewing both neutral films and films intended to cause stress. As a support to the universality of facial expressions, it was shown that both groups expressed almost exactly the same facial expressions when watching the films alone. However, differences were noted between the groups when the films were viewed with an authority figure present. The Japanese controlled their facial expressions more than the Americans. In particular, negative emotions were covered with a smile. These results are indicative of social display rules, meaning that in some contexts, facial expressions may be more controlled in some cultures than in others.

Tsai and colleagues [TCDFBP02] found more *similarities* in facial expressions than *differences* between European Americans and Hmong [4] Americans in facial expressions of relived emotions. The participants consisted of 50 Hmong American and 48 European American undergraduate students. The participants were asked to relive an extreme emotional experience after receiving a label and a description of six emotions: *anger, disgust, happiness, love, pride* and *sadness.* The duration of the relived emotion was signalled by the participants by the continuous pressing of a button. After reliving each emotion, the participants were asked to report the intensity of the felt emotion and the level of how well they were able to relive the emotion. The videotaped facial behaviour was then scored by three qualified FACS coders. The coders were not aware of the particular intended emotion they were scoring. After scoring, only the target behaviours (i.e., the facial expressions that would correspond to the emotion being experienced) were examined. Statistical tests were applied to the FACS-coded facial behaviours for each emotion. The authors reported that to their surprise, the results showed no significant differences between the two cultures in facial displays for pride, love, disgust or sadness. The only difference found was in how often non-Duchenne[5] smiles occurred during happiness. The European Americans displayed more non-Duchenne smiles than the Hmong Americans. As no significant results were found, the next step would be to examine facial expressions in a natural setting instead of an acted one. It is possible that the way they express a particular emotion is the same, but as in the previous study by Friesen [Fri72] what differs is how they react to a particular situation, i.e. what emotional expressions they will show in response to a certain situation and what intensity they will express.

[4] An Asian ethnic group originating from the mountainous regions of southern China.

[5] A non-Duchenne smile is defined as an unfelt smile (as opposed to a Duchenne smile which is considered to be a felt, or natural smile). It is believed to serve a social function and may be used to mask negative emotions [SH98][TCDFBP02].

Bodily expressions

In the realm of bodily expressions, Matsumoto and Kudoh carried out two studies designed to examine cross-cultural differences between Japanese and Americans in judging posture according to a set of semantic dimensions [KM85][MK87]. Based on other research [BNS75][MK83] (as cited in [MK87]), Kudoh and Matsumoto asserted that differences reported between Japanese and Americans are almost always due to status being a more important aspect of the Japanese culture than the American culture. *"With respect to postures, the status relationship between two interactants can be a primary dimension through which the semantic dimensions of each other's postures are interpreted"* [MK87]. Matsumoto and Kudoh's first study [KM85] investigated judgments on a corpus of verbal posture expressions from Japanese participants. Using the same methodology, the second study [MK87] investigated judgments from American participants. Using a principal-component factor analysis, the researchers hypothesised that the same factors would be extracted from the two sets of participants, but that the factor order would differ between the two cultures. The same corpus of 40 posture expression stimuli, created by a separate group of 372 Japanese university students [KM85], was used in both studies. The authors argued that even though the corpus was developed in Japan, the posture descriptions were not culture-specific. To create the corpus, the Japanese students were asked to provide written descriptions of postures from situations encountered in everyday life. Descriptions that did not correspond to posture terms (that occurred due to the free-form nature of the instructions) were discarded, leaving a set of 40 posture descriptions (refer to Table 2.9 for examples). Participants recruited to judge the written posture expressions consisted of 686 Japanese in [KM85] and 145 Americans in [MK87]. The participants were asked to rate the posture expressions on a five-point Likert scale for each of 16 semantic differential scale items (e.g., tense-relaxed, relieved-anxious, dominant-submissive, etc). The results showed that same three factors (*self-fulfillment, interpersonal positiveness* and *interpersonal consciousness*) were extracted in both studies, yet as predicted, in a different order for each culture, indicating differences in the importance of status between the two cultures (i.e., more important for the Japanese than the Americans). The order for the Japanese is as previously listed, whereas for the Americans, Factors I and II were reversed, with *interpersonal positiveness* as Factor I and *self-fulfillment* as Factor II. Testing the similarity of the factorial configurations by calculating coefficients of congruence [Har60] between the Japanese study and the American study, the results showed that indeed, Factor I of the Japanese study is basically the same as Factor II of the American study, and Factor II of the Japanese study is the same as Factor I of the American study. While the authors found cultural differences as expected, they also asserted that generalisability of the studies needs to be expanded and questioned whether cultural differences would be found with posture stimuli instead of verbal descriptions of postures.

Table 2.9: Examples of the posture expression descriptions used in [KM85][MK87].

Posture expression examples	
1. Hanging one's head	6. Shaking a fist
2. Leaning back	7. Putting one's hands together
3. Arms akimbo	8. Drooping one's shoulders
4. Bowing one's head	9. Squaring one's shoulders
5. Leaning forward	10. Crossing one's arms

Although not conclusive, the results of the studies presented may indicate a need for taking culture into account in various aspects of affect recognition research, such as labelling, and how affect is both expressed and perceived by members of different cultures. In particular, computational models for the recognition of affect may benefit from an understanding of the role culture plays in how humans express and perceive affect from nonverbal communication modalities.

2.5 Automatic Affective Recognition Systems

Automatic affect recognition systems (summarised in Table 2.10) have focused mainly on using facial expressions [PR00a] [BLF^{+}05][RYD94] and voice [Oud03][KF07] [YSLB03][NNT99]

[CK98][DPW96][LNP02][Pet99] as the input modality. Only recently have systems been built that centre on the automatic recognition of bodily expressions mono-modally [CMR+04] [KKVB+05][PLRC02][BR07] and multi-modally [KPI04][KBP07][GP07]. Similar to the behavioural studies discussed in Section 2.3, most automatic recognition systems, independent of modality, rely on corpora that have been acted. Furthermore, these systems also rely on the actors' labels to provide the ground truth. As will be explained in the following section, depending on the software application and its goal, issues exist in creating affect recognition systems that rely on actor information.

Facial expression recognition

A survey of several automatic facial expression recognition systems is presented in Pantic and Rothkrantz [PR00a]. All of the systems presented, many of which are listed in the upper part of Table 2.10, are centred on the recognition of the six basic emotion categories defined by Ekman and Friesen [EF75]. The testing results for the analysis of static images of acted facial expressions ranges from 73% to 91% for the systems listed.

A study by el Kaliouby and Robinson [eKR04] is also listed. In this study, El Kaliouby and Robinson present an automatic recognition system for determining subtle, complex affective states (*agreeing, concentrating, disagreeing, interested, thinking,* and *unsure*) instead of basic emotions, using a combination of facial expressions with head positions along the three rotational axes, roll, pitch, and yaw. Results of their system are well above chance level (16.67%), ranging between 64.5% and 88.9% accuracy. As the authors consider head gestures and facial expressions combined, they do not discuss the performance rates of the head gestures separately. This information could be very interesting to understand computationally how much importance to place on head position.

The last study presented in the automatic facial expression recognition section is aimed at recognising facial indicators of pain [ALC+09]. In this study, the authors investigated automatic recognition from video sequences labelled at two levels, frame-by-frame and across the entire sequence. Their question was, how should datasets be labelled for the auto-

Table 2.10: Automatic affective recognition systems for facial expressions, voice, and body. Basic = anger, disgust, fear, happiness, sadness, surprise; PCA = principal component analysis; LDA = linear discriminant analysis; MDC = minimum distance classifier; BP = back-propagation; RPROP = resilient propagation; SVM = support vector machine; MLB = Maximum Likelihood Bayes classifier; KR = Kernel Regression; k-NN = k nearest neighbour; MLP = Multilayer perceptron; diff = different; GP = Gaussian process; * = recognition rate for posture modality alone; F = Frame-level labelling; S = Sequence-level labelling; B = biased; U = unbiased.

Modality	**Reference**	**Affective states**	**Acted/ non-acted**	**No. & type of stimuli**	**Ground truth**	**Method**	**Accuracy**
Face	[ECT98]	(7) basic + neutral	25 actors	200 images	actor	PCA, LDA	74%
	[HNvdM98]	(7) basic + neutral	25 actors	175 images	actor	Elastic graph matching	73%
	[HH97]	(6) basic	15 actors	90 images	actor	PCA & MDC	84.5%
	[LBA99]	(7) basic + neutral	9 actors	193 images	actor	PCA & LDA 75%	
	[PC96]	(7) basic + neutral	12 actors	84 images	actor	BP	86%
	[PR00b]	(6) basic	8 actors	265 images	actor	Expert system rules	91%
	[ZLSA98]	(7) basic + neutral	9 actors	213 images	actor	RPROP	90%
	[eKR04]	(6) mental states	30 actors	164 videos	actor	Several	64%-89%
	[ALC^{+}09]	(2) pain, no pain	21 non-actors	69 videos	both	SVM + decision rule	82% (F), 77% (S)
Voice	[DPW96]	(4) angry, fear, happy, sad	5 actors	1250 utterances	actor	MLB, KR, k-NN	70%
	[LNP02]	(2) negative & non-negative	non-actors	1220 utterances	observer	LDC, k-NN, SVM	73.5%
	[Pet99]	(2) agitation & calm	30 actors	700 utterances	actor	k-NN, BP	73%-77%
	[YSLB03]	(3) hot anger, neutral & sad, happy	8 actors	2433 utterances	actor	SVM, k-NN, NN	57%
	[YSLB03]	(15) separate affective states	8 actors	2433 utterances	actor	SVM, k-NN, NN	8.7%
Body	[CMR^{+}04]	(4) anger, fear, grief, joy	5 actors	20 videos	actor	decision tree	35.6%
	[PLRC02]	(2) angry, neutral	26 actors	1560 arm movements	actor	MLP	32.5% efficiency
	[KKVB^{+}05]	(4) anger, fear, joy, sad	5 actors	40 point-light	actor	5 diff classifiers	62%-93%
	[BR07]	(4) angry, happy, sad + neutral	30 actors	1200 movements	n/a	SVM	50%(B), 81%(U)
Multimodal	[KPI04]	(3) levels of interest	8 non-actors	262 multimodal	observer	NN	55%*
	[KBP07]	(2) pre- or not pre-frustration	24 non-actors	24 multimodal	during task	k-NN, SVM, GP	79%
	[GP07]	(6) 4 basic + anxiety, uncertainty	4 actors	27 face, 27 body videos	actor	BayesNet	91% & 94%

matic detection of pain? This is an interesting question when considering the time, effort and resource costs involved in building training and testing sets for automatic modelling [ALC$^+$09][GP06]. Choosing to label data at the frame-level sometimes means that a smaller dataset must be used. However, labelling data at the sequence level requires less time by observers, meaning that a larger dataset can be created. At both levels, support vector machines (SVMs) were trained after which pain prediction was determined firstly by summing the SVM output scores for pain for the entire sequence. Secondly, a simple decision rule was developed by varying a threshold for pain vs. no pain. While the results, unsurprisingly according to the authors, showed higher recognition rates for labelling at frame-level (82%) than sequence-level (77%), *"more interestingly, the classifier trained with coarser (sequence-level) labels performs significantly better than 'random chance' when tested on individual frames"* [ALC$^+$09], indicating that sequence-level labelling may provide enough information for creating automatic affect recognition systems.

Voice recognition

Automatic voice recognition systems are presented in the second section of Table 2.10. In Yacoub et al. [YSLB03], a comparison of recognition accuracies for three different machine learning techniques in recognising acted affect from voice was carried out. They considered a set of 15 affective states which included both basic and non-basic emotions. In the very best testing case, a 94% recognition rate was achieved for distinguishing between only two states, *hot anger* and *neutral.* This result is not entirely surprising since hot anger is highly emotional while neutral is devoid of emotion. Accuracy rates began to decline when more emotions were added. For example, recognition decreased significantly to 57% when considering three classes, *neutral* and *sad* as one class, *hot anger*, and *happy.* Recognition for the entire set of 15 affective states was extremely low at a mere 8.7%, however this is still above chance (6.7%). Other affective voice recognition systems reported accuracies ranging from 60% to 77% for two or four affective states [DPW96][LNP02][Pet99].

Bodily expression recognition

The majority of today's affective recognition systems of body posture and movement (presented in the second to last section of Table 2.10) are based on LMA [vL71], and have focused on extracting emotion information from dance sequences [PPKW04][CTV02] [CMR$^+$04] [KOIH04].

Camurri and colleagues [CMR$^+$04][CLV03] examined cues and features involved in emotion expression in dance for four affective states, *anger, fear, grief* and *pride.* The corpus consisted of the same dance performed by five dancers four times each with a different affective state expressed each time. The video clips were then processed, removing facial information, and a set of motion cues (i.e., the amount of detected motion, a measure of the amount of the body's contraction/expansion, a measure of upward movement, the direction and length of the motion trajectories, and the velocity and acceleration of the motion trajectories) was extracted based on Laban's Theory of Effort [vL63]. Decision trees were chosen to build and test automatic recognition models. Testing was carried out using five testing sets extracted from the data. The results for the best decision tree model built on testing data ranged between 31% and 46% with the average across the four emotions attaining a mere 40% correct classification. The authors pointed out that even though these results seem low, the recognition rate for each emotion was well above chance level. The authors also assert that, in order to interpret automatic classification rates, they should *"be considered with respect to the rates of correct classification from spectators who have been asked to classify the same dances"* [CMR$^+$04]. In this case, the recognition rate for the human observers was only 56%. The recognition of fear was the worst for the model built on the test data, achieving below chance level classification rates. Fear was most often misclassified as anger. This is an intriguing result because body movement was used as opposed to static postures, and as postulated by Coulson [Cou04], dynamic information may help to increase recognition rates of fear in particular. Other automatic misclassifications occurred between joy and anger, and grief and joy. The misclassification of grief as joy is interesting given the authors' examination of the quality of motion feature, which showed joy movements to be very fluid and grief movements to be quite the opposite.

Another dance movement based automatic recognition system is that of Kapur et. al. [KKVB$^+$05]. In this study, a mix of professional and non-professional dancers was employed to enact four basic emotions (anger, fear, joy and sadness) through their body movements with no constraints placed on them. The body movements were recorded with a Vicon MX$^{\mathrm{TM}}$motion capture system. A human perception study was implemented to validate the affective dance expressions. Using a forced-choice experimental design, the human observers correctly classified 93% of the 40 point-light dance movements. Next, a set of dynamic features based on velocity and acceleration was extracted from the numeric motion capture data. The feature data was then used to build automatic recognition models with five different machine learning classifiers. The automatic recognition rates varied between 62% to 93% depending on the testing method used (i.e., 10 fold cross-validation and Leave One Subject Out (LOSO)).

The work involving LMA to describe the body for building affect recognition systems is of interest, however, dance movements are exaggerated and purposely geared toward conveying affect. Body movements and postures that occur during day-to-day human interactions and activities are typically more subtle and not overtly emotionally expressive.

Turning to non-dance-based automatic bodily expression recognition, Pollick and colleagues [PLRC02] carried out a study in which they compared automatic recognition model performance with human recognition performance for affect recognition to examine differences between the two in recognising different movement styles; are human observers able to make use of the available movement information. 1560 point-light displays of arm movements representing knocking, lifting and waving actions (26 actors) for two affective states, angry and neutral were used as the experimental stimuli. 1248 movements were used to train the neural network and 312 movements were used as the testing set for both the human observers and the neural network. Eighteen observers who were not aware of the purpose of the study were asked to view the movements and judge the affect displayed as neutral or angry. The average *d'* value achieved was 1.43. Using the same testing set of movements, the average *d'* achieved by the neural network was around 3 - almost twice that of the human observers. These results indicate that the system was able to discriminate between affective states more consistently than the human observers.

Bernhardt and Robinson [BR07] have built affect recognition models for non-stylised, acted knocking motions using Pollick et al's motion capture database [MPP06]. They considered three basic emotion categories (*angry, happy* and *sad*) and *neutral*. They make the point that not only is affect readily seen in body movement, but individual idiosyncrasies are also noticeable, which can make classification more difficult. To handle these differences, after segmenting the motions, personal biases were subtracted by taking an average over all the motions and removing that from the motion features. SVMs were used to build recognition models using a LOSO cross-validation method. After training, the classifier was tested on the motion samples from a single actor (approximately 3.5% of the total set of samples). The results showed a 50% recognition rate for the motions without removing the personal biases, while recognition significantly increased to 81% using the unbiased motions. To validate the performance of their recognition models, the results were compared with the results of the human observers from Pollick et al's study [PPBS01] aimed at affect recognition of knocking motions from point-light displays and video conditions since the same corpus was used. The results indicated that the recognition models built in Bernhardt and Robinson's [BR07] study achieved recognition rates (50% for biased and 81% for unbiased) similar to the humans of Pollick et al's [PPBS01] study (59% when viewing point-light stimuli and 71% when viewing video stimuli). Based on these results, Bernhardt and Robinson concluded that *"even humans are far from perfect at classifying affect from non-stylised body motions"*, suggesting that creating a 100% accurate affect recognition system is unlikely given that humans are not 100% accurate.

Multimodal recognition

The last section of Table 2.10 lists multimodal automatic affect recognition systems. Several of these systems include body posture or movement information as one of the modalities examined. Two of these systems have been designed by Picard's group at MIT [KMP01][KPI04] [KBP07]. Focused on non-acted affect, their system models a more complete description of the body, attempting to recognise three discrete levels of a child's interest (high interest, low interest and taking a break, e.g., 'refreshing') [KPI04] and self-reported frustration [KBP07] from postures detected through the implementation of a chair embedded with pressure sensors (shown in Figure 2.7), facial expressions, and task performance while the child used a computer to solve a puzzle. Their postures were defined by a set of eight high-level (coarse-grained) posture features (i.e., leaning forward, slumping back, sitting on the edge). Of the three types of input examined, in [KPI04] the highest recognition accuracy was obtained for posture activity (55.1%) over game status (33%) and individual Facial Action Units[6] (32.8%-49.7%). Accuracy rates for posture alone as an input modality for recognising frustrated were not reported in [KBP07]. A potential issue exists with using a chair to sense posture. It means that the recognition situations are limited to specifically seated contexts. Technologies today are ubiquitous; not limited to only seated situations. Furthermore, as the posture description is dependent on seated postures, important information from the body may be missing. For instance, at the time of their research in 2004, there were no features to describe the position of the head, hands or feet. More recently however, in 2007, while still employing a posture sensing chair, Picard's group added head position (shown by [Wal98][KdSBB06][eKR04] to be an important feature for discriminating between affective states) and velocity to the list of recognised features in the new version of the system aimed at recognising learner frustration (e.g., frustrated or not frustrated) [KBP07].

The automatic recognition system of Gunes and Piccardi [GP07] is bi-modal, recognising video sequences of facial expressions and upper-body gestures. They examined the automatic recognition performance of each modality separately before fusing (comparing feature level fusion with decision level fusion) information from the two modalities into a single system. They obtained video recordings of face and body expressions for six affective states (anger, anxiety, dusgust, fear, happiness and uncertainty) from four actors. For both modalities,

[6] A Facial Action Unit is *"a visually distinguishable and anatomically based unit of facial muscle movement"* [TCDFBP02].

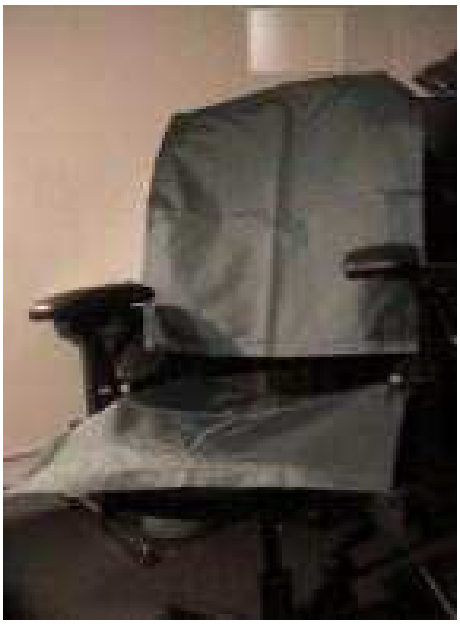


Figure 2.7: The sensor chair used in [KMP01][KPI04][KBP07]. To detect posture, it uses two sheets of force sensitive resistors. One sheet lays across the seat and one sheet lays across the back. From [KMP01].

the affective expressions to be enacted were scripted according to cues discussed in the studies of Coulson [Cou04] (for body expressions) and Burgoon et al. [BJM^{+}05] (for facial expressions). After extracting and reducing a feature vector for each modality separately, machine learning classifiers were applied to build automatic recognition models. The results reported in the article were for BayesNet from which the best results were obtained. As mono-modal systems, the automatic recognition performance was highest for the upper body sequences (89.90%), compared to the facial expression sequences (76.40%). The authors attributed this outcome to the fact that facial movements are much smaller in comparison to the upper body movements defined, and that even though high resolution video was used, it may not be sufficient enough for perfect recognition. The results on two bi-modal systems were presented: one with feature level fusion (94.02%) and one with decision level fusion (91.10%). The issue with the systems presented, both mono- and bi-modal, is that the expressions for each affective state, for each modality, were scripted, and therefore the high automatic recognition rates are not surprising. A question that needs to be addressed now is, what happens when spontaneous, unscripted expressions are used, both for acted and non-acted expressions?

As evidenced by the results presented throughout this section and listed in Table 2.10, there are significant variations between the studies that makes it difficult to compare them properly. The following section begins to tackle some of the issues that affect the creation and evaluation of automatic affect recognition systems.

2.6 Methodological Issues in Creating and Evaluating Affect Recognition Systems

This section focuses on two main issues that require consideration for creating affective recognition systems. First, how should the ground truth of affective expressions be determined? Second, what type of affective expression corpora should be used and how to obtain them?

Ground truth labelling

How should the ground truth of affective expressions be determined? What should be considered the *ground truth label*, the observer's judgment or the expresser's label? The problem is to know exactly what the expresser truly feels. Furthermore, if the affective expressions are naturalistic, i.e., non-acted, there may be no 'known' label.

In both acted and non-acted situations, self-report methods are commonly used in psychology based research for labelling an expresser's affective state. These techniques may be implemented either during the task or post-task. During the task, pop-up screens or 'talk-aloud' techniques may be implemented for asking participants about their affective state either at pre-determined points or any point at which the participants feel a change in their affective state.

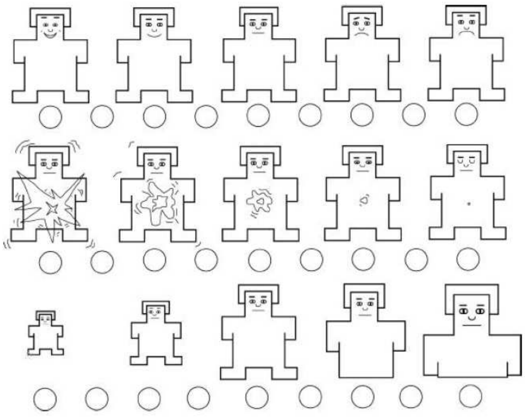


Figure 2.8: Self-Assessment Manikin on which valence, arousal and dominance scores are recorded. From [BL94].

Post-task methods include questionnaires or interviews. Questionnaires may be either verbal or visual. Verbal questionnaires consist of written statements, open-ended questions, or a list of items to be rated. Commonly used list-based questionnaires include *The Multiple Affect Adjective Check List (MAACL)* [ZL65] and the *Positive Affect/Negative Affect Measure (PANAS)* [WCT88]. Linguistic issues and cross-cultural compatibility of terms can be a problem with verbal self-report measures in particular. The use of visual self-report methods solve this issue as they employ graphical representations of affective states instead of affect terms. The *self-assessment manikin (SAM)* [Lan80], depicted in Figure 2.8, is an example of a visual self-report method in which users are instructed to rate a series of graphical characters displaying different levels of valence, arousal, and dominance.

While self-report techniques may provide a fairly quick way to gather affect response information, issues and limitations exist. A limitation with during task self-report methods is that they interrupt the expresser and may alter her/his experience, thus tainting the information obtained. A limitation with post-task self-report methods is that they depend on the expresser's memory. Picard points out that *"self-reported feelings at the end of a task are notoriously unreliable"* [KBP07]. The person may not remember what s/he was feeling at a specific moment or s/he may provide a re-appraisal of the situation, therefore important information may be lost. The person may be too embarrassed to view and/or judge their own behaviour. In a recent paper by Afzal and Robinson [AR09] in which they carried out a data collection exercise for naturalistic data, the authors attempted to obtain post-task self-report labels of facial expression. The encoders (i.e., the experiment participants from which the facial expression data was recorded) were asked to view 20 second intervals of video recordings of themselves and assign an emotion label from a predefined list of labels. The self-report task was abandoned because many of the encoders rushed through the labelling process, embarrassed at watching themselves. Instead, the authors chose to obtain labels for the facial expression data from external coders.

A more appropriate path to determining the ground truth of affective expressions may be to conduct human perception studies and base the ground truth label on agreement levels obtained from groups of observers. By averaging agreement over repeated subsets of the observers, some of the errors may be eliminated. The group of observers would depend on the application into which the affect recognition software is to be integrated. For instance, for a video game situation, the observers could be non-experts in order to create a system to take the place of a human interaction partner. For an eLearning situation, determining the ground truth according to expert observers may be more advantageous. Indeed, teachers are used to determine ground truth labels in the study measuring levels of interest described in the previous section [KPI04].

Another labelling method being employed in the artificial intelligence [FH05][Doy04] and machine learning [Yan09] fields (and more recently in affective computing) is preference learning. In the field of automatic affect recognition it is used to construct computational models of affect based on users' preferences. To this aim, human observers are asked to view two stimuli and express their preference for one stimuli over the other. In the case of affect recognition, the preference is for one affective label over another *("pairwise emotional preferences"* termed *"comparative affect analysis"* [Yan09]). For the automatic recognition process, the stimuli is not assigned a single label, but instead is labelled as a set of preferences for one particular label over another.

Affective expression corpora

An issue surrounding affective expression corpora is whether to use acted or non-acted corpora. *Acted* affective corpora are signified as actions that have been deliberately and knowingly expressed, whereas *non-acted* or naturalistic affective corpora are expressions that have been expressed naturally, without intention for the experimental procedure. The longstanding argument about acted vs. non-acted affective corpora is about the reliability of using acted stimuli for studying emotion/affect perception [Rus94][Wal98]. The early affective databases were acted or posed and focused on face and voice [JL02][KS00][LH97][YSLB03]. The difficulty and necessity of obtaining naturalistic, non-acted stimuli has been discussed for more than two decades, being described as *"one of the perennial problems in the scientific study of emotion"* [WS86]. Using material obtained from actors who are explicitly instructed to express specific emotions or affective states is considered to be unnatural and contrived/artificial [SSB$^+$04]. The research trend is now on naturally occurring affective expressions (yet, they are still focused mainly on facial expressions) [ALC$^+$09][AR09].

If acted data is used, the issue that arises is whether to use professional actors, or people who do not have any theatrical training, often referred to as non-professional actors. Using non-professionals may help to eliminate the possible exaggeration (i.e., overacting) that is said to occur with professionals due to training [MPP06], which limits variations in intensity levels. Using non-professional actors, it is also thought that a larger variety of affective expressions may be obtained [FT05][MPP06]. However, there are still issues with using non-professional actors; the affective expressions may still be stereotypical or perhaps not at all expressive. Moreover, the non-professional actors may not be aware if/when they use a modality different to the one being studied.

If non-acted data is used, the issue that arises is how to obtain the data. It has been argued that emotions expressed in a lab setting may not be very natural [SSB$^+$04]. However, it is also recognised that it is very difficult to obtain affective expressions as they occur in real-life situations due to ethical considerations, etc. Recognising that there are issues with systems trained on acted and/or exaggerated data which are intended for everyday situations, new corpora have been created for naturalistic data [BFH$^+$03]. However, currently available naturalistic affective corpora still are focused mainly on dialog and speech [ADK$^+$02] [BHS$^+$04][Cam02][DCCC03][FSS$^+$00].

As described in Sections 2.3 and 2.5, until recently, much of the research on body posture has focused on dance, often using video recordings of ballets and other dance performances for analysis of affective behaviour. This means that research groups aiming to examine more natural, day-to-day affective bodily behaviours are required to create their own corpora. In comparison, although this is not a comprehensive list, there are several databases of affective facial expression corpora available, such as:

- Pictures of Facial Affect [EF08].
- The Binghamton University 3D Facial Expression Database [YWS$^+$06], found online at [YWS$^+$08].
- The Japanese Female Facial Expression (JAFFE) Database [KLG08].
- The MMI Face Database [PVRM05], found online at [PVRM08].
- The Mind Reading DVD [BCGWH08].
- The Psychological Image Collection at Stirling (PICS) [pic08].
- The International Affective Picture System [Cen08].
- The Cohn-Kanade AU-Coded Face Expression Database [KCT00].

The HUMAINE project[7] [Hum08] aims to build a repository for researchers to upload affective databases as well as creating/providing an affective database of its own. While the repository is still in its early stages, five of the six available databases are focused on facial expressions and speech and voice. Only one database, created by the candidate, is focused specifically on body posture and movement. However, as discussed in Sections 2.3 and 2.5, Pollick and colleagues [MPP06] have created an affective motion capture database comprising affective actions of walking, knocking, lifting and throwing, which is available

[7] *"EU-funded network of excellence HUMAINE is currently making a co-ordinated effort to come to a shared understanding of the issues involved [in emotion-oriented computing], and to propose exemplary research methods in the various areas."*

for use. Coulson's [Cou04] computer-generated whole body postures are also available for research purposes [PP07].

While there are other freely available motion capture databases, they are not affect-based. They include:

- Vanrie and Verfaillie [VV04]: Contains 22 stylised actions, such as chopping, driving, rowing, etc.
- Shipley and Brumberg [SB08]: Contains 14 stylised actions, such as running, walking, karate kick, crawling, frisbee throw, etc. A markerless technique is used. A digital video camera is used to record the actions and then *x, y* coordinates are hand-labelled for each video frame.
- Hodgins [Hod08]: Contains general categories such as, human interaction, interaction with the environment, locomotion, physical activities and sports, etc.
- MOCAPDATA.COM [Yam08]: Contains actions such as tennis, baseball, pain, soccer, etc.

It is clear from the above list that while motion capture databases exist that involve the whole body, many of them comprise stylised actions of everyday activities, not affective expressions (whether acted or non-acted). Based on the variety of recent research presented throughout this chapter, it is apparent that providing databases of affective, whole body postures and movements, acted and non-acted could reduce (if not eliminate) the time-consuming task of developing a new corpus for each research endeavour, allowing researchers to focus on the main goal of understanding and automating affect recognition.

2.7 Research Links

The purpose of this chapter has been to demonstrate the novelty of the research presented in this thesis by providing both a background to the field of affect recognition and explaining the areas that still lack sufficient research as well as issues involved in other research approaches. The research presented in this thesis can be broken down into several research questions, which aim to address the issues that exist in the affective computing field. One research question is:

- *How should affective expressions be described?*

It was shown in Section 2.1 that most affect expressions and recognition research focuses on classification according to one of two schools of thought, discrete emotion categories and affective dimensions. There is no consensus about which approach is 'best'. It has been asserted that emotion science could benefit from empirical research examining whether *"both approaches can be integrated in a sensible way to provide a comprehensive understanding of affect"* [Fox08]. While the integration of affective states and dimensions is becoming an important issue, it remains important to examine what type of affective information can be recognised from nonverbal communication channels. In this vein, it has been shown that discrete emotion labels of experienced affect, facial expressions and arm movement typically do in fact fall in a circumplex structure on a two dimensional affective space. However, this has not yet been examined with body posture. This brings the discussion to the next research question:

- *What role does bodily information play in affect recognition, and is it a reliable nonverbal communication modality from which affect in humans can be recognised and modelled?*

In Section 2.2 it was shown that the majority of affective computing research has investigated affective facial expressions. Only recently has attention turned toward affective bodily expressions. Behavioural science research has shown that body posture is more important for nonverbal communication than was previously thought. In fact, it has been found that some affective states may be better recognised from the body than the face.

Section 2.3 explained that there have been a number of behavioural science studies aimed at understanding body posture and movement cues that may contribute to the recognition of specific affective states. However, there is still little systematic research in the direction of posture, with no well-established models for body posture and movement. A low-level posture description approach is taken in this thesis as it may allow for adaptability to almost any context, over that of a high-level approach.

- *Are there universal aspects to affect expression and recognition or are they affected by human factors?*

An overview was given in Section 2.4 of the research investigating whether or not cultural differences exist in how affect is expressed and perceived. It was pointed out that the majority of existing research on this topic has focused on affective facial expressions, leaving affective bodily expressions still a novel area to study. As it is accepted that people differ in how they express and perceive affect due to a variety of factors such as culture, personality, gender, etc., further research investigating these factors remains necessary in order to create affect recognition systems.

- *Is it feasible and advantageous to build automatic recognition systems for recognising affective body posture?*

In Section 2.5, it was observed that the majority of affective recognition systems have been built on facial expressions and voice, with less emphasis placed on systems built to recognise body posture. The systems that have focused on bodily expressions typically have focused on body movement and dance. One goal of affective computing research is to create multimodal recognition systems. However, in order to do so, the affective information available in individual modalities should be examined first. For body posture for instance, it must first be determined if and how well affect can be recognised by both humans and systems. This thesis aims to provide this knowledge for body posture.

From the investigation of automatic affect recognition systems, it was concluded that some significant problems exist in how these systems are created. These methodological problems were investigated in Section 2.6, to address the two remaining research questions:

- *How can the ground truth of affective expressions be established?*

It was found that the majority of the automatic recognition systems discussed in Section 2.5 relied on ground truth labels provided by the individual who displayed the affective expression. This approach may be problematic because, it is difficult to know for sure that the person really expressed what s/he intended to express or that it was expressed through the modality being investigated. The approach taken in this thesis is to use human observers to determine the ground truth of the affective postures instead of the expressers themselves. The approach and its rationale are discussed in detail in Chapter 3.

- *What type of affective data should be modelled? Acted or non-acted?*

The argument is that acted affective expressions appear artificial and contrived. Although investigating non-acted affective expressions is more desirable, obtaining them is more difficult, often due to ethical issues and environmental constraints. Acted affective expressions are used as a typical starting point for much of the affect recognition research. The idea is that if it is not possible to obtain 'reasonable levels' of recognition with acted affective expressions, then it is unlikely that non-acted affective expressions will be any easier to recognise as non-acted expressions tend to be more subtle.

The work presented in the remaining chapters of this thesis seeks to implement an approach that was devised to exploit these gaps in the current research. Details of the approach are described in the following chapter.

Chapter 3

Research Hypotheses and Methods

As discussed in Chapter 2, many qualitative psychological studies have been carried out to understand what type of bodily cues may be attributed to specific affective states while only recently has research started to focus on a quantifiable relationship between affect and its expression through body posture [dM89][Cou04][SR06]. The main hypothesis of this thesis is that affect, both discrete categories (e.g., happy, frustrated, etc.) and levels of affective dimensions (e.g., arousal, valence, etc.), can be recognised from whole body posture, and that computational models for affective body postures can be grounded into a low-level description of posture. The hypothesis is split into two sub-hypotheses. One, that human observers can recognise affect from posture at above chance levels. Two, that automatic models can be built that achieve recognition levels comparable to benchmarks computed based on human observers. Building such benchmarks is necessary when the ground truth is not available.

Rationale for the approach

An incremental approach is used to prove the hypotheses as purported to be necessary by Wallbott [Wal98]. The first step is to examine the recognition of basic emotions from acted postures. This situation was chosen as the first step based on the discussion presented in Chapter 2 that it is a typical starting point for affect recognition research. The second step is to examine the recognition of non-basic affective states from non-acted postures in a natural situation. A video game scenario was chosen as the natural situation for two reasons. One, research focused on affective information from video game players has recently become a hot topic. Two, it is thought that as the video game players are not made aware of the true purpose of the study, it will be possible to obtain true bodily expressions of affect. In these first two steps, the postures to be investigated are manually extracted. The third and final step in this research project is to examine automatic non-basic affective state recognition offline in a run-time situation using sequences of non-acted static postures that are not manually extracted. The aim is to demonstrate how a recognition system could be integrated into existing software situations.

For the reasons discussed in Chapter 2, human observers are used for evaluating the affective postures instead of the person who displayed the affective posture. Using a non-acted situation means that there is no pre-existing ground truth, thus one needs to be established. Once the ground truth has been established, the levels of observer agreement (within and between observers) and later, the performance of the recognition models must be computed. Chance level agreement has been chosen as the metric for validating the level of agreement between observers as it is the typical metric currently used in the affective computing field. A random repeated sub-sampling validation method is used to create benchmarks based on the human observers. These benchmarks are used to evaluate the performance of the automatic recognition models. Benchmarks that are created based on the recognition rates of human observers have been set as the target in this research, as will be discussed in Section 3.2.4. Described in more detail later in the chapter, the automatic recognition models are also built using a random repeated sub-sampling method to test generalisability to new observers and 10-fold cross-validation to test generalisability to new postures. As it is not possible to obtain an infinite number of samples, these validation methods help to ensure that the automatic recognition model performances are an estimation of the population and not simply an estimation of the population sample.

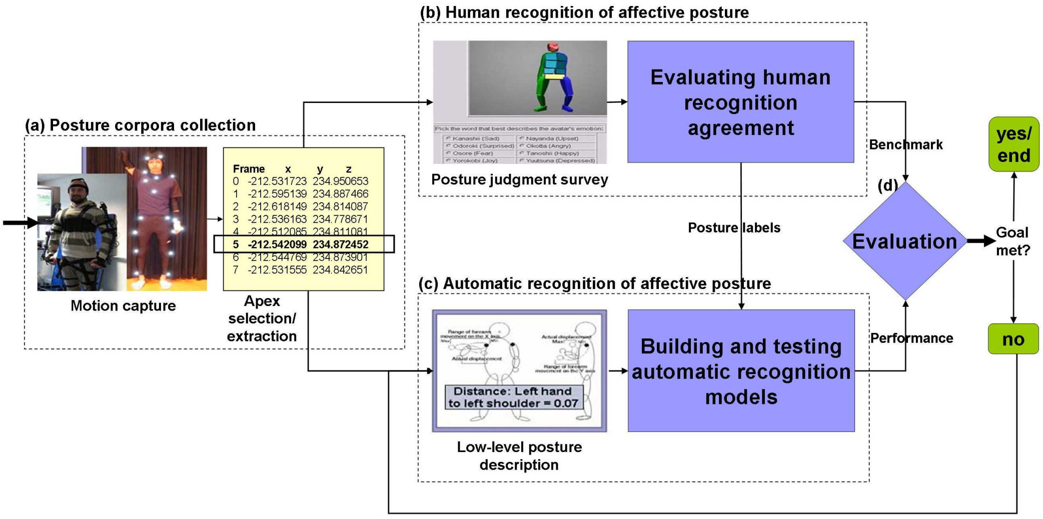


Figure 3.1: The general method used to investigate the research problem. (a) shows the first step: corpora collection. (Permission to publish the photos has been granted.) (b) outlines the procedure used to determine human recognition of affect. (c) demonstrates the procedure used for building and testing automatic recognition models of affect. (d) is the validation procedure of comparing recognition model performance to the human observer benchmarks.

The remainder of the chapter explains the approach taken (shown in Figure 3.1) to investigate the problems described and is organised as follows. Figure 3.1(a) and Section 3.1 explains the first step: collecting postural display data for assessing affect recognition by human observers and for building and testing automatic recognition models. The procedure for examining human recognition of affect is outlined in Section 3.2 and shown in Figure 3.1(b). The low-level posture description procedure (shown in the left side of Figure 3.1(c)) is explained in Section 3.3, and an analysis of how the posture description performs in discriminating between affective states and affective dimensions is described in Section 3.4. The method used to build automatic affective posture recognition models is represented in Figure 3.1(c) and explained in Section 3.5. Figure 3.1(d) represents the evaluation of the automatic recognition models by comparing the benchmarks developed from the human observers' recognition rates with the recognition rates of the models and highlighting specific misclassifications.

3.1 Posture Corpora

3.1.1 Motion capture data collection

As the overall hypothesis of this thesis is that affect can be recognised from whole body posture, the first step is to collect postural display data. As shown in Figure 3.1(a), motion capture systems are used for the collection of postural information instead of other tracking methods such as vision-based techniques. The main influencing factor for this decision is accuracy. Using motion capture systems, a precise numeric representation of the body in a 3D space can be more easily obtained. The use of numeric data allows for the humans to be represented in degrees of human form [MPP06]. Another influencing factor is privacy. Motion capture systems allow for complete anonymity. The use of anonymous data is an advantage for many types of potential research and commercial applications, from healthcare to video games. The individuals being recorded may wish to remain unidentifiable to other patients or video game players.

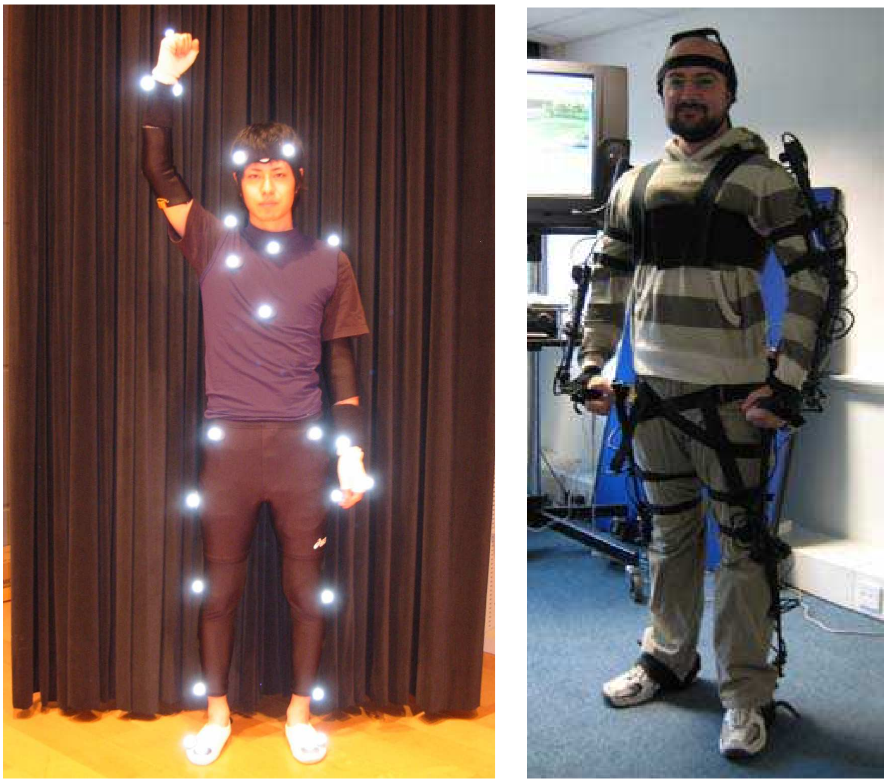
A Vicon MX series digital optical motion capture system[1] and a Gypsy 5 electro-mechanical motion capture system[2] are used. Two different motion capture systems are used due to the requirements of each study. For example, the Vicon system was readily available at the university in which the acted postures were collected. Furthermore, as the study is acted, it is not crucial that the capturing take place in a strictly natural environment. However, in the non-acted study naturalness of the environment is more of a requirement. Both the Vicon and the Gypsy systems were available for the non-acted postures study,

[1] http://www.vicon.com/, retrieved November 2007

[2] http://www.animazoo.com/, retrieved November 2007

however the fact that the Vicon system is installed inside a CAVE$^{\text{TM}}$-like system [CNSD93] lessened the naturalness of the setting making the Gypsy system more desirable.

The participants motion captured with the Vicon system are dressed in a lycra suit to which lightweight reflective markers are affixed to various joints and body segments as depicted in Figure 3.2(a). Eight infrared cameras, mounted in a circular configuration above the motion capture space, track and reconstruct the markers in a 3D space. The Gypsy 5 motion capture system is a full-body exoskeleton (refer to Figure 3.2(b)) comprising potentiometers located at the actor's joints and joined together by lightweight plastic rods. Two gyroscopes, one for the lower body and one for the upper body, calculate the rotational direction of each section of the exoskeleton.



(a) (b)

Figure 3.2: (a) An example of the markers used for motion capture with the Vicon system. (b) Animazoo's Gypsy 5 exoskeleton motion capture suit. Permission to publish the photos has been granted

Each motion capture system has different advantages and disadvantages. The Vicon system has been used by many industry leaders, such as Nintendo, Sony, BMW, Toyota, etc. The markers of the Vicon system are small and lightweight, thus they are less bulky than the exoskeleton configuration of the Gypsy system. However, an advantage of the electro-mechanical attribute of the Gypsy system means that there is no marker occlusion as there is with the Vicon. The Gypsy system is highly portable allowing it to be used in almost any setting, indoors and outdoors, making it feasible for use in real applications, whereas the Vicon system is more immobile, due to camera placement and set-up constraints. One disadvantage from which both motion capture systems suffer is the inability to record detailed hand and finger positions. While the importance of detailed hand gestures is acknowledged, adding this level of detail is outside the scope of the thesis.

For both systems, a manually labelled configuration model is created for each motion capture participant. The *configuration model* is used to define the arrangement and size of the individual's body. Its purpose is to fit each of the motion captures to the size and form of the individual's body. Although a configuration model is not essential, the advantage of using this model is that it allows for a more precise measurement of each body. However, in some situations it may be more desirable to use a default configuration model. For example, a video game scenario may not require the precise measurements of the individual and a general bodily representation may be sufficient; whereas in a healthcare situation such as physiotherapy, obtaining the individual's precise bodily measurements could be advantageous for tailoring the therapy.

Motion capture participants for the acted study are referred to as *actors* hereafter because, although they are not professional actors, they are explicitly asked to act out specific emotions through whole body postures. Motion capture participants for the non-acted study are referred to as *players* hereafter since a video game scenario is used.

3.1.2 Stimulus identification

After collecting the motion capture information, for the studies presented in Chapters 4 and 5, static postures are manually extracted from the original motion capture data because not all of the motion capture frames correspond to affective postures. A *static posture* corresponds to a single frame of motion capture data, and is considered the *apex*, or most expressive instant of the movement. The postures are used for investigating both human recognition and automatic recognition of affective postures. The study in Chapter 6 does not employ manual extraction of the postures but instead, sequences of static postures are automatically extracted to build a testing set for investigating how well the recognition system performs.

Once identified, the stimuli for all three studies are built in two ways. First, as a set of posture images to be used in a posture judgment survey for human recognition of affective posture as shown in Figure 3.1(b). The stimulus preparation procedure is described in the following section. Second, as a vector of low-level postural information, i.e., numeric descriptions of the configuration of the body, as shown in Figure 3.1(c), used to build, train and test automatic affective posture recognition models.

3.2 Human Recognition of Affect from Posture

The purpose of this section is to explain the approach taken to evaluate the hypothesis that human observers can recognise affect (i.e., discrete categories and levels of affective dimensions) from whole body posture at above chance levels. Figure 3.1(b) provides an overview of the approach and Figure 3.3 shows the approach in detail. Details of the process are discussed in the remainder of the section.

3.2.1 Ground truth labelling

To test human performance in recognising affect from posture, the ground truth must be built; labels need to be assigned to the affective postures. The view taken in this research

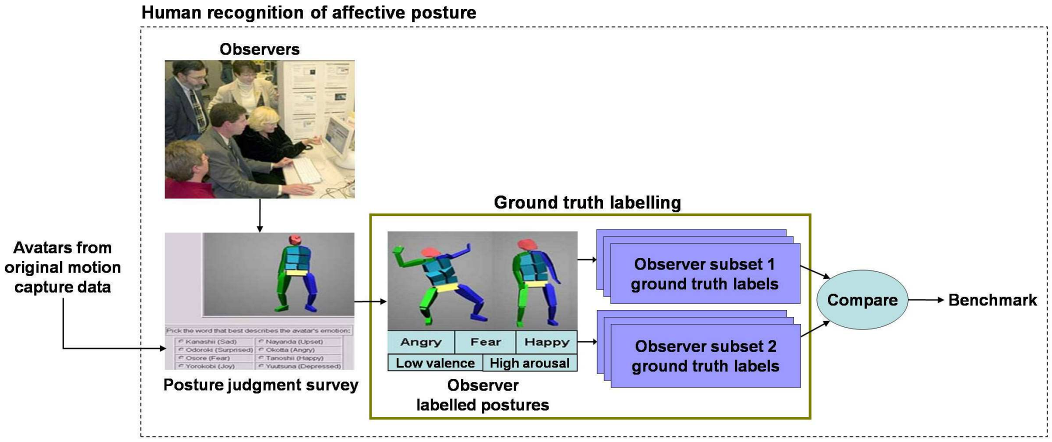
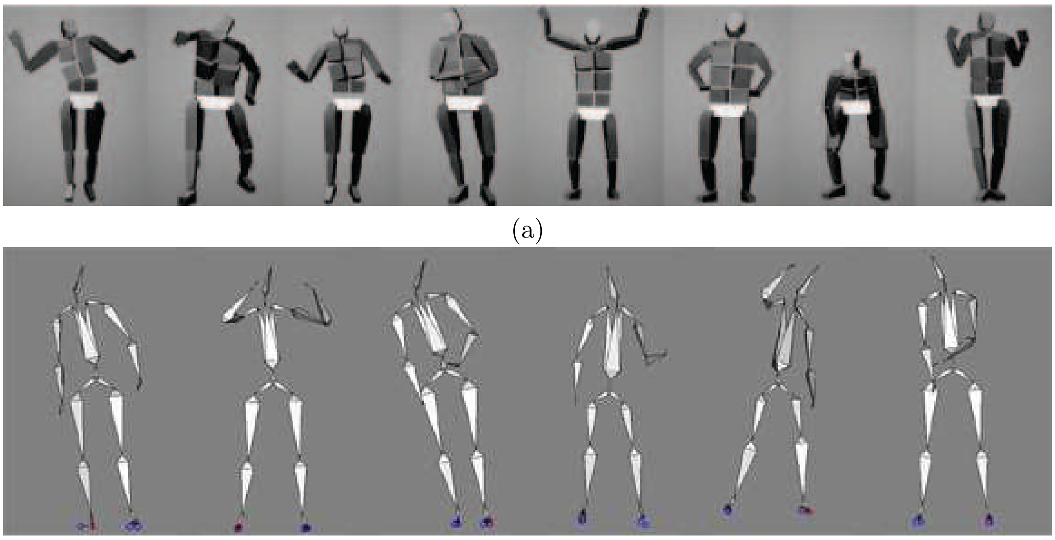

Human recognition of affective posture
Observers
Avatars from original motion capture data
Pick the word that best describes the avatar's emotion:
Kanashii (Sad)
Nayanda (Upset)
Odoroki (Surprised)
Okotta (Angry)
Osore (Fear)
Tanoshii (Happy)
Yorokobi (Joy)
Yuutsuna (Depressed)
Posture judgment survey
Ground truth labelling
Angry
Fear
Happy
Low valence
High arousal
Observer labelled postures
Observer subset 1 ground truth labels
Observer subset 2 ground truth labels
Compare
Benchmark

Figure 3.3: Expands the human recognition of affect box from Figure 3.1(b). Observers are recruited to judge the affective state and dimension levels of a set of postures. The set of observers are divided into subsets 10 times (i.e., trials) and a ground truth is assigned to each posture in each subset. The agreement between subsets is computed and a benchmark is calculated as the average agreement between the subsets across the 10 trials

project is that there is no inherent ground truth affective state label or dimension level that can be attached to the postures. Labels that could be assigned by the actors and players are not considered for the reasons described in Section 2.6 of Chapter 2. In the acted postures study, the actors' labels and dimension ratings are not used because the actors may not portray what they intend to portray as they are not professional actors. In the non-acted postures study, the players are not used to label their own postures because *"self-reported feelings at the end of a task are notoriously unreliable"* [KBP07], and it is not feasible to stop the players during the gaming session to ask them their current affective state. Furthermore, because the complete affective state is expressed through the combination of a variety of modalities in the non-acted scenario in particular, it is difficult for the players to be aware through which modality affect was expressed [RF99], or if their bodies were expressing their true feelings. Thus, the approach used in this research is to build ground truth labels from outside observers' judgments of the postures using posture judgment surveys. The survey participants are referred to as *observers* hereafter.

3.2.2 Stimulus preparation

The posture judgment surveys consist of static posture images of a humanoid avatar. The avatars are created from the extracted postures described in Section 3.1, i.e., the original motion capture data. The procedure used to extract postures from the Vicon data is described in Chapter 4 and the Gypsy data extraction procedure is presented in Chapters 5 and 6. The postures are presented in the frontal view. Avatar examples are shown in Figure 3.4. Avatars are used instead of human photos in order to create a faceless, genderless, non-culturally specific 'humanoid' in an attempt to eliminate bias. Using avatars, observers are not affected by facial expressions, as the focus is on how posture alone conveys the desired impression. Use of the face could confound the observers' evaluations because it would not be possible to discern which channel of information is used to judge the expression.



(a)

(b)

Figure 3.4: Examples of the affectively expressive avatars constructed from motion capture data. The avatars in (a) were created with the data from the Vicon motion capture system. The avatars in (b) were created with the data from the Gypsy motion capture system

3.2.3 Survey procedure

The posture judgment surveys are conducted online. The set of postures are presented in a randomised order and observers are asked to associate either an affective state label or affective dimension levels to each. For the affective category surveys, a forced-choice design is chosen over an open-ended design which requires the observers to freely choose labels. Russell [Rus94] reports the issue of using open-ended designs by citing a 1953 study by Frijda [Fri53] that used an open-ended format. Instead of emotion labels, the majority of the responses given were situation descriptions, which makes it difficult to assess the agreement between observers. A seven point Likert scale is used for the affective dimensions surveys. The observers are asked to judge each posture according to four affective dimension scales: *valence* (displeasure/pleasure), *arousal* (calm/excited), *potency* (control), and *avoidance* (avoid/attend to). Valence, arousal, and potency (also referred to as dominance) have been chosen based on psychological research throughout the last century which asserts that these three dimensions cover the majority of affect variability [Wun07][OST57][Dav64][MR74]. Avoidance is chosen as a fourth dimension because it could provide important information in a variety of contexts. The output of the posture judgment surveys are sets of labelled postures. Specific details about each survey will be presented in Chapters 4, 5 and 6.

3.2.4 Procedure for measuring human agreement performance

This section explains the process used to measure the performance of human observer agreement on recognising affect from whole body posture. Four main procedures are carried out: i) to determine a ground truth for each posture; ii) to assess agreement within and between observers; iii) to assess the reliability within and between the observers' agreement levels; iiii) to build benchmarks to be used for evaluating automatic recognition model performances.

For each survey, let $P = \{p_1, ..., p_r\}$ be the set of postures used in the survey; let $L = \{l_1, ..., l_n\}$ be the set of labels available in a forced choice survey; let $D = \{d_1, ..., d_4\}$ be the set of dimensions evaluated in an affective dimensions survey; let $O = \{o_1, ..., o_m\}$ be the set of observers that participate in the survey.

In the forced choice surveys (the affective state labels case), observers assign a label to each posture. For each observer o_k, posture p_i and label l_j, the evaluation function is defined as $eval_{cat}(o_k, p_i, l_j)$ such that $eval_{cat}(o_k, p_i, l_j) = 1$ if the observer o_k has assigned the label l_j to p_i; otherwise $eval_{cat}(o_k, p_i, l_j) = 0$.

In the affective dimensions surveys, observers evaluate each posture p_i over each dimension d_g in set D on a seven point rating scale. For each observer o_k, posture p_i and dimension d_g the evaluation function is hence defined as $eval_{dim}(o_k, p_i, d_g) = c$ where c is a value from 1 to 7 (the rating scale).

- **Posture ground truth**

To measure the performance of affective posture recognition by human observers, for the affective categories cases, for each posture p_i and a set of observers O, the ground truth $gtl(p_i, O)$ is defined as the label l_j for which $freq(p_i, l_j) > freq(p_i, l_t)$ for each $t = \{1, ..., n\}$ and $t \neq j$ where

$$freq(p_i, l_j) = \frac{1}{m} \sum_{k=1}^{m} eval_{cat}(o_k, p_i, l_j) \tag{3.1}$$

where m is the number of observers in O. $freq(p_i, l_j)$ ranges between [0,1] and the value for $eval_{cat}(o_k, p_i, l_j)$ can be either 0 or 1. If $freq(p_i, l_j) = freq(p_i, l_t)$, then the ground truth label is randomly selected between the two labels l_j and l_t.

For the affective dimensions cases, for each p_i, the ground truth rating, $gtm(p_i, O) = \text{median}\{eval_{dim}(o_k, p_i, d_g) : k = 1, ..., m\}$.

- **Observer agreement**

Within observers agreement: The within observers agreement on the set of postures P is computed for a set of observers O. For the affective categories cases, within observers agreement is computed across the set of labels L and for each individual label l_j. The within observers agreement $W.AgrLabel(l_j, O)$ for a label l_j is defined as

$$W.AgrLabel(l_j, O) = \frac{1}{r}\sum_{i=1}^{r} freq(p_i, l_j, O) \tag{3.2}$$

where r is the number of postures and $freq(p_i, l_j, O)$ is the frequency of use for the label l_j for a posture p_i as defined in Equation (3.1). $W.AgrLabel(l_j, O)$ ranges between [0,1]. A within observers agreement $W.AgrLabel(L, O)$ across the set of labels L is defined as follows

$$W.AgrLabel(L, O) = average(W.AgrLabel(l_1, O), ..., W.AgrLabel(l_n, O)) \tag{3.3}$$

where $W.AgrLabel(L, O)$ ranges between [0,1].

Between observers agreement: Given two sets of observers, O_1 and O_2 and a posture p_i, the agreement between the two sets of observers for p_i is 1 if $gtl(p_i, O_1) = gtl(p_i, O_2)$, otherwise 0. The between observers agreement $B.AgrLabel(O_1, O_2)$ is defined as the average of the agreement between the two sets of observers O_1 and O_2 across the set of postures P.

For the set of affective dimensions D, between observers agreement $B.Agr(d_g, O_1, O_2)$ is not considered binomial because a rating scale is used and the distance between two ratings is important. Therefore, between observers agreement in the dimensions cases is defined as

$$B.Agr(d_g, O_k) = 1 - \left(\frac{\sum_{i=1}^{r} |gtm(p_i, O_1) - gtm(p_i, O_2)|}{r * (7-1)}\right) \tag{3.4}$$

where r is the number of postures and $7 - 1 =$ the degrees of freedom in the rating scale and $B.Agr(d_g, O_k)$ ranges between [0,1].

- **Observer agreement reliability**

To assess the reliability of the observers' judgments, Cohen's kappa [Coh60] and Fleiss' kappa [Fle71] are computed for the categorical data and Cronbach's α [Cro51] is computed for the dimensional data. Kappa coefficients were chosen because they are well-known, well-used statistical methods for assessing inter-rater reliability. *"Kappa is intended to give the reader a quantitative measure of the magnitude of agreement between observers"* [VG05].

Cohen's kappa is computed between two subsets of observers. Fleiss' kappa, a variation of Cohen's kappa, takes into account multiple raters. Thus, it is computed within an entire set of observers O. The equations for both Fleiss' and Cohen's kappa can be written in the same way. Kappa (K)

$$K = \frac{(A - A_e)}{(1 - A_e)} \tag{3.5}$$

where A is the agreement and A_e is the agreement expected by chance. The kappa ranges between [-1,1] with a kappa of 1 showing perfect agreement, a kappa of 0 showing the agreement that would be expected by chance alone and a kappa of -1 showing total contradiction. Landis and Koch [LK77] provided a breakdown for interpreting the ratings, depicted in Table 3.1.

Table 3.1: Interpreting the kappa ratings.

Kappa	**Interpretation**
< 0	No agreement
0.0 - 0.20	Slight agreement
0.21 - 0.40	Fair agreement
0.41 - 0.60	Moderate agreement
0.61 - 0.80	Substantial agreement
0.81 - 1.00	Almost perfect agreement

For the dimensional data, Cronbach's α is cited as the most accepted statistical method for computing the reliability of scale data [Fie05]. Cronbach's α is defined as

$$\alpha = \frac{H^2\overline{Cov}}{\Sigma q_{item}^2 + \Sigma Cov_{item}} \tag{3.6}$$

where H is the number of items squared and multiplied by the average covariance between the items and then divided by the sum of all the item variances and item covariances [Fie05]. Cronbach's α ranges between [0,1]. The higher the α level, the more consistent the observers ratings. Cronbach's α is computed for both the within and between observers cases for the dimensional data.

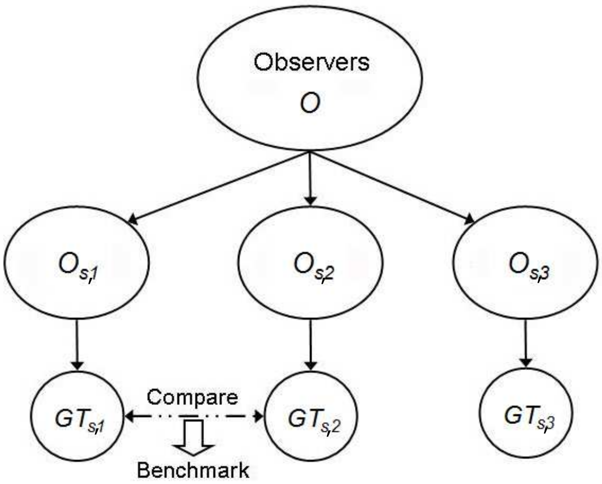


Figure 3.5: The method for creating the benchmarks of human recognition. The set of observers O, is randomly divided into 3 disjoint subsets, $O_{s,1}$, $O_{s,2}$ and $O_{s,3}$, repeated for 10 trials, i.e., $s = 1, ..., 10$. For each O_{sk}, ground truth labels $gtl(p_i, O_{sk})$ or ground truth ratings $gtm(p_i, O_{sk})$ are assigned to each posture p_i in the set of postures P. $O_{s,1}$ is compared with $O_{s,2}$ to determine a benchmark for human recognition of affect from whole body posture

- **Creating the benchmarks**

To create benchmarks, the questions: *what does accuracy mean?* and *what is acceptable or sufficiently 'high' performance?* need to be answered. While not the perfect method, affective computing research typically reports that a system performs well if the accuracy rate is **above chance level**. The issue is that, depending on the application, chance level recognition may not be sufficient. Instead, what is needed is a way to measure the recognition rate acceptable for a particular application so that the user's experience with the affective technology improves rather than degrades. For example, in an affective technology aimed to act as a counselor, basing automatic recognition performance on recognition levels obtained by untrained or non-empathic human observers would not be sufficient. However, in a video game scenario, achieving a human observer level of performance may be a sufficient target recognition rate as it has been shown that people prefer to play games together [LLCBB08].

A random repeated sub-sampling method is used for creating the benchmarks of human recognition performance (depicted in Figure 3.5). This method is used in an effort to obtain performance rates that may reflect a real population. Repeated sub-sampling helps to ensure replicability, i.e., that the results are not limited to a particular partitioning instance [Fin72]. Resampling methods are used in situations where it is not possible to obtain an infinite number of examples (e.g., observer judgments) [VL89].

For the affective categories and the affective dimensions separately, shown in Figure 3.5, a set of observers O is randomly divided into three disjoint subsets, $O_{s,1}$, $O_{s,2}$ and $O_{s,3}$, repeated for 10 trials, $s = 1, ..., 10$. For each O_{sk}, the respective ground truth $gtl(p_i, O_{sk})$ or $gtm(p_i, O_{sk})$ (depending on the case, i.e., labels in the former and dimensions in the latter) is assigned to each posture p_i. For each trial s, a between observers agreement is computed between observer subsets $O_{s,1}$ and $O_{s,2}$ as previously stated in the 'Observer agreement' bullet point. A benchmark for the affect recognition model evaluation is computed as the average across the 10 trials. The benchmarks are used later to evaluate the recognition performance of the automatic recognition models built with the reserved third subset of observers $O_{s,3}$, which will be discussed in Section 3.5.

3.3 Low-Level Posture Description

Each posture corresponds to a single frame of motion capture data as previously explained in Section 3.2.2. For each selected frame, i.e., a posture p_i, a vector of low-level posture features, $F_i = \{f_{i1}, ..., f_{iu}\}$ (shown in the 'low-level posture description' box of Figure 3.1(c) and Figure 3.6 explained in Section 3.5) describing the configuration of the posture is built. These low-level features describe the posture configuration in terms of distances between joints and angles between body segments. The use of these low-level features allows for a general description of the postures displayed. A major strength of the low-level posture description approach is that it is general, meaning that it is independent of affective state and context, provided that the postures offer the information necessary to recognise the affective state. Context information could be added to bias the recognition without having to change the way the posture description is computed, but this is outside the scope of the thesis. The posture feature computation details are provided in Chapters 4 and 5 separately.

3.4 Low-Level Posture Description Analysis

To determine the discriminative power of the low-level posture description, a feature analysis is carried out using an analysis of variance (ANOVA) approach. Specifically, ANOVA is used to evaluate how informative and useful each of the low-level features in the description is for distinguishing between postures associated to different affective states and between postures associated to different dimension ratings. The ANOVAs are computed using SPSS 16 [Fie05].

3.5 Automatic Recognition of Affect from Posture

This section describes the analysis of the low-level posture description and the approach taken to examine whether automatic recognition models of affective posture can be built that are able to achieve recognition levels similar to the benchmarks set by the human observers. Figure 3.1(c) provides an overview of the approach and Figure 3.6 illustrates the details of the approach. The complete process is discussed in the remainder of the section.

3.5.1 Model Creation and Evaluation

After the low-level posture description has been computed and the features have been analysed, automatic recognition models are built that map the low-level posture descriptions into labels describing the affective state or affective dimension levels conveyed by each posture.

The goal is not to evaluate or modify the learning algorithm itself or to define new algorithms; this is outside the scope of the thesis. Instead the goal is to test automatic recognition performance and evaluate the low-level posture description. A multi-layer perceptron (MLP) with a back-propagation algorithm [Hay99] is implemented using Weka 3.6 [HFH$^+$09]. An MLP was chosen because it is often used in affective computing and thus may

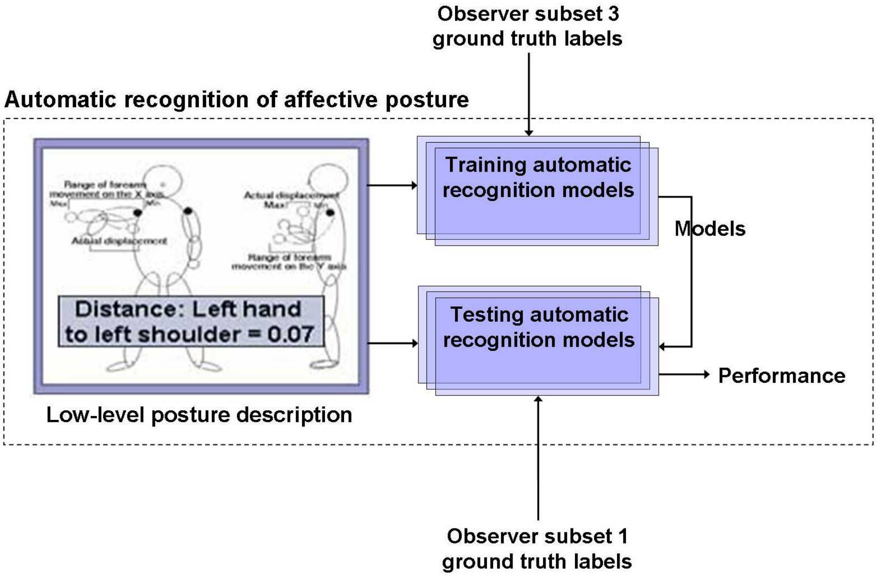

Observer subset 3
ground truth labels
Automatic recognition of affective posture
Range of forearm movement on the X axis
Max
Min
Actual displacement
Actual displacement
Max
Min
Range of forearm movement on the Y axis
Distance: Left hand to left shoulder = 0.07
Low-level posture description
Training automatic recognition models
Models
Testing automatic recognition models
Performance
Observer subset 1
ground truth labels

Figure 3.6: Expands the automatic recognition of affect box in Figure 3.1(c). The vector of low-level posture features F_i is computed for each posture p_i. $O_{s,3}$ is used to train the automatic recognition models which are then tested against $O_{s,1}$. This procedure is repeated for each of the 10 trials. $O_{s,1}$ and $O_{s,3}$ are defined in Figure 3.5

be considered a benchmark-setting method. It was also chosen for its ability to effectively handle discrete categorical data as well as continuous data. An MLP learns by iteratively processing a set of training samples and comparing the network's prediction for each sample with the known class label. Using the MLP, the automatic recognition models are tested for their ability to generalise in two ways: i) to new observers and ii) to new postures. Each generalisation procedure is outlined in the following paragraphs.

Generalising to new observers

Due to the lack of a definitive affective ground truth, one goal is to test how well the recognition models can generalise to new observers. This testing procedure is depicted in Figure 3.7 and uses observer subsets $O_{s,1}$ and $O_{s,3}$ previously defined in Section 3.2.4. Keep in mind that the set of postures P is the same in all three subsets. The automatic

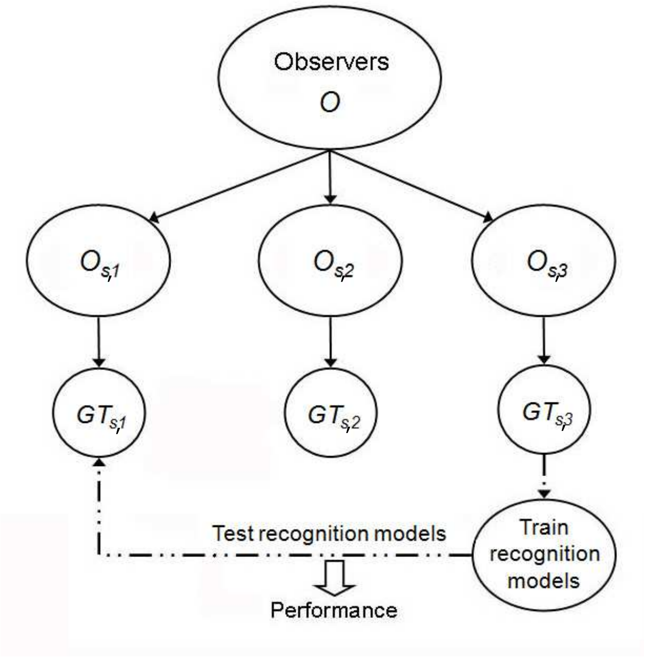


Figure 3.7: The method for testing and evaluating automatic recognition models' ability to generalise to new observers. $O_{s,1}$ and the previously unused $O_{s,3}$, defined in Section 3.2.4 and shown in Figure 3.5, are used. $O_{s,3}$ is used to train recognition models which are then tested against $O_{s,1}$. This procedure is repeated for each of the 10 trials

recognition models are built and trained on $O_{s,3}$ and then tested using $O_{s,1}$. The testing procedure is repeated for each of the 10 trials. To evaluate the automatic recognition models' performance, the average of the automatic recognition rates computed across the 10 trials is compared to the benchmark computed in Section 3.2.

Generalising to new postures

Another goal of testing is to assess how well the automatic recognition models can generalise to new postures as the set of postures would not remain static, but instead would continue to grow were the models integrated into existing software applications. Testing posture generalisation is achieved using 10 fold cross-validation. The procedure is depicted in Figure 3.8. Each posture p_i is associated with a ground truth label $gtl(p_i, O)$ or rating $gtm(p_i, O)$.

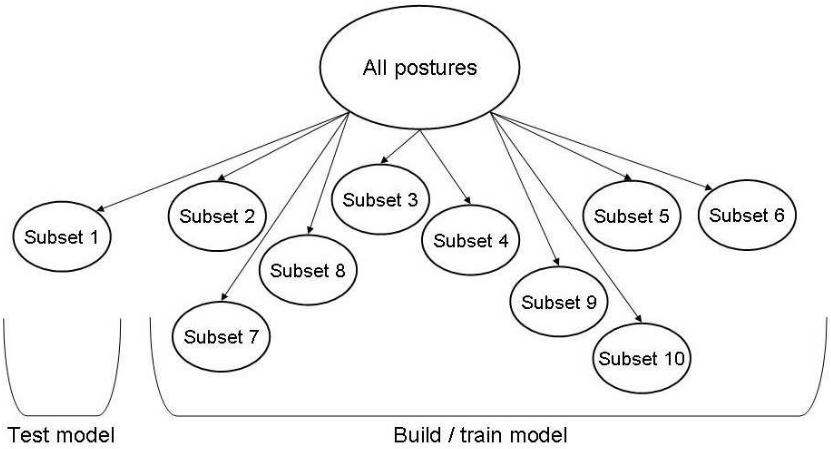


Figure 3.8: The set-up for 10 fold cross-validation. The postures are assigned ground truth labels and automatically divided into 10 subsets. A model is built and trained on 9 of the subsets and then tested against the 10th subset. The process is automatically repeated until each subset has been tested against a trained model

Next, the postures are randomly divided into 10 equal subsets. Automatic recognition models are built and trained using nine of the subsets and then the tenth subset is used for testing the models. The procedure is automatically repeated until all of the subsets have been used for testing. The recognition performance is the average across the 10 subsets.

Chapter 4

Case Study 1: Modelling Acted Basic Emotions

The goals of the acted postures study are to assess whether acted postural expressions of basic emotions can be recognised by human observers and automatic recognition models. The recognition of acted displays of emotion was chosen as the first case study because it is the typical starting point for a lot of research aimed at recognising affect from nonverbal communication modalities [DCCS$^+$07]. Perceiving basic emotions is accepted as easier than perceiving more subtle, non-stereotypical, complex affective states [eKR04].

Section 4.1 reports on the posture data collection. Sections 4.2, 4.4 and 4.5 address the recognition (at human and automatic levels) of discrete emotion categories from posture. Section 4.3 explains the low-level posture features used for describing posture in this study. The second part of the chapter reports on the recognition of levels of affective dimensions from posture in the same format as the emotion category recognition. The chapter ends with a summary in Section 4.9.

4.1 Posture Corpora

4.1.1 Motion capture data collection

The first step in assessing if basic emotions can be recognised from acted postures was to collect a set of postural data. As described in Section 3.1 of Chapter 3, the Vicon motion capture system [Vic07] was used in this study. Three-dimensional motions were recorded while actors enacted four emotions, *angry, fear, happy* and *sad* through bodily expressions. These emotions were chosen on the basis that they are included in the set of basic emotions defined by Ekman and Friesen [EF75]. After explaining the general purpose and goal of the study, the actors were dressed in the lycra suit to which the markers were affixed. The motion capture sessions were carried out at a Japanese university that does not require written consent. Next, a *configuration model* was created using a single frame (i.e., a static instance) of motion capture data.

4.1.2 Actors

Thirteen actors: 10 Japanese (three females and seven males between the ages of 18 and 22), two Sri Lankans (one female and one male both 28 years of age) and one American female research assistant, age 55, were recruited for participation. The Japanese and Sri Lankan actors were computer science students. The actors were asked to perform their own idea of the four different emotions through bodily expression. Each emotion was expressed by each actor four times (4 x 13 x 4 = 208). No constraints were placed on the actors in how they performed the affective postures. The affective expressions were represented by contiguous frames describing the position of the 32 markers in the 3D space.

4.1.3 Stimulus identification

Twenty-six postures were discarded due to data post-processing problems, leaving a set of 182 affective postures P for experimental use. Once the affective postures were collected, it was necessary to manually locate the *apex* instant of the postures to be used as the study stimuli. The apex postures corresponded to the actors' evaluation of the most emotionally expressive instant of each motion performed.

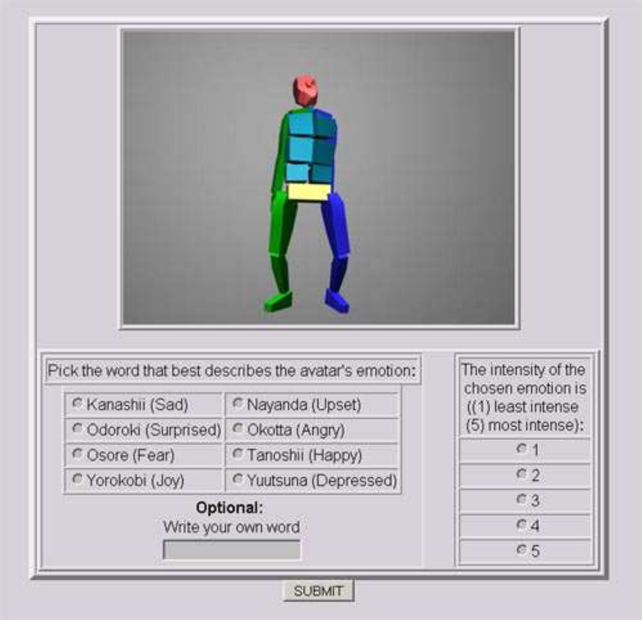
4.2 Human Recognition of Basic Emotion Categories from Posture

The goal of this section is to examine the extent to which human observers can recognise basic emotion categories from posture. As of yet there are no recognised benchmarks for evaluating recognition rates. Thus, chance level is typically considered the target rate for human recognition of affect. Using the repeated sub-sampling method outlined in Chapter 3, benchmarks computed on the observers' agreements will be used for evaluating the performance of automatic recognition models discussed later in the chapter.

4.2.1 Survey Procedure

To create the stimuli for the human recognition of affect, the original motion capture data was imported into 3D Studio Max [Aut08] and mapped to the default 3D faceless humanoid avatar. Static posture images were rendered for each apex posture. An online survey was conducted using these avatar stimuli. The aim was to obtain judgments on the affective postures in order to assign a ground truth emotion category to each posture as defined in Chapter 3, Section 3.2.4.

A forced-choice experimental design was implemented. For each page (one posture per page), observers were asked to choose an emotion label to represent the posture displayed. The set of labels $L = \{angry, fear, happiness, sadness\}$. Two nuances were considered for each label in L to make an eight word list, i.e., angry, upset (angry); fearful, surprised (fear); happy, joy (happy); and sad, depressed (sad). The purpose was to offer the observers a larger variety of emotion options, but remain focused on the four discrete categories. The set of posture stimuli used was reduced from the original 182 static posture images. Some of the



Pick the word that best describes the avatar's emotion:

○ Kanashii (Sad)	○ Nayanda (Upset)
○ Odoroki (Surprised)	○ Okotta (Angry)
○ Osore (Fear)	○ Tanoshii (Happy)
○ Yorokobi (Joy)	○ Yuutsuna (Depressed)

Optional:
Write your own word

The intensity of the chosen emotion is ((1) least intense (5) most intense):

○ 1
○ 2
○ 3
○ 4
○ 5

Figure 4.1: An example of the emotion category evaluation survey.

postures were similar to each other, therefore to reduce the amount of time required by the observers, many of the very similar looking postures were excluded, yielding 108 postures, i.e., $|P| = 108$. An example of the emotion category posture judgment survey can be seen in Figure 4.1.

Observers

The posture judgment survey was completed by a set of 87 observers. Thirty-three Japanese O_{JA} ranging in age from 18 to 25. The majority of these observers were computer science university students. Twenty-seven Sri Lankans O_{SL} ranging in age from 23 to 30 and 27 Caucasian Americans O_{US} ranging in age from 28 to 60. While the educational level of the Sri Lankan and American observers was similar, the educational background and career status were more varied. As the Sri Lankan participants were educated in English, the survey was presented in English and Japanese only.

Table 4.1: An overview of the within observers agreement results for all observers combined O and each group O_{JA}, O_{SL}, O_{US} separately. The $W.AgrLabel(L, O)$ across L is listed in the second column and the $W.AgrLabel(l_j, O)$ for each l_j is listed in the last four columns. The number of p_i per l_j per $O, O_{JA}, O_{SL}, O_{US}$ is noted in parentheses

Observers	$W.AgrLabel(L, O) * 100$	$W.AgrLabel(l_j, O) * 100$			
		Angry	**Fear**	**Happy**	**Sad**
3 groups combined	61%	57% (33)	66% (26)	57% (29)	72% (24)
Japanese	67%	63% (34)	71% (33)	61% (23)	72% (18)
Sri Lankan	62%	54% (35)	69% (23)	60% (26)	66% (24)
American	64%	61% (31)	64% (28)	58% (26)	73% (23)

4.2.2 Overview of the Survey Data

Before computing the benchmarks for this study, the aim was to get a general overview of the survey data; the measures of within observers agreement $W.AgrLabelLabel(l_j, O)$ for each label $l_j : j = 1, ..., 4$ (as defined in Equation (3.2)) and within observers agreement $W.AgrLabel(L, O)$ across the set of labels L (as defined in Equation (3.3)) were computed.

Observer-observer agreement

A ground truth label $gtl(p_i, O)$ from L was assigned to each posture p_i in P for the observers from the three sets of observers combined O and for each group individually O_{JA}, O_{SL}, O_{US}. The results can be seen in Table 4.1. The Table lists the results for the three groups combined O in the first row and the results within each group of observers O_{JA}, O_{SL}, O_{US} separately in the last three rows. The columns list the $W.AgrLabel(L, O)$ across the set of labels L first and the $W.AgrLabel(l_j, O)$ for each emotion label l_j in the remaining four columns. Chance agreement was 25% since four emotion categories were considered. For the entire set of observers O, $W.AgrLabel(L, O)$ across the the set of labels L and $W.AgrLabel(l_j, O)$ for each label l_j were both well above chance level. The same results were obtained for each set of observers O_{JA}, O_{SL}, O_{US} separately as hypothesised.

The overall results for the three groups combined O are reported in Figure 4.2. Each posture p_i is represented by a pie chart showing the frequency of use $freq(p_i, l_j)$ (as defined

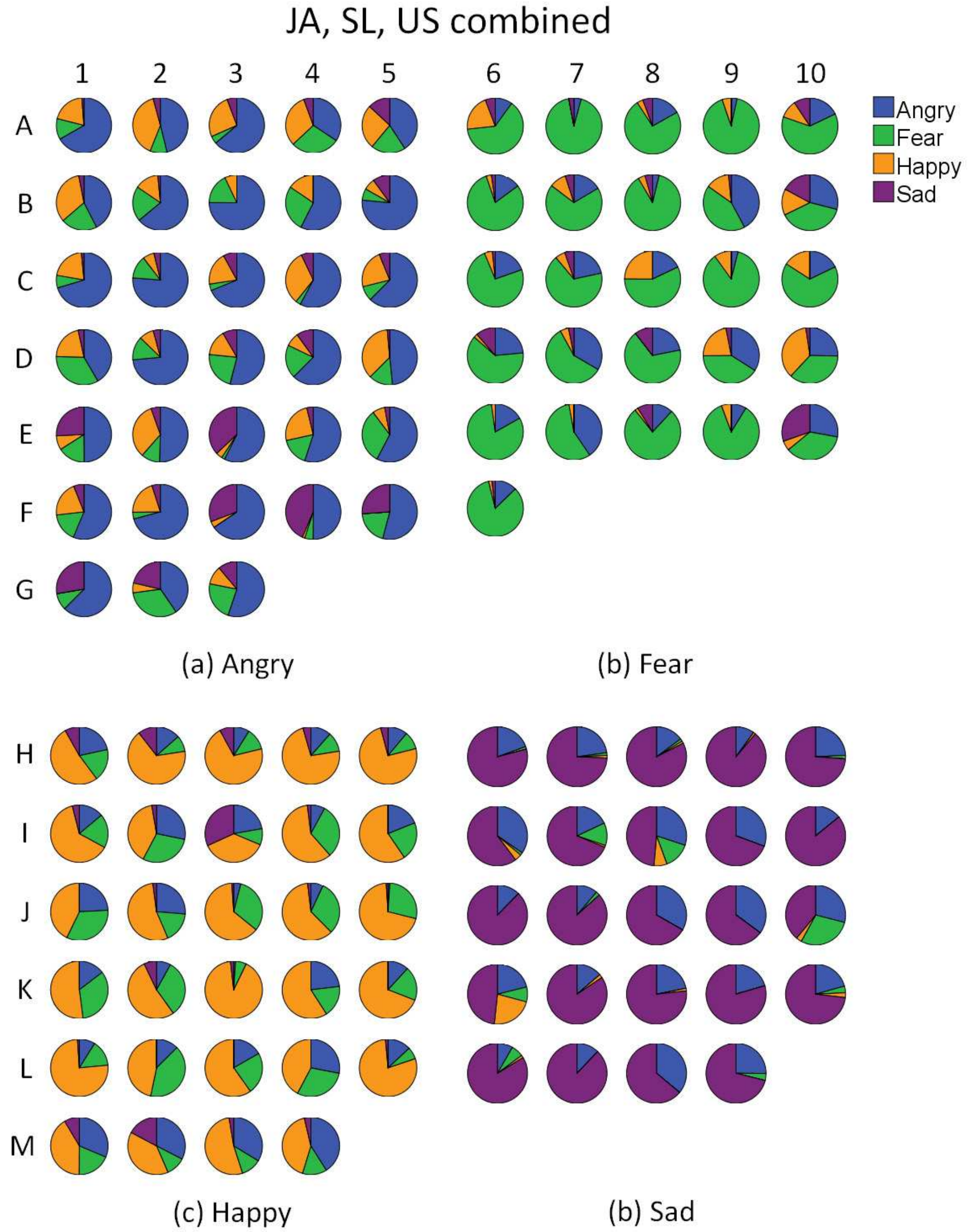


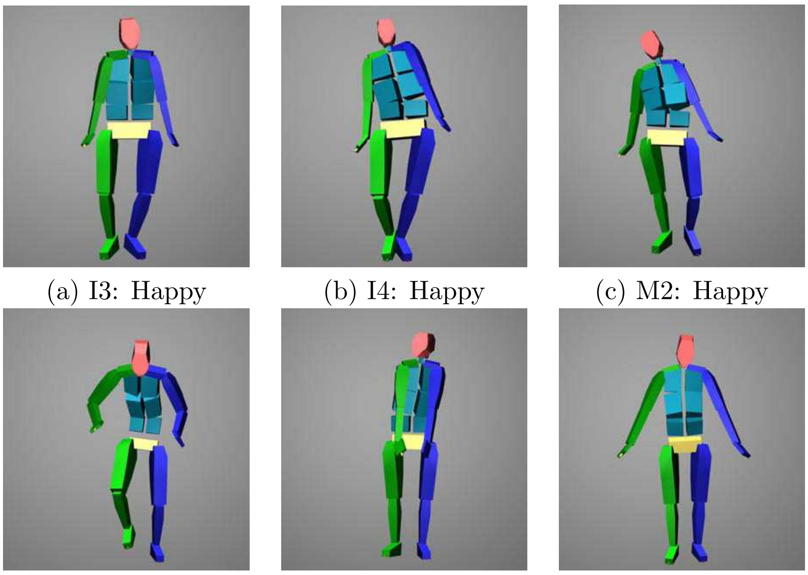
Figure 4.2: Each pie chart indicates the frequency of use $freq(p_i, l_j)$ for each emotion label $l_j : j = 1, ..., 4$ for each posture p_i for the set of postures P for the three sets of observers combined O. The column numbers and the row letters allow specific postures to be easily identified and located when referenced in the text or in other Figures

in Equation (3.1)) of each label $l_j : j = 1, ..., 4$. The pie charts are grouped according to their most frequent label, i.e., the ground truth label $gtl(p_i, O)$ associated to the corresponding posture p_i.

The individual results for each of the three groups O_{JA}, O_{SL}, O_{US} are reported in Figures 4.4, 4.5 and 4.6. In this case, the pie charts are grouped according to the order presented in Figure 4.2 rather than the ground truth assigned by an individual group of observers. This is simply to facilitate the comparison between the overall results and the individual group results. In order to easily locate individual postures, the rows are labelled with letters and the columns are labelled with numbers. For example, the posture located in the third row of the third column corresponds to position C3. Ordering the postures in this way allows for easy identification of specific posture differences between the three groups separately and combined. These reference positions will be used throughout the rest of the emotion categories sections of this Chapter for the discussions on individual postures.

- **All observers combined:** As listed in the first row of Table 4.1, the emotion categories l_j with the lowest within observers agreement $W.AgrLabel(l_j, O)$ were angry and happy. As seen in Figure 4.2(a), the angry labelled postures, there was little consensus as to the second most frequent label. For the happy labelled postures (Figure 4.2(a)), the confusion was with more activated types of emotions such as anger or fear. The second most frequent label was rarely confused with sad with the exception of posture I3. In fact, this posture (Figure 4.3(a)) somewhat resembles other sad labelled postures with the head just starting to bend forward and the arms stretched alongside the body. The difference is that the wrists are bent, extending the hands laterally which resembles some of the lower frequency (i.e., ambiguous) happy postures such as I4 and M2 (Figure 4.3(b) and (c), respectively).

The ambiguities with the fear labelled postures also occurred mainly with the more activated types of emotions, angry and happy, with the exception of sad for a few postures, such as E10 and B10. In the case of posture E10 (Figure 4.3(d)), it does appear to have angry and sad qualities according to the postures ground truth labelled as such, e.g., the



(a) I3: Happy

(b) I4: Happy

(c) M2: Happy

(d) E10: Fear

(e) J10: Sad

(f) K6: Sad

Figure 4.3: Posture examples for the three groups of observers combined O. (a)-(c) show low frequency happy postures; (d) shows a low frequency fear posture (e) and (f) show low frequency sad postures. The letter and number pair under each posture image refers to the location of that posture in the pie charts in Figures 4.2, 4.4, 4.5 and 4.6

bent elbows of angry and the head bent forward.

For the sad labelled postures, the ambiguities occurred for angry with the exception of J10 and K6. These postures, with the head turned and the body tilted slightly to the side, resembles fear labelled postures. In the case of posture K6 (Figure 4.3(f)), while the body is straight and the head is bent forward, the arms are somewhat extended laterally, which does not resemble the other sad labelled postures.

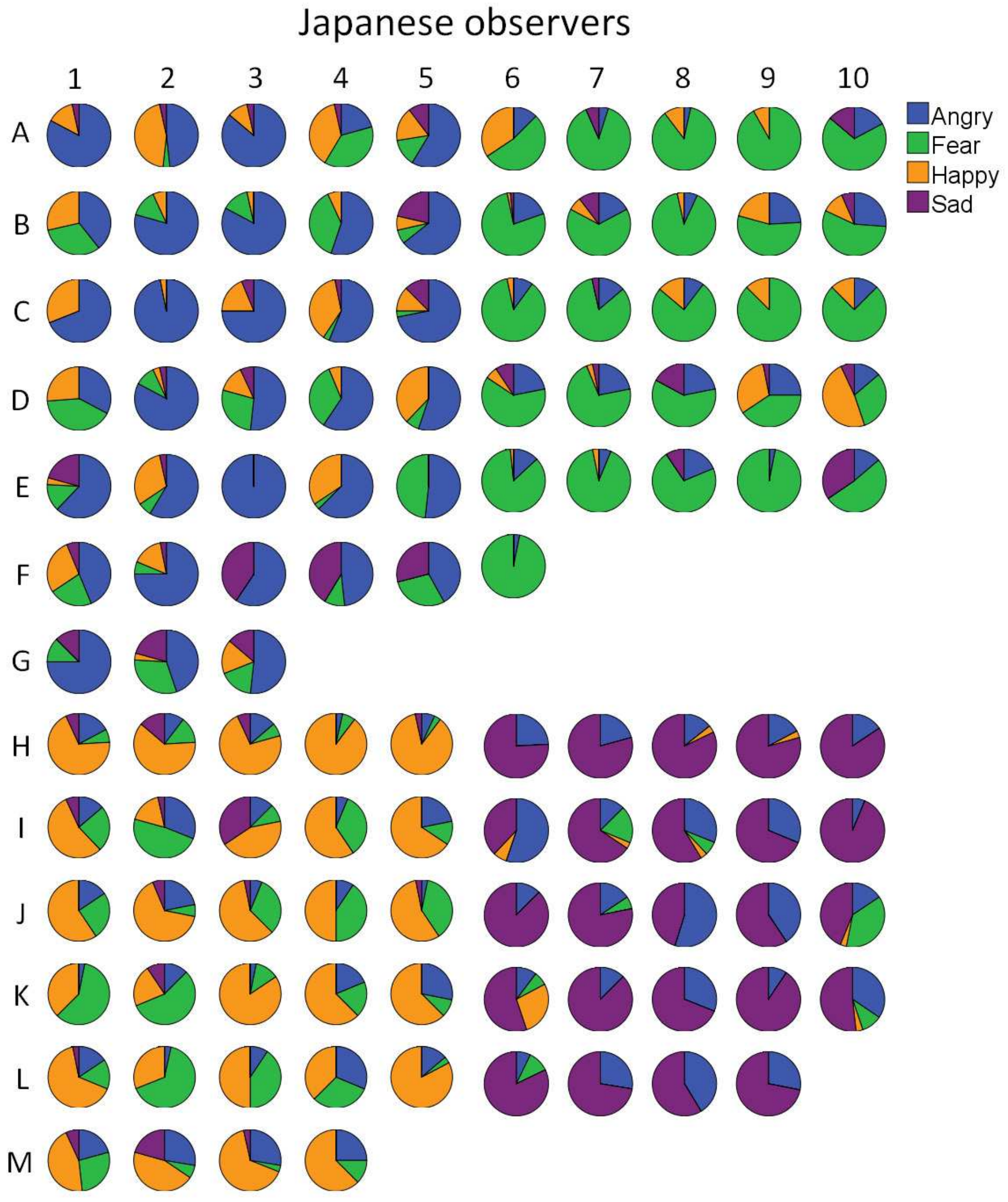


Figure 4.4: Each pie chart indicates the frequency of use $freq(p_i, l_j)$ for each emotion label $l_j : j = 1, ..., 4$ for each posture p_i for the set of postures P for the Japanese observers O_{JA}. The column numbers and the row letters allow specific postures to be easily identified and located when referenced in the text or in other Figures

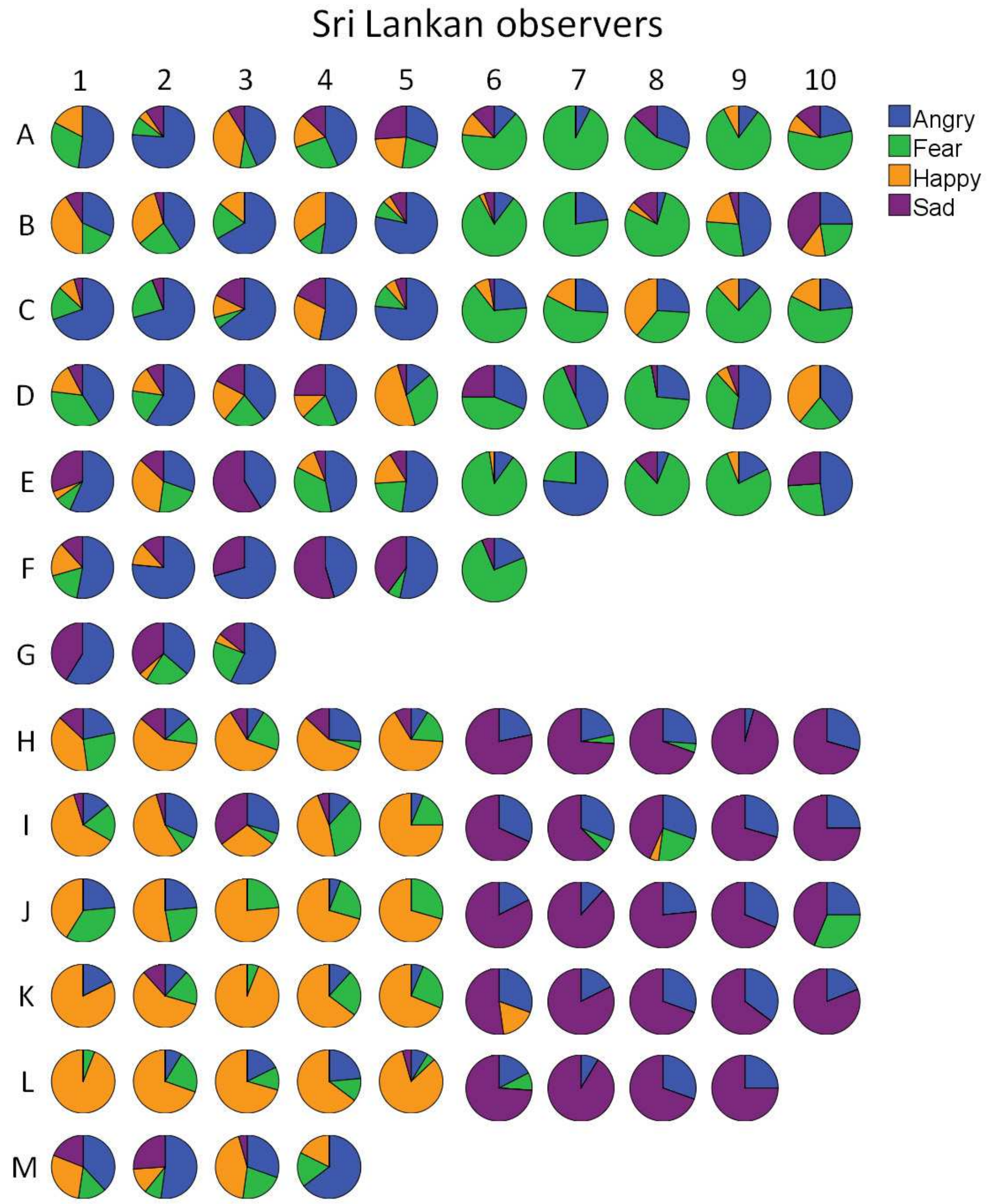


Figure 4.5: Each pie chart indicates the frequency of use $freq(p_i, l_j)$ for each emotion label $l_j : j = 1, ..., 4$ for each posture p_i for the set of postures P for the Sri Lankan observers O_{SL}. The column numbers and the row letters allow specific postures to be easily identified and located when referenced in the text or in other Figures

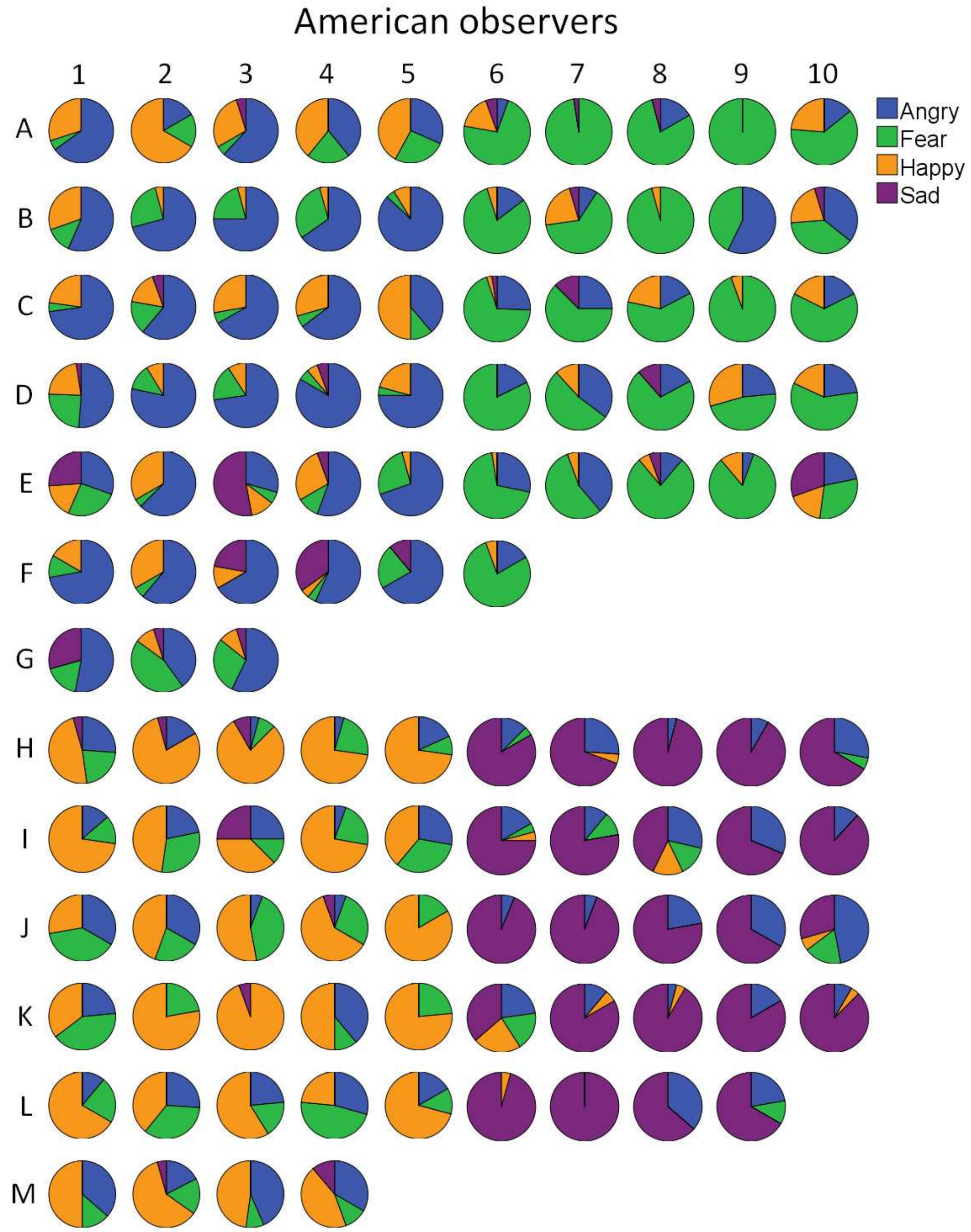
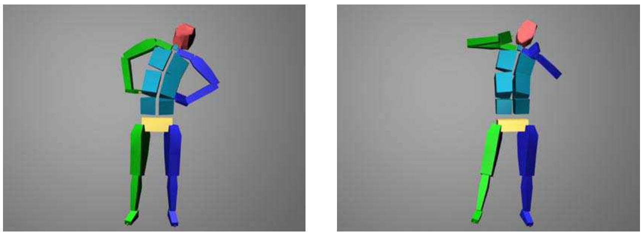


Figure 4.6: Each pie chart indicates the frequency of use $freq(p_i, l_j)$ for each emotion label $l_j : j = 1, ..., 4$ for each posture p_i for the set of postures P for the American observers O_{US}. The column numbers and the row letters allow specific postures to be easily identified and located when referenced in the text or in other Figures



(a) A2 (b) A3

Figure 4.7: The postures discussed in comparing O_{JA}, O_{SL}, O_{US}. (a) Angry for O_{SL} but mainly happy for O_{JA} and O_{US}; (b) obtained high frequencies of both angry and happy according to O_{SL} but high frequencies of mainly angry only according to O_{JA} and O_{US}. A2 and A3 refer to the location of the postures in Figures 4.3, 4.4, 4.5 and 4.6.

- **Individual observer groups:** Looking at the pie charts for the individual groups of observers O_{JA}, O_{SL}, O_{US}, Figures 4.4, 4.5 and 4.6, it can be seen that posture A2 (Figure 4.7(a)) obtained a high frequency of use for happy for the Japanese O_{JA} (Figure 4.4) and the Americans O_{US} (Figure 4.6), but the posture was mainly angry for the Sri Lankans O_{SL} (Figure 4.5). In fact, it was one of the four postures with the highest frequency for angry for the Sri Lankans, which may indicate a difference in how anger is perceived from bodily expressions by the Sri Lankans as opposed to the Japanese or the Americans. However, posture A3 (Figure 4.7(b)), which has some configurational similarities to posture A2, obtained a much higher frequency of use for angry for the Japanese (86%). This posture looks more like the other angry postures for all three observer groups combined O.

4.2.3 Creating Benchmarks

To create benchmarks for the human recognition of basic emotions from posture, the first step was to assess observer reliability. Fleiss' kappa was computed for each group of observers O_{JA}, O_{SL}, O_{US} separately as well as for all observers combined O. For the individual groups, Fleiss' kappa was highest for the Japanese O_{JA} (0.459 = moderate agreement), with the Americans O_{US} second (0.415 = moderate agreement) and the Sri Lankans O_{SL} last (0.356 = fair). Fleiss' kappa for all observers combined O was 0.397 (fair agreement).

The second step was to create the benchmarks that will be used to evaluate the performance rates of the automatic recognition models. Using the method described in Chapter 3, Section 3.2.4, the analysis was carried out in two ways: i) between the three groups of observers, with each group considered to be a subset O_{JA}, O_{SL}, O_{US}; and ii) with the three groups combined O and using the random repeated sub-sampling method to partition the observers into three disjoint subsets $O_{s,1}$, $O_{s,2}$ and $O_{s,3}$.

Observer agreement reliability results: Between the three groups

Inter-observer agreement reliability was measured to test the consistency between the three groups of observers. First, a ground truth label $gtl(p_i, O_{sk})$ was assigned to each posture p_i for each group of observers O_{JA}, O_{SL}, O_{US}. Next, the between observers agreement $B.AgrLabel(O_{s,1}, O_{s,2})$, i.e., the average agreement between two subsets of observers (as explained in Chapter 3, Section 3.2.4), Cohen's kappa [Coh60], the 95% confidence interval and the strength of agreement [LK77] between all pairs of the three observer groups were computed. The results, listed in Table 4.2, show substantial levels of inter-observer agreement reliability for three group pairs, indicating excellent agreement beyond chance [BCMS99].

Table 4.2: The $B.AgrLabel(O_{s,1}, O_{s,2})$ and the inter-observer agreement reliability between each pair of observer groups. The resulting benchmark (used to evaluate automatic recognition model performance) is listed in the last column.

Inter-observer agreement reliability between the 3 groups					
Observer ID	$B.AgrLabel(O_{s,1}, O_{s,2}) * 100$	**Kappa**	**95% CI**	**Strength**	**Benchmark**
JA * SL	72.22%	0.627	0.513, 0.741	Substantial	76.54%
JA * US	80.56%	0.739	0.639, 0.839	Substantial	
SL * US	76.85%	0.690	0.582, 0.798	Substantial	

Observer agreement reliability results: The three groups combined

Inter-observer agreement reliability was also measured to test the consistency between observers of the three groups of combined O. Ten trials (i.e., $s = 1, ..., 10$) were created using the random repeated sub-sampling procedure described in Chapter 3. Each trial com-

Table 4.3: The $B.Agr(L, O_{s,1}, O_{s,2})$ and the inter-observer agreement reliability (i.e., Cohen's kappa) between $O_{s,1}$ and $O_{s,2}$ for the 10 trials. The resulting benchmark to be used in evaluating the performance of the automatic recognition models is listed in the last column.

Inter-observer agreement reliability between the 3 groups					
Trial	$B.AgrLabel(O_{s,1}, O_{s,2}) * 100$	**Kappa**	**95% CI**	**Strength**	**Benchmark**
1	85.29%	0.804	0.712, 0.896	Substantial	84.80%
2	83.33%	0.777	0.681, 0.873	Substantial	
3	86.27%	0.816	0.726, 0.906	Almost perfect	
4	81.37%	0.752	0.652, 0.852	Substantial	
5	89.22%	0.856	0.776, 0.936	Almost perfect	
6	84.31%	0.789	0.693, 0.885	Substantial	
7	83.33%	0.777	0.681, 0.873	Substantial	
8	85.29%	0.803	0.711, 0.895	Substantial	
9	81.37%	0.752	0.652, 0.852	Substantial	
10	88.24%	0.843	0.759, 0.927	Almost perfect	

prised three disjoint subsets $O_{s,1}$, $O_{s,2}$ and $O_{s,3}$ of 29 observers. For each trial, the between observers agreement $B.AgrLabel(O_{s,1}, O_{s,2})$ and Cohen's kappa coefficient were computed between $O_{s,1}$ and $O_{s,2}$. The results are listed in Table 4.3. Each row constitutes a trial and lists the $B.Agr(L, O_{s,1}, O_{s,2})$, Cohen's kappa, the 95% confidence interval and the strength of agreement [LK77]. The strength of agreement was interpreted as substantial and almost perfect for all 10 trials, which can be taken to mean excellent agreement beyond chance [BCMS99].

Benchmarks

To set the benchmarks of human recognition of basic emotion from body posture, the same three-way data splits were maintained: i) between the three groups O_{JA}, O_{SL}, O_{US} and ii) for all the observers combined O. The benchmark between the three groups is listed in the last column of Table 4.2. It is the average across the between observers agreements $B.Agr(L, O_{s,1}, O_{s,2})$ for all pairs of observer groups, 76.54% ($SD = 2.35\%$).

The benchmark for all the observers combined O is listed in the last column of Table 4.3. It is the average of the between observers agreements $B.Agr(L, O_{s,1}, O_{s,2})$ across the 10 trials, 84.80% ($SD = 2.63\%$). Comparing the between observers agreements achieved in this research with those found in other research, comparable or higher between observers agreement levels were obtained here. For instance, 52% agreement was obtained by Camurri and colleagues [CLV03] for acted dance motions of three basic emotions (anger, joy and fear). 68% agreement was obtained between observers on acted affective walking movements on three basic emotions (anger, joy and sad) in a study by Crane and Gross [CG07].

4.2.4 Discussion

The results obtained for the within observers agreement for the three groups combined O and O_{JA}, O_{SL}, O_{US} separately showed that the emotion categories with the highest agreement were fear and sad. The result for sad is not surprising as sad is often very well recognised by human observers [Cou04][KKVB$^+$05]. This may be especially true when considering the limited list of labels from which the observers had to choose.

The agreement for fear was interestingly high considering other research in which the observer agreement for fear was less than other basic emotions, such as anger and happiness [Cou04]. Righart et al's [VdSRdG07] results also indicated a difficulty in recognising fear from static posture. In fact, the authors claimed that it was *"the most difficult emotion to recognise in a forced-choice paradigm."* However, Coulson [Cou04] reported that the neural systems involved in detecting motion become activated when viewing, which may affect the viewing of static images as well as motion. This is referred to as 'implied bodily action' and is a survival mechanism processed by the brain - not a conscious process [vHMGdG07]. Atkinson et al [ADGY07] assert that static form information of posture, i.e., configurational cues, plays a bigger part in recognising fear than kinematics.

4.3 Low-Level Posture Description

The role of this section is to explain how the low-level posture description is determined. Towards building automatic recognition models of affective posture, the numerical description of the postures must be obtained. The upper body is the main focus of the posture description as determined by preliminary results in this research project indicating that the upper body is used most in standing postural displays of emotion and affect [KFTBB03]. For each posture p_i, the posture description comprises a set of 24 low-level configuration features $F_i = \{f_{i1}, ..., f_{i24}\}$ [BBK03]. The features and the parameters to compute them are listed in Table 4.4 and illustrated in Figure 4.8. The computed features are angles between segments and distances between joints.[1] Before the features are computed, the *x, y, z* positions of each joint of the posture (collected by motion capture) are rotated so that the axis between the hips is aligned with the x axis of the 3D space and perpendicular to the y axis, with the front of the body facing in a positive direction. Each computed feature $f_{iw} : w = 1, ..., 24$ is normalised to [0,1] to take into account an actor's body size.

The 'distance' features (V6 to V23 in Table 4.4) describe a pseudo distance between two joints on each axis separately, i.e., vertically, laterally and frontally. For example, V6, V8 and V10 describe the normalised position of the hand with respect to the shoulder along the z axis (vertical extension), the x axis (lateral extension) and the y axis (frontal extension), respectively. Refer to Figure 4.8. Hence, for each individual axis *x, y, z,* a distance feature f_{iw} between a joint jp_1 and a joint jp_2 is computed as follows

$$f_{iw} = 1 - \left(\frac{b_{iw} - v_{iw}}{max\ range_{iw}} \right) \tag{4.1}$$

where b_{iw} is the maximum value that a joint jp_1 could reach on that particular axis with respect to the current position of jp_2, v_{iw} is the current position of the body joint jp_1 on that axis, normalised with respect to $max\ range_{iw}$, i.e., the maximum range that the joint jp_1 can cover according to each actor's body and with respect to the joint jp_2. For example, for the vertical extension of the hand (jp_1) with respect to the shoulder (jp_2) (V6),

b_{iw} = length of the fore arm + length of the upper arm + z position of the shoulder

v_{iw} = z position of the wrist

$max\ range_{iw}$ = (length of the upper arm + length of the fore arm) $*$ 2

[1] Initially, a comparison was carried out to examine the differences between 1D, 2D and 3D features. The performance results were similar. It was decided to keep mainly the 1D and 2D features as it allows for different points of view of the posture to be simulated (outside the scope of the thesis).

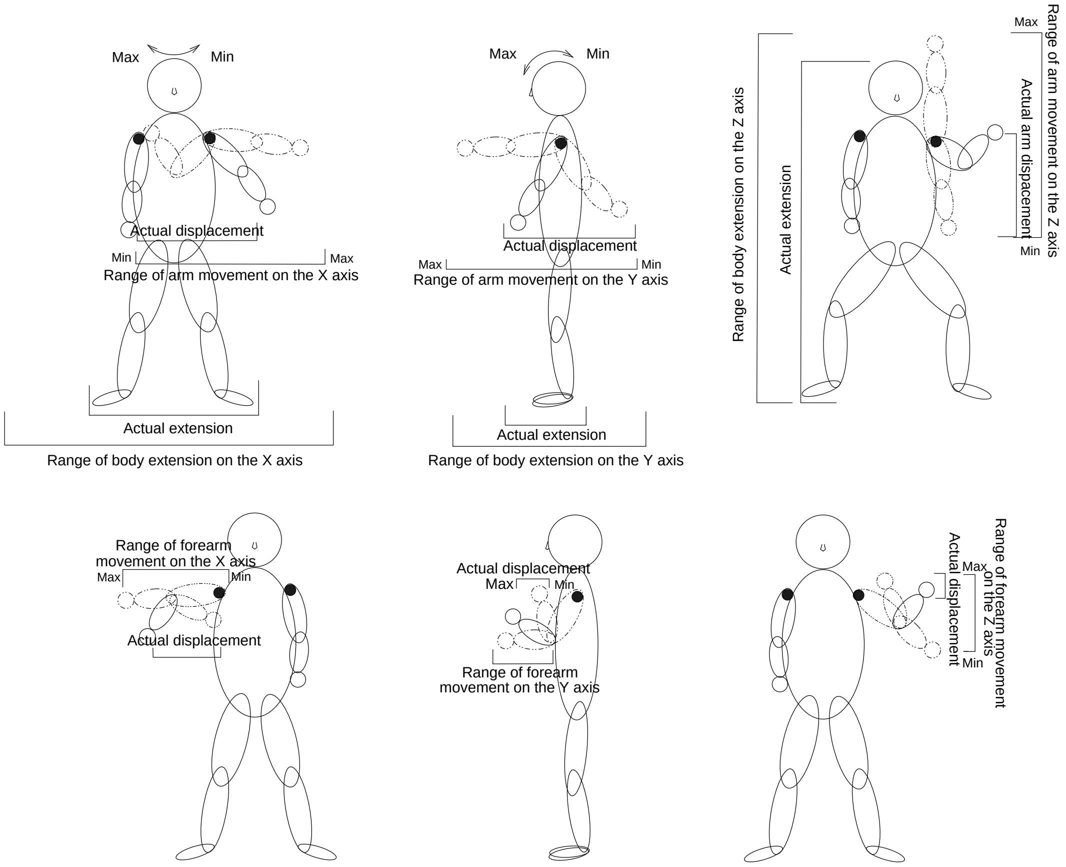

Max
Min
Actual displacement
Min
Max
Range of arm movement on the X axis
Actual extension
Range of body extension on the X axis
Max
Min
Actual displacement
Max
Min
Range of arm movement on the Y axis
Actual extension
Range of body extension on the Y axis
Range of body extension on the Z axis
Actual extension
Max
Range of arm movement on the Z axis
Actual arm dispacement
Min
Range of forearm movement on the X axis
Max
Min
Actual displacement
Actual displacement
Max
Min
Range of forearm movement on the Y axis
Range of forearm movement on the Z axis
Actual displacement
Max
Min

Figure 4.8: Examples of the low-level posture description features. The upper part shows the rotation and bending of the head, and the extension of the arms along the three axes. The lower part shows the pseudo distance between the hands and the elbows and the elbows and the shoulders along the three axes

Table 4.4: The set of low-level posture features used. The Code column lists the individual low-level posture features. v_{iw} = the current body joint position; b_{iw} = the maximum value of a joint; $maxrange_{iw}$ = the maximum range possible according to an actor's body. The following short-cuts are used: Orient: Orientation, Dist: Distance, L: Left, R: Right, B: Back, F: Front, Sho: Shoulder, Shos: Shoulders, LG: length, FA: fore arm, UA: upper arm

Code	**Posture features** f_{iw}	v_{iw}	b_{iw}	$max\ range_{iw}$
V4	$Orient_{XY}$: B.Head - F.Head axis	xy direction B. & F. Head	45	135
V5	$Orient_{YZ}$: B.Head - F.Head axis	yz direction B. & F. Head	30	80
V6	$Dist_z$: R.Hand - R.Sho	z R. Wrist	LG FA + LG UA + z R. Sho	(LG FA + LG UA) $*$ 2
V7	$Dist_z$: L.Hand - L.Sho	z L. Wrist	LG FA + LG UA + z L. Sho	(LG FA + LG UA) $*$ 2
V8	$Dist_y$: R.Hand - R.Sho	y R. Wrist	LG FA + LG UA + y R. Sho	LG FA $*$ 2 + LG UA $*$ 2/3
V9	$Dist_y$: L.Hand - L.Sho	y L. Wrist	LG FA + LG UA + y L. Sho	LG FA $*$ 2 + LG UA $*$ 2/3
V10	$Dist_x$: R.Hand - L.Sho	x R. Wrist	LG FA + LG UA + x R. Sho	LG FA + LG UA + LG btwn L. & R. Shos
V11	$Dist_x$: L.Hand - R.Sho	x L. Wrist	x L. Sho $-$ LG FA $-$ LG UA	LG FA + LG UA + LG btwn L. & R. Shos
V12	$Dist_x$: R.Hand - R.Elbow	x R. Wrist	LG FA + x R. Elbow	LG FA $*$ 2
V13	$Dist_x$: L.Hand - L.Elbow	x L. Wrist	LG FA + x L. Elbow	LG FA $*$ 2
V14	$Dist_x$: R.Elbow - L.Sho	x R. Elbow	LG UA + x R. Sho	LG UA $*$ 3/2
V15	$Dist_x$: L.Elbow - R.Sho	x L. Elbow	LG UA + x L. Sho	LG UA $*$ 3/2
V16	$Dist_z$: R.Hand - R.Elbow	z R. Wrist	LG FA + z R. Elbow	LG FA $*$ 2
V17	$Dist_z$: L.Hand - L.Elbow	z L. Wrist	LG FA + z L. Elbow	LG FA $*$ 2
V18	$Dist_y$: R.Hand - R.Elbow	y R. Wrist	LG FA + y R. Elbow	LG FA
V19	$Dist_y$: L.Hand - L.Elbow	y L. Wrist	LG FA + y L. Elbow	LG FA
V20	$Dist_y$: R.Elbow - R.Sho	y R. Elbow	LG UA + y R. Sho	LG UA $*$ 2
V21	$Dist_y$: L.Elbow - L.Sho	y L. Elbow	LG UA + y L. Sho	LG UA $*$ 2
V22	$Dist_z$: R.Elbow - R.Sho	z R. Elbow	LG UA + z R. Sho	LG UA $*$ 2
V23	$Dist_z$: L.Elbow - L.Sho	z L. Elbow	LG UA + z L. Sho	LG UA $*$ 2
V24	$Orient_{XY}$: Shos axis	xy direction R. & L. Shos	35	60
V25	$Orient_{XZ}$: Shos axis	xz direction R. & L. Shos	35	60
V26	$Orient_{XY}$: Heels axis	xy direction R. & L. Heels	60	120
V27	$3D - Dist$: R.Heel - L.Heel	LG btwn R. & L. Heels	LG hip $*$ 2	LG hip $*$ 2

It should be noted that the ranges described do not correspond to the entire kinematically possible range of what the actor can do. Instead, they correspond to the range of movement that was likely to occur given the scenario. In the case that a wider range of movement did in fact occur, the values are capped between [0,1]. The same is true for the orientation features described in the next paragraph.

For the orientation features, i.e., the rotation and bending of the head (V4 and V5), the rotation and inclination of the shoulders (V24 and V25) and the rotation of the heels (V26), are computed. The normalised segment connecting two body points v_{i1} and v_{i2} are computed on each 2D cartesian plane, separately (e.g., *xy, xz, yz*) as follows

$$f_{iw} = \frac{dir(v_{i1}, v_{i2}) + b_{iw}}{max\ range_{iw}} \tag{4.2}$$

where $dir(v_{i1}, v_{i2})$ is the direction of the segment with respect to the first dimension of the plane (e.g., for a 2D plane *xy*, it is the angle between the x axis and the segment), b_{iw} is the minimum angle that that body segment may portray and $max\ range_{iw}$ is the maximum range that that body segment can cover. For example, for the bending of the head in the plane *yz* (V5), $b_{iw} = 30$ and $max\ range_{iw} = 80$.

Given the 3D space *x, y, z*, and the position of the left heel (hee_1) and the right heel (hee_2) in this space, the 3D distance between the heels (V27) is computed as follows

$$f_{iw} = \frac{3D\ Euclidean\ Distance(hee_{i1}, hee_{i2})}{max\ range_{iw}} \tag{4.3}$$

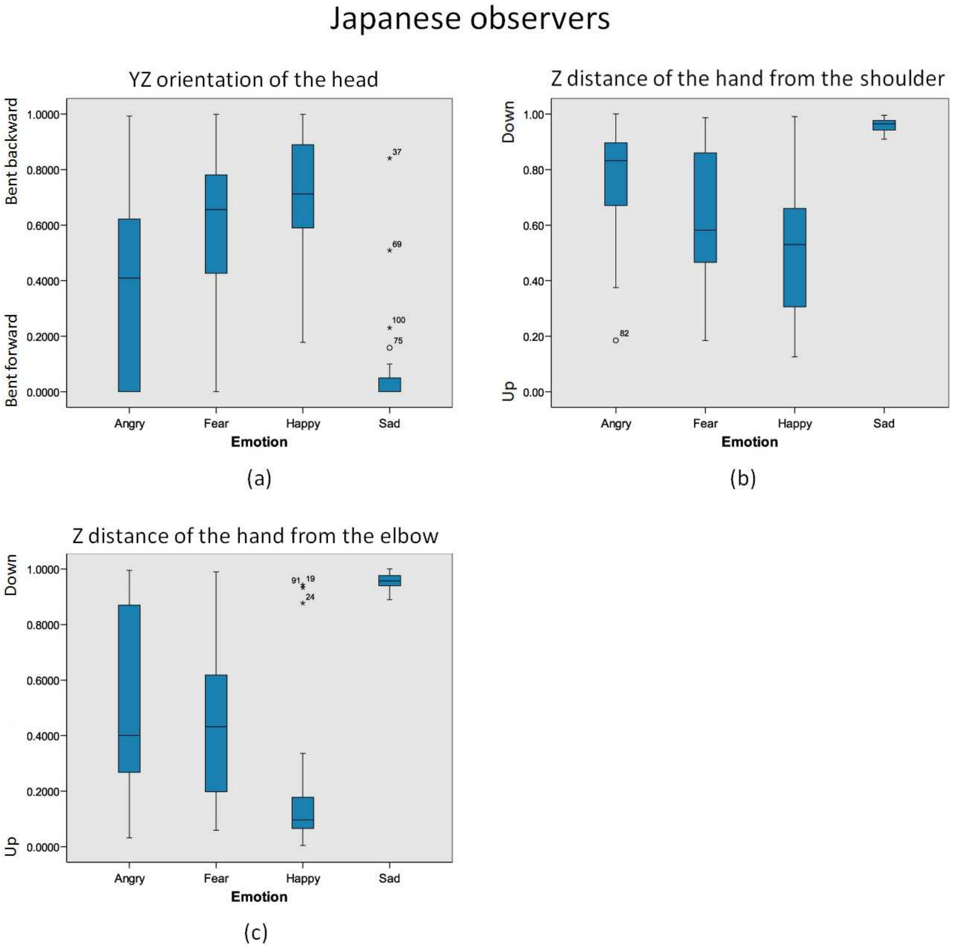
4.4 Low-Level Posture Description Analysis

The low-level posture description was analysed to evaluate the discriminative power of each low-level feature. To do so, each low-level feature f_{iw} was subjected to one-way ANOVAs for the set of emotion labels L for each of the three observer groups O_{JA}, O_{SL}, O_{US} separately. The results are summarised in Tables 4.5, 4.6 and 4.7, respectively. Listed in the first column of each Table are the low-level features shown to be important for discriminating between emotions with the significance level shown in the last column. The means for each emotion label l_j are shown in the middle four columns. The superscript letter pairs listed with the means denote significant differences between those pairs of emotion labels according to Tamhane's T2 post hoc comparisons, implemented for unequal variances (verified using Levene's test of homogeneity of variance). Boxplots are used to highlight and discuss the results. The outliers in the boxplots are illustrated with circles (i.e., the datum is 1.5 times the interquartile range) and asterisks (i.e., the datum is 3 times the interquartile range). The number beside each outlier indicates the row of the file in which the datum (i.e., posture p_i) can be found.

4.4.1 Japanese observers

The ANOVA results for the Japanese observers are depicted in Table 4.5. Significant differences were obtained for 20 of the 24 low-level posture features. For the majority of the features (V5-V9, V14, V16-V23), the main differences occur between sad and the other emotion categories. For V5, the forward/backward bending of the head (the boxplot in Figure 4.9(a)), the most interesting significant differences occur for fear with sad, and happy with sad. It is noticeable that for fear and happy, the head is generally straight up or bent backward slightly, whereas for sad (apart from four outliers) the head is almost bent forward as far as possible. For the vertical distance of the hands from the shoulders (V7), shown in Figure 4.9(b), even though differences are seen between angry and happy, the most noticeable result is that sad is significantly different from the other emotion categories. Other vertical differences occur for the distance of the hands from the elbows (V16), shown in Figure 4.9(c), particularly between happy and the other three emotions, and between sad and the other three emotions. For sad, the arm is extended straight down, close to body, whereas for happy, the arm is raised over the head.

Examples of the typical postures that the Japanese assigned to each emotion category are shown in Figure 4.10. Angry is associated to postures in which the head is bent forward



(a)

(b)

(c)

Figure 4.9: Examples of low-level posture description features with significant differences between emotions for the Japanese observers. (a) The forward/backward bending of the head (V5); (b) The vertical distance of the hand from the shoulder (V7); (c) The vertical distance of the hand from the elbow (V16)

Table 4.5: The low-level posture description features which reached significance between the emotions for the Japanese observers (one-way ANOVAs with df = 3 (emotions)). For each feature f_{iw}, *a-e* pairs indicate the significant differences between means according to Tamhane's T2 *post-hoc* comparisons.

Japanese observers					
	Means for affective states				
Low-level feature	**Angry**	**Fear**	**Happy**	**Sad**	p
V5 - $Orientation_{YZ}$: B.Head - F.Head axis	.36^{abc}	.62^{ad}	.71^{be}	.10^{cde}	.000
V6 - $Distance_z$: R.Hand - R.Shoulder	.67^{ab}	.61^{c}	.47^{ad}	.96^{bcd}	.000
V7 - $Distance_z$: L.Hand - L.Shoulder	.77^{abc}	.63^{ad}	.53^{be}	.96^{cde}	.000
V8 - $Distance_y$: R.Hand - R.Shoulder	.61^{a}	.61^{b}	.64^{c}	.86^{abc}	.000
V9 - $Distance_y$: L.Hand - L.Shoulder	.70^{a}	.61^{b}	.65^{c}	.88^{abc}	.000
V10 - $Distance_x$: R.Hand - L.Shoulder	.56^{a}	.49	.42^{ab}	.58^{b}	.005
V11 - $Distance_x$: L.Hand - R.Shoulder	.60^{a}	.58^{b}	.41^{abc}	.58^{c}	.001
V12 - $Distance_x$: R.Hand - R.Elbow	.69^{ab}	.59^{c}	.42^{acd}	.57^{bd}	.000
V13 - $Distance_x$: L.Hand - L.Elbow	.74^{ab}	.63	.47^{a}	.58^{b}	.000
V14 - $Distance_x$: R.Elbow - L.Shoulder	.32^{a}	.29^{b}	.32^{c}	.50^{abc}	.004
V15 - $Distance_x$: L.Elbow - R.Shoulder	.94^{ab}	.83^{ac}	.96^{cd}	.83^{bd}	.000
V16 - $Distance_z$: R.Hand - R.Elbow	.54^{ab}	.44^{cd}	.22^{ace}	.96^{bde}	.000
V17 - $Distance_z$: L.Hand - L.Elbow	.49^{abc}	.35^{ad}	.23^{be}	.68^{cde}	.000
V18 - $Distance_y$: R.Hand - R.Elbow	.70^{a}	.63^{b}	.69^{c}	.89^{abc}	.003
V19 - $Distance_y$: L.Hand - L.Elbow	.73^{a}	.69^{b}	.67^{c}	.91^{abc}	.002
V20 - $Distance_y$: R.Elbow - R.Shoulder	.43^{a}	47^{b}	.48^{c}	.59^{abc}	.023
V21 - $Distance_y$: L.Elbow - L.Shoulder	.49	.43^{a}	.49	.58^{a}	.037
V22 - $Distance_z$: R.Elbow - R.Shoulder	.77^{a}	.75^{b}	.67^{c}	.96^{abc}	.000
V23 - $Distance_z$: L.Elbow - L.Shoulder	.87^{a}	.77^{b}	.71^{c}	.95^{abc}	.000
V27 - $3D - Distance$: R.Heel - L.Heel	.42	.36^{a}	.57	.56^{a}	.005

slightly, the arms are bent slightly, and the hands remain close to the body between shoulder and hip height. This is somewhat different to features denoting angry in Coulson's study [Cou04]. The difference is that Coulson found angry attributed to postures with a backward bending head and arms held frontal. The similarity between the two studies is that arms are somewhat raised. Wallbott [Wal98] also identifies angry to postures with arms in front of the body.

Postures that signify fear for the Japanese are similar to Coulson's findings. The fear posture configuration shows a slightly backward bent head, with the arms somewhat frontal and lateral and bent at the elbow. Happy postures for the Japanese are also similar to

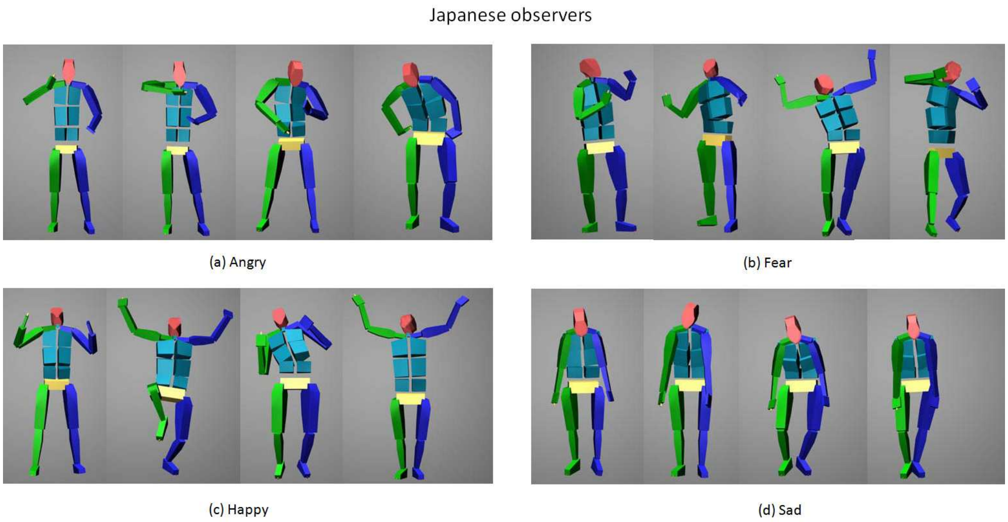


Figure 4.10: Avatar examples representing typical postures according to the posture description analysis for the four emotion categories according to the Japanese observers. (a) Angry; (b) Fear; (c) Happy; (d) Sad

Coulson's study. In both, happy is identified as having a backward bent head (more than for fear), with arms raised to shoulder height or over head. The difference between the two is that Coulson found straight arms to be indicative of happy, whereas the arms are bent at the elbows for the Japanese. Similar to both Coulson's study and the Japanese results, elated joy in Wallbott's study claims a backward bending head. However, different from Coulson's study and the Japanese results, Wallbott found the arms stretched out frontally to be indicative of elated joy.

Sad in both this thesis and Coulson's study is characterised by a head bent far forward and arms straight, extended down along the body. The results are similar the postural configuration of both sadness and shame in Wallbott's study which is characterised by a collapsed upper body.

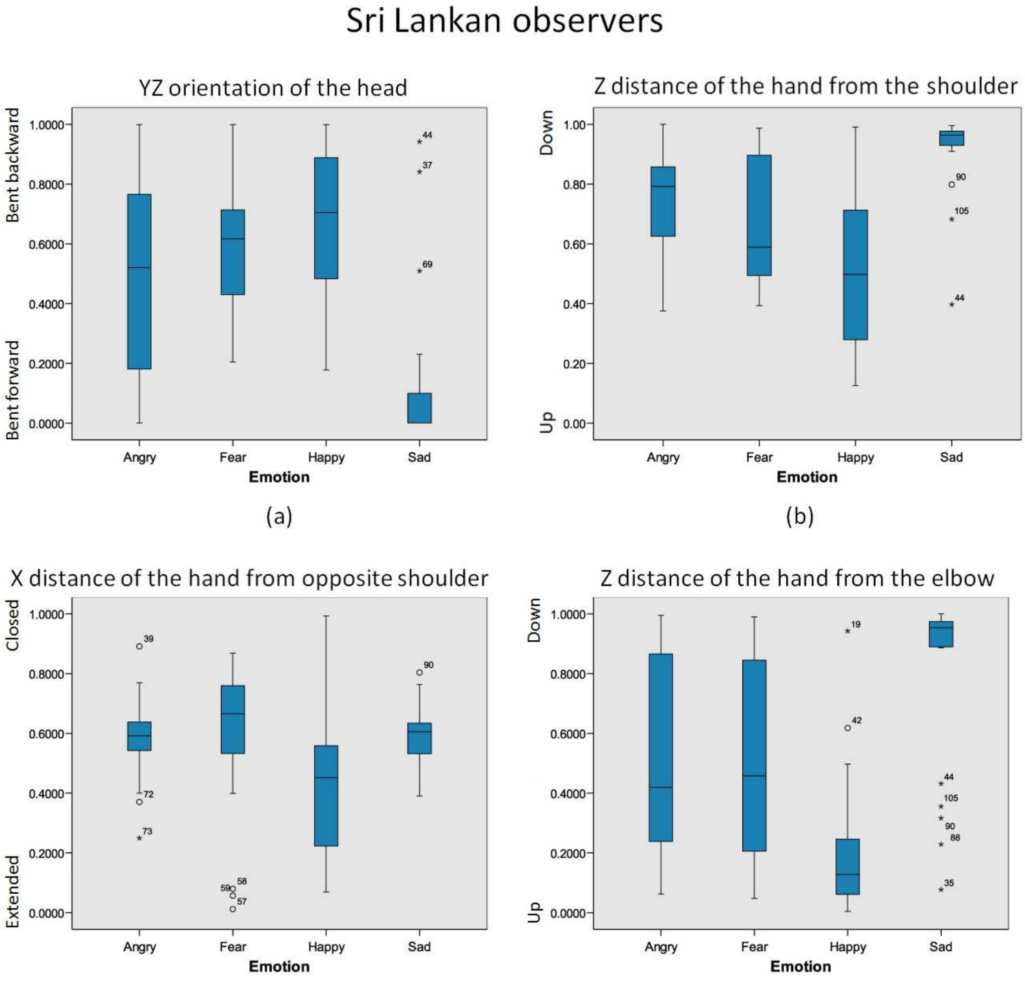
4.4.2 Sri Lankan observers

The results for the Sri Lankan observers are depicted in Table 4.6. Significant differences were obtained for 19 of the 24 low-level posture features. Shown in the boxplot in Figure 4.11(a), for the forward/backward bending of the head (V5), a very forward head bend is seen for sad, with sad significantly different from the three other emotions. The height of the hands in relation to the shoulders (V7), shown in Figure 4.11(b), and the elbows (V16), shown in Figure 4.11(d), is significant for distinguishing happy from the other emotions and sad from the other emotions. The arm is raised for happy, whereas the arm is extended down for sad. The lateral extension of the hands (V11), shown in Figure 4.11(c) is also important for happy. In fact, the hands are very laterally extended.

Examples of the typical postures that the Sri Lankans assigned to each emotion category are shown in Figure 4.12. In general, the Sri Lankans associate angry and fear to a wide variety of postures. No distinct head position indicates angry. A more defined head position is seen for fear, ranging from no bend to slightly bent backward. This result is similar to Coulson's [Cou04] findings, however the backward bending of the head was more pronounced in Coulson's study.

Happy postures for the Sri Lankans are similar to Coulson's findings. In both, the happy postures are demonstrated by arms that are mainly slightly vertical and lateral with a backward bending head. The difference between the two studies is that Coulson specifies that there is no bend of the elbows. The backward bending of the head for the Sri Lankans corresponds to Wallbott's [Wal98] findings, however Wallbott found the arms stretched out frontal instead of vertical and lateral.

Sad for the Sri Lankans is depicted by a closed body, i.e., the arms are straight down with little frontal or lateral positioning, and the head is very bent forward. The configuration of sadness for Coulson and Wallbott is similar to the results found here.



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

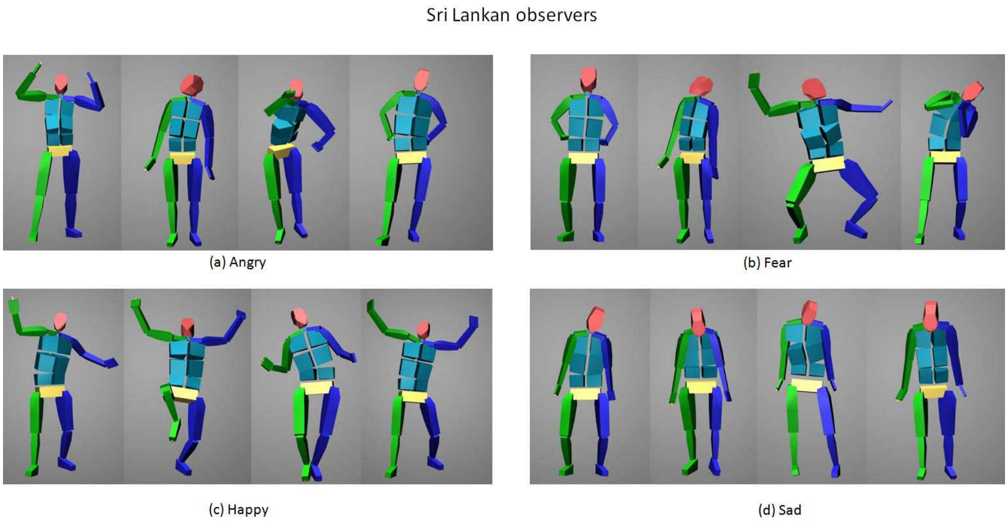
Figure 4.11: Examples of low-level posture description features with significant differences between emotions for the Sri Lankan observers. (a) The forward/backward bending of the head (V5); (b) The vertical distance of the hand from the shoulder (V7); (c) The lateral distance of the hand from the opposite shoulder (V11); (d) The vertical distance of the hand from the elbow (V16)

Table 4.6: The low-level posture description features which reached significance between the emotions for the Sri Lankan observers (one-way ANOVAs with df = 3 (emotions)). For each feature f_{iw}, *a-e* pairs indicate the significant differences between means according to Tamhane's T2 *post-hoc* comparisons.

Sri Lankan observers					
	Means for affective states				
Low-level feature	**Angry**	**Fear**	**Happy**	**Sad**	p
V5 - $Orientation_{YZ}$: B.Head - F.Head axis	$.48^{ab}$	$.61^{c}$	$.68^{ad}$	$.11^{bcd}$	.000
V6 - $Distance_z$: R.Hand - R.Shoulder	$.68^{ab}$	$.66^{cd}$	$.42^{ace}$	$.89^{bde}$	.000
V7 - $Distance_z$: L.Hand - L.Shoulder	$.74^{ab}$	$.68^{c}$	$.51^{ad}$	$.92^{bcd}$	.000
V8 - $Distance_y$: R.Hand - R.Shoulder	$.58^{a}$	$.63^{b}$	$.64^{c}$	$.82^{abc}$	.000
V9 - $Distance_y$: L.Hand - L.Shoulder	$.62^{a}$	$.66^{b}$	$.66^{c}$	$.85^{abc}$	.000
V10 - $Distance_x$: R.Hand - L.Shoulder	$.53^{a}$	$.54^{b}$	$.37^{abc}$	$.61^{c}$	.000
V11 - $Distance_x$: L.Hand - R.Shoulder	$.59^{a}$	.60	$.43^{ab}$	$.59^{b}$	.002
V12 - $Distance_x$: R.Hand - R.Elbow	$.64^{a}$	$.66^{b}$	$.39^{abc}$	$.62^{c}$	.000
V13 - $Distance_x$: L.Hand - L.Elbow	$.73^{ab}$	.65	$.47^{a}$	$.61^{b}$	.000
V14 - $Distance_x$: R.Elbow - L.Shoulder	$.32^{a}$	$.31^{b}$	$.25^{c}$	$.49^{abc}$	.000
V15 - $Distance_x$: L.Elbow - R.Shoulder	$.94^{a}$	.85	.92	$.85^{a}$	.003
V16 - $Distance_z$: R.Hand - R.Elbow	$.53^{ab}$	$.49^{cd}$	$.20^{ace}$	$.83^{bde}$	.000
V17 - $Distance_z$: L.Hand - L.Elbow	$.43^{ab}$	$.36^{c}$	$.26^{ad}$	$.65^{bcd}$	.000
V18 - $Distance_y$: R.Hand - R.Elbow	$.66^{a}$	$.67^{b}$	$.67^{c}$	$.87^{abc}$	.004
V19 - $Distance_y$: L.Hand - L.Elbow	$.66^{a}$	$.76^{b}$	$.67^{c}$	$.89^{abc}$	.000
V20 - $Distance_y$: R.Elbow - R.Shoulder	.42	.46	.50	.55	.046
V22 - $Distance_z$: R.Elbow - R.Shoulder	$.78^{a}$	$.78^{b}$	$.60^{c}$	$.94^{abc}$	.000
V23 - $Distance_z$: L.Elbow - L.Shoulder	$.86^{a}$	$.83^{b}$	$.66^{abc}$	$.93^{c}$	.000
V27 - $3D - Distance$: R.Heel - L.Heel	$.39^{a}$	.40	.52	$.55^{a}$	.046

4.4.3 American observers

The results for the American observers are depicted in Table 4.7. Significant differences are obtained for 20 of the 24 low-level posture features. Sad is most characteristically different from the other emotions. The majority of the significant differences occur between sad and the other three emotions. Angry and happy each have only one feature for which significant differences are obtained against the other emotions and fear does not have any. In the case of the forward/backward bending of the head (V5), for sad, the head is extremely bent forward, shown in Figure 4.13(a). Sad is significantly different from the other emotions. Angry is also significantly different from the other emotions, but with greater variation in



Sri Lankan observers

(a) Angry

(b) Fear

(c) Happy

(d) Sad

Figure 4.12: Avatar examples representing typical postures for the four emotion categories according to the Sri Lankan observers. (a) Angry; (b) Fear; (c) Happy; (d) Sad

the amount of head bend, ranging from a slight bend backward to very bent forward. Shown in Figure 4.13(b), the lateral distance of the hand from the elbow (V12), the hand is more laterally open for happy than for the other emotions.

Examples of the typical postures that the Americans assigned to each emotion category are shown in Figure 4.14. Completely opposite to angry postures found in Coulson's [Cou04] study, angry for the Americans is shown with a head bent forward somewhat and some lateral opening of the arms (similar to Wallbott's [Wal98] findings), typically shown with the elbows bent.

While fear is indicated by a wide variety of postures, some postures are shown with a slightly backward bending head, elbows bent with hands raised to around shoulder height, similar to Coulson's findings. For other fear postures, when there is lateral extension of the arms, the hands typically are also raised over the head.

Happy postures are also diverse for the Americans. The majority of the happy postures show a backward bending head (similar to Wallbott's and Coulson's studies), some lateral opening of the arms, the elbows often bent and hands raised to either shoulder height or

Table 4.7: The low-level posture description features which reached significance between the emotions for the American observers (one-way ANOVAs with df = 3 (emotions)). For each feature f_{iw}, *a-e* pairs indicate the significant differences between means according to Tamhane's T2 *post-hoc* comparisons.

American observers					
	Means for affective states				
Low-level feature	**Angry**	**Fear**	**Happy**	**Sad**	p
V5 - $Orientation_{YZ}$: B.Head - F.Head axis	$.42^{abc}$	$.63^{ad}$	$.71^{be}$	$.06^{cde}$	.000
V6 - $Distance_z$: R.Hand - R.Shoulder	$.65^{a}$	$.60^{b}$	$.53^{c}$	$.91^{abc}$	.000
V7 - $Distance_z$: L.Hand - L.Shoulder	$.74^{a}$	$.64^{b}$	$.57^{c}$	$.94^{abc}$	.000
V8 - $Distance_y$: R.Hand - R.Shoulder	$.58^{a}$	$.61^{b}$	$.66^{c}$	$.83^{abc}$	.000
V9 - $Distance_y$: L.Hand - L.Shoulder	$.65^{a}$	$.62^{b}$	$.67^{c}$	$.86^{abc}$	.000
V10 - $Distance_x$: R.Hand - L.Shoulder	.54	.50	$.42^{a}$	$.60^{a}$	.001
V11 - $Distance_x$: L.Hand - R.Shoulder	$.62^{a}$	.57	$.43^{ab}$	$.58^{b}$	.001
V12 - $Distance_x$: R.Hand - R.Elbow	$.68^{a}$	$.60^{b}$	$.42^{abc}$	$.62^{c}$	.000
V13 - $Distance_x$: L.Hand - L.Elbow	$.77^{ab}$	.63	$.48^{a}$	$.60^{b}$	.000
V14 - $Distance_x$: R.Elbow - L.Shoulder	$.29^{a}$	$.30^{b}$	$.32^{c}$	$.49^{abc}$	.001
V15 - $Distance_x$: L.Elbow - R.Shoulder	$.93^{a}$	.85	$.94^{b}$	$.85^{ab}$	.004
V16 - $Distance_z$: R.Hand - R.Elbow	$.53^{a}$	$.39^{b}$	$.33^{c}$	$.85^{abc}$	.000
V17 - $Distance_z$: L.Hand - L.Elbow	$.45^{a}$	$.33^{b}$	$.29^{c}$	$.66^{abc}$	.000
V18 - $Distance_y$: R.Hand - R.Elbow	$.66^{a}$	$.64^{b}$	$.70^{c}$	$.87^{abc}$	.004
V19 - $Distance_y$: L.Hand - L.Elbow	$.70^{a}$	$.70^{b}$	$.70^{c}$	$.88^{abc}$	.007
V20 - $Distance_y$: R.Elbow - R.Shoulder	$.40^{a}$	.46	.50	$.56^{a}$	.016
V21 - $Distance_y$: L.Elbow - L.Shoulder	$.45^{a}$	$.44^{b}$	.50	$.58^{ab}$	.019
V22 - $Distance_z$: R.Elbow - R.Shoulder	$.74^{a}$	$.76^{b}$	$.68^{c}$	$.95^{abc}$	.000
V23 - $Distance_z$: L.Elbow - L.Shoulder	$.85^{a}$	$.80^{b}$	$.71^{c}$	$.95^{abc}$	.000
V27 - $3D - Distance$: R.Heel - L.Heel	.40	.38	.55	.52	.031

over head. The verticality of the arms is similar to Coulson's study, however, as stated in Section 4.4.2, Coulson specifies straight arms, i.e., no bend at the elbows.

Sad is demonstrated with a slight frontal positioning of the arms. Other than that, sad postures appear closed vertically and laterally, which is the same as the postures associated with the results for the Japanese and the Sri Lankans presented here, as well as for both Coulson's and Wallbott's studies.

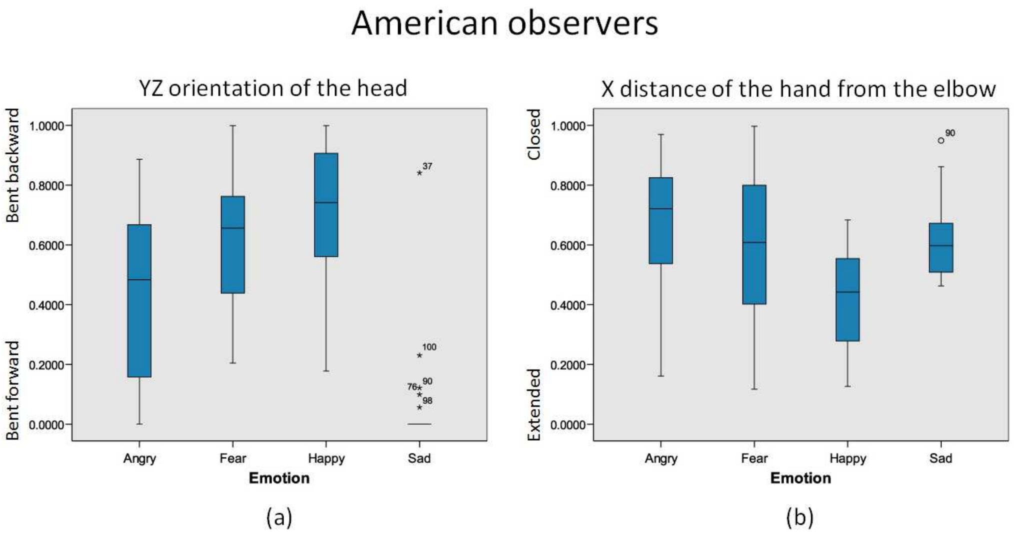


Figure 4.13: Examples of low-level posture description features with significant differences between emotions for the American observers. (a) The forward/backward bending of the head (V5); (b) The lateral distance of the hand from the elbow (V12)

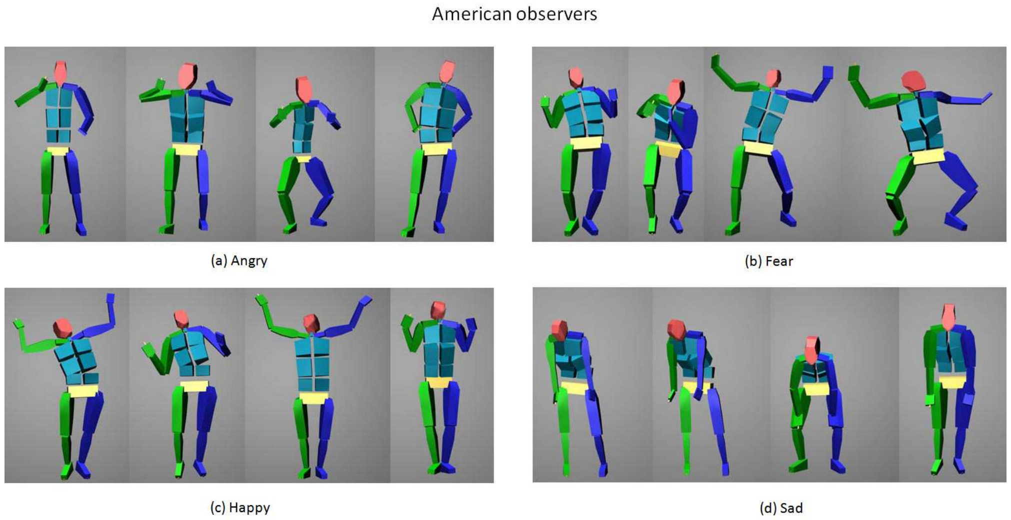


Figure 4.14: Avatar examples representing typical postures for the four emotion categories according to the American observers. (a) Angry; (b) Fear; (c) Happy; (d) Sad

4.5 Automatic Recognition of Basic Emotions from Posture

Armed now with a better understanding of how human observers attribute emotion to acted postures, automatic recognition models can be built, tested and evaluated against the human recognition benchmarks that were defined in Section 4.2.3. The hypothesis is that the automatic recognition models can achieve accuracy rates similar to the benchmarks set by the human observers. The models were tested for their ability to generalise to new observers (Figure 3.7) and to new postures (Figure 3.8) as explained in Chapter 3. To build the automatic recognition models, each static posture p_i was associated with a vector of low-level posture features $F_i = \{f_{i1}, ..., f_{i24}\}$ listed in Table 4.4 and a ground truth label $gtl(p_i, O)$.

The topology of the MLP used in this study consists of three layers: one input, one hidden and one output. The number of nodes in the input layer corresponds to the number of features used. The number of nodes in the hidden layer corresponds to the number of input nodes divided by two. The output layer contains four nodes, one node corresponding to each emotion label. 10,000 epochs were used to train the recognition models.

4.5.1 Generalising to New Observers

As with the human observer agreement examination, automatic recognition models were built i) between the three groups of observers O_{JA}, O_{SL}, O_{US}, and ii) with the disjoint subsets created from all the observers combined O. For the recognition models built on the three separate groups of observers, the models were trained on one observer group (e.g., O_{JA}) and tested on a second observer group (e.g., O_{SL}), leaving the third group (e.g., O_{US}) out. This process continued until models were built and tested for all combinations of the three groups. The results are listed in the 'automatic recognition' total (fourth) column of Table 4.8 and shown in Figure 4.15. The average recognition across all six trials was 77.16% ($SD = 3.5\%$).

Table 4.8: The testing and evaluation results between observer groups O_{JA}, O_{SL}, O_{US} for generalising to new observers. Columns 2 and 3 list the observer groups used to train and test the models. Column 4 lists the automatic recognition model performances for each trial. The remaining four columns list the recognition rates for each emotion label l_j for each trial.

Trial	**Train set**	**Test set**	**Auto. rec. total**	**Angry**	**Fear**	**Happy**	**Sad**
1	JA	SL	74.07%	67.65%	60.61%	78.26%	100%
2	JA	US	82.41%	76.47%	75.76%	91.3%	94.44%
3	SL	JA	73.15%	65.71%	86.96%	69.23%	75%
4	SL	US	77.78%	68.57%	82.61%	76.92%	87.5%
5	US	JA	79.63%	83.87%	89.29%	76.92%	65.22%
6	US	SL	75.93%	77.42%	67.86%	73.08%	86.96%
Average			77.16%	73.28%	77.18%	77.62%	84.85%
SD			3.5%	7.08%	11.27%	7.48%	12.77%

Table 4.9: The testing and evaluation results for the 10 trials created from all observers combined O for generalising to new observers. Column 2 lists the automatic recognition model performances for each trial. The remaining four columns list the recognition rates for each emotion label l_j for each trial.

Trial	**Auto rec. total**	**Angry**	**Fear**	**Happy**	**Sad**
1	86.27%	82.8%	74.1%	95.7%	95.7%
2	87.25%	76.7%	88 %	88.5%	100 %
3	86.27%	83.9%	74.1%	95.7%	95.2%
4	88.24%	78.1%	100 %	96.2%	81.8%
5	81.37%	73.3%	80.8%	82.6%	91.3%
6	83.33%	73.3%	77.8%	91.7%	95.2%
7	84.31%	72.4%	90.9%	80.6%	100 %
8	85.29%	81.5%	81.5%	88.9%	90.5%
9	91.18%	86.2%	87.5%	92.6%	100 %
10	87.25%	81.5%	91.3%	86.2%	91.3%
Average	86.08%	78.97%	84.6%	89.87%	94.1%
SD	2.73%	4.92%	8.38%	5.52%	5.67%

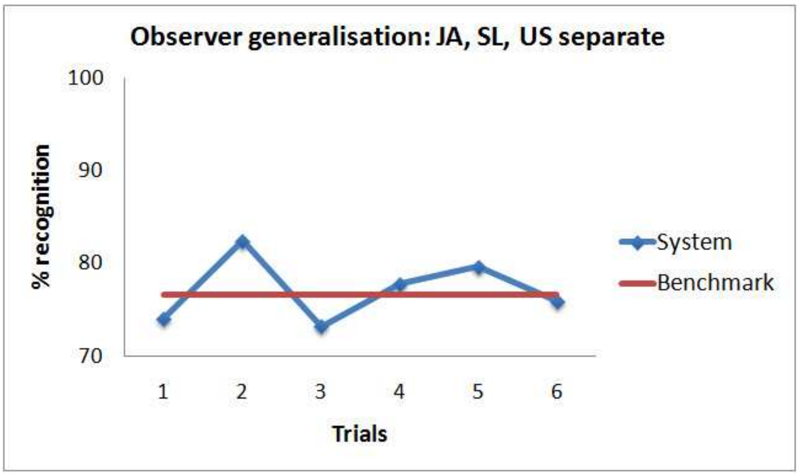


Figure 4.15: Observer generalisation results for the three groups separate compared with the benchmark

For the recognition models built on the 10 trials created from all observers combined O, the models were trained with the previously unused subset $O_{s,3}$ data and tested with subset $O_{s,1}$ with $s = 1, ..., 10$. The results are listed in the second column of Table 4.9 and shown in Figure 4.16. The average recognition across all 10 trials was 86.08% ($SD = 2.73\%$). The last four columns of both Tables list the recognition rates for the individual emotion categories.

Evaluation and Discussion

In building automatic models for generalising to new observers, the goal was to achieve recognition rates equivalent to the human observer benchmarks. Recall from Section 4.2.3, the benchmark computed between the observer group pairs (e.g., O_{JA} and O_{SL}) was 76.54%. For the set of 10 trials carried out with the subsets $O_{s,1}$ and $O_{s,3}$ created from all the observers combined O, the benchmark was 84.80%.

For the automatic recognition models built between the observer group pairs separately (Table 4.8 and Figure 4.15), it can be seen that three of the six models achieve recognition rates better than the benchmark. While three recognition models do not outperform the benchmark, they achieve recognition rates just below the benchmark. The two models that

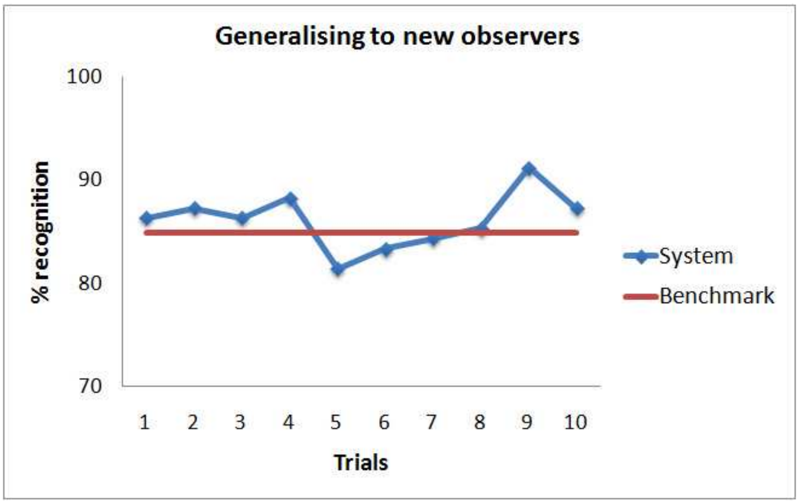
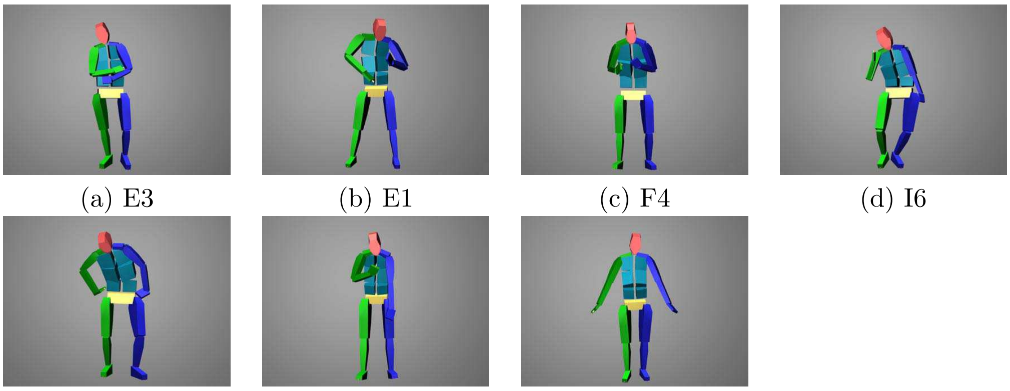


Figure 4.16: Observer generalisation results for the three groups combined across the 10 trials compared with the benchmark

achieve the highest recognition performances are trials 2 (the model trained on the Japanese O_{JA} and tested with the Americans O_{US}) and 5 (the model trained on the Americans O_{US} and tested with the Japanese O_{JA}). These results may indicate that the Japanese and the Americans are more similar in the features they consider important than the Sri lankans are with these cultures. However, this is not the case according to Hofstede's cultural dimensions [Hof06].

According to the individualism dimension, the Sri Lankans and the Japanese have quite similar ratings (37 and 43, respectively). This places them on the collectivistic side of the scale, meaning that typically there is a fairly strong feeling of belonging to a group for these cultures. On the other hand, the Americans rank significantly higher (91) on this dimension, placing them firmly on the individualistic side of the scale. The people in individualistic societies are deemed to have looser ties to others, and pride themselves on being different from one another. According to the uncertainty dimension, it is the Sri Lankans and the Americans who are very similar (42 and 43, respectively). These ratings place the Sri Lankans and the Americans on the uncertainty accepting side of the scale, indicating societies in which it is acceptable to have ideas and beliefs that are different from



(a) E3 (b) E1 (c) F4 (d) I6

(e) J8 (f) K10 (g) K6

Figure 4.17: The postures that were misclassified as sad by the six recognition models built with the observer group pairs when the ground truth label $gtl(p_i, O)$ was angry. The letter and number pair under each posture image refers to the location of that posture in the pie charts in Figures 4.2, 4.4, 4.5 and 4.6

other members in the society. The Japanese are considered drastically different (92), placing them on the uncertainty avoiding side of the scale. People brought up in cultures that rank high in the uncertainty dimension will feel more uncomfortable in uncertain, novel situations and attempt to avoid them as much as possible.

Looking at the second to last row of both Tables which shows the average recognition rate for the recognition models and for the individual emotion categories, it can be seen that the lowest recognition rate occurred for angry and the highest recognition rate occurred for sad. For the automatic recognition models using $O_{s,1}$ and $O_{s,3}$, sad was almost never misclassified.

For the six trials in which the three observer groups were tested against each other, e.g., O_{JA} and O_{SL}, when sad was misclassified, it was almost always misclassified as angry. Examples of the postures that were misclassified as either angry or sad are illustrated in Figure 4.17. In all of these postures, the head is bent forward which was a typical characteristic of sad postures for all of the human observer agreement analyses as well as for sad in the case of the Japanese and the Americans. The postures presented in the figure also display one or both arms bent at the elbow, which was found to be indicative of angry for all three observer groups. For some of the postures, the arm position is slightly frontal which was a common position for angry postures in Coulson's study [Cou04]. The exception to the arm configuration can be seen in Figure 4.17(g) in which the arms are straight and extended slightly laterally, giving the body a somewhat open appearance which is more indicative of happy for the human observers.

4.5.2 Generalising to New Postures

To test the models' ability to generalise to novel postures, an automatic recognition model was built for each observer group O_{JA}, O_{SL}, O_{US} separately, and one model was built for all the observers combined O. The results are summarised in Table 4.10. The second column lists the recognition rate for the four emotion categories combined and the remaining columns list the recognition rates for the individual emotion categories. The results for all observers combined O are also illustrated in the receiver operator characteristic (ROC) curves in Figure 4.18. ROC curves demonstrate the relationship between the percentage of true positives (sensitivity) and the percentage of false positives (1-specificity).

Table 4.10: The testing results for generalising to novel postures for each group of observers O_{JA}, O_{SL}, O_{US} (the first three rows) and all three groups combined O (the last row). Column 2 lists the automatic recognition model performances for each observer group. The remaining 4 columns list the recognition rates for each emotion label l_j for each observer group.

Observers	**Auto Rec. total**	**Angry**	**Fear**	**Happy**	**Sad**
JA	76.85%	76.47%	63.64%	86.96%	88.89%
SL	62.96%	62.86%	39.13%	69.23%	79.17%
US	66.67%	64.52%	50%	69.23%	86.96%
All combined	73.53%	62.07%	57.69%	84.62%	95.24%

Evaluation and Discussion

In evaluating the automatic recognition results presented in Table 4.10, most noticeable is that the lowest recognition rate occurred for fear for all of the models, with fear for the Sri Lankan model lowest at 39.13%. The confusion matrix for the Sri Lankan model

Table 4.11: The confusion matrix for generalising to novel postures for the Sri Lankans

Sri Lankans

Angry	**Fear**	**Happy**	**Sad**	← *classified*
22	5	5	3	**Angry**
11	**9**	2	1	**Fear**
4	3	**18**	1	**Happy**
2	3	0	**19**	**Sad**

Table 4.12: The confusion matrix for generalising to novel postures for the Americans

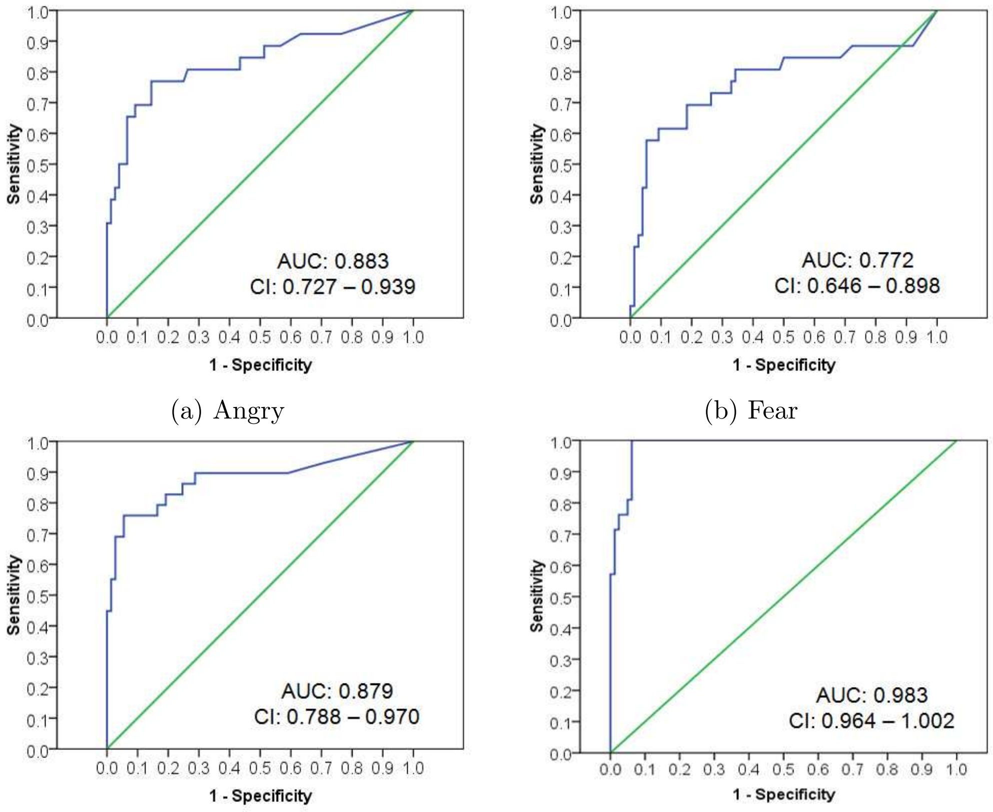
Americans

Angry	**Fear**	**Happy**	**Sad**	← *classified*
20	6	2	3	**Angry**
5	**14**	8	1	**Fear**
1	6	**18**	1	**Happy**
2	1	0	**20**	**Sad**

Table 4.13: The confusion matrix for generalising to novel postures for the Japanese

Japanese

Angry	**Fear**	**Happy**	**Sad**	← *classified*
26	3	1	4	**Angry**
7	**21**	5	0	**Fear**
1	1	**20**	1	**Happy**
0	1	1	**16**	**Sad**



(a) Angry

(b) Fear

(c) Happy

(d) Sad

Figure 4.18: The ROC curves for the four emotion categories for the automatic recognition model built for all observers combined for generalising to novel postures. AUC = area under the curve; CI = 95% confidence interval. (a) Angry; (b) Fear; (c) Happy; (d) Sad

(refer to Table 4.11) reveals that fear postures were most often misclassified as angry. For the American model, only 50% of the fear postures were correctly classified. While some fear postures were misclassified as angry (similar to the Sri Lankan model), according to the confusion matrix (refer to Table 4.12), most of the fear misclassifications occurred for happy. For the Japanese model, a higher percentage of fear postures were correctly classified 63%, however the same type of misclassifications can be seen - fear postures misclassified as angry and happy as revealed by the confusion matrix (refer to Table 4.13). Although high levels of agreement were obtained between the human observers for fear postures, this has not been the case for automatic posture generalisation models.

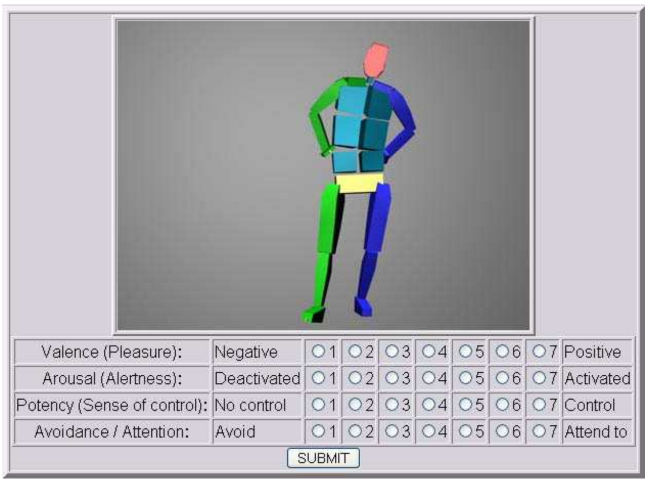
The automatic recognition model results are similar to the findings of Camurri et al [CMR$^+$04] in which fear dance movements were most often misclassified as anger, as opposed to Kapur et al [KKVB$^+$05] with fear dance movements most often misclassified as sad. Also in Camurri et al's study, another common classification error was fear misclassified as happy. As discussed in Section 4.2.2, humans may be able to imply movement when viewing static images of fear postures. However, different from Camurri et al's and Kapur et al's studies, dynamic information is not part of the low-level posture description in this thesis.

4.6 Human Recognition of Affective Dimensions from Posture

In addition to basic emotion category recognition, affect has been shown to be recognised according to levels of affective dimensions, such as valence and arousal. The remainder of this case study focuses on the recognition of affective dimensions from acted posture.

4.6.1 Survey Procedure and Observers

Similar to the emotion category survey, an online survey was conducted to obtain judgments on a set of affective postures. 111 postures were chosen from the 182 postures collected and described in Section 4.1, i.e., $|P| = 111$. Each posture was presented on a separate page in a randomised order. For each page, the observers were asked to rate each posture p_i according to a seven-point Likert scale for a set of four affective dimensions $D = \{valence, arousal, potency, avoidance\}$. An example of the affective dimension posture judgment survey can be seen in Figure 4.19. 10 observers (six females and four males) participated. Given the high agreement results achieved for each cultural group for the emotion categories and that the same linguistic issues do not exist in the use of affective dimensions, similar positive results in the affective dimension analysis may be expected. Therefore, the culture of the observer was not addressed. However, this remains an interesting question that could be addressed in future work.



Valence (Pleasure):	Negative	○1	○2	○3	○4	○5	○6	○7	Positive
Arousal (Alertness):	Deactivated	○1	○2	○3	○4	○5	○6	○7	Activated
Potency (Sense of control):	No control	○1	○2	○3	○4	○5	○6	○7	Control
Avoidance / Attention:	Avoid	○1	○2	○3	○4	○5	○6	○7	Attend to

Figure 4.19: An example of the affective dimensions posture judgment survey

4.6.2 Overview of the Survey Data

A general overview of how the set of observers O judged the affective postures was obtained as the first step in the analysis. To this aim, each affective dimension d_g was examined separately and is represented as a series of bar charts. Due to space constraints, the complete set of bar charts can be found in Appendix C. Each row is a posture p_i and each column is an affective dimension $d_g : g = 1, ..., 4$. The x-axis shows the rating scale (i.e., 1 to 7) and the y-axis shows the number of evaluations obtained for each rating in the scale. The number to the right of the posture image is used as a posture identifier to allow for easy identification for the discussions that take place in the remainder of the chapter. Presented in this section are some of the most interesting results.

An initial examination of the postures according to the arousal dimension reveals that many postures obtained high arousal ratings, i.e., 5 to 7, while very few obtained low arousal ratings. These results may be due to the use of an acted scenario, from which the resulting expressions tend to be more exaggerated than if a natural, non-acted scenario were used. Looking at the postures to which high arousal ratings were attributed, two main types of posture configurations can be seen and are represented in Figures 4.20 and 4.21. In Figure

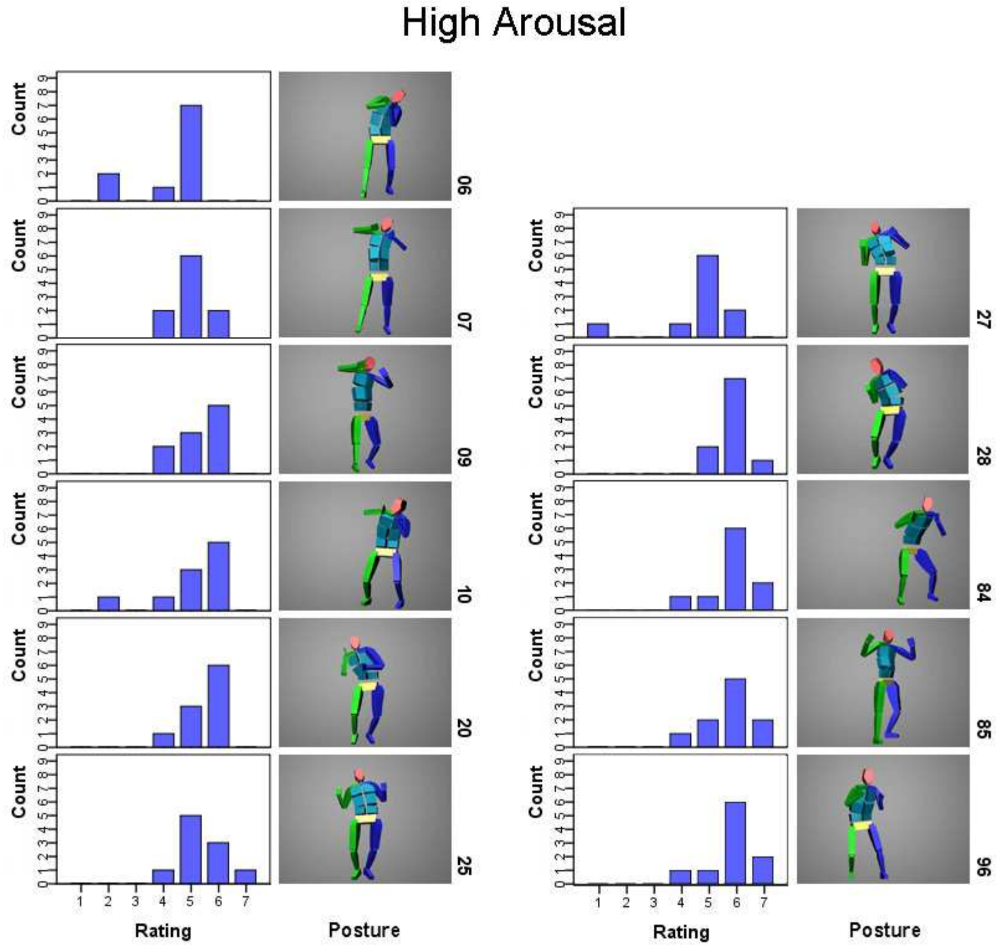


Figure 4.20: Postures rated as high arousal. For each bar chart, the x-axis shows the rating scale (i.e., 1 to 7) and the y-axis shows the number of evaluations obtained for each rating in the scale. The number to the right of the posture image is used as a posture identifier to allow for easy location of the posture in Appendix C

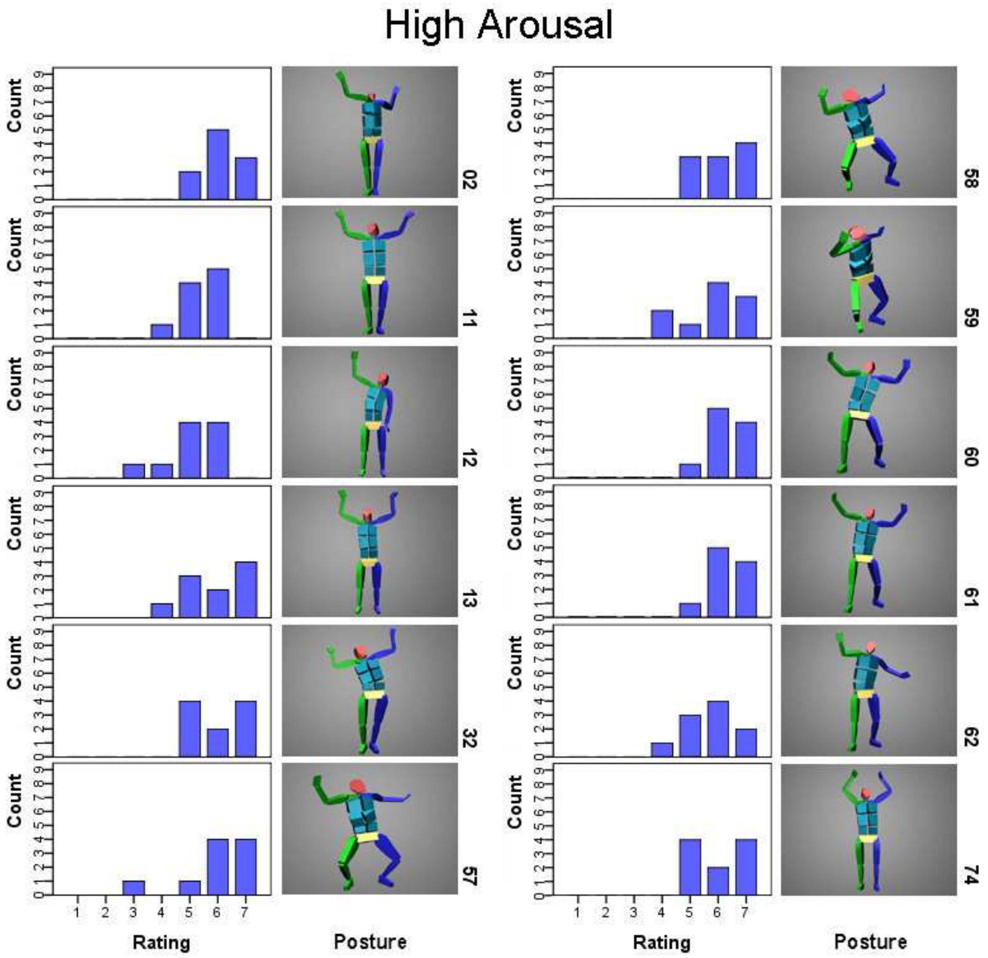


Figure 4.21: Postures rated as high arousal. For each bar chart, the x-axis shows the rating scale (i.e., 1 to 7) and the y-axis shows the number of evaluations obtained for each rating in the scale. The number to the right of the posture image is used as a posture identifier to allow for easy location of the posture in Appendix C

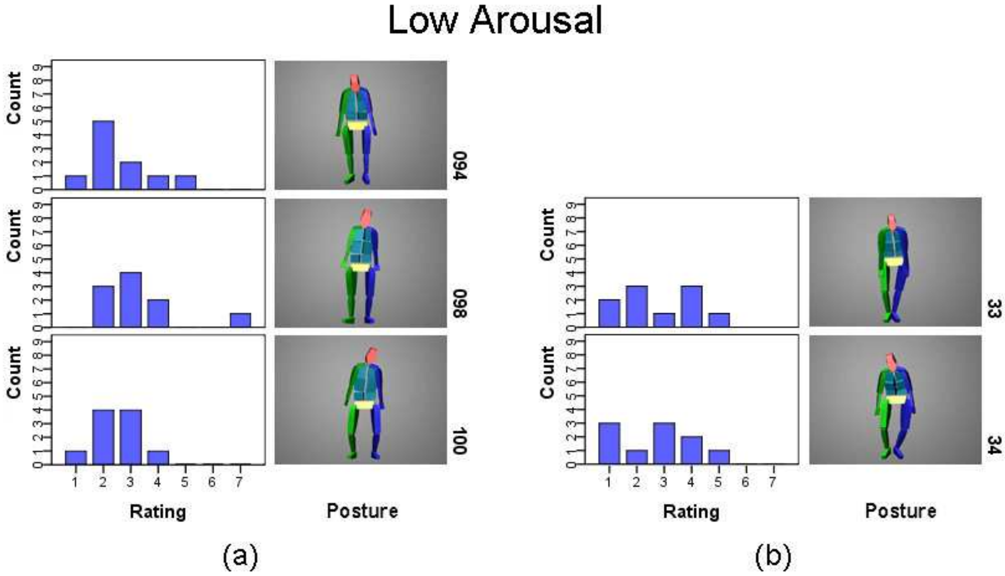


Figure 4.22: Postures rated as low arousal. For each bar chart, the x-axis shows the rating scale (i.e., 1 to 7) and the y-axis shows the number of evaluations obtained for each rating in the scale. The number to the right of the posture image is used as a posture identifier to allow for easy location of the posture in Appendix C

4.20, at least one of the arms is bent at the elbow with the hand raised and held near the face and the head is either straight or slightly bent backward. The other type of high arousal postures, in Figure 4.21, typically have arms bent at the elbow and the hands raised above head height, and the head is bent backward.

There were far fewer postures to which low arousal ratings were associated, however the configurations of those postures are very similar as illustrated in Figure 4.22. The arms are straight, extended down along the side of the body and the head is bent forward, shown in Figure 4.22(a). Two other postures, shown in Figure 4.22(b), achieved less consistent, yet overall low ratings as well. The postural configuration is similar to the postures in Figure 4.22(a), with the addition of the upper body bent forward.

An examination of the valence dimension reveals results similar to the arousal dimension. The postures to which high valence ratings were achieved were a mixture of the two types of

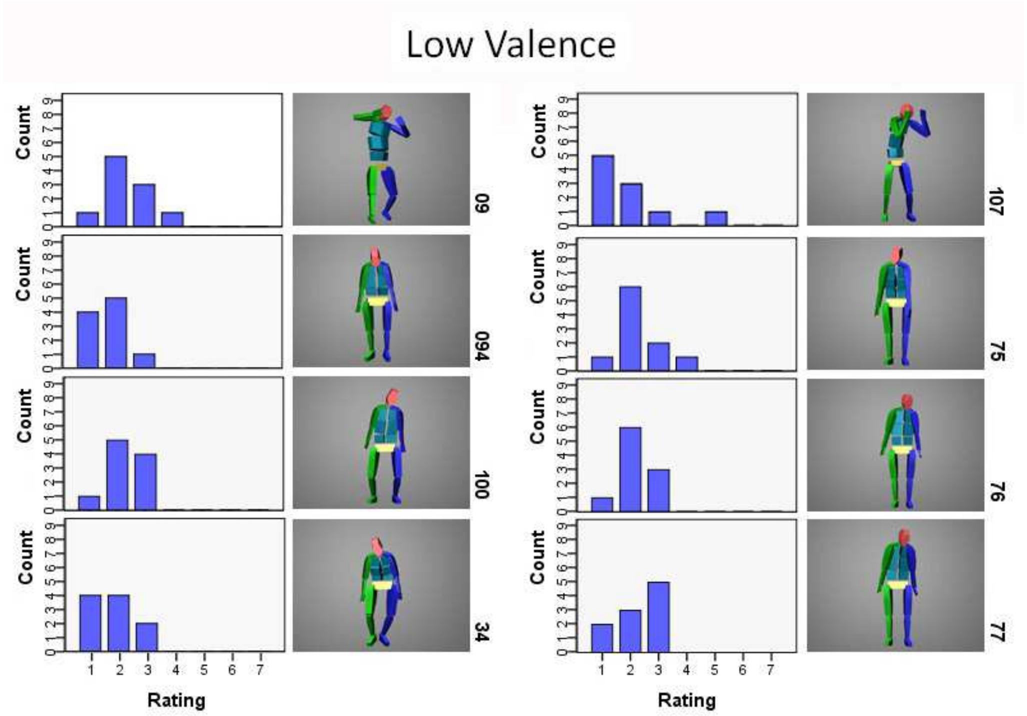


Figure 4.23: Postures rated as low valence. For each bar chart, the x-axis shows the rating scale (i.e., 1 to 7) and the y-axis shows the number of evaluations obtained for each rating in the scale. The number to the right of the posture image is used as a posture identifier to allow for easy location of the posture in Appendix C

high arousal postures. In the case of low valence, several postures were assigned low ratings. These postures are depicted in Figure 4.23 and fall into two main configuration types. The postures in the last three rows of the Figure are all similar with the head bent forward and the arms extended straight down along the body. These postures also obtained mainly low ratings on the arousal dimension. Illustrated in the first row of the Figure is the second type of low valence postures. These postures obtained high ratings on the arousal dimension and the postural configuration represents this difference. These postures are distinctly different from the other low valence postures with arms bent at the elbows and the hands brought up to face level in a somewhat protective pose.

An investigation of the avoidance and potency dimensions reveals few postures that were

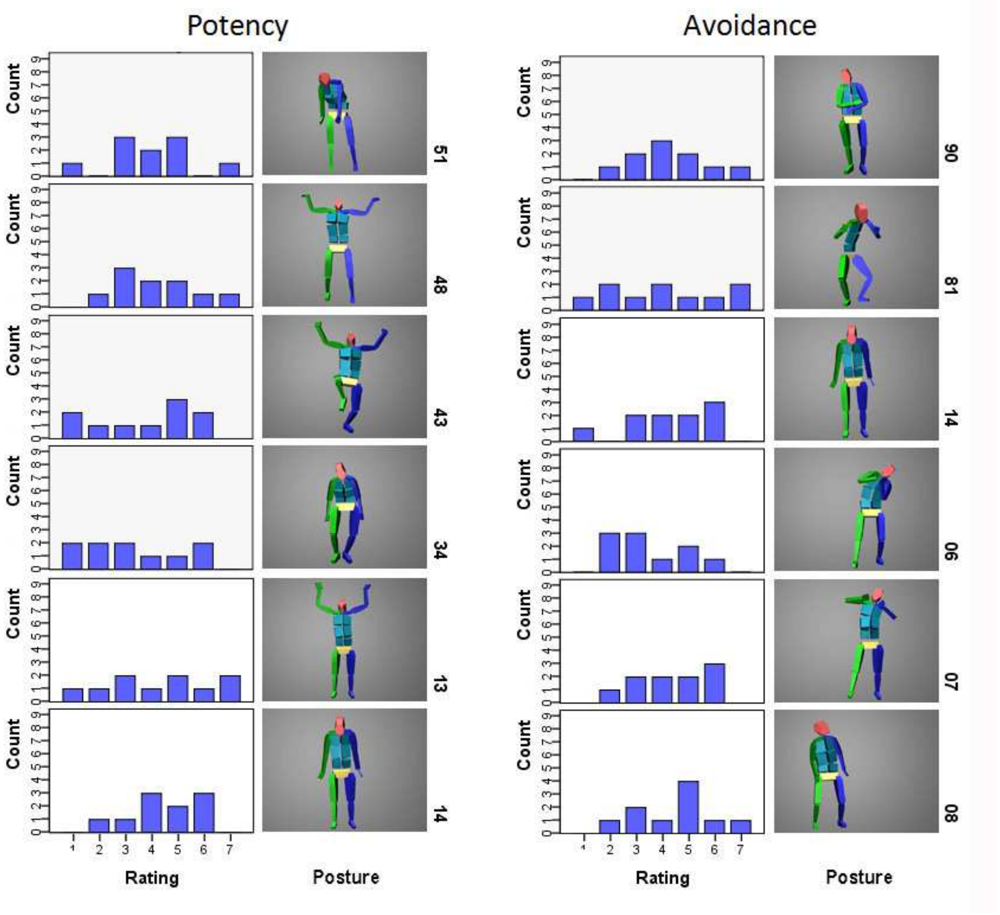


Figure 4.24: Potency and avoidance postures with ratings spread across the rating scale. For each bar chart, the x-axis shows the rating scale (i.e., 1 to 7) and the y-axis shows the number of evaluations obtained for each rating in the scale. The number to the right of the posture image is used as a posture identifier to allow for easy location of the posture in Appendix C

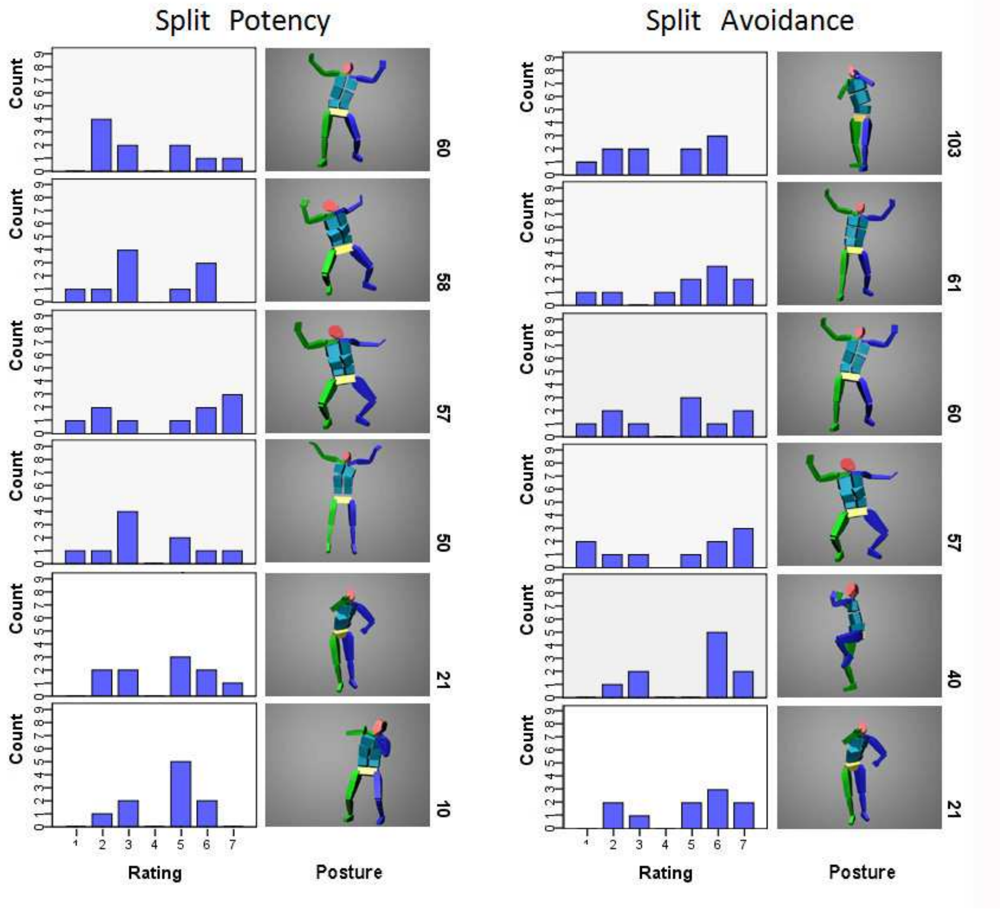


Figure 4.25: Potency and avoidance postures with ratings at both ends of the rating scale causing a bimodal distribution. For each bar chart, the x-axis shows the rating scale (i.e., 1 to 7) and the y-axis shows the number of evaluations obtained for each rating in the scale. The number to the right of the posture image is used as a posture identifier to allow for easy location of the posture in Appendix C

rated consistently at either the low or high ends of the scale. Instead, there appears to be a lot of ambiguity for many postures, with ratings spread across almost the entire range of the scale (refer to Figure 4.24) or with ratings split between low and high, causing a bimodal distribution (refer to Figure 4.25).

4.6.3 Creating Benchmarks

The purpose of this section is to define the benchmarks for human recognition of dimensions of affect from whole body postures. These benchmarks define the target recognition rates for the automatic recognition models discussed later in the chapter.

Observer agreement reliability results

The first task is to examine the level of consistency in how the observers rated the postures. Cronbach's α was computed for the entire set of observers together O for each affective dimension $d_g : g = 1, ..., 4$ separately. The affective dimension d_g with the highest strength of agreement was *arousal* with an α of 0.885. The *valence* dimension was second with an α of 0.807. Observer agreement reliability was lowest for *potency* ($\alpha = .531$) and *avoidance* ($\alpha = .522$). Recall from Chapter 3 that an α of 0.7 and above generally indicates high reliability [Fie05].

Given the low Cronbach's α levels of the potency and avoidance dimensions, a principal components analysis (PCA) was subsequently carried out to assess the results more concretely and determine whether or not to continue with the analysis of these two dimensions. The purpose was to investigate whether the observers had used different constructs when evaluating the postures according to these two dimensions. For each dimension d_g, PCA was applied to a 111 x 7 x 10 matrix which has 111 cases (the postures) and 10 variables (the observers). The values in the matrix correspond to the ratings (on the 7 point scale) given by each of the 10 observers for the 111 postures. It is expected that if the dimension is defined by only one construct (i.e., the meaning assigned to the evaluated dimension) and all the observers considered that dimension to have that meaning, the PCA for that dimension would return only one component with an eigenvalue > 1 (Kaiser criterion [Kai60], i.e., a component with an eigenvalue > 1 represents at least one variable). If this is not the case, the ratings from the observers may relate to different non-comparable constructs. Furthermore, an observer o_k was considered highly loaded if the absolute component loading value was ≥ 0.50. The results for the two dimensions are discussed separately.

• **Potency:** The PCA analysis of the potency ratings revealed that 4 components had an eigenvalue > 1 (i.e., each of these 4 components represents at least 1 observer's ratings) and different observers load onto different components. Furthermore, the 4 components together accounted for only 58.85% of the variance, indicating that the term *potency* was used with several different meanings. Five observers loaded onto the first component which covered 22.02% of the variance. Two observers loaded onto the second component which covered 14.71% of the variance. Two observers loaded onto the third component which covered 11.97% of the variance. One observer loaded onto the fourth component which covered 10.15% of the variance. To conclude, given that not all of the observers loaded onto a single component, this signifies that the observers seemed to be using different interpretations of the term potency.

• **Avoidance:** The PCA analysis of the avoidance ratings revealed that 3 had an eigenvalue > 1 (i.e., each of these 3 components represents at least 1 observer's ratings) and different observers load onto different components. Furthermore, the 3 components together accounted for only 51.47% of the variance, indicating that the term *avoidance* was used with several different meanings. Three observers loaded onto the first component which covered 22.54% of the variance. Two observers loaded onto the second component which covered 17.52% of the variance. One observer loaded onto the third component which covered 11.42% of the variance. This result shows that more than half of the observers have not been used, further indicating that the observers may have been using different interpretations of the term avoidance.

Given the lack of agreement amongst O for $d_{potency}$ and $d_{avoidance}$, using a unique value c for each posture p_i to train and evaluate automatic recognition models is not viable. One possible interpretation is that these two dimensions are more grounded on the dynamic characteristics of the gesture, rather than the configurational characteristics. Based on these results, it was decided to eliminate further examination of the potency and avoidance dimensions.

Benchmarks

To set the benchmarks of human recognition of affective dimensions from body posture, random repeated sub-sampling was used to create a set of trials, i.e., $s = 1, ..., 10$. For each trial, the set of observers O was split into three disjoint subsets $O_{s,1}$, $O_{s,2}$ and $O_{s,3}$. $O_{s,1}$ and $O_{s,2}$ each contained three observers and $O_{s,3}$ contained two observers. For each subset, a ground truth rating $gtm(p_i, O_{sk})$ was assigned to each posture p_i for each dimension d_g. The between observers agreement $B.Agr(d_g, O_{sk})$, defined in Equation (3.4), and Cronbach's α were computed between $O_{s,1}$ and $O_{s,2}$ for each of the 10 trials, for the valence and arousal dimensions separately. For each dimension $d_g : g = valence, arousal$, the benchmark is computed as the average between observers agreement $B.Agr(d_g, O_{sk})$ across the 10 trials. For valence the benchmark is 83.11%, $SD = 1.66\%$ and for arousal the benchmark is 86.80%, $SD = 1.3\%$. The results are listed in Table 4.14.

4.6.4 Discussion

The overview of the data revealed that there was little consistency in how the set of observers O seemed to rate the set of postures P for the potency and avoidance dimensions. The survey data overview was better for the valence and arousal dimensions. Why the poorer results for the potency and avoidance dimensions? It is possible that observers have more difficulty understanding what is meant by these terms. Another possibility is that the addition of other information, such as context or additional modalities could help increase between observers agreement $B.Agr(d_g, O_{sk})$ and reliability levels. These possibilities may be investigated in future work. Based on the PCA results, it was decided to discontinue further analysis on the potency and avoidance dimensions at this time.

Table 4.14: The benchmarks computed for the affective dimensions. The $B.Agr(d_g, O_{sk})$ are listed in column 3 and Cronbach's α levels are listed in column 4. The benchmarks, listed in the last column, is the average across all 10 $B.Agr(d_g, O_{sk})$ levels obtained for that affective dimension.

Human recognition benchmarks

Affective dim	**Trial**	$B.Agr(d_g, O_{sk}) * 100$	**Cronbach's** α	**Benchmark**
Valence	1	81.83%	0.601	**83.11%**
	2	85.89%	0.668	
	3	83.48%	0.691	
	4	82.88%	0.539	
	5	83.63%	0.758	
	6	80.38%	0.592	
	7	80.93%	0.641	
	8	83.78%	0.686	
	9	84.53%	0.654	
	10	83.78%	0.614	
Arousal	1	87.54%	0.781	**86.80%**
	2	86.34%	0.834	
	3	86.94%	0.818	
	4	85.43%	0.799	
	5	88.14%	0.814	
	6	87.84%	0.788	
	7	88.89%	0.86	
	8	84.99%	0.716	
	9	85.43%	0.684	
	10	86.49%	0.783	

For the arousal dimension, high levels of between observers agreement $B.Agr(d_g, O_{sk})$ and reliability were found. The benchmark for arousal set at 86.8%; higher than the benchmark for valence (83.11%). According to the INDSCAL findings of the study by Paterson et al [PPS01] in which head and arm movements for affective drinking and knocking motions were mapped to an affective space, arousal was identified as the first dimension of a 2D affective space, accounting for 70% of the variance. Valence was identified as the second dimension, accounting for 17% of the variance. These results may indicate that arousal is more easily identified from bodily expressions than valence. Indeed, findings by Clavel et al [CPM$^+$09] appear to validate that assumption. In their study, face only and posture only levels of arousal and valence of an affective virtual agent were judged by observers. The results showed that arousal was correctly perceived from posture more than valence was.

4.7 Low-Level Posture Description Analysis

To assess the discriminative power of the low-level posture description for distinguishing between ratings of valence and arousal, the same set of features $F = \{f_{i1}, ..., f_{i24}\}$ described in Section 4.3 was used. Each low-level posture feature f_{iw} was subjected to one-way ANOVAs for each affective dimension d_g separately. The results are presented in Tables 4.15 and 4.16. As with the previous ANOVA results discussed for the set of emotion labels L, listed in the first column of the following Tables are the low-level features shown to be important for discriminating between levels of the affective dimensions with the significance level shown in the last column. The means for each affective dimension d_g rating c are shown in the middle seven columns. The superscript letter pairs listed denote significant differences between those rating pairs according to Tamhane's T2 *post-hoc* comparisons implemented for unequal variances (verified using Levene's test of homogeneity of variance). Reported in this section is a discussion on the most interesting results of each affective dimension, illustrated with a boxplot and corresponding avatar examples.

4.7.1 Valence

The results for the valence dimension are listed in Table 4.15. Significant differences were obtained for 14 of the 24 low-level posture features. For the majority of the posture features, the main differences occur between scale rating 7 and ratings 2 to 4 and sometimes between ratings 7 and 1 or 5. The posture features that achieved the most interesting differences are depicted in the boxplots in Figure 4.26. For V10, the lateral extension of the hand from the opposite shoulder (Figure 4.26(a)), the significant differences occur between rating 7 and ratings 1 to 4. The rating 7 postures show hands that are laterally extended far away from the opposite shoulder, whereas the postures that achieved ratings of 1 to 4 are less laterally distant. Examples of the typical postures for these ratings are represented in Figure 4.27.

For V15, the lateral extension of the elbow from the opposite shoulder (Figure 4.26(b)), the significant differences occur between ratings 1 to 3 and ratings 6 and 7. The postures

Table 4.15: The low-level posture description features which reached significance between the ratings for valence (one-way ANOVAs with df = 6 (ratings)). For each feature, *a-e* pairs demonstrate the significant differences between means according to Tamhane's T2 *post-hoc* comparisons.

	Means for Valence dimension ratings							
Low-level feature	**1**	**2**	**3**	**4**	**5**	**6**	**7**	p
V5 - $Orient_{YZ}$: B.Head - F.Head axis	.24	.33^{a}	.29^{b}	.49	.61	.63	.76^{ab}	.002
V6 - $Dist_z$: R.Hand - R.Shoulder	.72	.73^{a}	.73^{b}	.75^{c}	.52^{d}	.50	.23^{abcd}	.000
V7 - $Dist_z$: L.Hand - L.Shoulder	.75^{a}	.75^{b}	.79^{c}	.80^{d}	.63^{e}	.50	.26^{abcde}	.000
V10 - $Dist_x$: R.Hand - L.Shoulder	.59^{a}	.51^{b}	.61^{c}	.56^{d}	.44	.43	.23^{abcd}	.000
V11 - $Dist_x$: L.Hand - R.Shoulder	.48	.58	.62^{a}	.61^{b}	.50	.42	.28^{ab}	.000
V12 - $Dist_x$: R.Hand - R.Elbow	.66	.60^{a}	.72^{bc}	.60^{d}	.57	.39^{b}	.27^{acd}	.000
V13 - $Dist_x$: L.Hand - L.Elbow	.57	.60	.71^{a}	.69^{b}	.60	.52	.33^{ab}	.002
V14 - $Dist_x$: R.Elbow - L.Shoulder	.41	.32^{a}	.40^{b}	.41^{cd}	.21^{c}	.37	.10^{abd}	.000
V15 - $Dist_x$: L.Elbow - R.Shoulder	.94	.84^{ab}	.88^{cd}	.87^{ef}	.94	1.0^{ace}	1.0*bdf*	.008
V16 - $Dist_z$: R.Hand - R.Elbow	.63	.65^{ab}	.53^{c}	.59de	.39	.17^{ad}	.10^{bce}	.001
V17 - $Dist_z$: L.Hand - L.Elbow	.56	.46^{a}	.45^{b}	.49^{c}	.38^{d}	.18	.09^{abcd}	.000
V19 - $Dist_y$: L.Hand - L.Elbow	.89^{a}	.86^{b}	.77	.66^{ab}	.72	.61	.79	.020
V22 - $Dist_z$: R.Elbow - R.Shoulder	.80^{a}	.80^{b}	.87^{c}	.86de	.63^{df}	.75	.35^{abcef}	.000
V23 - $Dist_z$: L.Elbow - L.Shoulder	.80	.84^{a}	.91^{b}	.90^{c}	.76^{d}	.69	.41^{abcd}	.000

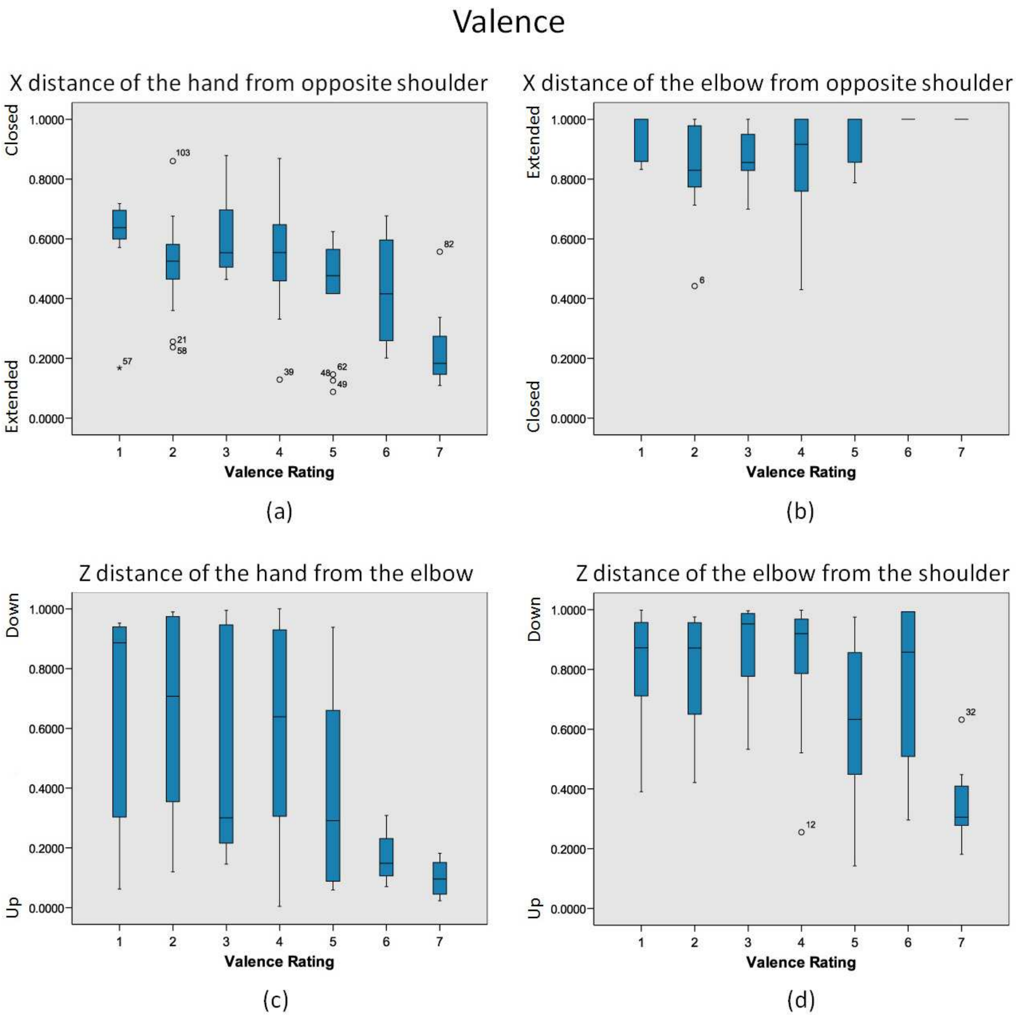


Figure 4.26: Examples of low-level posture description features with significant differences between scale ratings for valence. (a) The lateral extension of the hand from the opposite shoulder (V10); (b) The lateral extension of the elbow from the opposite shoulder (V15); (c) The vertical extension of the hand from the elbow (V16); (d) The vertical extension of the elbow from the shoulder (V22)

rated 6 and 7 have elbows that are very laterally extended away from the opposite shoulder, as illustrated by the examples in Figure 4.28(d)-(f). These postures differ from the low rated, i.e., 1 to 3, postures in that the elbow distance from the opposite shoulder has a wider range. However, the elbow is still fairly extended. Posture examples are shown in Figure 4.28(a)-(c).

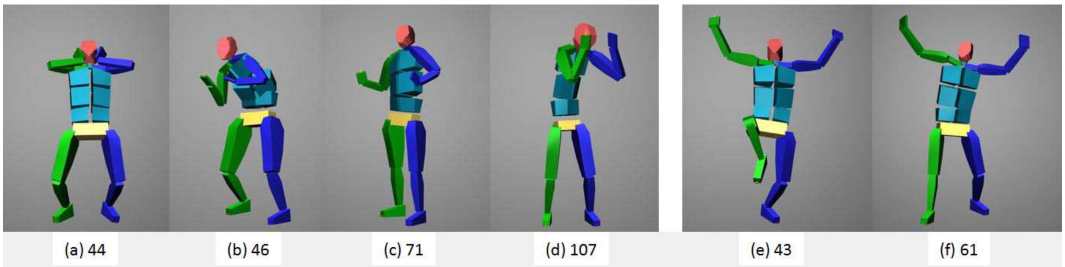


Figure 4.27: Avatar examples representing typical postures for the significant differences of V10, the lateral extension of the hand from the opposite shoulder. (a)-(d) Representing ratings 1 to 4 are postures 44, 46, 71 and 107; (e) and (f) Representing rating 7 are postures 43 and 61. Recall from Section 4.6.2 that the numbers refer to the posture position in the bar charts located in Appendix C for easy cross-referencing

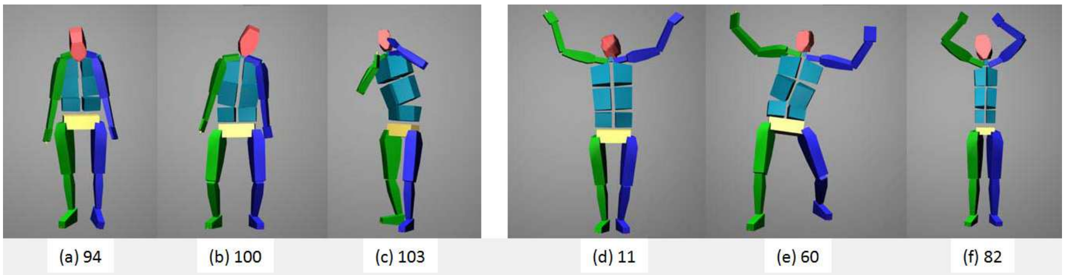


Figure 4.28: Avatar examples representing typical postures for the significant differences of V15, the lateral extension of the elbow from the opposite shoulder. (a)-(c) Representing ratings 1 to 3 are postures 94, 100 and 103; (d)-(f) Representing ratings 6 and 7 are postures 11, 60 and 82. Recall from Section 4.6.2 that the numbers refer to the posture position in the bar charts located in Appendix C for easy cross-referencing

The vertical distance of the hand from the elbow (V16) is depicted in boxplot (c) of Figure 4.26. For this feature the significant differences occur between ratings 2 to 4 and ratings 6 and 7. The range of vertical distance for the postures with ratings 2 to 4 is quite

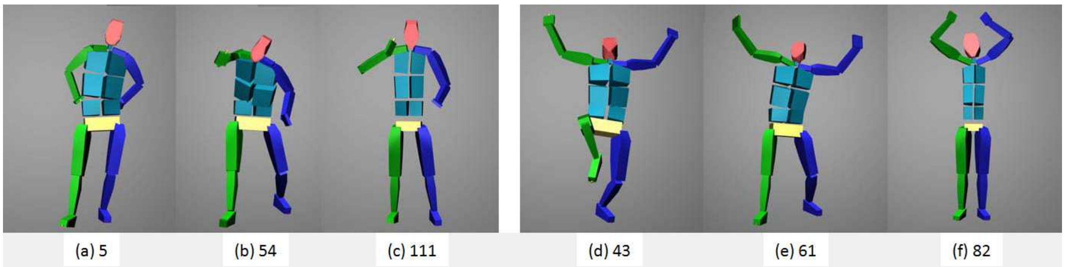


Figure 4.29: Avatar examples representing typical postures for the significant differences of V16, the vertical extension of the hand from the elbow. (a)-(c) Representing ratings 2 to 4 are postures 5, 54 and 111; (d)-(f) Representing ratings 6 and 7 are postures 43, 61 and 82. Recall from Section 4.6.2 that the numbers refer to the posture position in the bar charts located in Appendix C for easy cross-referencing

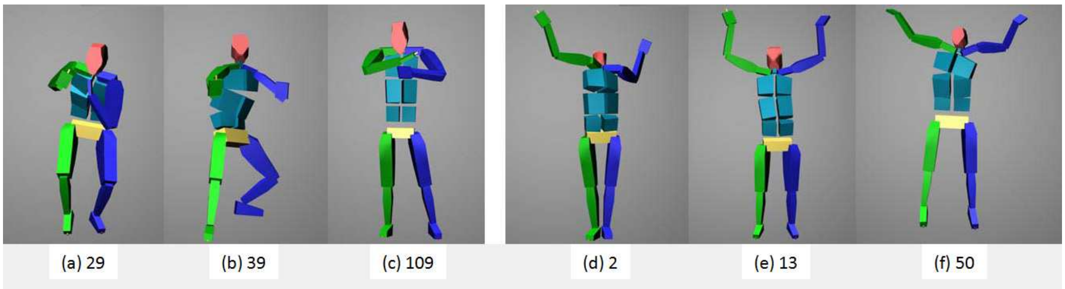


Figure 4.30: Avatar examples representing typical postures for the significant differences of V22, the vertical extension of the elbow from the shoulder. (a)-(c) Representing ratings 1 to 5 are postures 29, 39 and 109; (d)-(f) Representing rating 7 are postures 2, 13 and 50. Recall from Section 4.6.2 that the numbers refer to the posture position in the bar charts located in Appendix C for easy cross-referencing

vast (i.e., the distance ranges from about 0.2 to 0.9). Posture examples are shown in Figure 4.29(a)-(c). It is noticeable that the hand height position in relation to the elbow changes from below the elbow, to the same height and above. Conversely, the range of vertical distance for the postures with ratings of 6 and 7 is very narrow, from about 0.03 to 0.25, and it can be seen in the posture examples (Figure 4.29(d)-(f)) that the hand is consistently raised above the elbow.

The final posture feature to examine for the valence dimension is the vertical distance of the elbow from the shoulder (V22), illustrated in the boxplot in Figure 4.26(d). In this case, the significant differences occur mainly between rating 7 and ratings 1 to 5. The postures with ratings 1 to 5 show the elbow just below shoulder height, typically raised to about chest height, whereas the elbow in the 7 rating postures is vertically level with, or raised above the shoulder. Examples of the 1 to 5 rating postures are depicted in Figure 4.30(a)-(c) and 7 rating postures are shown in (d)-(f).

4.7.2 Arousal

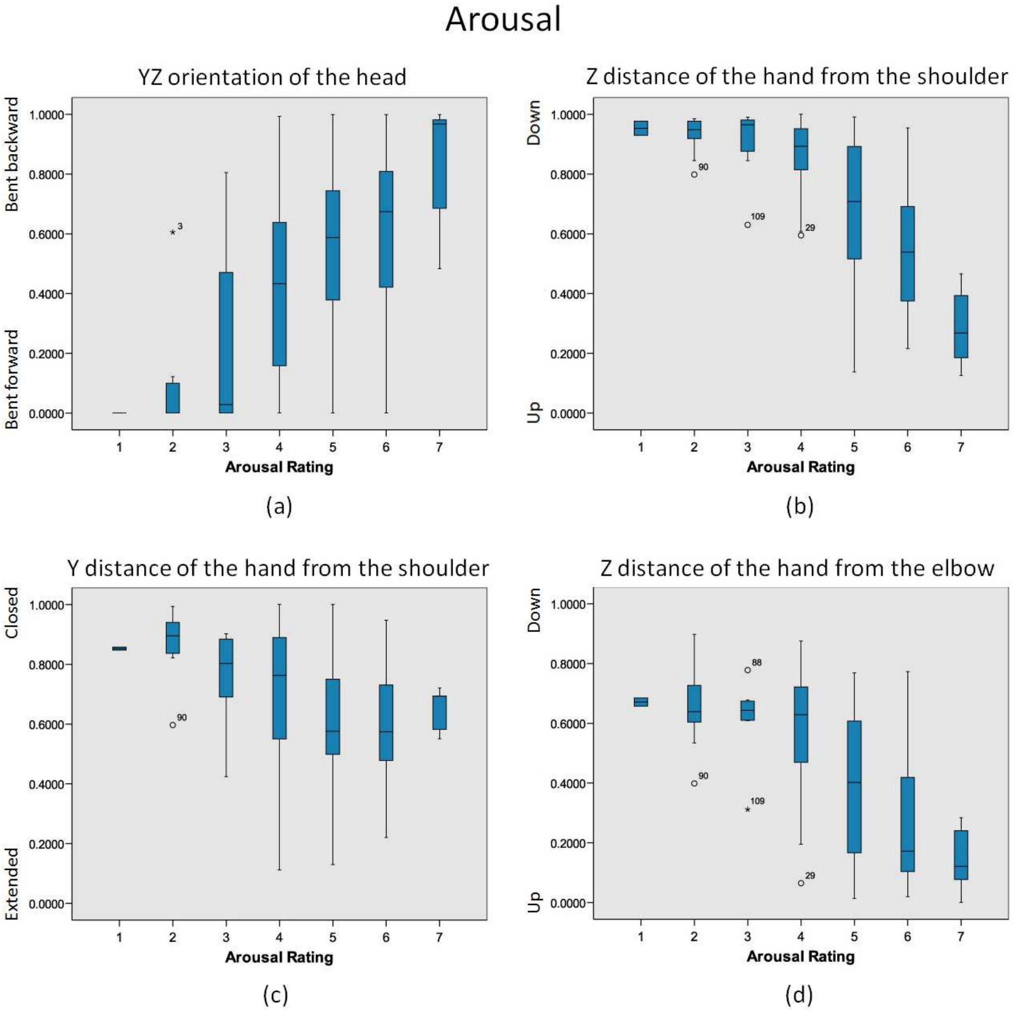
The results for the arousal dimension are listed in Table 4.16. Significant differences were obtained for 14 of the 24 low-level posture features. The posture features that achieved the most interesting differences are depicted in the boxplots in Figure 4.31. For the forward/backward bending of the head, V5 (Figure 4.31(a)), the significant differences occur between ratings 1 and 2 against ratings 4 to 7. The postures with ratings 1 and 2 have a head that is almost completely bent forward, whereas the head in the postures with ratings between 4 and 7 range from somewhat bent forward to completely bent back as the rating increases. Posture examples can be seen in Figure 4.32.

The next posture feature to examine is V7, the vertical distance of the hand from the shoulder (Figure 4.31(b)). Significant differences occur between ratings 5 to 7 against ratings 1 to 4. According to the boxplot, the hands in postures with ratings 1 to 4 are extended almost completely down at the side of the body, with little variation in hand position. Examples can be seen in Figure 4.33(a)-(c). There is much greater variation in vertical hand position for postures with rating 5, ranging from fairly down to waist height. The hand height increases further with ratings 6 and 7, ending with the hand raised up higher than the shoulder as demonstrated by the examples in Figure 4.33(d)-(f).

Another posture feature exhibiting interesting significant differences between scale ratings is the frontal extension of the hand from the shoulder, V8 (Figure 4.31(c)). Differences occur between rating 1 against ratings 5 and 6, and between rating 2 against ratings 5 to 7. For postures with ratings 1 or 2, the body is fairly closed and the hands typically remain close to the body, not extended in front. Examples of these postures can be seen in Figure

Table 4.16: The low-level posture description features which reached significance between the ratings for arousal (one-way ANOVAs with df = 6 (ratings)). For each feature, *a-e* pairs demonstrate the significant differences between means according to Tamhane's T2 *post-hoc* comparisons.

	Means for Arousal dimension ratings							
Low-level feature	**1**	**2**	**3**	**4**	**5**	**6**	**7**	p
V5 - $Orient_{YZ}$: B.Head - F.Head axis	$.00^{abcd}$	$.08^{efgh}$	.23	$.41^{ae}$	$.55^{bf}$	$.57^{cg}$	$.82^{dh}$	.000
V6 - $Dist_z$: R.Hand - R.Shoulder	$.96^{abcd}$	$.92^{efg}$	$.87^{hi}$	$.80^{ajk}$	$.65^{belm}$	$.48^{cfhjl}$	$.26^{dgikm}$	.000
V7 - $Dist_z$: L.Hand - L.Shoulder	$.95^{abc}$	$.93^{def}$	$.91^{ghi}$	$.87^{jkl}$	$.70^{adgjm}$	$.56^{behk}$	$.29^{cfilm}$	.000
V8 - $Dist_y$: R.Hand - R.Shoulder	$.85^{ab}$	$.87^{cde}$	.76	.71	$.60^{ac}$	$.59^{bd}$	$.65^{e}$	.001
V9 - $Dist_y$: L.Hand - L.Shoulder	$.92^{abc}$	$.88^{de}$	.79	$.75^{a}$	$.63^{bd}$	$.62^{ce}$	.75	.000
V10 - $Dist_x$: R.Hand - L.Shoulder	.62	$.60^{a}$	.62	$.56^{b}$	.54	$.41^{ab}$	.33	.000
V14 - $Dist_x$: R.Elbow - L.Shoulder	.55	$.44^{ab}$	$.48^{cd}$	$.42^{ef}$	$.37^{g}$	$.21^{ace}$	$.10^{bdfg}$	.000
V16 - $Dist_z$: R.Hand - R.Elbow	$.95^{abcd}$	$.88^{efg}$	$.79^{hi}$	$.71^{ajk}$	$.47^{bel}$	$.29^{cfhj}$	$.10^{dgikl}$	.000
V17 - $Dist_z$: L.Hand - L.Elbow	$.67^{abc}$	$.66^{def}$	$.62^{ghi}$	$.59^{jkl}$	$.38^{adgj}$	$.27^{behk}$	$.14^{cfil}$	.000
V18 - $Dist_y$: R.Hand - R.Elbow	.86	$.94^{abc}$	$.90^{de}$	$.68^{a}$	$.70^{bd}$	$.61^{ce}$	.69	.004
V19 - $Dist_y$: L.Hand - L.Elbow	.85	$.94^{abc}$	$.90^{de}$	$.72^{af}$	$.66^{bdg}$	$.68^{ceh}$	$.97^{fgh}$	.000
V22 - $Dist_z$: R.Elbow - R.Shoulder	$.97^{a}$	$.94^{bc}$	$.93^{d}$	$.88^{e}$	$.79^{b}$	$.74^{acde}$	.40	.000
V23 - $Dist_z$: L.Elbow - L.Shoulder	$.98^{abc}$	$.93^{de}$	$.94^{fg}$	$.93^{hi}$	$.83^{aj}$	$.73^{bdfh}$	$.40^{cegij}$	.000
V27 - $3D - Dist$: R.Heel - L.Heel	.56	.56	.40	.56	.49	.35	.25	.022



(a)

(b)

(c)

(d)

Figure 4.31: Examples of low-level posture description features with significant differences between scale ratings for arousal. (a) The forward/backward bending of the head (V5); (b) The vertical distance of the hand from the shoulder (V7); (c) The frontal extension of the hand from the shoulder (V8); (d) The vertical height of the hand in relation to the elbow

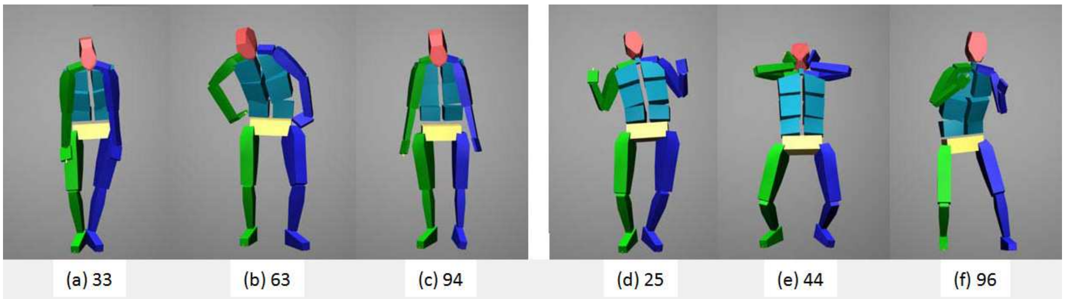


Figure 4.32: Avatar examples representing typical postures for the significant differences of V5, the forward backward bending of the head. (a)-(c) Representing ratings 1 and 2 are postures 33, 63 and 94; (d)-(f) Representing rating 7 are postures 25, 44 and 96. Recall from Section 4.6.2 that the numbers refer to the posture position in the bar charts located in Appendix C for easy cross-referencing

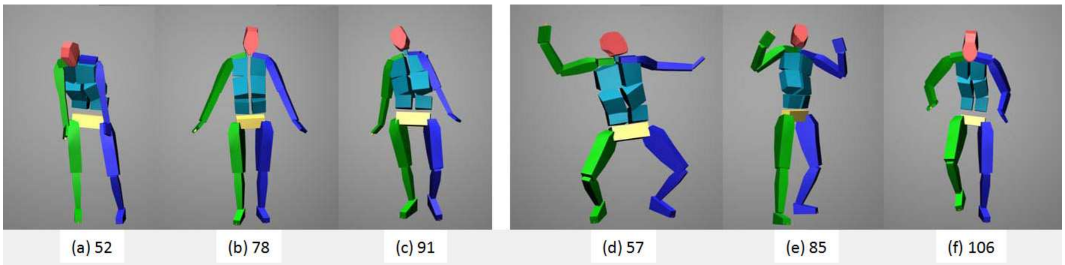


Figure 4.33: Avatar examples representing typical postures for the significant differences of V7, the vertical distance of the hand from the shoulder. (a)-(c) Representing ratings 1 to 4 are postures 52, 78 and 91; (d)-(f) Representing ratings 5 to 7 are postures 57, 85 and 106. Recall from Section 4.6.2 that the numbers refer to the posture position in the bar charts located in Appendix C for easy cross-referencing

4.34(a)-(c). For postures with ratings 5 to 7, the range of frontal extension of the hands is between slightly in front to about halfway in front, e.g., the elbows bent and the hands in front of the body. Examples are presented in (d)-(f) of Figure 4.34.

The last posture feature to examine for arousal is V17, the vertical distance of the hand from the elbow (Figure 4.31(d)). For this feature, like most of the other arousal and valence features examined, the lower ratings are significantly different from the higher ratings. In this case, the significant differences occur between ratings 1 to 4 and ratings 5 to 7. According to the boxplot, the hand height ranges from being lower than the elbow (ratings 1 to 3) to

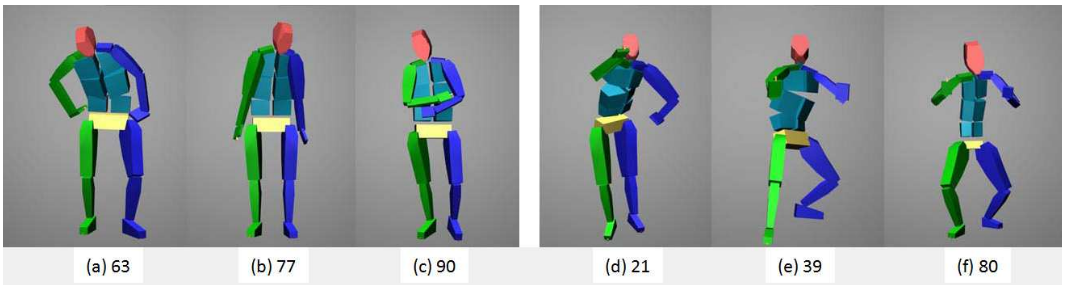


Figure 4.34: Avatar examples representing typical postures for the significant differences of V8, the frontal extension of the hand from the shoulder. (a)-(c) Representing ratings 1 and 2 are postures 63, 77 and 90; (d)-(f) Representing ratings 5 to 7 are postures 21, 39 and 80. Recall from Section 4.6.2 that the numbers refer to the posture position in the bar charts located in Appendix C for easy cross-referencing

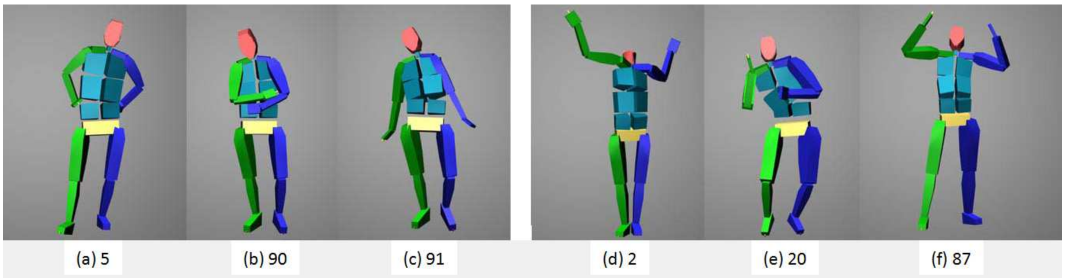


Figure 4.35: Avatar examples representing typical postures for the significant differences of V17, the vertical distance of the hand from the shoulder. (a)-(c) Representing ratings 1 to 4 are postures 5, 90 and 91; (d)-(f) Representing ratings 5 to 7 are postures 2, 20 and 87. Recall from Section 4.6.2 that the numbers refer to the posture position in the bar charts located in Appendix C for easy cross-referencing

level with or slightly higher than elbow height (rating 4). Examples are shown in (a)-(c) of Figure 4.35. This differs from ratings 5 to 7 in which the hand tends to be raised higher than the elbow. This is especially true for rating 7 in which the hand height appears to be the highest. Examples are shown in (d)-(f) of Figure 4.35.

4.8 Automatic Recognition of Affective Dimensions from Posture

Automatic recognition models were built and tested for valence and arousal dimensions separately. Testing was conducted to assess the models' ability to generalise to both new observers and new postures. To build the automatic recognition models, each posture p_i was associated with the median as the ground truth rating $gtm(p_i, O)$ for each affective dimension d_g, and a vector of 24 low-level posture description features $F_i = \{f_{i1}, ..., f_{i24}\}$ listed in Table 4.4.

The topology of the MLP used in this study consists of three layers: one input, one hidden and one output. The number of nodes in the input layer corresponds to the number of features used. The number of nodes in the hidden layer corresponds to the number of input nodes divided by two. The output layer contains one node. 10,000 epochs were used to train the recognition models.

4.8.1 Generalising to New Observers

To test the ability to generalise to new observers, automatic models were trained with the previously unused observer subset three $O_{s,3}$ and tested with $O_{s,1}$ for each of the 10 trials. The results are summarised in Table 4.17. The average recognition rate across the 10 trials for valence was 83.89% (SD = 1.96%) and for arousal was 86.92% (SD = 1.37%).

4.8.2 Generalising to New Postures

Automatic recognition models were also tested for their ability to generalise to novel postures. Each posture p_i was associated with its corresponding feature vector F, and the median rating was assigned as the ground truth $gtm(p_i, O)$. The models were built using an MLP with 10 fold cross-validation. The percentage of recognition achieved for the valence dimension was 79.29% (SD = 1.7%) and for the arousal dimension it was 80.35% (SD = 1.5%).

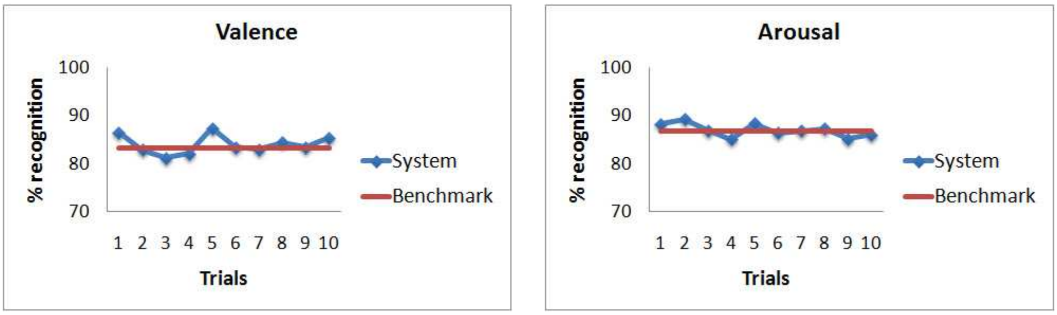
Table 4.17: The performance rates for the automatic recognition models that generalise to new observers for the valence and arousal dimensions.

Generalising to new observers				
Affective dim.	**Trial**	**Auto rec.**	**Average**	**SD**
Valence	1	86.39%	**83.89%**	**1.96%**
	2	82.77%		
	3	81.18%		
	4	81.97%		
	5	87.35%		
	6	83.31%		
	7	82.87%		
	8	84.38%		
	9	83.3 %		
	10	85.37%		
Arousal	1	88.22%	**86.92%**	**1.37%**
	2	89.25%		
	3	86.92%		
	4	85.04%		
	5	88.32%		
	6	86.33%		
	7	86.74%		
	8	87.19%		
	9	85.16%		
	10	86 %		

4.8.3 Evaluation and Discussion

The goal of the affective dimension automatic recognition models for generalising to new observers is to achieve recognition rates comparable to the benchmarks that were set by the human observers. The results are illustrated in Figure 4.36. For both dimensions, while not all of the automatic models' recognition rates surpass the benchmark, they achieve a similar percentage of recognition. Indeed, the average across the automatic recognition models of all 10 trials for each dimension d_g is slightly higher than the benchmark, indicating that these models could replace a human interaction partner in this scenario.

In the models built for generalising to novel postures for each affective dimension d_g, there was little difference in performance between the two dimensions, with arousal performing only slightly better than valence, similar to the observer generalisation results. A study by Zeng et al [ZTP$^+$05] which examined the automatic recognition of valence and arousal from



(a) (b)

Figure 4.36: Observer generalisation results for (a) valence and (b) arousal, and the observer benchmarks for each

a combination of acted facial expressions and vocal prosody also obtained slightly higher recognition for arousal (87%) over valence (84%). Because arousal equates to activity, i.e., from calm to excited, it may be more obvious from bodily expressions, especially stereotypical, exaggerated acted expressions. As discussed and shown throughout this chapter, the postures range from the appearance of very little movement, i.e., a closed body, to a lot of implied movement. As the study presented in the next chapter examines non-acted bodily expressions, it will be interesting to see if similar differences in the recognition of valence and arousal are obtained.

4.9 Chapter Summary

This chapter presented experiments designed to investigate the recognition of basic emotion categories and affective dimensions from acted body posture information. It was hypothesised that human observers could achieve above chance agreement levels when attributing basic emotions and levels of affective dimensions to static images of a faceless humanoid avatar. It was also hypothesised that automatic recognition models could be grounded on a set of low-level posture features and achieve recognition rates similar to benchmarks computed based on the overall agreement levels of the human observers.

Posture data was collected using a Vicon motion capture system. Participants were recruited to enact four emotions through bodily expressions, *angry, fear, happy* and *sad.* No constraints were placed on the actors in how they performed the affective postures. A set of affectively expressive avatars was created from the original motion capture data. The static postures chosen corresponded to the apex instants of the motions as identified by the actors.

Next, a different group of participants, the observers, was asked to judge static posture images according to two nuances of the four different basic emotion categories, angry (angry, upset), fear (fear, surprise), happy (happy, joy) and sad (sad, depressed), and levels of four affective dimensions, (*valence, arousal, potency* and *avoidance*) to each posture p_i.

To analyse human agreement rates, within observers agreement and inter-observer agreement reliability were ascertained for the observers' judgments first. As hypothesised, the within observers agreement levels $W.AgrLabel(L, O)$ were well above chance level for the set of emotion labels L. For the affective dimensions, an overview of the survey data and the Cronbach's α scores computed within observers indicated good consistency between observers' judgments for the valence and arousal dimensions but little consistency between observers' judgments for the for the potency and avoidance dimensions. Based on these results, these two dimensions were eliminated from further analysis. Next, a set of benchmarks was defined from the observers' judgments by computing the average of the between observers agreements $B.AgrLabel(O_{s,1}, O_{s,2})$ for the set of affective labels L and the between observers agreements $B.Agr(d_g, O_{s,1}, O_{s,2})$ for the valence and arousal dimensions.

The next section explained the low-level information used to describe the postural displays after which an anlysis of the low-level posture description was carried out to assess how disciminative each feature f_{iw} was for distinguishing between affective states and affective dimensions. The results for the basic emotions showed significant differences between most of the low-level posture features for each of the observer groups O_{JA}, O_{SL}, O_{US} and that specific low-level features could be used for distinguishing between emotions as reflected in associated avatar examples. The ANOVA results for the affective dimensions also revealed significant differences for many of the low-level features with many of the differences occurring between the higher ratings (i.e., 5 to 7) against the other ratings.

In the second half of the analysis for both the affective state labels and the affective dimensions, automatic recognition models were built and tested to determine if performance rates similar to the benchmarks could be obtained (observer generalisation) using repeated sub-sampling validation. For both the emotion categories and the affective dimensions, the recognition rates for many of the observer generalisation trials outperformed the benchmarks set by the human observers. Indeed, the average across the emotion category models and the affective dimension models did in fact slightly outperform the benchmarks.

Automatic recognition models were also tested for their ability to generalise to novel postures (using 10 fold cross-validation). In the case of the emotion labels, the models performed slightly worse than the observer generalisation emotion recognition models. The results were attributed to the misclassification of fear postures as either angry or happy. Similar to the observer generalisation affective dimension recognition models, a slightly higher recognition rate was obtained for the arousal model than for the valence model. It was conjectured that arousal may be more easily identifiable from acted bodily expressions than valence as arousal denotes activity. The discussions throughout this chapter have remarked that there appear to be differences between the more activated types of emotion represented through posture and the less activated types of emotion.

The results of this study have shown that acted bodily expressions can be recognised at above chance levels by human observers and that automatic recognition models can achieve performance rates similar to benchmarks computed on the human observers' agreement levels. Presented in the following chapter, the second step in the incremental approach is to examine the recognition of more subtle affective states and affective dimensions from non-acted postures in a natural situation (i.e., a video game scenario).

Chapter 5

Case Study 2: Modelling Non-Acted Affect in a Video Game Scenario

The goal of the non-acted postures study is to assess whether non-basic affective states and levels of affective dimensions can be recognised from non-acted displays of body posture. As with the acted postures study, this goal is evaluated in two directions, through human observer recognition and automatic model recognition. In the acted postures study, it was seen that stereotypical postures, purposely performed to express basic emotion categories could be recognised at above chance levels by humans and that automatic recognition models outperform the benchmarks computed. However, in most daily situations, these exaggerated affective displays do not occur and the prototypic emotions are much more infrequent [RC03]. An examination of non-acted postures in which non-basic emotions and affective dimensions are depicted is a natural next step in affective computing research [LNP02]. Hence, it is necessary to test whether the hypothesis holds in a natural situation. To this end, a video game context was chosen. The video game players interact with and control the game using a controller geared to elicit whole body movements of the player. The players are completely unaware of the true nature of the study in an aim to ensure naturally expressed affect.

Several steps were involved in carrying out this study and are reflected in the organisation of this chapter. First, in Section 5.1, posture corpora were collected using motion capture from which postures were manually extracted for use in both human and automatic recognition of affect. Section 5.2, 5.4 and 5.5 explain the human observer recognition of affective state categories and the benchmark set from the results. Section 5.3 describes the low-level description of posture. Similar to Chapter 4, the second half of this chapter reports on the recognition of levels of affective dimensions from non-acted posture in the same section layout as the affective state recognition. The chapter ends with a summary in Section 5.9.

5.1 Posture Corpora

5.1.1 Motion capture data collection

The first step in assessing if non-basic affective states and affective dimensions can be recognised from non-acted postures is to obtain postural data. As described in Section 3.1 of Chapter 3, the Gypsy 5 motion capture system [Ani07] was used in this study to numerically record the body motions of players during video game play.

The players were given an information sheet to read and a consent form to sign if they chose to proceed with the experiment. After signing the consent form, two photos were taken of each player in order to create the configuration model specific to that person. The photos were taken while the player was standing inside a wireframe cube which helps define the volume of space around the body. Refer to Figure 5.1 for an example. One front-facing photo and one right-facing photo were used. The photos were loaded into the AutoCal software (which comes with the Gypsy suite of software). Marker points first were positioned over the corners of the cube to define the space. Next, a set of marker points labelled according to specific body joints and areas of the body were positioned on the images of the player,

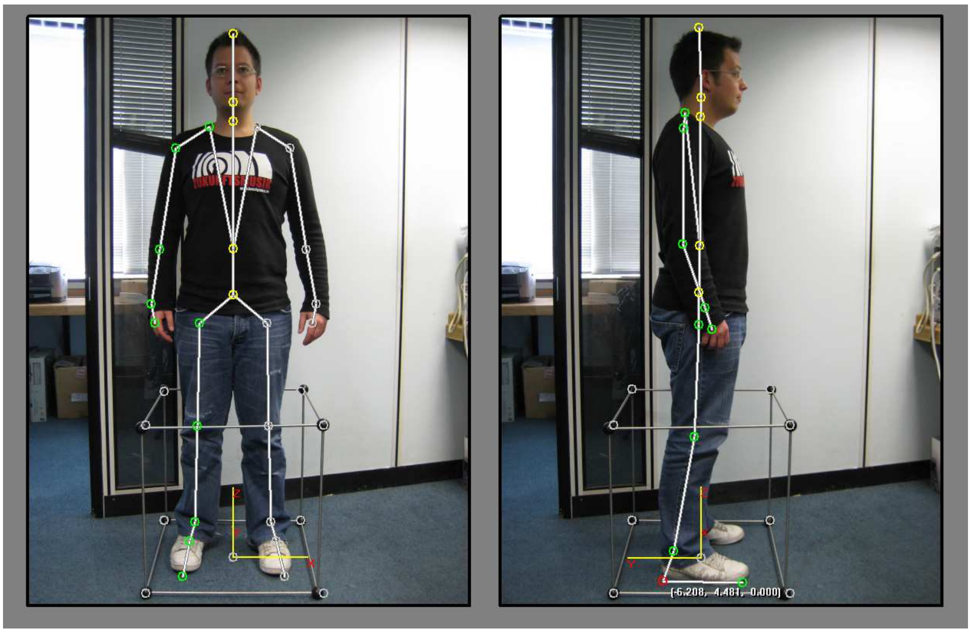


Figure 5.1: An example of how the configuration model is created. Permission to publish the photo has been granted.

thus creating a player-specific model. When finished, the player was fitted with the Gypsy suit and asked to remain still facing north for the calibration process. Motion capture began when the player started to play a sports game with the Nintendo Wii$^{\text{TM}}$.

5.1.2 Players

Eleven players, six females and five males, ranging in age from 20 to 30, were recruited for participation. The players were asked to play sports games with the Wii for up to 30 minutes and have their body motions recorded while wearing the Gypsy 5 motion capture suit. The players were free to choose between tennis, bowling, baseball and golf and could switch games at any time during play. To add a human-human interaction element to the scenario, I communicated with the players throughout the gaming session. It was believed that this interaction would help create a more comfortable and natural atmosphere for the players.

5.1.3 Stimulus identification

After collecting the motion capture data, the apex instants of the motion capture files were manually located. Due to the nature of the non-acted study, a player-defined apex instant does not exist nor was there a definitive static posture for many of the motions. Thus, sections of each motion capture file in which affect was displayed needed to be located first. The non-game play windows are defined as *replay windows* because they are the points during the gaming session in which the player views a replay of the point that was just played. While the player may have been experiencing affective states throughout the entire motion capture session, it is only the replay windows that are considered relevant in this research project. The reasoning is that the different types of actions (i.e., game play versus non-game play) may require different training and testing sets as the actions involved in actual game play have an effect on how affect is expressed.

Three university students were recruited as novice coders. They were asked to locate the start and end frames of the replay windows which they felt contained affective bodily expressions. The coders also provided potential affective state labels for these sections to obtain a list of possible affective states to be used in the forced-choice posture judgment survey described in the next section. The prospective labels are listed in Table 5.1. The labels are grouped according to the affective state that was ultimately chosen for the survey, partially determined from an article by Lazzaro [Laz04] which describes some of the typical affective states associated with general game play.

Table 5.1: The affective and cognitive state labels identified by the coders. The labels are grouped according to the label used in the study.

Concentrating	**Defeated**	**Frustrated**	**Triumphant**
Determined	Bored	Angry	Confident
Focused	Defeated	Frustrated	Excited
Interested	Give up/Sad	Frustrated/Angry	Excited/Motivated
	Sad		Happy/Excited
	Tired		Happy
			Victory

The motion capture files were viewed through the graphical visualisation software CobraView, part of the Gypsy motion capture suite of software. After the coders annotated the replay windows, I viewed the annotated replay windows of each file and chose what appeared to be the apex instant of the motion contained within that window. The apex instant of each affective display was selected as a single static posture stimulus. Static postures were chosen from 12 motion capture files with $|P| = 103$. Posture examples are shown in Figure 3.4(b).

5.2 Human Recognition of Non-Basic Affect from Posture

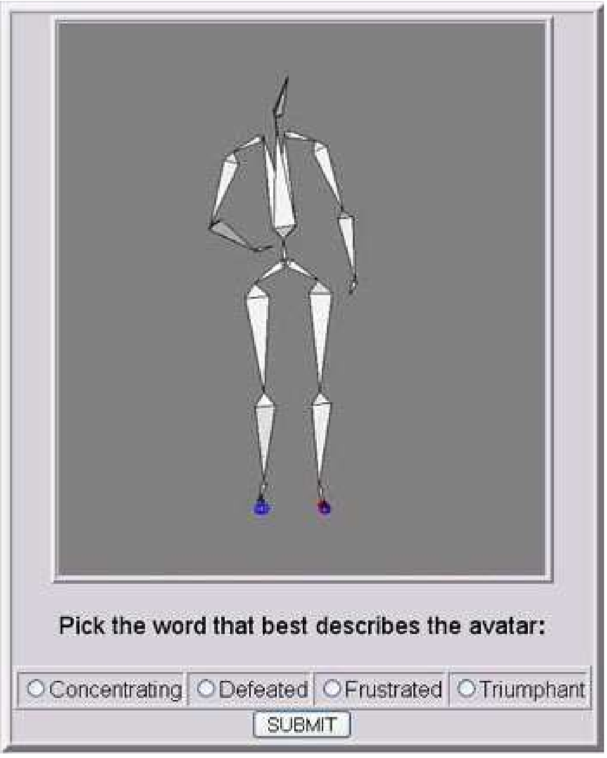
The goal of this section is to examine the extent to which human observers can recognise non-basic affective states from static images of non-acted whole body postures. As previously stated, as of yet there are no recognised benchmarks for evaluating human recognition rates, thus chance level is considered the target, as it is the current metric used in affective computing. A benchmark computed on the observers' agreements will be used as the benchmark for evaluating the performance of the automatic recognition models discussed in Section 5.4.

5.2.1 Survey Procedure

To create the stimuli for the human recognition of affect, the motion capture files were viewed in CobraView. A static image was rendered for each apex posture p_i. An online survey was conducted using the avatar stimuli. The aim was to obtain judgments on the set of affective postures P to associate ground truth affective state labels $gtl(p_i, O)$ to each posture p_i.

A forced-choice experimental design was used. A set of affective labels was defined as $L = \{concentrating, defeated, frustrated, triumphant\}$. The observers were asked to view the set of postures P and associate a possible affective state label l_j in L to each posture p_i

in P. The posture judgment survey itself was carried out using two classification methods: an online task as a series of webpages and a card sorting task [RM97]. These two different methods were used in an attempt to reduce potential boredom experienced by the observers. The tasks were performed on the entire set of postures P. For the online posture judgment survey, the stimuli were presented in a randomised order, one posture per webpage. An example webpage can be seen in Figure 5.2.



Pick the word that best describes the avatar:

Figure 5.2: An example of the posture judgment survey for the non-acted, non-basic affective states study.

For the card sorting task, the posture images were printed in gray scale, one per card (5.5 cm x 6.5 cm), and given to the observers in a random order. The observers were asked to categorise the postures according to the set of affective states L. No other constraints were placed on the observers and they were allowed to take as much time as they felt necessary to classify the postures.

Observers

The posture judgment survey was completed by a set of 8 observers: five males and three females between the ages of 23 and 31. Each observer made five evaluations (8 observers x 5 evaluations = 40 evaluations) on the entire set of postures P: one online evaluation and four card sorting evaluations. At least 12 hours elapsed between evaluations. The rationale for collecting multiple evaluations from a small set of observers was that it may help to eliminate differences in how individuals interpret an affective label. According to a study by Picard and colleagues [PVH01], multiple judgments collected from a single observer may be more consistent because individuals may have different definitions for, or ways of interpreting affective states. This may be especially true for more subtle, non-basic affective states which may be subject to even greater variation in interpretation. In the remainder of the affective categories study in this chapter, O is defined as a set containing 40 observation evaluations, i.e., $|O| = 40$.

5.2.2 Overview of the Survey Data

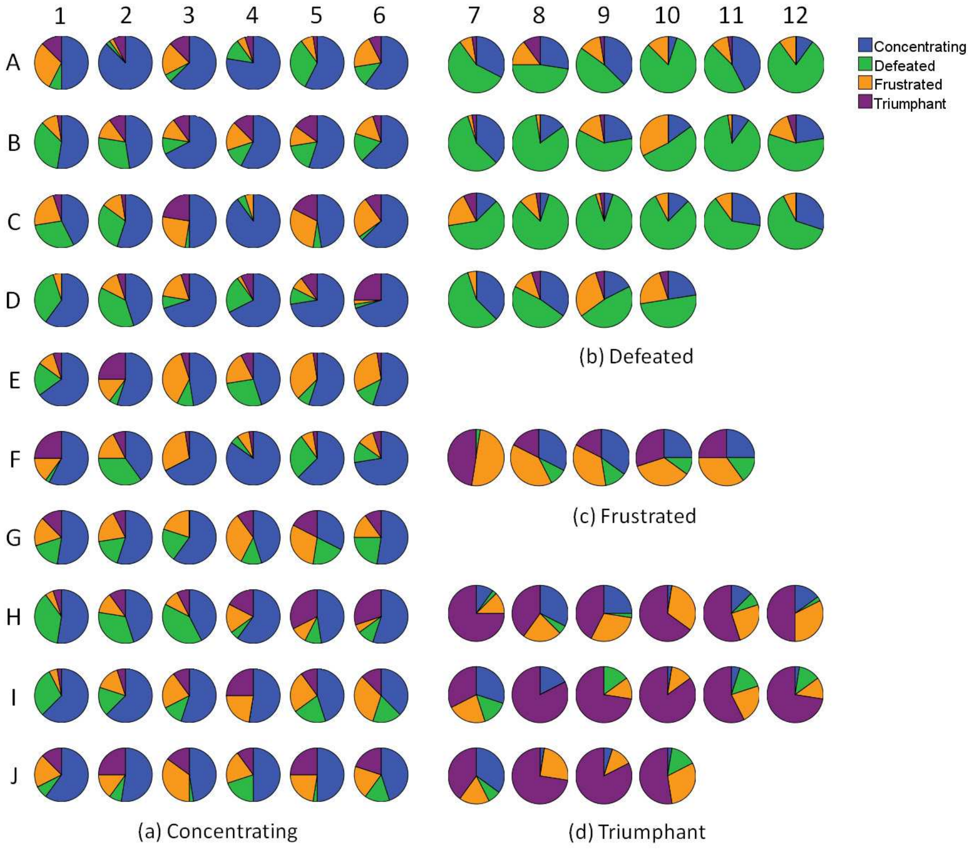
To gain a general overview of the survey data, the most frequent label across the 40 evaluations was taken as the ground truth label $gtl(p_i, O)$ for each p_i. As a result, 60 postures were labelled as concentrating, 22 as defeated, 5 as frustrated and 16 as triumphant.

The within observers agreement $W.Agr(L, O)$ across the set of labels L (as defined in Equation (3.3)), and the within observers agreement $W.Agr(l_j, O)$ for each affective label l_j (as defined in Equation (3.2)), is listed in Table 5.2 and represented in the pie charts illustrated in Figure 5.3. Similar to the agreement graphs shown in Chapter 4, Section 4.2, each pie chart represents the frequency of use $freq(p_i, l_j)$ for each affective state label l_j for each posture p_i. The pie charts (i.e., postures) are organised according to ground truth label $gtl(p_i, O)$. In order to easily locate individual postures, the rows have been labelled with letters and the columns have been labelled with numbers.

The within observers agreement $W.Agr(l_j, O)$ for each affective label l_j, is above chance

Table 5.2: An overview of the agreement between the set of observers O to classify the set of postures P. The first column lists the $W.Agr(L, O)$ across the set of labels L and the remaining four columns list the $W.Agr(l_j, O)$ for each affective label l_j. The number of postures per affective state l_j is noted in parentheses.

$W.Agr(L, O) * 100$	$W.Agr(l_j, O) * 100$			
	Concentrating	**Defeated**	**Frustrated**	**Triumphant**
58%	57% (60)	64% (22)	39% (5)	61% (16)



(b) Defeated

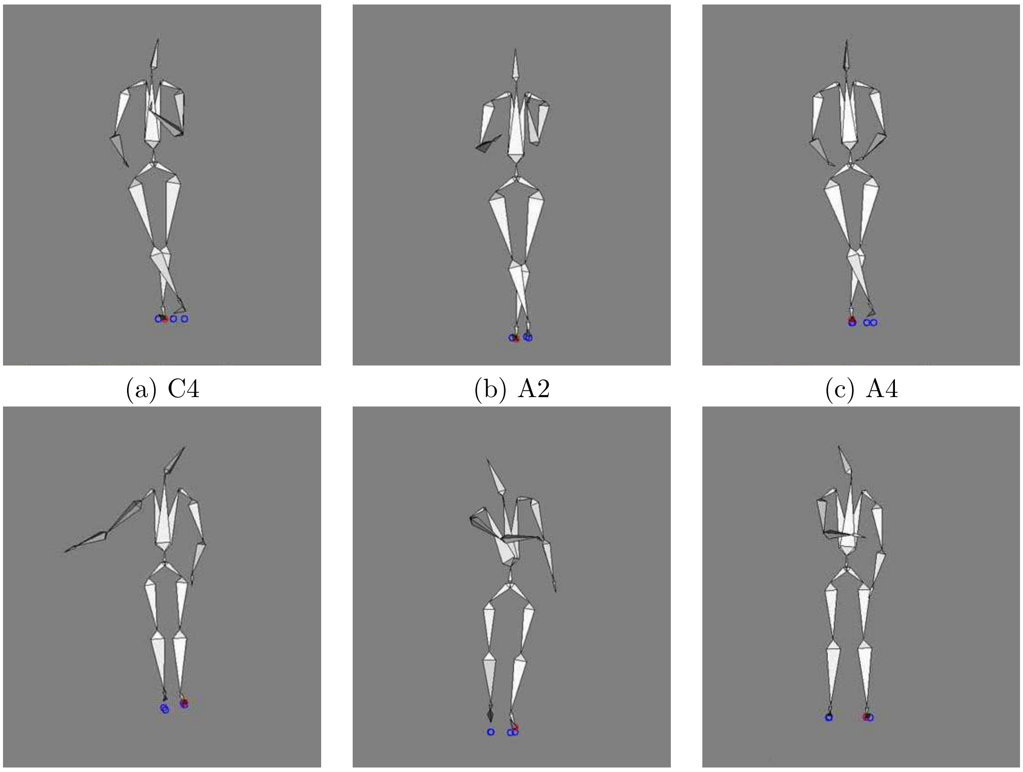
(c) Frustrated

(a) Concentrating

(d) Triumphant

Figure 5.3: Each pie chart indicates the frequency of use $freq(p_i, l_j)$ for each emotion label $l_j : j = 1, ..., 4$ in L for each posture p_i in P according to O. The column numbers and the row letters allow specific postures to be easily identified and located when referenced in the text or in other Figures

level of 25% (considering four affective state categories). The concentrating category contains considerably more postures (60) than the other categories. One possible reason for this outcome could be due to some of the postures not fitting into the affective labels available. At the end of the classification task, some of the observers informally reported that they often used this category when they felt that the posture being evaluated did not fit into any of the other categories (i.e., concentrating seems to have been interpreted as neutral).



(a) C4

(b) A2

(c) A4

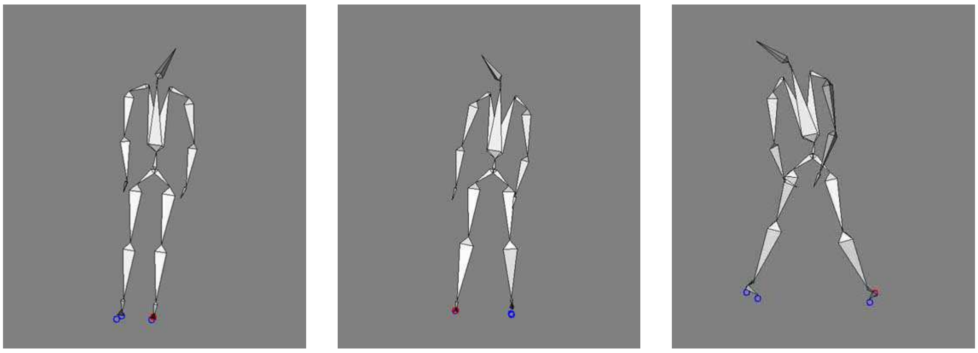
(d) G5

(e) I6

(f) F2

Figure 5.4: (a)-(c) the 3 postures with the highest percentage of agreement for concentrating; (d)-(f) the 3 postures with the lowest percentage of agreement for concentrating. The letter/number pairs refer to the location of the postures in Figure 5.3

The postures that achieved the highest frequency of use for concentrating within the concentrating postures, shown in Figure 5.4(a)-(c), were C4 (90%), A2 (87.5%) and A4 (77.5%). Evaluating the postures visually from the 2D image, the 'typical' concentrating postures all exhibit legs crossed at the heels, arms bent at the elbows with hands in front of the body held around waist and torso height, and the head straight, i.e., no bending or tilting. The postures with the lowest frequency of use for concentrating, shown in Figure 5.4(d)-(f), are G5 (32.5%), I6 (37.5%) and F2 (40%). These postures all have tilted heads, one arm relaxed down along the body and the other arm raised to around torso height with the hand extended.

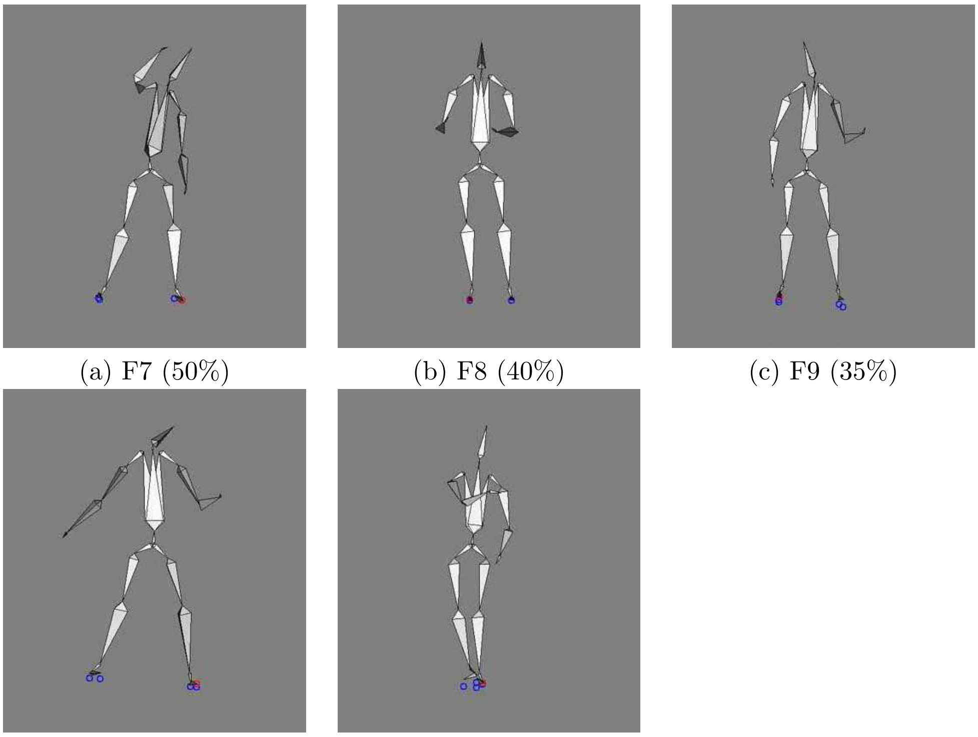


(a) C9 (b) B11 (c) C8

Figure 5.5: (a)-(c) the 3 postures with the highest frequency of use for defeated. The letter/number pairs refer to the location of the postures in Figure 5.3

The defeated category contains 22 postures. Within this category, there is almost no disagreement for triumphant. Instead, disagreement occurs with concentrating, the other less activated type of affective state. The postures with the highest frequency of use for defeated were C9 (90%), B11 (87.5%) and C8 (82.5%), shown in (a)-(c) of Figure 5.5, respectively. All three of these postures have heads tilted to the side and the arms extended down along the body. C8, the posture with the third highest, is the most different from the other two high frequency defeated postures with the feet more than shoulder width apart and the torso tilted sideways slightly.

Frustrated is the category with the most disagreement in labelling. Only five postures obtained frustrated as the ground truth label $gtl(p_i, O)$. The frustrated postures can be seen in Figure 5.6(a)-(e), ordered from highest to lowest frequency of use. The configuration of the frustrated postures appears to be quite different from the postures in the other categories. Most noticeable is that they appear more animated than the concentrating or defeated



(a) F7 (50%)

(b) F8 (40%)

(c) F9 (35%)

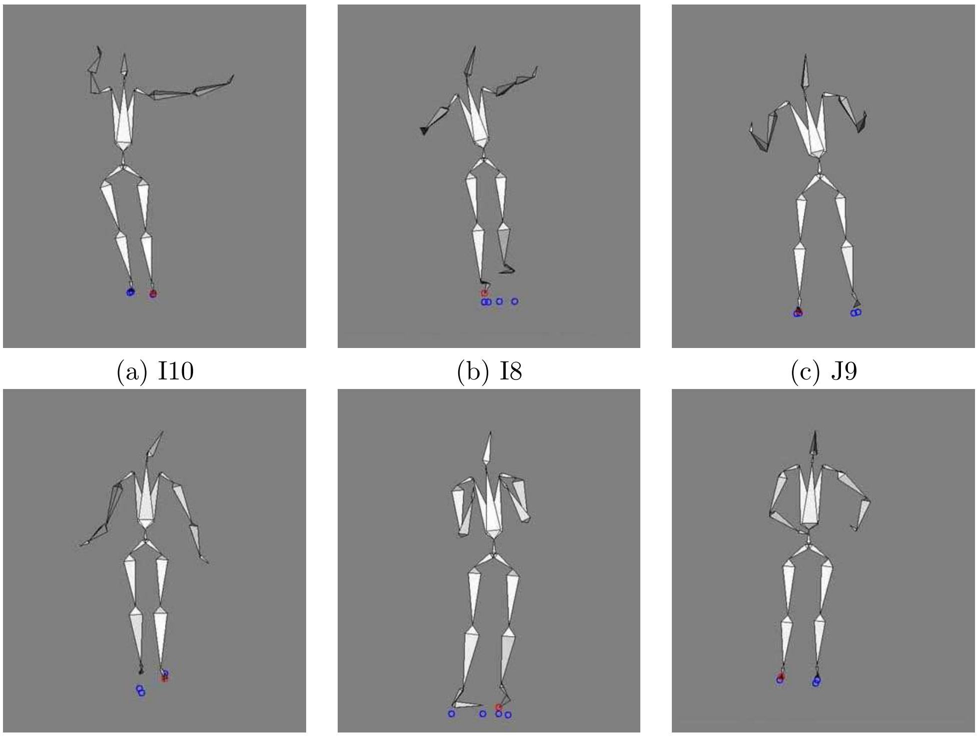
(d) F10 (35%)

(e) F11 (35%)

Figure 5.6: The 5 frustrated postures. The letter/number pairs refer to the location of the postures in Figure 5.3

postures. For all five postures, at least one arm is bent at the elbow and raised frontally or laterally. The feet range from close together to more than shoulder width apart.

Sixteen postures were assigned triumphant as the label l_j with the highest frequency of use $freq(p_i, O)$. The triumphant postures with the highest frequency of use, I10 (85%), I8 (82.5%) and J9 (82.5%), are shown in Figure 5.7(a)-(c). The postures with the lowest frequency of use for triumphant, I7 (32.5%), H8 (40%) and J7 (40%), are shown in Figure 5.7(d)-(f). I10 and I8 both have the arms raised to around shoulder height and extended laterally and/or frontally - appearing very animated. J9 appears less animated with the arms extended slightly laterally, bent at the elbows and the hands raised frontally to around shoulder height. J9 is similar to the low frequency of use triumphant posture, H8, except that the arms are not laterally extended, but instead the elbows remain close to the body.



(a) I10

(b) I8

(c) J9

(d) I7

(e) H8

(f) J7

Figure 5.7: (a)-(c) the 3 postures with the highest frequency of use for triumphant; (d)-(f) the 3 postures with the lowest frequency of use for triumphant. The letter/number pairs refer to the location of the postures in Figure 5.3

In the other two low frequency triumphant postures, I7 and J7, the hands are at waist height, with a slight bend in the elbows. Disagreement within the triumphant category occurs mainly with frustrated, the other activated affective state, and with concentrating for a smaller number of postures.

5.2.3 Creating Benchmarks

To create benchmarks for the human recognition of affective states from posture, the first step was to assess observer agreement reliability. Fleiss' kappa, computed for O, achieved 0.391, equating to the upper end of the range for 'fair' agreement. The second step was to create the benchmarks that will be used to evaluate the performance rates of the automatic recognition models discussed in the second half of this chapter.

Table 5.3: The $B.AgrLabel(O_{s,1}, O_{s,2})$ and the inter-observer agreement reliability (i.e., Cohen's kappa) between $O_{s,1}$ and $O_{s,2}$.

Inter-observer agreement reliability

Trial	$B.AgrLabel(O_{s,1}, O_{s,2}) * 100$	**Kappa**	**95% CI**	**Strength**	**Benchmark**
1	62.14%	0.436	0.313, 0.559	Moderate	66.70%
2	52.43%	0.295	0.158, 0.432	Fair	
3	70.87%	0.542	0.411, 0.673	Moderate	
4	69.90%	0.523	0.392, 0.654	Moderate	
5	76.70%	0.623	0.498, 0.748	Substantial	
6	76.70%	0.600	0.471, 0.729	Moderate	
7	63.11%	0.462	0.339, 0.585	Moderate	
8	59.22%	0.372	0.233, 0.511	Fair	
9	67%	0.470	0.329, 0.611	Moderate	
10	68.93%	0.497	0.362, 0.632	Moderate	

Inter-observer agreement reliability was measured to test the consistency between subsets of observers. Ten trials (i.e., $s = 1, ..., 10$) were created using the random repeated subsampling procedure described in Chapter 3. Each trial comprised three disjoint subsets $O_{s,1}$, $O_{s,2}$ and $O_{s,3}$. $O_{s,1}$ and $O_{s,2}$ contained 15 evaluations (by three different observers) and O_{s3} contained 10 evaluations (by two different observers). For each trial, the between observers agreement $B.AgrLabel(O_{s,1}, O_{s,2})$ was computed between $O_{s,1}$ and $O_{s,2}$. O_{s3} will be used in the second half of this chapter to train automatic recognition models of non-basic affective states. The results are listed in Table 5.3. Each row constitutes a trial and lists the $B.AgrLabel(O_{s,1}, O_{s,2})$, Cohen's kappa, the 95% confidence interval and the strength of agreement [LK77]. The strength of agreement is mostly 'moderate' across the 10 trials, which can be taken to mean good agreement beyond chance [BCMS99]. This is what was expected, given that low inter-observer agreement reliabilities are an acknowledged outcome of employing naturalistic data [CDCC05][ADBM05] as cited in [AR09]. The benchmark obtained is 66.70% ($SD = 7.64\%$) and is listed in the last column of the Table. Recall from Chapter 3 that the benchmark is calculated as the average agreement across the 10 trials. Similar to the within observers agreement $W.AgrLabel(l_j, O)$ on each label l_j and the overview of the survey discussed in Section 5.2.2, the lowest agreement levels occurred for frustrated postures across the 10 trials as was expected.

5.2.4 Discussion

Examining the results of the overview of the data, the $W.AgrLabel(l_j, O)$ for each label $l_j : j = 1, ..., 4$ was above chance level, thus outperforming the target rate. The highest agreement levels were seen for defeated and triumphant. It is possible to conjecture that they are the two most strongly opposite affective states being studied. Also, defeated could be considered part of the sad emotion family and triumphant part of the happy emotion family. Taking this view, the results are similar to the Chapter 4 within observers agreement $W.AgrLabel(sad, O)$ which obtained the highest agreement of the four emotion categories. In Coulson's study [Cou04], for the frontal view postures, sad and happy were attributed to the highest number of postures where observer concordance rates between 50% to 80%. In Kapur et al.'s study [KKVB$^+$05], sad and joy also obtained the highest levels of observer agreement at 95% and 99% respectively.

$W.AgrLabel(frustrated, O)$ was the lowest of the four categories. However, only five postures obtained a ground truth label $gtl(p_i, O)$ of frustrated. This result may signify several things. First, that dynamic information, such as direction and force of movement, may be necessary for identifying frustrated, similar to Coulson's research [Cou04] on fear. For instance, there may be similarities in static representations of frustrated and triumphant and knowing the direction may solve these ambiguities as triumphant movements may contain more upward movement and frustrated could contain more forceful downward or repetitive movements. A similar situation between happy and angry acted postures was resolved with the addition of dynamic features in a study by the candidate [KFBB05], not included in this thesis.

Second, that the video game situation considered did not elicit a frustrating type of experience. However, as described in the 'Survey Procedure' section, as the experimenter, I remained in the room, interacting with the players, and detected frustration in several players through vocal content. Thus, it may be that posture alone is not enough to discern frustration and the addition of another modality (such as voice) is required. For instance, high levels of human-human agreement (72%) were obtained for detecting a combination of annoyance and frustration from speech utterances [ADK$^+$02].

5.3 Low-Level Posture Description

As explained in Chapter 3, each posture p_i corresponds to a single frame of motion capture data. This data is used to build the low-level description of the configuration of posture from the Gypsy motion capture data. Each extracted frame of motion capture data, comprising 3D positions of the body was considered to be the set of low-level posture configuration features $F_i = \{f_{i1}, ..., f_{i41}\}$. Refer to Figure 5.8 to see a map of the posture configuration features (rotations) at the initialisation (i.e., neutral) pose, defined by the potentiometer placement and the additional features determined from the player configuration model.

The potentiometer information is transformed into Euler coordinates automatically at the time of recording. Although not kinematically possible, the range of movement for each feature is from between 0 and 360 degrees, positive and negative, with 0 as the neutral position (i.e., no movement occurred). To adjust for plausible human movement, the possible (human) range of movement for each feature was assessed and then normalised to [0,1]. Table 5.4 lists the low-level posture configuration features computed by the motion capture system and the ranges considered for normalisation.

For some of the joints, the range of one direction of the movement, e.g., the forward movement, was greater than the range for the opposite direction of the movement, e.g., the backward movement. Therefore, each portion of the range of movement (e.g., [30,0] = range of backward movement of the hip and (0,-55] = range of forward movement of the hip) was transformed independently to ensure that 0 remained the neutral position. To do this, each original value v_{iw} is transformed according to the following rule

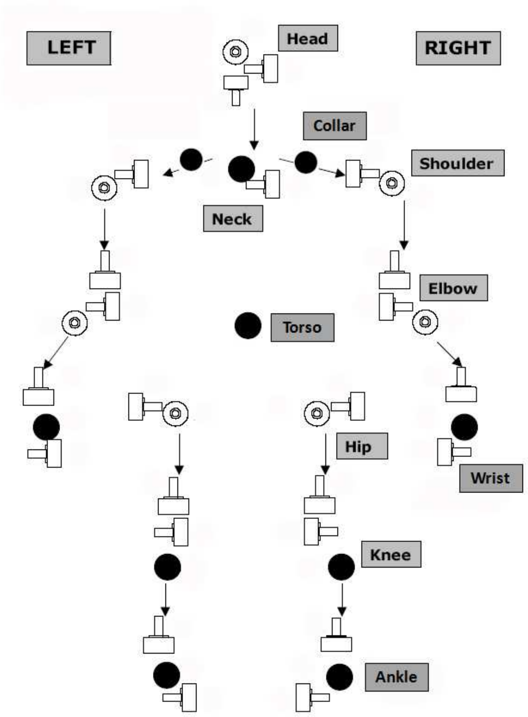


Figure 5.8: The potentiometer map and the additional features determined from the player configuration model at the initialisation pose. The data recorded comprises the set of low-level posture configuration features in the non-acted postures study

If $v_{iw} < 0$

$$f_{iw} = \frac{1}{2} * \frac{v_{iw} - b_{iw}}{e_{iw} - b_{iw}} \tag{5.1}$$

If $v_{iw} > 0$

$$f_{iw} = \frac{1}{2} * \frac{v_{iw} - b_{iw}}{e_{iw} - b_{iw}} + 0.5 \tag{5.2}$$

where f_{iw} is the normalised value of the feature, v_{iw} is the original value from the motion capture data, b_{iw} is the start of the range, and e_{iw} is the end of the range. In this way, 0.5 corresponds to a neutral position. The two ranges are weighted differently somewhat on the

Table 5.4: The set of low-level posture configuration features used in the non-acted postures study and the ranges of each for normalisation

Low level posture features and normalisation ranges			
Features	Z	X	Y
	Begin to **End**	**Begin** to **End**	**Begin** to **End**
Left hip	22 to -22	30 to -55	20 to -20
Left knee	–	55 to -9	–
Left collar	4 to -4	6 to -3	4 to -4
Left shoulder	170 to -40	45 to -110	90 to -50
Left elbow	8 to -135	8 to -55	1 to -90
Left wrist	55 to -55	45 to -40	90 to -90
Right hip	-22 to 22	30 to -55	-20 to 20
Right knee	–	55 to -9	–
Right collar	-4 to 4	6 to -3	-4 to 4
Right shoulder	-170 to 40	45 to -110	-90 to 50
Right elbow	-8 to 135	8 to -55	-1 to 90
Right wrist	-55 to 55	45 to -40	-90 to 90
Torso	35 to -35	-15 to 55	-26 to 26
Neck	-15 to 15	-18 to 18	-22 to 22
Head	-50 to 50	-65 to 65	-55 to 55

basis of the effort and feasibility of performing that particular movement.

For non-directional rotation, a further transformation was applied. The z and y rotations of the head, neck and torso features were considered non-directional, meaning that the head, for example, turned to the left was the same as the head turned to the right. This was accomplished with the following decision rule

$$if(f_{iw} \geq 0.5), \quad then \quad f_{iw} = 2(1 - f_{iw}), \quad else \quad f_{iw} = 2f_{iw} \tag{5.3}$$

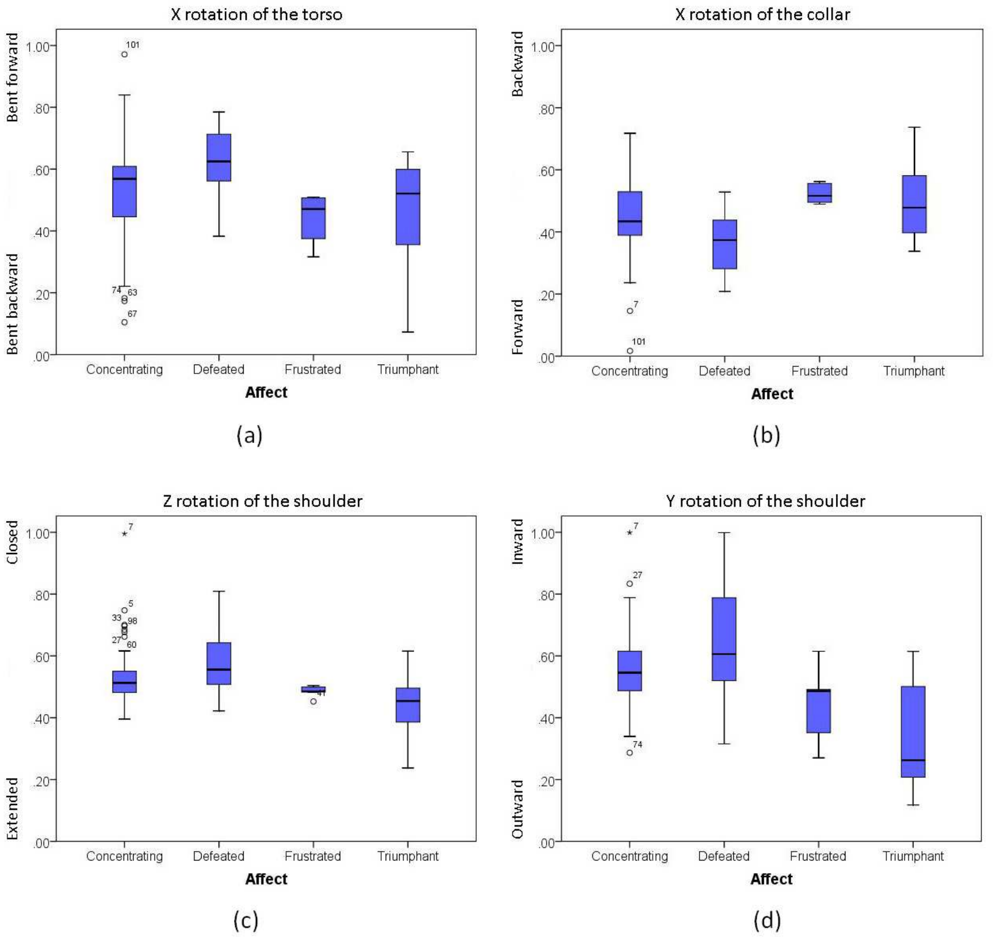
5.4 Low-Level Posture Description Analysis

To evaluate the discriminative power of the set of low-level posture configuration features F_i for distinguishing between the affective states examined, each feature f_{iw} was subjected to a one-way ANOVA with Tamhane's T2 post hoc comparisons implemented for unequal variances. Each posture p_i was associated with the ground truth label $gtl(p_i, O)$ defined according to the set of observers O and a vector of the low-level posture features F_i. The results are summarised in Table 5.5. The first column lists the features shown to be important for discriminating between affective states and the significance level is listed in the last column. The means for each affective state label l_j are listed in the middle four columns. The superscript letter pairs associated with the means signify the significant differences between those affective state label pairs according to the post hoc comparisons. Boxplots depict some of the most interesting results.

Table 5.5: The low-level posture description features which reached significance between the affective states (one-way ANOVA with df = 3 (emotions)). For each feature f_w, *a-e* pairs indicate the significant differences between means according to Tamhane's T2 *post-hoc* comparisons Rot. = rotation, L. = left, R. = right

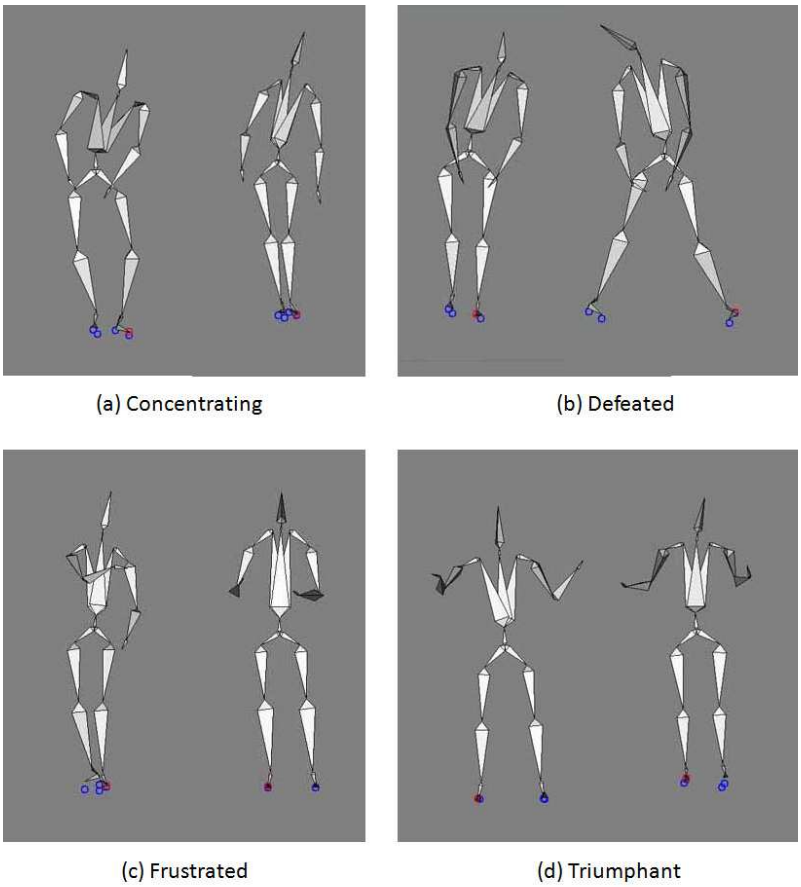
	Means for affective states				
Low-level feature	**Concentrating**	**Defeated**	**Frustrated**	**Triumphant**	p
x rot. torso	$.53^{a}$	$.63^{abc}$	$.44^{b}$	$.48^{c}$	.003
x rot. L. collar	$.45^{ab}$	$.36^{acd}$	$.52^{bc}$	$.49^{d}$	.001
y rot. L. collar	.53	.51	.41	.46	.039
x rot. R. collar	$.45^{ab}$	$.36^{acd}$	$.52^{bc}$	$.49^{d}$	.001
y rot. R. collar	$.47^{ab}$	$.49^{acd}$	$.59^{bc}$	$.54^{d}$	.039
z rot. L. shoulder	$.53^{ab}$	$.58^{cd}$	$.48^{ac}$	$.44^{bc}$	.000
y rot. L. shoulder	$.55^{a}$	$.65^{b}$	.44	$.34^{ab}$	.000
y rot. R. shoulder	$.50^{a}$	$.61^{b}$	.34	$.32^{ab}$	.000
z rot. L. elbow	$.66^{a}$	$.54^{ab}$	.66	$.73^{b}$	.003
x rot. L. elbow	$.75^{a}$	$.62^{ab}$	.83	$.84^{b}$	.011
y rot. L. elbow	$.64^{a}$	$.52^{ab}$	.62	$.72^{b}$	.001
z rot. R. elbow	$.57^{a}$	$.50^{b}$	.66	$.75^{ab}$	.000
x rot. R. elbow	$.67^{a}$	$.55^{b}$	.77	$.86^{ab}$	.001
y rot. R. elbow	$.13^{ab}$	$.03^{ac}$	.28	$.46^{bc}$	.000
x rot. L. wrist	.58	$.47^{a}$	.59	$.68^{a}$	.004
x rot. R. wrist	.49	.48	.52	.63	.005
z rot. neck	.77	.65	.84	.71	.017

Even though a standing scenario was chosen, it is interesting to note that the important low-level posture description features are mainly the arms and upper body. This could indicate that the majority of the movement really is upper body for the type of scenario used. Looking more closely at the results, it can be seen that significant differences occurred for the x rotation of the torso (the degree of forward or backward bending of the body) between the more 'active' affective states (frustrated and triumphant) and the less 'active' states (concentrating and defeated). This difference is evident in boxplot (a) depicted in Figure 5.9. Looking at the avatars in Figure 5.10, it can be seen that the body is slightly bent forward in the concentrating and defeated postures, whereas the body remains upright in the frustrated and triumphant postures.



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

Figure 5.9: Examples of the low-level posture description features with significant differences between affective state labels (a) the x rotation of the torso; (b) the x rotation of the collar; (c) the z rotation of the shoulder; (d) the y rotation of the shoulder



(a) Concentrating

(b) Defeated

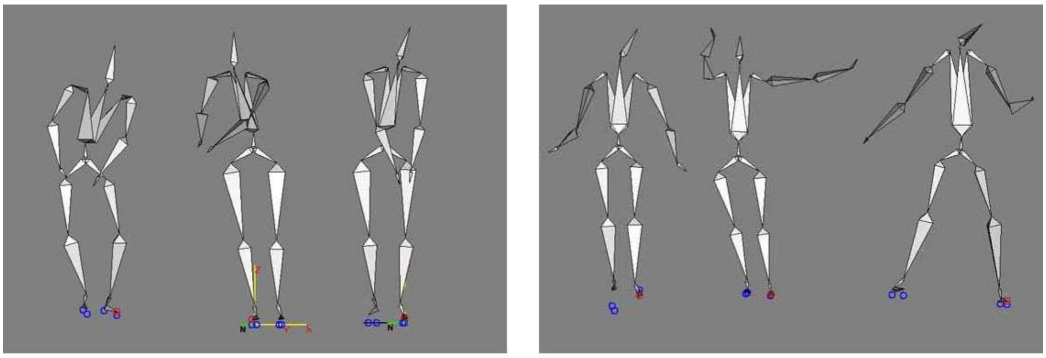
(c) Frustrated

(d) Triumphant

Figure 5.10: Avatar examples demonstrating the x rotation of the torso differences between the affective states.

In the case of the x rotation of the collar (the degree of forward slumping or backward straightening of the collar) significant differences occurred between concentrating and defeated against frustrated, and defeated against triumphant. The boxplot in Figure 5.9(b) and the avatars in Figure 5.11 illustrate these differences. The concentrating and defeated avatars in Figure 5.11(a) have shoulders that are slumped forward more than the frustrated and triumphant avatars in Figure 5.11(b).

Significant differences were also obtained for the z and y rotations of the shoulders, shown in boxplots (c) and (d) of Figure 5.9. For the z rotation of the shoulder (the lateral



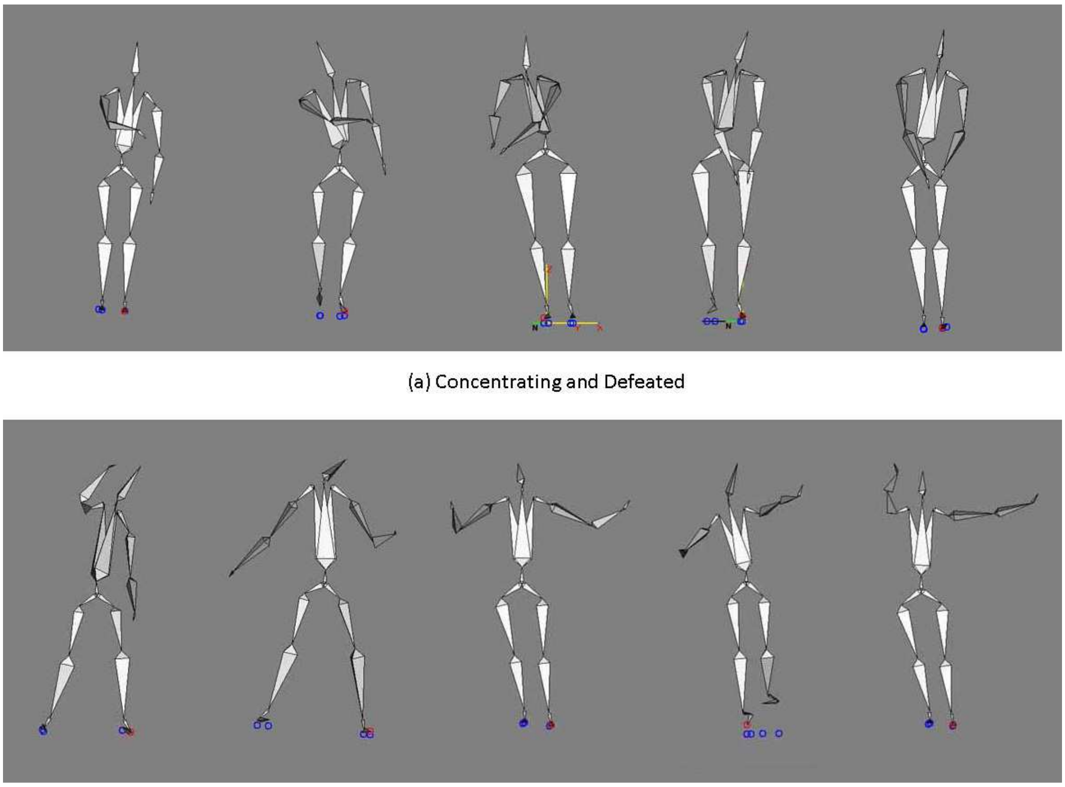
(a) Concentrating and Defeated

(b) Frustrated and Triumphant

Figure 5.11: Avatar examples demonstrating the differences of (a) concentrating and defeated against (b) frustrated and triumphant for the x rotation of the collar

and vertical extension of the arm at the shoulder), the significant differences occurred with concentrating and defeated against frustrated and triumphant. The arms are raised and extended laterally, i.e., open (refer to the avatar examples in Figure 5.12(b)). This differs from the concentrating and defeated postures in which the arms are closed, i.e., extended slightly diagonally across the body (refer to the avatar examples in Figure 5.12(a)).

For the y rotation of the shoulder (the rolling of the shoulder causing a lateral to and from frontal movement), the significant differences occurred with concentrating and defeated against triumphant. Similar to the z rotation of the shoulders, for the y rotation, much more 'movement' is implied in the triumphant postures with the shoulders squared back and open (refer to the avatar examples in Figure 5.13(c)), whereas the arms are much more closed for the concentrating and defeated postures; the shoulders are rounded inward (refer to the avatar examples in Figure 5.13(a) and (b)).



(a) Concentrating and Defeated

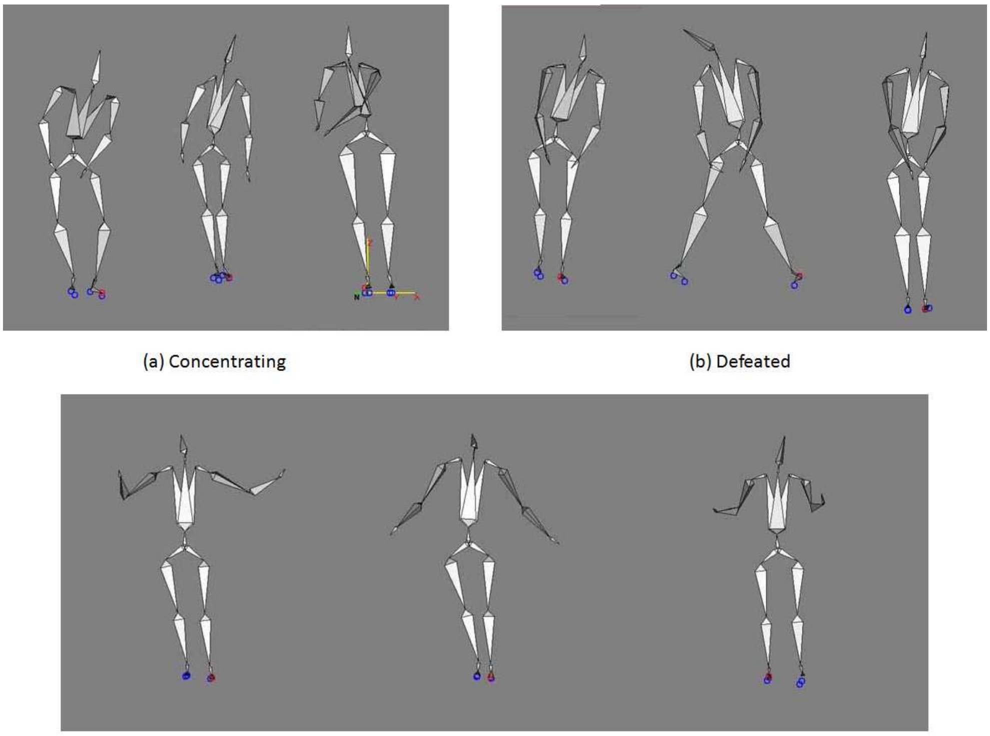
(b) Frustrated and Triumphant

Figure 5.12: Avatar examples demonstrating the differences for (a) concentrating and defeated against (b) frustrated and triumphant for the z rotation of the shoulders

5.5 Automatic Recognition of Non-Basic Affect from Posture

The next task was to build and evaluate automatic recognition models of the subtle affective states from posture. As described in Chapter 3, Section 3.3.2, recognition models were evaluated for their ability to generalise to: i) new observers and ii) new postures. The input for creating the models was the vector of low-level posture features, $F = \{f_1, ..., f_{41}\}$ and a non-basic affective state label $gtl(p_i, O)$, for each static posture p_i.

The topology of the MLP used in this chapter consists of three layers: one input, one



(a) Concentrating

(b) Defeated

(c) Triumphant

Figure 5.13: Avatar examples demonstrating the differences for (a) concentrating and (b) defeated against (c) triumphant for the y rotation of the shoulders

hidden and one output. The number of nodes in the input layer corresponds to the number of features used. The number of nodes in the hidden layer corresponds to the number of input nodes divided by two. The output layer contains four nodes, one node corresponding to each affective state label. 10,000 epochs were used to trained the recognition models.

5.5.1 Generalising to New Observers

To test the automatic recognition models' ability to generalise to new observers, $O_{s,3}$ for each of the 10 trials defined in Section 5.2 was used to build and train the recognition models. Each model was tested with the corresponding $O_{s,1}$. The results are summarised in Table 5.6. The average recognition across the 10 trials was 59.22% (SD = 11.8%). The

Table 5.6: The automatic recognition model testing and evaluation results for generalising to new observers.

Generalising to new observers

Trial	Sys tot	Concentrating	Defeated	Frustrated	Triumphant
1	48.54%	43.2%	58.1%	40%	48%
2	48.54%	61.9%	50%	25%	37.9%
3	62.14%	67.2%	93.8%	0%	76.9%
4	70.87%	80.7%	75%	20%	52.4%
5	45.63%	39.2%	51.5%	0%	62.5%
6	42.71%	37.1%	45%	25%	76.9%
7	75.73%	83.9%	50%	50%	92.3%
8	67.96%	83.6%	78.6%	0%	65%
9	66.99%	77.6%	63.2%	.09%	73.3%
10	63.11%	78.8%	53.3%	0%	61.5%
Avg	59.22%	65.32%	61.85%	16.01%	64.67%
SD	11.8%	18.95%	15.76%	18.82%	15.98%

automatic recognition performances for concentrating, defeated and triumphant are not very different from the set benchmark (66.7%). Instead, the performances on frustrated are very low, similar to results obtained by Zeng et al [ZTL$^+$04] for bimodal prosody and facial expression recognition. This was to be expected given the very small data set available and the low level of agreement among the observers. Hence, the frustrated postures were removed due to lack of training set and a new set of automatic recognition models was built. Figure 5.14 illustrates the differences between the average recognition performances with and without the frustrated postures and the benchmark. Excluding the performances on frustrated, the average recognition across the 10 trials increases to 69.89% ($SD = 9.87\%$).

Evaluation and Discussion

The goal of the observer generalisation models was to achieve classification rates equivalent to the human observer benchmarks. Has the goal been achieved? In automatic facial expression recognition research, it has been acknowledged that recognising basic emotions is easier than recognising non-basic affective states [eKR04]. However, the automatic recognition models presented in this testing achieved performance rates comparable to the bodily expression

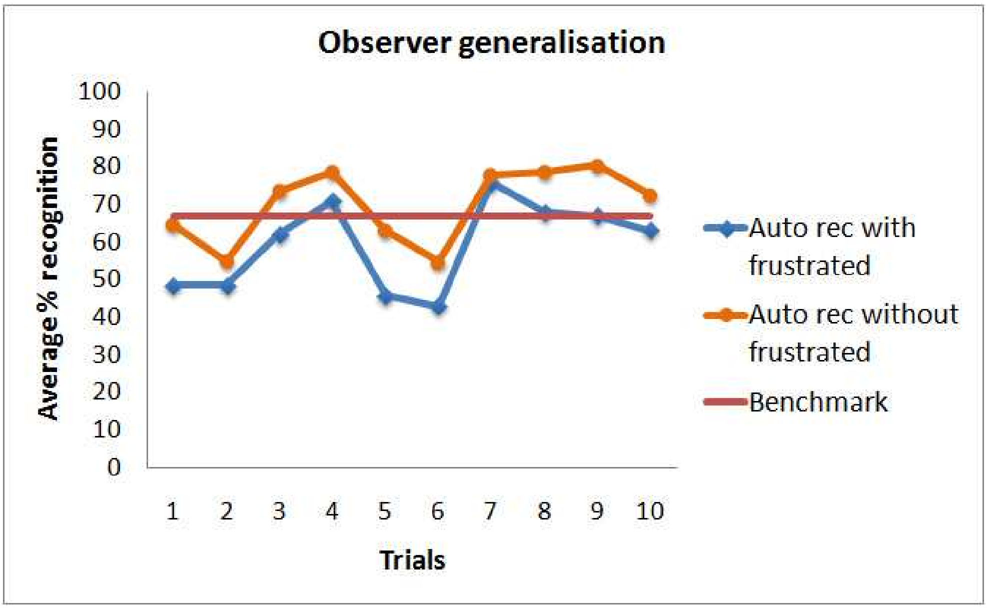


Figure 5.14: Observer generalisation across the 10 trials compared with the benchmark. The average values for each trial have been computed with and without the recognition rate for frustrated

recognition systems presented in Table 2.10, such as Bernhardt and Robinson [BR07] and Kapur et al [KKVB$^+$05], which recognised acted, basic emotion categories and relied on an actor-defined ground truth.

In examining specific misclassifications by the automatic recognition model that includes frustrated, it was found that one posture was misclassified in every trial. This posture (F11 according to Figure 5.3) was predicted as frustrated by the recognition model in four trials but the ground truth label $gtl(p_i, O)$ was triumphant. In four other trials, the opposite misclassification occurred. Comparing the automatic recognition results with the human observer agreement results, posture F11 was ground truth labelled as frustrated (35%) but triumphant and concentrating frequencies were close behind at 25%. These results highlight the confusion that occurred for the observers in distinguishing between frustrated and triumphant.

Without reliable agreement levels by humans in recognising frustrated from body posture, it is difficult to test a recognition model's performance for frustrated. It was suggested in the discussion on human observer agreement (Section 5.2.4) that the integration of information from other modalities such as voice might be necessary. Indeed, a face and prosody bi-modal recognition system achieved a recognition rate of 83.56% for frustration [SCGH06], and D'Mello et al [DCSG06] found that frustration was recognised with high precision levels from dialogue sessions with an intelligent tutoring system.

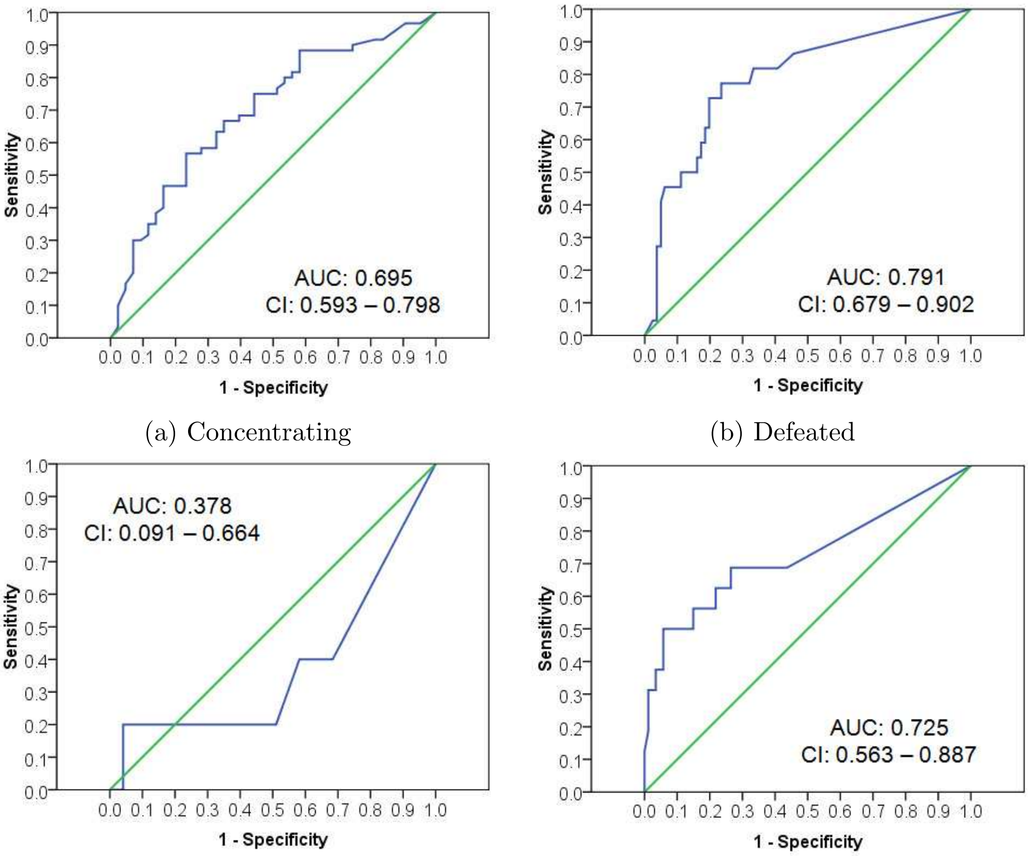
5.5.2 Generalising to New Postures

Automatic recognition models were also tested for their ability to generalise to novel postures as outlined in Chapter 3 and shown in Figure 3.8. To do so, each posture p_i was associated with a ground truth label $gtl(p_i, O)$ assigned by the set of observers O and the low-level posture feature vector F. An automatic recognition model was built and tested using the MLP with 10 fold cross-validation. The model achieved a recognition rate of 59.22%. The model's performance on each affective state label $l_j : j = 1, ..., 4$ is listed in the first row of Table 5.7 and illustrated in the ROC curves in Figure 5.15. Again, due to the lack of sufficient frustrated labelled postures, the recognition model is unable to classify these postures.

Similar to the observer generalisation described in Section 5.5.1, the five frustrated postures were removed and a new recognition model was built. This model achieved an increased recognition rate of 66.33%. The recognition rate for each affective state label is listed in the second row of Table 5.7.

Table 5.7: The recognition model results for generalising to novel postures. The total recognition rate is listed in the first column and the recognition rates for each affective state label l_j are listed in the remaining columns.

Generalising to new postures					
Rec. model	**Total**	**Concentr.**	**Defeated**	**Frustrated**	**Triumphant**
Includes frustrated	59.22%	66.7%	59.1%	0%	50%
Excludes frustrated	66.33%	70%	59.1%	–	62.5%



(a) Concentrating

(b) Defeated

(c) Frustrated

(d) Triumphant

Figure 5.15: The ROC curves for the four affective state labels for the automatic recognition model built for generalising to novel postures. AUC = area under the curve; CI = 95% confidence interval. (a) Concentrating; (b) Defeated; (c) Frustrated; (d) Triumphant

Evaluation and Discussion

The confusion matrices for each recognition model are shown in Table 5.8 (includes frustrated) and Table 5.9 (excludes frustrated). As can be seen in both matrices, the majority of the concentrating postures were correctly classified. This is not surprising given that the concentrating category contained the highest number of postures. In both recognition models however, some concentrating postures were misclassified as defeated and some as triumphant. Again, these results may be attributed to the set of observers O classifying

Table 5.8: The confusion matrix for generalising to novel postures (includes frustrated)

Concentrating	**Defeated**	**Frustrated**	**Triumphant**	← *classified*
40	14	1	5	**Concentrating**
9	13	0	0	**Defeated**
2	0	0	3	**Frustrated**
6	0	2	8	**Triumphant**

Table 5.9: The confusion matrix for generalising to novel postures (excludes frustrated)

Concentrating	**Defeated**	**Triumphant**	← *classified*
42	11	7	**Concentrating**
9	13	0	**Defeated**
5	1	10	**Triumphant**

postures as concentrating when they felt none of the other categories were appropriate. Thus, the concentrating category acted as a type of neutral. The results are similar to the labelling results of Ang et al [ADK$^+$02] in which the task was to classify subtle, naturally occurring speech utterances into non-basic affective states (such as annoyance, tired etc.). However, the majority of the utterances were classified as neutral.

An examination of the misclassification results for the model that included the frustrated postures reveals that triumphant postures were misclassified as frustrated and frustrated postures were misclassified as triumphant. Even though it has already been established that there were not enough frustrated postures for the automatic model to clearly define its classification rules for this affective state, the results add some further verification that frustrated and triumphant may share similar static posture features. Indeed, referring back to the boxplots presented in Figure 5.9, it can be seen that the feature ranges for the frustrated postures are more condensed than, but included in the feature ranges for triumphant. This indicates that the triumphant postures may be more animated (i.e., bigger amplitude) than the frustrated postures, but that the configurations are in fact similar.

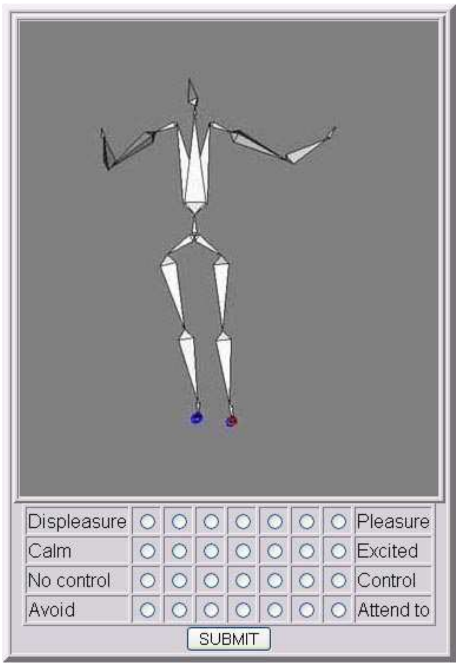
5.6 Human Recognition of Affective Dimensions

Human recognition of levels of affective dimensions of non-acted whole body postures were tested using the same process that was used for human recognition of affective states as reported in the first half of this chapter. The remainder of this case study describes the investigation on the recognition of affective dimensions from non-acted postures.

5.6.1 Survey Procedure and Observers

Similar to the affective state posture judgment survey described in the first half of this chapter, an online survey was conducted to obtain judgments on the non-acted affective postures to determine levels of affective dimensions. The set of 103 posture images used for the affective state judgment survey was reduced by removing several of the very similar looking postures. The result was a set of 94 postures. This was done because of the time required to judge the postures on the four affective dimension scales. The feedback received by two observers that acted as pilot participants was that the survey, conducted as one session, took too long and their interest level in completing the task had significantly waned by the end. To remedy this problem, the final survey was split into three sessions. The recruited observers were asked to complete one session. They were able to complete subsequent sessions after a break if they so chose.

Each posture was presented on a separate page in a randomised order. For each page, the observers were asked to rate each posture p_i in P according to a seven-point Likert scale for a set of four affective dimensions $D = \{valence, arousal, potency, avoidance\}$. An example of the affective dimension posture judgment survey can be seen in Figure 5.16. A set of 30 observers (17 females and 13 males), completed at least one session each. These were combined to create a set of 15 evaluations on the complete set of postures P.



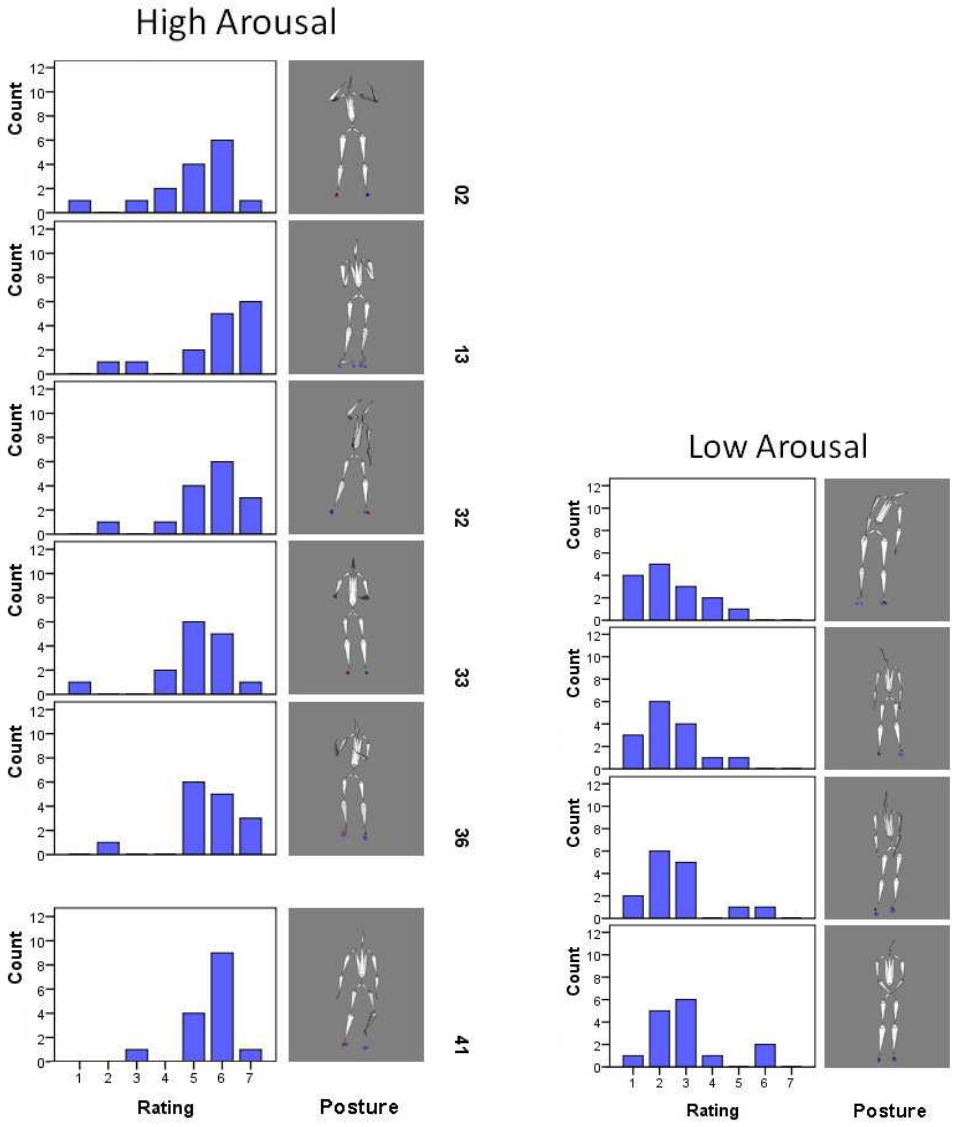
Displeasure	○	○	○	○	○	○	○	Pleasure
Calm	○	○	○	○	○	○	○	Excited
No control	○	○	○	○	○	○	○	Control
Avoid	○	○	○	○	○	○	○	Attend to

Figure 5.16: Example of the affective dimensions posture judgment survey

5.6.2 Overview of the Survey Data

The first step in the analysis was to gain an overview of how the set of observers O judged the affective postures. To this aim, each affective dimension d_g was examined separately and is represented as a series of bar charts. Due to space constraints, the complete set of bar charts can be found in Appendix F. Each row is a posture p_i and each column is an affective dimension $d_g : g = 1, ..., 4$. The x-axis shows the rating scale (i.e., 1 to 7) and the y-axis shows the number of evaluations obtained for each rating in the scale. The number to the right of the posture image is used as a posture identifier to allow for easy identification for the discussions that take place in the remainder of the chapter. Some of the most interesting results are presented in this section.

An examination of the arousal dimension revealed that the majority of the evaluations gave at least 20 of the postures (e.g., postures 1, 2, 4, 13, 16, 32, 33, 36, 37, 41, 68, 70, 71, 75, 80, 84, 85, 87, 89, 91, 93) ratings of 5 or above, i.e., high arousal. A sampling



(a)

(b)

Figure 5.17: (a) High arousal postures, i.e., evaluations of rating 5 and above; (b) Low arousal postures, i.e., evaluations of rating 3 and below. For each bar chart, the x-axis shows the rating scale (i.e., 1 to 7) and the y-axis shows the number of evaluations obtained for each rating in the scale. The number to the right of the posture image is used as a posture identifier to allow for easy location of the posture in Appendix F

of these postures is depicted in Figure 5.17(a). What is interesting to note is that all of these postures have either one or both arms bent at the elbows and/or raised over the head, except for posture 41. In this case, the high arousal ratings could be accounted for by the bent knee and slightly raised leg. Conversely, the configuration of the body of the postures that achieved mainly ratings of 3 or below, i.e., low arousal, appears more closed. Refer to Figure 5.17(b) for examples. For instance, the arms are extended straight down along the body (slightly frontal for three of the postures) or crossed at the wrists. Furthermore, the head and the upper body are somewhat bent forward or tilted to the side.

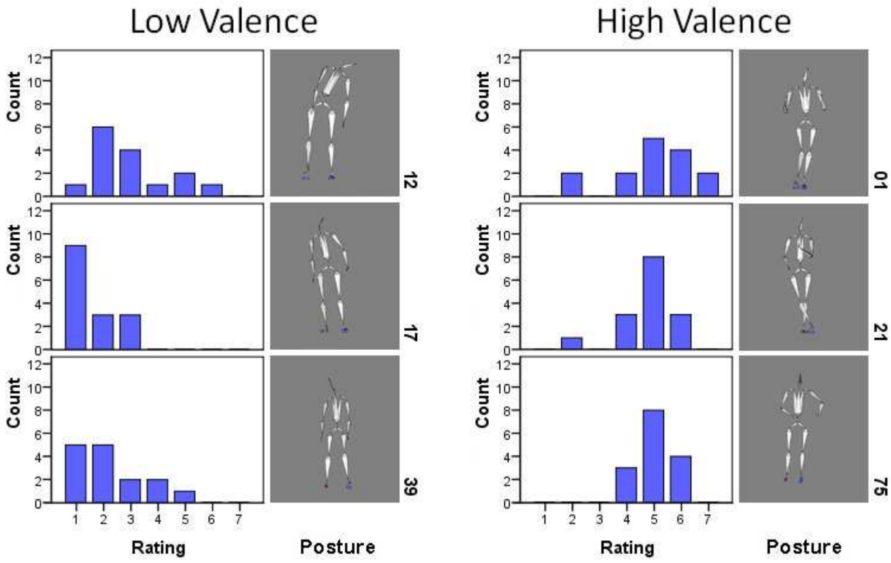


Figure 5.18: Low valence postures, i.e., evaluations of rating 3 and below are in the first column. High valence postures, i.e., evaluations of rating 5 and above are in the second column. For each bar chart, the x-axis shows the rating scale (i.e., 1 to 7) and the y-axis shows the number of evaluations obtained for each rating in the scale. The number to the right of the posture image is used as a posture identifier to allow for easy location of the posture in Appendix F

An examination of the valence dimension also revealed a distinction between low and high levels of valence postures (refer to Figure 5.18 for examples), although to a lesser degree than for arousal. The configuration of the body in the low valence postures (the first column of Figure 5.18) shows arms to be fairly straight and extended down, the head tilted to the side and the feet about shoulder width apart. This stance is contrasted with the high valence postures (the second column of Figure 5.18) in which the elbows are bent (to varying degrees across each posture), the head is almost straight up, not tilted, and the feet are much closer together or crossed in one instance.

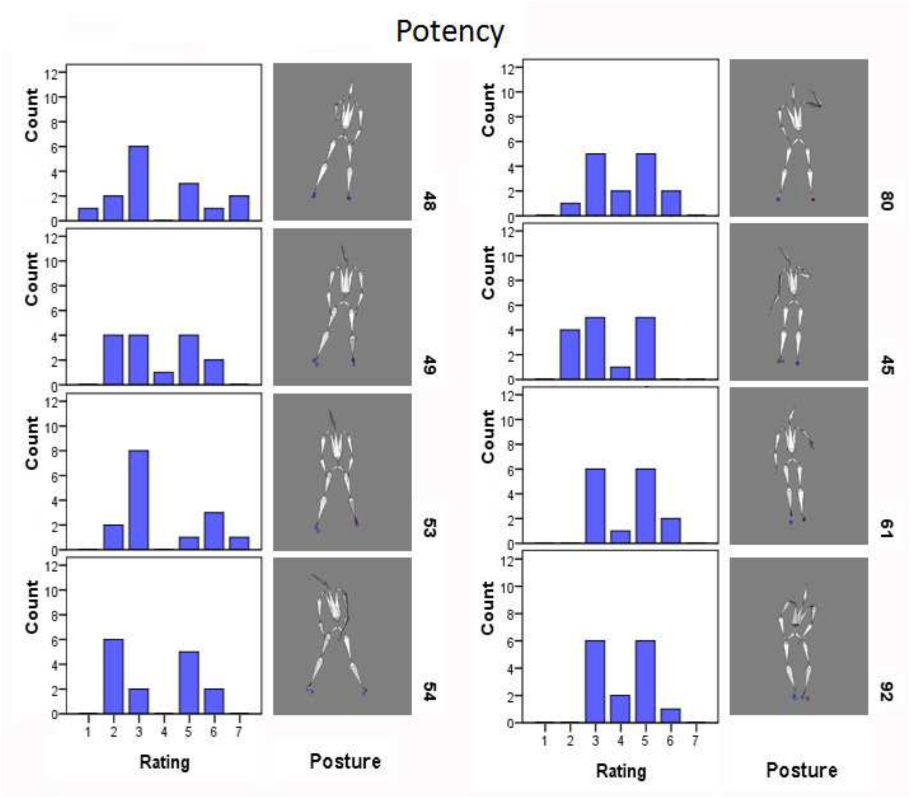


Figure 5.19: Examples of postures with split ratings for potency. For each bar chart, the x-axis shows the rating scale (i.e., 1 to 7) and the y-axis shows the number of evaluations obtained for each rating in the scale. The number to the right of the posture image is used as a posture identifier to allow for easy location of the posture in Appendix F

The potency dimension is not as often investigated as the valence and arousal dimensions. The overview of the survey data for potency revealed that there was nearly an equal number of evaluations for both low and high ratings on each posture p_i, causing a bimodal distribution (refer to Figure 5.19 for examples). This result is similar to the results of the affective dimensions investigation presented in Chapter 4.

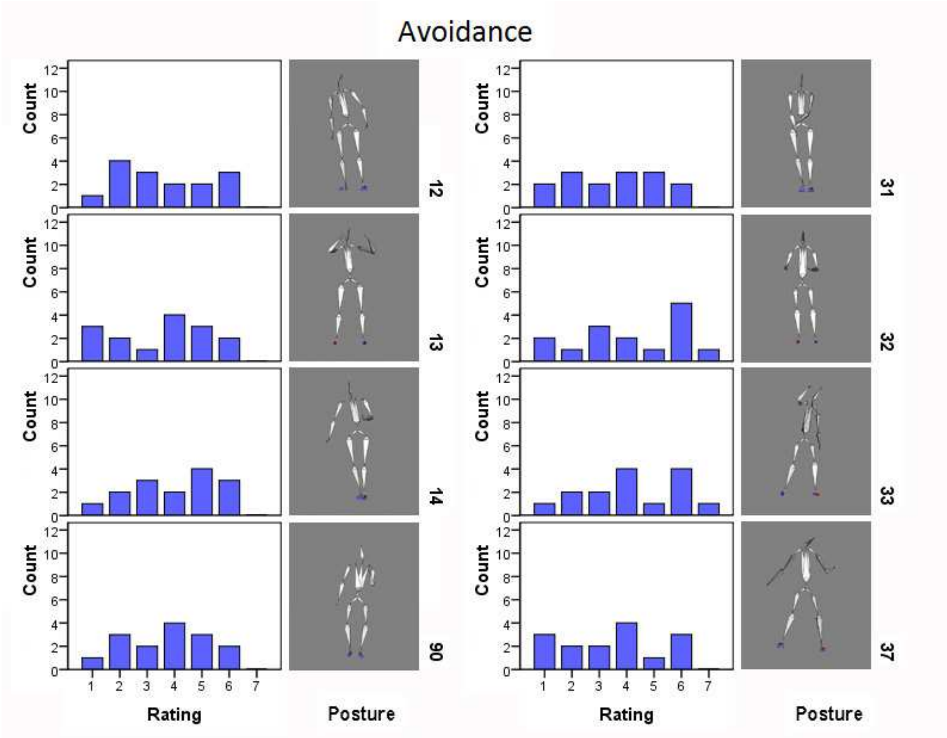


Figure 5.20: Examples of postures rated for the avoidance dimension. For each bar chart, the x-axis shows the rating scale (i.e., 1 to 7) and the y-axis shows the number of evaluations obtained for each rating in the scale. The number to the right of the posture image is used as a posture identifier to allow for easy location of the posture in Appendix F

In the case of the avoidance dimension, there were no clearly defined 'low' or 'high' avoidance postures as rated by the observers. Instead, for the majority of the postures, the observers' judgments were spread across the entire rating scale, as shown in Figure 5.20.

5.6.3 Creating Benchmarks

The purpose of this section is to define the benchmarks for human recognition of dimensions of affect from whole body postures. These benchmarks define the target recognition rates for the automatic recognition models discussed later in the chapter.

Observer agreement reliability results

To create a benchmark of human recognition of affective dimensions, the first step was to assess the consistency in how the observers rated the postures. Cronbach's α was computed for O for each affective dimension $d_g : g = 1, ..., 4$ separately. Recall from Chapter 3 that the higher the α, the more consistent the ratings. The affective dimension with the highest strength of agreement between the observers was arousal with an α of 0.816. The valence dimension was second with an α of 0.628; potency third (α = 0.575) and avoidance last (α = 0.423).

Given the low Cronbach's α levels of the potency and avoidance dimensions, it was decided to eliminate further examination of these two dimensions. Because not all observers evaluated all postures, the PCA conducted in Chapter 4 was not possible here.

Benchmarks

To set the benchmarks of human recognition of affective dimensions from non-acted posture, the same method used throughout this thesis was implemented, random repeated sub-sampling to create 10 trials, i.e., $s = 1, ..., 10$. For each trial, the set of observers O was split into three disjoint subsets $O_{s,1}$, $O_{s,2}$ and $O_{s,3}$ of five observers each. For each subset, a ground truth rating $gtm(p_i, O_{sk}$ was assigned to each posture p_i for each dimension d_g. The between observers agreement $B.Agr(d_g, O_{sk})$, defined in Equation (3.4), and Cronbach's α were computed between $O_{s,1}$ and $O_{s,2}$ for each of the 10 trials, for the valence and arousal dimensions separately. For each dimension $d_g : g = \{valence, arousal\}$, the benchmark is computed as the average between observers agreement $B.Agr(d_g, O_{sk})$ across the 10 trials. For valence the benchmark is 84.41%, SD = 1.14 and for arousal it is 87.37%, SD = 1.68. The results are listed in Table 5.10. The between observers agreement $B.Agr(d_g, O_{sk})$ ranges from 83% to 89%, which seems high compared to the equivalent testing in the affective state labels, Section 5.2.3. However, the between observers agreement $B.AgrLabel(O_{s,1}, O_{s,2})$ in the affective state labels testing was coded as binomial, whereas the between observers agreement $B.Agr(d_g, O_{sk})$ in the affective dimensions testing takes into account distances between judgment ratings.

Table 5.10: The benchmarks computed for the affective dimensions. The $B.Agr(d_g, O_{sk})$ are listed in column 3 and Cronbach's α levels are listed in column 4. The benchmarks are listed in the last column.

Human recognition benchmarks				
Affective dim	**Trial**	$B.Agr(d_g, O_{sk}) * 100$	**Cronbach's α**	**Benchmark**
Valence	1	84.93%	0.327	**84.41%**
	2	84.57%	0.399	
	3	84.22%	0.432	
	4	85.28%	0.501	
	5	86.17%	0.534	
	6	84.57%	0.437	
	7	84.57%	0.427	
	8	81.91%	0.453	
	9	84.57%	0.406	
	10	83.33%	0.330	
Arousal	1	89.18%	0.739	**87.37%**
	2	86.52%	0.702	
	3	87.59%	0.706	
	4	85.99%	0.696	
	5	89.36%	0.787	
	6	85.46%	0.658	
	7	89.18%	0.801	
	8	84.57%	0.598	
	9	87.77%	0.707	
	10	88.12%	0.778	

5.6.4 Discussion

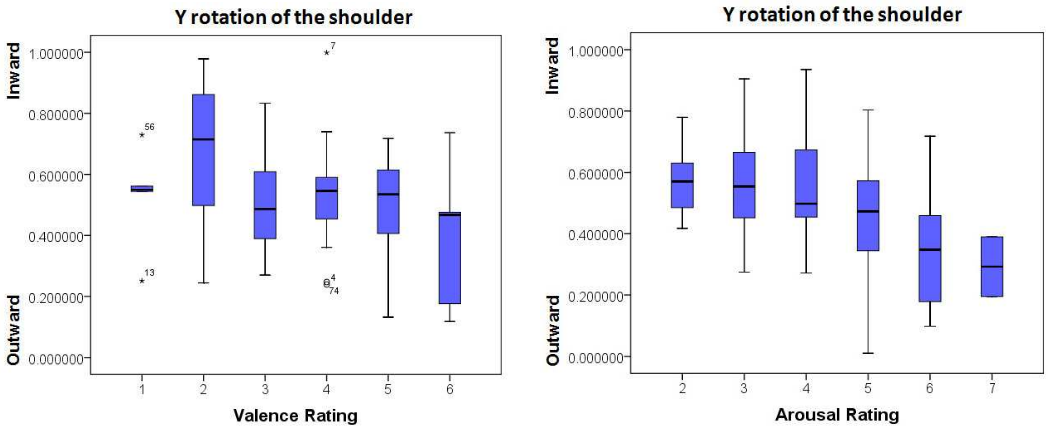
The overview of the data revealed that there was little consistency in how the set of observers O seemed to rate the set of postures P for the potency and avoidance dimensions, while the overview of the data was better for the valence and arousal dimensions. These are the two dimensions that are most typically examined. Reasons for the poor performance on the potency and avoidance dimensions could be due to the use of static information or the subtlety of many of the postures. Another reason could be that potency, i.e., the player's control over the situation, and avoidance, i.e., level of approach or withdrawal, are not apparent from the video game situation used. Finally, the poor consistency could be due to the experimental procedure used for the survey. Only very short descriptions of the meaning of each dimension were provided, thus it may be that the observers interpreted these descriptions differently. While the terms used to describe valence and arousal were quite standard, the terms used to describe potency and avoidance may not have been as easily understood.

5.7 Low-Level Posture Description Analysis

To assess the discriminative power of the low-level posture description for distinguishing between ratings of valence and arousal, the same set of features $F = \{f_{i1}, ..., f_{i41}\}$ described in Section 5.3 was used. Each low-level posture feature f_{iw} was subjected to one-way ANOVAs for each affective dimension d_g separately. The results are summarised in Tables 5.11 and 5.12. Listed in the first column of both Tables are the low-level features shown to be important for discriminating between levels of the dimensions (i.e., 1 to 7) with the significance level shown in the last column. The means for each affective dimension d_g rating c are shown in the middle columns. Bonferroni post hoc comparisons were implemented here as opposed to Tamhane's T2, due to equal variances (verified using Levene's test of homogeneity of variance). The superscript letter pairs listed with the means denote significant differences between those dimension level pairs according to the Bonferroni post hoc comparisons. Reported in this section is a discussion on the most interesting results of each affective dimension d_g, illustrated with a boxplot and corresponding avatar examples. The boxplots can be seen in Figure 5.21.

5.7.1 Valence

The results for the valence dimension are listed in Table 5.11. Significant differences were obtained for six of the 41 low-level posture features. The boxplot in Figure 5.21(a) shows the results for the y rotation of the shoulder. For this feature, the significant differences



(a) Valence

(b) Arousal

Figure 5.21: Examples of low-level posture description features with significant differences between scale ratings for each affective dimension. (a) Valence: y rotation of the left shoulder; (b) Arousal: y rotation of the right shoulder

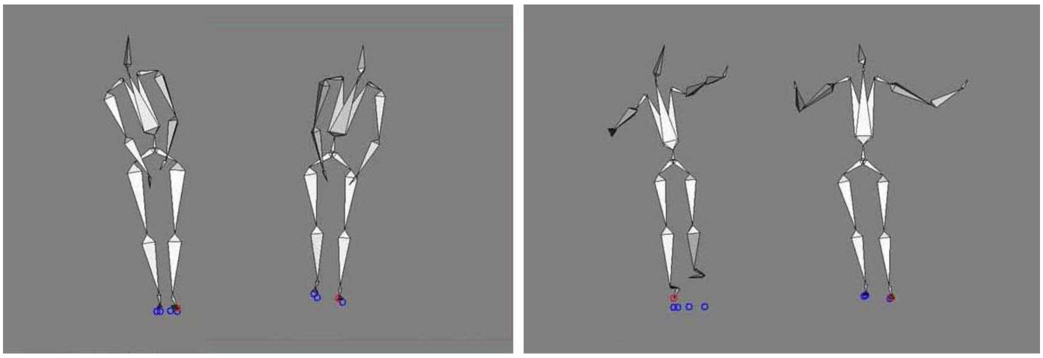
occurred for scale rating 2 against rating 6. The shoulders of postures rated 2 tend to be rotated inward, toward the body, whereas the shoulders of postures rated 6 appear rotated outward. Examples can be seen in Figure 5.22.

Table 5.11: The low-level posture description features which reached significance between the ratings for valence (one-way ANOVA with df = 5 (ratings). There were no ratings of 7.). For each feature f_{iw}, *a-e* pairs demonstrate the significant differences between means according to Bonferroni *post-hoc* comparisons. Rot. = rotation, L. = left, R. = right

	Means for valence ratings						
Low-level feature	**1**	**2**	**3**	**4**	**5**	**6**	p
y rot. L. collar	.55	.47	.57^a	.52	.45^a	.55	.008
y rot. R. collar	.45	.53	.43^a	.48	.55^a	.45	.008
z rot. L. shoulder	.49	.59^a	.51	.54	.52	.41^a	.041
y rot. L. shoulder	.53	.68^a	.51	.53	.51	.39^a	.051
x rot. L. wrist	.63	.50	.51	.58	.62	.71	.052

5.7.2 Arousal

The results for the arousal dimension are listed in Table 5.12. Significant differences were obtained for seven of the 41 low-level posture features. The boxplot in Figure 5.21(b) shows



(a) Rating 2: Postures 26 and 76

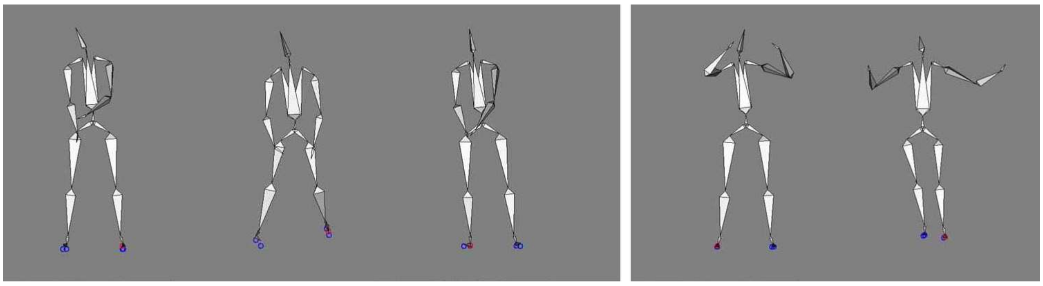
(b) Rating 6: Postures 68 and 70

Figure 5.22: Avatar examples representing postures for the significant differences of the y rotation of the shoulder for valence. (a) Representing rating 2 are postures 26 and 76; (b) Representing rating 6 are postures 68 and 70. Recall from Section 5.6.2 that the numbers refer to the posture position in the bar charts located in Appendix F for easy cross-referencing

the results for the y rotation of the shoulder. For this feature, the significant differences occurred for ratings 2 through 4 against rating 7. According to the boxplot, the shoulders for postures with ratings 2 through 4 generally are only very slightly turned in and the rating 7 postures are turned outward slightly. While only two postures achieved a rating of 7, in both cases the shoulders are indeed turned outward. Examples can be seen in Figure 5.23.

Table 5.12: The low-level posture description features which reached significance between the ratings for arousal (one-way ANOVA with df = 5 (ratings). There were no ratings of 1.). For each feature f_{iw}, *a-e* pairs demonstrate the significant differences between means according to Bonferroni *post-hoc* comparisons. Rot. = rotation, L. = left, R. = right

	Means for arousal ratings						
Low-level feature	**2**	**3**	**4**	**5**	**6**	**7**	p
y rot. L. shoulder	$.61^{a}$	$.64^{bc}$	$.57^{d}$	.51	$.42^{b}$	$.18^{acd}$	.000
z rot. R. shoulder	.68	.68	.84	.81	.84	.86	.052
y rot. R. shoulder	$.57^{a}$	$.55^{b}$	$.56^{c}$	.45	$.34^{abc}$	.29	.001
y rot. R. elbow	.09	.09	.08	.19	.30	.60	.009
x rot. R. wrist	.45	.43	.57	.51	.56	.68	.036



(a) Ratings 2-4: Postures 6, 53, 24

(b) Rating 7: Postures 13 and 70

Figure 5.23: Avatar examples representing postures for the significant differences of the y rotation of the shoulder for arousal. (a) Representing ratings 2 to 4 are postures 6, 53, and 24; (b) Representing rating 7 are postures 13 and 70. Recall from Section 5.6.2 that the numbers refer to the posture position in the bar charts located in Appendix F for easy cross-referencing

5.8 Automatic Recognition of Affective Dimensions from Posture

Automatic recognition models were built and tested for the valence and arousal dimensions separately. Testing was conducted to assess the models' ability to generalise to both new observers and new postures. To build the automatic recognition models, each posture p_i was associated with the median as the ground truth rating $gtm(p_i, O)$ for each affective dimension d_g, and a vector of 41 low-level posture description features $F = \{f_1, ..., f_{41}\}$ listed in Table 5.4.

The topology of the MLP used in this study consists of three layers: one input, one hidden and one output. The number of nodes in the input layer corresponds to the number of features used. The number of nodes in the hidden layer corresponds to the number of input nodes divided by two. The output layer contains one node. 10,000 epochs were used to train the recognition models.

5.8.1 Generalising to New Observers

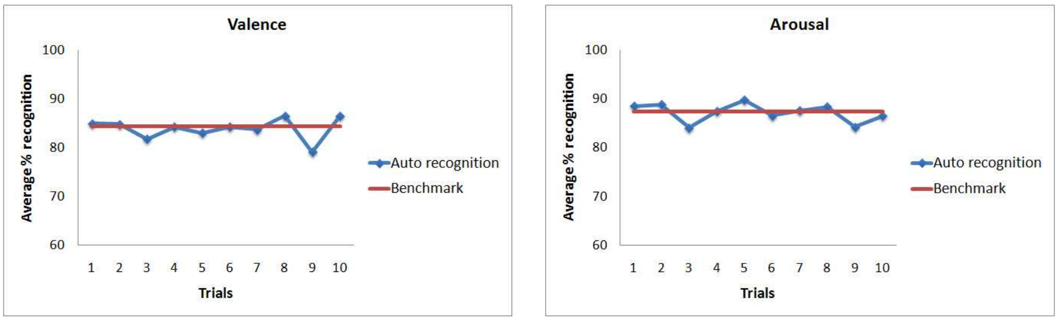
To test the ability to generalise to new observers, automatic models were trained with the previously unused observer subset three $O_{s,3}$ and tested with $O_{s,1}$ for each of the 10 trials. The results are summarised in Table 5.13. The average recognition model results across the 10 trials for valence was 83.86% (SD = 2.22%) and for arousal was 87.15% (SD = 1.88%).

Table 5.13: The performance rates for the automatic recognition models that generalise to new observers for the valence and arousal dimensions.

Generalising to new observers				
Affective dim.	**Trial**	**Auto rec.**	**Average**	**SD**
Valence	1	84.89%	**83.86%**	**2.22%**
	2	84.74%		
	3	81.73%		
	4	84.21%		
	5	82.98%		
	6	84.2 %		
	7	83.69%		
	8	86.52%		
	9	79.08%		
	10	86.51%		
Arousal	1	88.46%	**87.15%**	**1.88%**
	2	88.81%		
	3	84.04%		
	4	87.41%		
	5	89.72%		
	6	86.52%		
	7	87.59%		
	8	88.3 %		
	9	84.22%		
	10	86.46%		

5.8.2 Evaluation and Discussion

The goal of the automatic models for recognising levels of affective dimensions models from posture is to achieve performance rates comparable to the benchmarks that were set by the human observers. The results are illustrated in Figure 5.24. For both dimensions, while not all of the automatic models' recognition rates surpass the benchmark, they achieve a similar percentage of recognition. While this suggests that the goal has almost been achieved, use of the median resulted in very few postures at the extreme ends of the scale, making it difficult for the automatic recognition model to build rules for classifying those postures.



(a) (b)

Figure 5.24: Observer generalisation results for (a) valence and (b) arousal, and the observer benchmarks for each

In an examination of how specific postures were classified by the recognition model, the results were compared with the results of the affective state classification results presented in the first half of this chapter. The reasoning was that several research studies that examined affective dimensions classified affective information according to affective state labels and investigated how it fit into a 2D affective space [PPS01][Bre03][ZTP$^+$05]. Thus for each dimension d_g, each posture p_i was associated with the rating predicted by the automatic recognition model as well as the ground truth label $gtl(p_i, O)$ from the first half of the chapter. The set of postures P was then ordered according to affective state label l_j. The results revealed that all of the frustrated and triumphant postures were recognised as high arousal, as expected according to Russell's [Rus80] circumplex model shown in Figure 2.4 (Chapter 2). Furthermore, the majority of the triumphant postures were recognised as high valence. Examples of high arousal/high valence postures can be seen in Figure 5.25). Ideally, as frustrated is a negative state, the frustrated postures should be recognised as low valence. However, the results were split between low and high valence.

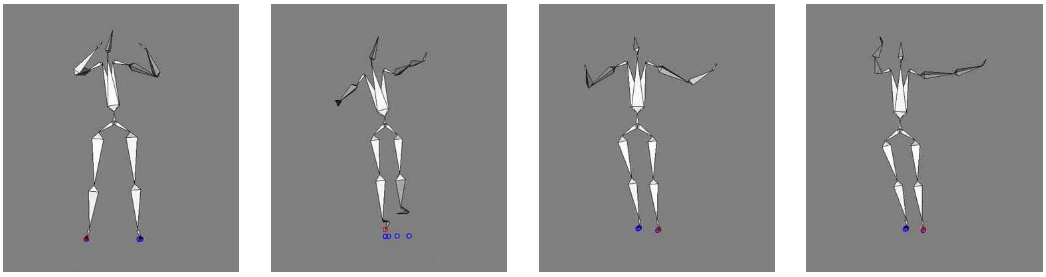


Figure 5.25: Examples of high arousal, high valence postures predicted by the recognition model

5.8.3 Generalising to New Postures

Automatic recognition models were also tested for their ability to generalise to novel postures. Each posture p_i was associated with its corresponding feature vector F, and the median rating was assigned as the ground truth $gtm(p_i, O)$. The models were built using an MLP with 10 fold cross-validation. The percentage of recognition achieved for the valence dimension was 84.2% (SD = 12.8%) and for the arousal dimension it was 82.9% (SD = 14.1%).

5.8.4 Evaluation and Discussion

In generalising to novel postures, there was little difference in performance between the two dimensions. The same result was found in the acted posture generalisation examination presented in Chapter 4, Section 4.8.3. However, the difference is that in the non-acted posture generalisation examination presented here, the performance for valence was slightly better than arousal. In Chapter 4 it was postulated that arousal may be easier to recognise than valence using acted expressions as they tend to be more exaggerated and the difference between activated (i.e., high arousal) and less activated (i.e., low arousal) types of emotions expressed through posture seems vast. Could the slight difference between the two studies presented in this thesis, i.e., valence and arousal recognised from body posture, be due to using acted postures in Chapter 4 and non-acted naturalistic (i.e., more subtle) postures in this chapter? Indeed, a study by Amir et al [AWH09] which obtained naturalistic, non-acted emotion expressions found that arousal was more difficult to judge than valence from a combination of affective speech and facial expression data.

5.9 Chapter Summary

This chapter presented experiments designed to investigate the recognition of subtle affective states and affective dimensions from non-acted body posture information. Similar to the acted postures study presented in Chapter 4, it was hypothesised that human observers could achieve above chance agreement levels (both within and between sets of observers O) when attributing subtle affective states and levels of affective dimensions to static images of a faceless humanoid avatar. It was also hypothesised that automatic recognition models could be grounded on a set of low-level posture information and achieve recognition rates similar to benchmarks computed from human observers.

Posture data was collected using a Gypsy motion capture suit of participants playing sports video games with the Nintendo Wii$^{\text{TM}}$. Because the players were unaware of the true purpose of the study, it is believed that their bodily expressions of affect were non-acted and unsolicited. After locating the affectively expressive sections of the motion capture data, a different set of participants, the observers, was asked to judge static posture images taken from the motion capture data by associating one of four affective state labels (*concentrating, defeated, frustrated* and *triumphant*) and levels of four affective dimensions (*valence, arousal, potency* and *avoidance*) to each posture.

To analyse human agreement rates, within observers' agreement and inter-observer agreement reliability were ascertained for the observers' judgments first. The results were found to be above chance level for the set of affective state labels L as hypothesised. For the affective dimensions, an overview of the survey data and the Cronbach's α scores computed within observers indicated good consistency between observers' judgments for the valence and arousal dimensions but little consistency between observers' judgments for the for the potency and avoidance dimensions. Based on these results, these two dimensions were eliminated from further analysis. Next, a set of benchmarks was defined from the observers' judgments by computing the average of the between observers agreements $B.AgrLabel(O_{s,1}, O_{s,2})$ for the set of affective labels L and the between observers agreements $B.Agr(d_g, O_{s,1}, O_{s,2})$ for the valence and arousal dimensions.

The next section explained the low-level information used to describe the postural displays after which an anlysis of the low-level posture description was carried out to assess how disciminative each feature f_w was for distinguishing between affective states and affective dimensions. The results showed significant differences between low-level upper body features for both the set of affective state labels L and the valence and arousal dimensions. For the affective states, the most significant differences typically occurred between the less active states (*concentrating* and *defeated*) and the more active states (*frustrated* and *triumphant*). Associated postures reflected these differences. For the valence and arousal affective dimensions, the significant differences between scale ratings were less distinct. For both dimensions, few postures were judged at either extreme of the scale.

In the second half of the analysis for both the affective state labels and the affective dimensions, automatic recognition models were built and tested to determine if performance rates similar to the benchmarks could be obtained (observer generalisation) using repeated sub-sampling validation. The results showed that the recognition rates for many of the trials were equal to or better than the benchmarks. In the case of the affective state labels, some of the classification problems that occurred were attributed to the very low number of postures for the frustrated category, meaning that the recognition model was unable to build classification rules for that category. An examination of specific recognition model misclassifications indicated some difficulty for the recognition model to distinguish between frustrated and triumphant. In the case of the affective dimensions, the recognition rates for the observer generalisation models performed similarly to the benchmarks, with the average across the recognition models for each dimension d_g less than 1% lower than the benchmarks.

Automatic recognition models were also tested for their ability to generalise to novel postures (using 10 fold cross-validation). In the case of the affective state labels, the recognition rates were found to be similar to recognition rates discussed in Chapter 2 for automatic recognition systems built with acted postures and considering basic emotion categories over the more difficult to recognise non-basic affective states. An investigation of the recognition rates for each affective state label l_j showed a confusion for the recognition model between concentrating and defeated postures. Furthermore, several postures from all labels were misclassified by the recognition model as concentrating. This was not unexpected considering the disproportionate number of concentrating postures over the number of postures for the other affective labels. In the case of the affective dimensions, the recognition rates for both valence and arousal were well above chance level with the valence model performing slightly better than the arousal model. The high recognition rate seems surprising given the subtle non-acted postures. This could be due to the use of the median as the ground truth label which meant that very few postures were rated at the extreme ends of the scale.

The next chapter outlines the third and final study in the incremental approach. This study aims to to evaluate how an affective posture recognition system performs when applied to *sequences* of non-acted static postures as if in a runtime situation.

Chapter 6

Case Study 3: Real Time Affective Posture Recognition

As the third step in the incremental approach, the aim of the final study presented in this chapter is to evaluate how the affective posture recognition system performs when applied to *sequences* of non-acted static postures as if in a runtime situation. In this study, the affective postures have not been manually extracted, as opposed to the previous two studies which examined single, apex instant postures that were explicitly, manually extracted. Another difference is that sequences of postures are analysed in this study. As this study builds on the non-acted postures study presented in Chapter 5, a video game scenario was selected in which the participants played tennis with the Nintendo Wii. Wii tennis was chosen based on informal discussions with the players from the non-acted study presented in Chapter 5. When asked to rank the Wii sports games according to enjoyment, the majority of the players ranked tennis as the most enjoyable.

The chapter is organised in the following manner. The affective posture recognition system is described in Section 6.1. Sections 6.2 to 6.5 explain the approach taken to test the system on a set of posture sequences. Section 6.2 explains the posture corpora collection which consists of a similar method described in Chapter 5. Section 6.3 discusses the posture judgment surveys which were conducted to build a training set and a testing set for investigative purposes. The low-level posture description is explained in Section 6.4. In Section 6.5, the system is tested and the results are reported and evaluated. A discussion is provided in Section 6.6 and the chapter ends with a summary in Section 6.7.

6.1 The Affective Posture Recognition System

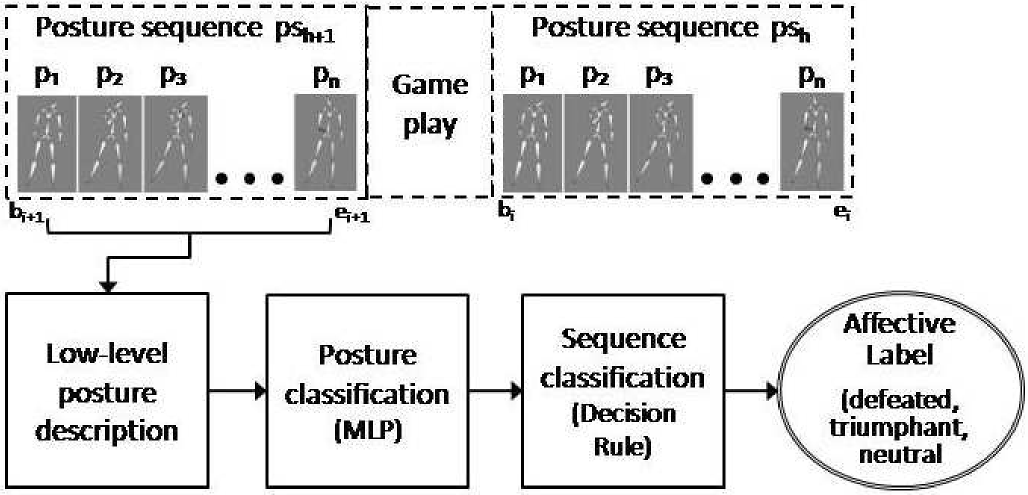

Posture sequence ps_{h+1}
p1 p2 p3 pn
Game play
Posture sequence ps_h
p1 p2 p3 pn
b_{i+1}
e_{i+1}
b_i
e_i
Low-level posture description
Posture classification (MLP)
Sequence classification (Decision Rule)
Affective Label
(defeated, triumphant, neutral

Figure 6.1: The affective posture recognition system. A vector of low-level posture features F_i is computed for each posture p_i in a posture sequence ps_h. The posture description F_i of each posture p_i is then sent to the MLP. A decision rule is applied to an entire sequence of the MLP output, after which the affective state label l_j for the posture sequence ps_h is determined

The affective posture recognition system is implemented as a combination of an MLP and a decision rule, similar to the approach used by Ashraf et al [ALC$^+$09] to automatically recognise pain from facial expressions. Refer to Figure 6.1 to see the flow of the system (which has already been trained with a set of postures P). A vector of low-level posture description features F_i is computed for each posture p_i in a posture sequence ps_h. A posture sequence corresponds to a replay window as defined in Chapter 5, Section 5.1.3. The posture description F_i of each posture p_i is then presented to the trained MLP and each individual posture p_i within the sequence is evaluated. Given a posture sequence $ps_h : h = 1, ..., g$, the output of the MLP for a posture p_i is a probability distribution for the set of labels $L = \{defeated, triumphant, neutral\}$ for each posture p_i of ps_h. For each posture sequence ps_h, cumulative scores $L_{sum}(l_j) : j = 1, ..., 3$ are computed as

$$L_{sum}(l_j) = \frac{1}{n} \sum_{i=1}^{n} q_{ij} \tag{6.1}$$

where q_{ij} is the output score for posture p_i and label l_j (i.e., the probability that a label l_j is used to label p_i of a sequence ps_h) and n is the number of postures with that label.

Next, a decision rule is applied to the normalised cumulative scores $L_{sum}(l_j)$ of the posture sequence ps_h. The decision rule is employed as a way to group the postures in each sequence and assign a label l_j to an entire sequence instead of individual postures only (as is the case with the MLP). The assumption is that only one affective state is expressed in each posture sequence ps_h, i.e., replay window. The decision rule is defined as follows

$$\begin{aligned}
&if(L_{sum}(\mathit{defeated}) < \mathit{threshold} \;\; \&\& \;\; L_{sum}(\mathit{triumphant}) < \mathit{threshold})\\
&\qquad \{\mathit{sequence\ label} = \mathit{neutral}\}\\
&else\ if(L_{sum}(\mathit{defeated}) > L_{sum}(\mathit{triumphant}))\\
&\qquad \{\mathit{sequence\ label} = \mathit{defeated}\}\\
&else\ if(L_{sum}(\mathit{defeated}) < L_{sum}(\mathit{triumphant}))\\
&\qquad \{\mathit{sequence\ label} = \mathit{triumphant}\}
\end{aligned} \tag{6.2}$$

where $L_{sum}(defeated)$ and $L_{sum}(triumphant)$ are the normalised cumulative scores for the defeated and triumphant affective state labels for the posture sequence ps_h. The *threshold* was experimentally defined by building ROC curves using the normalised cumulative scores for $L_{sum}(defeated)$ and $L_{sum}(triumphant)$ after which the coordinates of the curves were assessed. The point at which the true positive rate and the false positive rate were close to equal for the *defeated* ROC curve was chosen as the *threshold* [ALC$^+$09]. The affective state label l_j of the posture sequence ps_h is the output of the decision rule.

6.2 Posture Corpora

6.2.1 Motion capture data collection

As with the previous two studies, motion capture data of whole body posture and movement information was collected as the first step. The motion capture collection process was similar to the one presented in Chapter 5. The same Gypsy 5 motion capture system was used. After reading the information sheet and signing a consent form, a player-specific calibration model was created using the process reported in Chapter 5, Section 5.1.

In addition to having their body motions recorded with a motion capture suit, the players were also videotaped. The purpose of videotaping the sessions was so that the replay windows (described in Chapter 5) could be more easily located during analysis. This step can be easily automatised when the game is connected to the recognition system. The camera was placed at the back of the room and slightly to the side, allowing the video game and the player's upper body (from behind) to be recorded. The camera placement behind the players aimed to maintain as natural a setting as possible.

6.2.2 Players

Ten players were recruited for participation (three females) ranging in age from approximately 20 to 40. All players had little to no experience playing with the Nintendo Wii as experienced players have been shown to be less expressive with their body postures when they play to win [BBKP07][PBBvDN09]. The players were asked to play Wii tennis for at least 20 minutes and have their body motions recorded while wearing the Gypsy 5 motion capture suit. Again, as with the previous non-acted study, the players were unaware of the purpose of the study to ensure that any affective displays would be spontaneous and non-acted. In the previous study, as the experimenter, I communicated with the players in an aim to encourage a more natural atmosphere. In this study players were asked to come with a friend with whom they could interact during game play as it has been shown to increase affective output [RST^{+}05].

6.2.3 Replay window identification

After collecting the motion capture information, the files were annotated by locating the replay windows of each gaming session. Like the study presented in Chapter 5, the replay windows, considered the relevant information for this research, are the periods in the video game after which a point has been won or lost. The posture sequences are taken from these replay windows. The number of replay windows (i.e., posture sequences) can vary for each gaming session. Furthermore, the length of the posture sequence may also vary, meaning that the number of postures within a sequence is not fixed. The replay windows were manually located by viewing the video and motion capture data simultaneously. The reason for manually identifying the replay windows was that if the automatic recognition system being proposed here was integrated into an existing software application, the application itself would be able to signal the periods under investigation. The output of the identification process was a contiguous section of motion capture data.

6.3 Posture Judgment Surveys

Two separate posture judgment surveys were conducted in order to build separate training and testing sets for use in testing the affective posture recognition system. A separate group of observers was recruited for each survey.

6.3.1 Building the Training Set

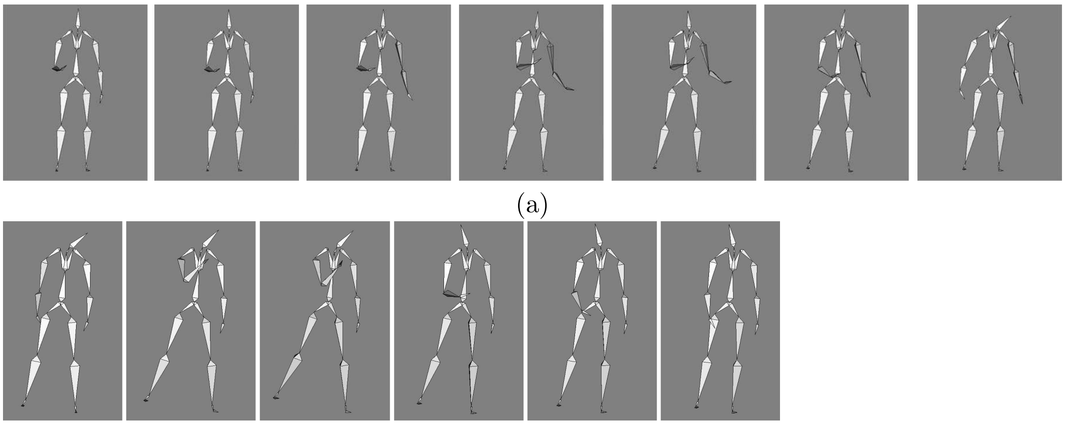
A new training set was built for this study because, although this study extends the previous non-acted study, the set of labels L has changed by removing *concentrating* and *frustrated* and adding *neutral.* Furthermore, the previous study focused on apex instants only and this study investigates the recognition of posture sequences which may include a wider variety of subtle, natural expressions. Therefore, additional non-apex postures were sought.

A manual extraction of the training set postures was used in order to ensure that a variety of posture configurations were considered. To build the training set, three postures were taken for each of 20 replay windows, yielding 60 postures. Three postures spanning an entire replay window were used to represent the entire movement for training the automatic recognition model considering that most of the movements were short in duration and/or contained little movement. The three postures were chosen as: i) a posture at the start of the replay window (as soon as game play stopped when a game point was won or lost); ii) a posture in the middle of the movement itself; iii) the apex of the movement.

A set of eight observers (i.e., $|O_{training}| = 8$) was recruited to judge the set of postures (i.e., $|P| = 60$) online, using the same procedure as the posture judgment surveys (for the discrete affective state categories) presented in Chapters 4 and 5, respectively. As with the previous studies, the posture order was randomised for each observer $o_k : k = 1, ..., 8$, who was asked to associate an affective state label $l_j : j = 1, ..., 3$ to each posture p_i one page at a time. For each posture p_i, the label l_j with the highest frequency of use (as defined in Chapter 3, Equation (3.1)) was determined to be the ground truth label $gtl(p_i, O_{training})$, yielding 10 *defeated*, 17 *triumphant* and 33 *neutral* postures. To create a more balanced training set, the *defeated* (22) and *triumphant* (16) postures from the Chapter 5 study were added to create a final training set of 98 postures: 32 *defeated*, 33 *triumphant* and 33 *neutral.*

6.3.2 Building the Testing Set

To build a testing set for automatic recognition, posture sequences were automatically extracted from the remaining replay windows (i.e., those not used to build the training set). Due to the high capture rate of the motion capture system (120 frames per second), the configuration of the body did not change significantly from one motion capture frame to the next. Therefore, it was not necessary to extract every posture within a replay window. Instead, every 40th posture was automatically extracted which allowed for a variety of postures that represented the entire movement within a sequence. Due to the differences in replay window length, the automatic extraction yielded posture sequences ranging from two to 40 frames in length. 836 posture frames across 75 posture sequences were extracted. Two posture sequence examples are shown in Figure 6.2.



(a)

(b)

Figure 6.2: Posture sequence examples - posture frames automatically extracted from two different replay windows of motion capture files

A set of five observers (three females) (i.e., $|O_{testing}| = 5$) was recruited to view these posture sequences and assign a single label l_j to each entire sequence ps_h, as opposed to labelling each individual posture p_i within a sequence. Each sequence was viewed as an animated clip of a simplistic humanoid avatar. This approach was adopted in order to determine the overall affective state of the player across the entire sequence as opposed to each particular posture instant.

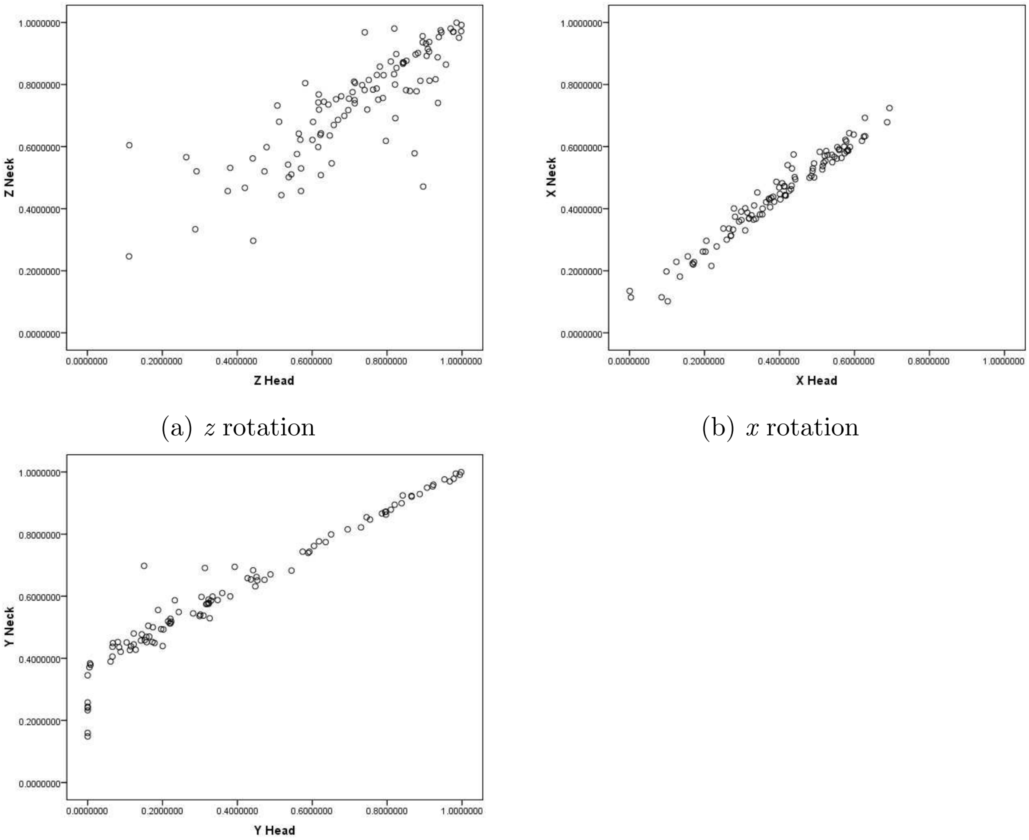
The posture sequences and the evaluation directions were emailed to the observers who agreed to take part. The entire task took approximately 1.5 hours to complete. The observers were instructed to take a break every 30 minutes in an effort to control for boredom. The set of posture sequences was randomly divided into four subsets, different for each observer.

As with the training set, to determine the ground truth $gtl(ps_h, O_{testing})$ of a sequence ps_h in the testing set, the label l_j with the highest frequency of use $freq(ps_h, l_j)$ was chosen for each posture sequence ps_h. The within observers agreement $W.AgrLabel(L, O_{testing})$ (as defined in Chapter 3, Equation (3.3)) was 66.67% with Fleiss' kappa reaching 0.162, indicating slight agreement. Possible reasons for the low consistency between the observers' judgments may be due to the subtlety of the posture sequences, the limited set of labels L from which to choose or the small set of observers $O_{testing}$ considered. There were 14 posture sequences ps_h with the defeated ground truth label, 8 triumphant, 39 neutral and 14 ties. As outlined in Chapter 3, in the case of ties, the ground truth $gtl(ps_h, O_{testing})$ was randomly selected between the tied labels.

6.4 Low-Level Posture Description

The low-level posture description method used was the same as the one presented in Chapter 5. Unfortunately however, the three head features had to be removed due to a hardware malfunction in which the head rod repeatedly became fixed in one position. The considerable use of the motion capture system from one study to the next may have caused excessive wear and tear on the hardware. An analysis was conducted to confirm that the neck features could account for the information provided by the head features. The Pearson correlation coefficient r results for each corresponding pair of head and neck features from previous accurate data showed high correlation ($df = 101, p < .01$); $r_z = .82$, $r_x = .98$, $r_y = .97$. The results are illustrated in the scatterplots of Figure 6.3.

A vector of low-level posture description features $F_i = \{f_{i1}, ..., f_{i38}\}$ was computed for



(a) z rotation

(b) x rotation

(c) y rotation

Figure 6.3: The Pearson correlation coefficient scatterplots of the head and neck features; (a) z rotation; (b) x rotation; (c) y rotation

each posture p_i of the training set and each posture p_i in each sequence ps_h of the testing set. The features are listed in Table 5.4 and include all but the three head features as explained in the previous paragraph.

6.5 Testing Results

The affective posture recognition system was trained with the training set of 98 postures described in Section 6.3.1 and tested with the testing set of 75 posture sequences described in Section 6.3.2. The percentage of correct automatic recognition achieved for the 75 posture sequences was 52%. It is noted that, as stated in Section 6.3, there were 14 posture sequences that were assigned two affective state labels with equal frequencies by the set of observers $O_{testing}$, thus the ground truth label $gtl(ps_h, O_{testing})$ was randomly determined between those two labels. Taking this into account and assessing correct recognition according to either of the two labels, the recognition rate increased to 57.33% which is still lower than the human observer overall agreement of 66.67%. However, it is well above chance level (33.33% considering three labels) and similar to the automatic recognition rates of the affective label testing on non-acted postures presented in Chapter 5.

To investigate the possible factors that may have contributed to the lower than anticipated system performance rate, a confidence rating task was carried out. A new observer o_6 was recruited to view the seven posture sequences for which the set of observers $O_{testing}$ had assigned two affective state labels with equal frequencies, and were also misclassified by the system. The task was repeated three times over three days (i.e., three trials). For each trial the seven posture sequence clips were presented in a randomised order with a different label l_j from the set of labels L. The observer o_6 was asked to view the animated posture sequences and provide a confidence rating on a scale from 1 (not confident) to 5 (very confident) that the label l_j corresponds to the expression portrayed by the sequence [AR09]. In the event of posture sequences with low confidence ratings (i.e., 1 or 2), o_6 was asked to provide an alternative label l_t from the remaining two. The results are listed in Table 6.1.

Looking at the Table and comparing it with the judgments from the group of observers and the system results, a few issues and limitations are highlighted. One of the most obvious findings is that some posture sequences seem to be ambiguous, partially due to the subtlety of natural, non-acted postures. For instance, ps_4 and ps_5 received confidence ratings of 3 or above no matter which label was presented with the sequence. This result may mean that no amount of system modifications or additional observer judgments may resolve the labelling issue for the sequences. It is possible that these sequences were not affective, but also did not fit into the *neutral* category. It is also possible that the affective state expressed was not in the set of affective state labels L considered.

Another issue highlighted by the confidence rating task may be the small observer sample size. The addition of the three confidence ratings would have changed the ground truth label $gtl(ps_h, O_{testing})$ of some posture sequences. For example, the label ties for ps_6 and ps_7 in particular could possibly have been resolved.

A third issue involves the training set of postures. It may not be well separated enough due to low agreement amongst the observers for some of the postures. Having a better defined, separable training set could help to increase the system's recognition rate. There are several possible ways to handle the training set issue. First and most simply, the low agreement postures could be removed. However, this could create an unbalanced training set, depending on the number of postures removed from each category. Therefore, additional high observer agreement postures would have to be found. Second, additional information for the training set postures could be gathered at the posture judgment level. For instance, in addition to choosing an affective state label l_j, the set of observers O could be asked to provide a confidence rating, an intensity level or a label ranking for each posture p_i. This information could be used to bias the computation of the most frequent labels by using the information as way to weight the evaluation of each observer. Third, expert coders could be enlisted to label the training set instead of non-expert coders. However, the goal of recruiting non-expert coders was to be able to create a system that acts like a lay person in order to simulate human-human interaction situations. Such an approach is useful in situations where the aim is not about creating technology to improve peoples' ability in recognising the affective state of their interlocutor or to substitute affect reader experts. Rather, the aim of this study was to create technology that can play the role of a generic companion such as a partner or an adversary in a computer game.

6.6 Discussion

Inter-observer agreement reliability on labelling the testing set of posture sequences was calculated using Fleiss' kappa. The results (0.162) indicated only slight agreement between the observers' judgments. While the result is clearly lower than desired, it is not unexpected given the use of naturalistic data [AR09]. *"Even for a human expert, it is difficult to define what constitutes an emotion"* [AR09]. The argument could be made that inter-observer agreement reliability may increase considerably if a larger set of observers O was enlisted. However, this is not necessarily the case. In the study by Afzal and Robinson [AR09], they recruited 108 non-expert coders to judge non-basic affective states from naturalistic expressions. The reliability rating was quite low at 0.2, indicating, like the study presented in this chapter, only slight agreement.

The affective posture recognition system achieved a recognition rate of 57.33%. While the recognition rate is well above chance level (33.33% considering three affective states) and similar to the automatic recognition results obtained in the affective category testing of Chapter 5, it is still lower than the human observer agreement target of 66.67%. However, the results are comparable to other automatic recognition systems of body movement or dance presented in Chapter 2, Table 2.10. For instance, the recognition system presented by Camurri et al [CMR$^+$04] reached a low recognition rate of 36.5% (chance level = 25%), considering acted dance movements of basic emotions. Indeed, the system performance was considerably lower than the 56% agreement between the observers. Similar to Camurri et al's study, the study by Kapur et al [KKVB$^+$05] also considers acted dance movements of basic emotions. However, using several different automatic classifiers, they achieved higher recognition rates (62%-93%), but most still below the observers' level of agreement (93%). The last bodily expression recognition system to discuss is that of Bernhardt and Robinson [BR07] which focused on the recognition of acted knocking expressions of basic emotions. The results achieved by the system presented in this thesis are comparable to Bernhardt and Robinson's biased system (i.e., personal idiosyncrasies not removed) 50%.

The approach taken in this thesis for building the affective posture recognition system was similar to that of Ashraf et al [ALC+09] in the use of an automatic modeling technique combined with a decision rule in order to determine an affective state l_j of a sequence of input (i.e., static postures in this thesis). The main disadvantage of the system is that the order of presentation of the individual postures in a sequence is not considered. Each posture p_i in a sequence ps_h is evaluated separately upon its presentation to the MLP. This issue could be addressed in future work by using a different classifier such as Hidden Markov Models that take into account temporal information.

6.7 Chapter Summary

The the study presented in this chapter provides an understanding of how to create affective recognition systems when the affective expressions are subtle and naturalistic. This chapter built on the non-acted postures study of Chapter 5 by examining a proof of concept for the automatic recognition of sequences of static postures that were not manually extracted. Similar to the previous two studies presented in Chapters 4 and 5, it was hypothesised that i) human observers could reach above chance agreement on the sequences and ii) that an automatic recognition system could achieve accuracy rates similar to the human observers.

An affective posture recognition system was built using a combination of an MLP and a decision rule defined in this research. The rationale is that the decision rule allows for the affective state label to be determined according to an entire posture sequence instead of to single static postures, as the MLP alone does.

The first step was to collect posture data in order to build training and testing sets for investigating the performance of the affective posture recognition system. Posture data was collected using a Gypsy motion capture suit of participants playing tennis with the Nintendo Wii$^{\text{TM}}$. The participants played the video game in pairs with a friend in an attempt to create more genuine affective expressions [LLCBB08]. A training set of posture data was built using a combination of postures from the motion capture session described in this chapter and the session presented in Chapter 5. A testing set of posture sequence data was determined by automatically extracting sequences of postures that occurred during the replay windows. Human observers labelled each posture sequence at sequence level as opposed to frame level as other research has shown that reliable results could be obtained [ALC$^+$09]. As hypothesised, the target rate of above chance level agreement was obtained.

The performance of the affective posture recognition system was lower than the target within observers agreement but still better than chance level. An evaluation of the results highlighted some of the potential causes of misclassification, such as ambiguity of the posture sequences and low observer agreement on training set postures. Possible solutions were considered.

Table 6.1: The 7 posture sequences for which 2 affective state labels were assigned with equal frequencies by $O_{testing}$ and the confidence rating assigned to each l_j for each ps_h from a new observer o_6. When the confidence rating was below 3, the observer o_6 provided an alternative label l_t from the set of labels L. Column 2 lists the l_j determined by the system (i.e., MLP plus the decision rule). The remaining columns list the results of each trial, i.e., each confidence rating task, with the l_j to be evaluated first, the confidence rating second, and the new label l_t third, repeated for each of the 3 trials. Tri = triumphant; Neu = neutral; Def = defeated

		Trial 1			**Trial 2**			**Trial 3**		
Posture sequence	**System result**	**Label**	**Confidence rating**	**New label**	**Label**	**Confidence rating**	**New label**	**Label**	**Confidence rating**	**New label**
ps1	Tri	Neu	4		Def	4		Tri	2	Neu
ps2	Neu	Def	1	Neu	Tri	2	Neu	Neu	5	
ps3	Neu	Tri	5		Def	1	Tri	Neu	5	
ps4	Neu	Tri	5		Def	3		Neu	4	
ps5	Neu	Tri	4		Def	3		Neu	3	
ps6	Neu	Tri	3		Def	4		Neu	1	Def
ps7	Neu	Def	4		Tri	2	Neu	Neu	3	

Chapter 7

Conclusions

This thesis is fundamentally concerned with and contributes to the field of affective computing. In particular, the research presented investigated a low-level description of body posture, proposed a method for creating benchmarks for evaluating affective posture recognition models, and provided an understanding of how posture is used to communicate affect. Chapter 2 provided both a background to affect recognition and explained the areas that still lack sufficient research which helped define the research questions and contributions presented in Chapter 1. An approach was devised for achieving the contributions (Chapter 3) which was implemented through a series of three case studies. Each case study was designed to examine the recognition of affective postures in two directions: i) through human observers and ii) through automatic recognition models.

The first case study investigated the recognition of *acted* postures of *basic* emotions and affective dimensions (Chapter 4). The second case study aimed to address the lack of automatic recognition systems focused on naturalistic expressions. Therefore, the study investigated the recognition of *non-acted* postures of *non-basic* affective states and affective dimensions (Chapter 5). Expanding on non-acted affective posture recognition, the third case study was designed to evaluate automatic affect recognition from posture *sequences* to examine how effectively the system may perform in a runtime situation when the affective postures arrive continuously and have not been manually selected (Chapter 6).

The remainder of this chapter is organised as follows. Section 7.1 discusses details of the main contributions made by the research presented in this thesis. The main findings of each study are highlighted according to each contribution. Potential directions for future work to extend this thesis are presented in Section 7.2. Section 7.3 ends the chapter with a brief summary.

7.1 Contributions

The investigation of a low-level description of posture

The main contribution of this thesis was the investigation of a low-level description of the configuration of posture. The purpose was to examine if a low-level description could provide enough information for recognising affect from posture. While behavioural science studies have examined the role played by specific features of the body and whether they are predictive of specific affective states, as discussed in Chapter 2, there has been less work in the computing fields to create computational models of affective body posture. The work that does exist has focused mainly on high-level posture features [KPI04][KBP07] or body movement such as dance [CMR$^+$04][KKVB$^+$05] or specific actions [BR07].

The posture description considered in this thesis focused on a low-level, static configuration of the body. Findings of neuropsychological and neurophysiological research indicate that configuration information can be instrumental in the recognition of biological motion [HH06][ADGY07]. A major strength of the low-level posture description approach adopted in this thesis was that it is general, meaning that it was independent of affective state or situational context. Hence, the posture description could be easily adapted without having to modify the way the posture description is computed.

The low-level posture description was evaluated in two directions. First, in Chapters 4 and 5, a statistical examination of the importance played by each feature was carried out to understand the information in body posture that can be recognised by human observers. In both chapters, the results showed that specific features were predictive of specific affective state categories and levels of valence and arousal dimensions. In general, the posture configurations attributed to the discrete basic emotions were in keeping with the results found by Coulson [Cou04] and Wallbott [Wal98] in particular. For instance, sad postures were characterised by a head bent forward and arms hanging down at the side of the body. However, there were also some differences. In Chapter 4, observers from three different cultures were considered and differences were noted in terms of which features each culture attributed to the emotion categories studied. For instance, angry for the Japanese was characterised by a forward bending head. The result differs from the results for the Sri Lankan observers in which no distinct head position was found to specify angry. The result is also different from angry postures in Coulson's study which were characterised by a backward bending head.

The results of the low-level posture description examination for the non-basic affective states were more difficult to validate with results from other studies because the same non-basic affective states have not been studied in this capacity. Similarly, in the case of valence and arousal, while research has investigated dynamic characteristics of body movement [PPS01], the majority of the research examining specific features that can be attributed to displays of these two affective dimensions has been carried out for facial expressions and speech [CMK+06][SCDC+01].

The second direction in which the low-level posture description was examined was the automatic recognition model level in Chapters 4-6. In each case study, the low-level posture features were used to build automatic recognition models. The results showed that the low-level posture description approach could be used to discriminate between both acted, stereotypical posture expressions (Chapter 4) and more subtle, non-acted posture expressions (Chapters 5 and 6) to levels similar to target benchmarks set in this thesis using a rigorous method detailed in Chapter 3 and discussed in the following contribution. These results demonstrate that automatic affective posture recognition systems can be built that act, i.e., recognise affect from posture, as well as a general human companion would.

A method for creating benchmarks for evaluating affective posture recognition

The research in this thesis proposed a new approach for creating benchmarks to evaluate automatic recognition models. The view taken was that there was no inherent ground truth label that could be assigned to the affective expressions. Although some studies have used observers' judgments to compute the ground truth [KPI04][LNP02], they do not address the variability that exists between the observers and the fact that the observers do not represent the entire population. I proposed to build the ground truth by taking into account only the observers' evaluations and used a repeated sub-sampling method to build the benchmarks in an attempt to increase the reliability for the observer population considered.

The benchmarks for evaluating automatic recognition models in affective computing research have typically been based on the actors' labels. A typical approach is to compare the agreement between the observers (i.e., how well the observers agreed with the actors) and the performance of the automatic recognition models [KKVB$^+$05][CMR$^+$04]. This approach was considered to be quite limited as the number of observers recruited is generally low due to the difficulty of gathering a large number of evaluations on the affective expressions over a large number of observers. Furthermore, most of the observers recruited are often from a narrow population. Thus, an approach was devised to account for this type of situation; a random repeated sub-sampling method was implemented to create the benchmarks. The purpose behind this method was to obtain performance rates that may reflect a real population. As discussed in Chapter 3, repeated sub-sampling helped to ensure replicability, i.e., that the results were not limited to a particular partitioning instance [Fin72].

Understanding how affect can be communicated through posture

The research presented in this thesis provided an understanding of how posture can be used to communicate affect. It was examined through human observers' judgments on acted and non-acted affective expressions and the statistical analysis of the low-level posture description as explained in the discussion in the first contribution. The knowledge that was gained can also be used by researchers in other areas such as affect synthesis to create embodied avatars, affectively expressive robots, etc. by informing the researchers about which low-level posture features indicate specific affective states. Indeed, the posture corpora and the results of this thesis are already being used to build an affective language for multimodal virtual agents [CPM+09].

7.2 Potential Directions for Future Work

The results of all three case studies indicated the effectiveness of creating automatic models for recognising affective body posture. In addition, the results of each study brought to light areas that could be extended. The first area is the low-level posture description. The second area is to devise a more comprehensive method for defining a ground truth of affective expressions by taking advantage of more of the information that is collected from the observers. The third area deals with the understanding of how affect can be communicated through posture. Finally, the fourth area is aimed at developing a more complex affective posture recognition system.

Refine the low-level posture description

The entire body was investigated on the basis that affect is expressed through a variety of configurations. A key question that is raised by the rapidly growing number of motion sensing controllers is whether information from the entire body is necessary. Could enough of the posture configuration be ascertained from the controller alone? In the case of the affective states (Chapter 5), features from the non-controller side of the body were found to be important. For instance, the ANOVA results found more information from the arm not holding the controller to be important. This was less the case with the levels of valence and arousal even though the postures were not always symmetrical, making this an interesting issue to examine further.

The low-level posture description was examined in two situations in this thesis: an acted scenario and a non-acted video game scenario. Both of these scenarios considered standing postures only. A main advantage of the low-level description is that it is general and independent of context, and applying the description to either standing or seated postures should not make a difference. An investigation of the posture description in these situations may help bring to light the existence of previously unforeseen issues or missing features. Indeed, the posture description was preliminarily examined in a publication by the candidate [KFBB05] (not included in the thesis) using a combination the of acted standing postures investigated in Chapter 4 and postures from a non-acted seated situation with 70% correct recognition achieved.

Due to limitations of the motion capture systems that were used in this research, features that are definitely missing are detailed hand and finger movements, such as fists and pointing. Clenched fists are a stereotypical feature of acted expressions of anger [MN89]. The ability to detect these features could reduce misclassifications that occur between angry and happy, and angry and fear [CMR+04]. To carry out this extension, data from a motion sensor glove could be integrated with the low-level posture description.

A more comprehensive method for defining ground truth

The method used in this research for defining the ground truth of the affective postures considers a single label assigned to each posture, i.e., the most frequent label or the median scale rating. Thus, by distilling the observer labels into a single observation, the labels provided by other observers who may also be considered important, are 'lost'. Taking advantage of this information may be especially necessary when dealing with non-acted, subtle affective expressions as it is possible for more than one affective state to occur at the same time [AR09].

The ground truth labelling method of the affective expressions could be extended to incorporate preference learning techniques [Yan09] as discussed in Chapter 2. When the observers have a choice between more than two labels, the observers' preference for all of the labels could be ranked or weighted according to the frequency of use of each affective label. In the case of the affective posture recognition system presented in Chapter 6, the weighted information could be applied at the decision rule level in order to bias the result.

Information about the observer could be used for building a more objective ground truth as a method for increasing its reliability. For instance, a person's ability to empathise with others is considered important for the recognition of another person's emotional state [ME72]. Furthermore, as stated in Chapter 1, other factors, such as the observer's age, gender or culture are also considered to affect the way a person perceives emotion [Pic98]. This information could be used to bias the computation of the most frequent labels by using the information as way to weight the evaluation of each observer.

Extend the understanding of affect communicated through posture

Observers from three different cultures were investigated in Chapter 4. As discussed in Section 7.1, some differences between the cultures were found about which posture features are indicative of different emotions. The results could lend evidence to support the idea of emotional 'dialects' as described by Elfenbein and Ambady [EA02]. They consider the idea that emotional expression is a universal language, and that different 'dialects' of that universal language exist across cultures. This is an important concept for affective computing in order to build effective affect recognition systems. As discussed in Chapter 2, the addition of information about how different cultures express and perceive affect has become more and more important in a number of real-life affective computing situations, such as embodied museum agents [KGKW05][LAJ05] and eLearning systems [DM06]. As systems replace humans, it is important that how they express and perceive non-verbal behaviours in a multi-cultural community is as natural as possible so that the user is not made uncomfortable.

Extend the real time affective posture recognition system

The affective posture recognition system presented in Chapter 6 was implemented to recognise affect from sequences of postures as they would arrive in a real time situation. The results were promising on a testing set of sequences with the system achieving a recognition rate of only 10% less than the target rate set according to the level of agreement achieved by the human observers. Reasons for the difference in recognition rates were postulated which highlighted ways in which the system could be refined and extended. For instance, the training set may not have been comprehensive enough. Indeed, several postures included in the training set achieved low observer agreement, signifying that these postures may be too ambiguous. Creating a better defined training set consisting of only postures for which observer agreement was high may help increase the system's performance.

Another extension of the affective posture recognition system could be the addition of context information. Context could include such things as the type of application into which the recognition system is integrated, since the system developed in this thesis is not intended to be specific only to a video game situation. Information about the user, such as the history of her interaction with the system [eK05] may also be included as a type of context.

7.3 Thesis Summary

The research presented in this thesis is centred in the rapidly growing field of affective computing and focused on the automatic recognition of affect. In order to create systems aimed at the recognition of affect, this thesis tested the power of body posture as a modality upon which affect recognition systems can be based. The main hypothesis was that affect, both discrete categories and affective dimensions, could be recognised from whole body postures using a low-level description of the body in both acted and non-acted situations by both human observers and automatic recognition models. A labelling and benchmark setting approach was devised around human observers, and the results showed that a recognition system could be built that is capable of performing to a level similar to a human interaction partner. The results also highlighted four exciting directions for future work which aim to extend the robustness and recognition reliability of the system: i) refining the low-level posture description; ii) devising a more comprehensive ground truth; iii) extending the understanding of how affect can be communicated through posture; and iv) extending the real time affective posture recognition system.

Appendix A

Publications

1. (Kleinsmith, et al., 2010): Kleinsmith, A., Bianchi-Berthouze, N., and Steed, A. (2010). *Automatic Recognition of Non-Acted Affective Postures.* Submitted to IEEE Transactions on Systems, Man, and Cybernetics Part B, February, 2010.

2. (Kleinsmith and Bianchi-Berthouze, 2010): Kleinsmith, A. and Bianchi-Berthouze, N. (2010). *Modelling Non-Acted Affective Posture in a Video Game Scenario.* Proceedings of the International Conference on Kansei Engineering and Emotion Research, KEER 2010.

3. (Kleinsmith and Bianchi-Berthouze, 2007): Kleinsmith, A. and Bianchi-Berthouze, N. (2007). *Recognizing affective dimensions from body posture.* LNCS: Proceedings of the Second International Conference on Affective Computing and Intelligent Interaction. pp. 48-58.

4. (Kleinsmith, et al., 2006): Kleinsmith, A., de Silva, P., and Bianchi-Berthouze, N. (2006). *Cross-cultural differences in recognizing affect from body posture.* Interacting with Computers. vol. 18, pp. 1371-1389.

5. (Kleinsmith, et al., 2006): Kleinsmith, A., Bianchi-Berthouze, N., and Berthouze, L. (2006). *An effect of gender in the interpretation of affective cues in avatars.* Pro-

ceedings of Workshop on Gender and Interaction: Real and Virtual Women in a Male World, held in conjunction with AVI 2006.

6. (Kleinsmith and Bianchi-Berthouze): Kleinsmith, A. and Bianchi-Berthouze, N. (2006). *Affective Posture Recognition: Human Factors and Modelling.* Proceedings of the Doctoral Consortium of British HCI.

7. (Kleinsmith et al., 2005): Kleinsmith, A., de Silva, P., and Bianchi-Berthouze, N. (2005). *Building user models based on cross-cultural differences in recognizing emotion from affective postures.* LNCS: Proceedings of the 10th International Conference on User Modeling. pp. 50-59.

8. (Kleinsmith, et al., 2005): Kleinsmith, A., de Silva, P., and Bianchi-Berthouze, N. (2005). *Grounding affective dimensions into posture features.* LNCS: Proceedings of the First International Conference on Affective Computing and Intelligent Interaction. pp. 263-270.

9. (de Silva et al., 2005): de Silva, P., Kleinsmith, A., and Bianchi-Berthouze, N. (2005). *Towards unsupervised detection of affective body posture nuances.* LNCS: Proceedings of the First International Conference on Affective Computing and Intelligent Interaction. pp. 32-39.

10. (Kleinsmith, et al., 2005): Kleinsmith, A., Fushimi, T., and Bianchi-Berthouze, N. (2005). *An incremental and interactive affective posture recognition system.* In Proceedings of the Workshop on Adapting the Interaction Style to Affective Factors, held in conjunction with UM'05.

11. (Bianchi-Berthouze and Kleinsmith): Bianchi-Berthouze, N. and Kleinsmith, A. (2003). *A categorical approach to affective gesture recognition.* Connection Science. vol. 15, pp. 259-269.

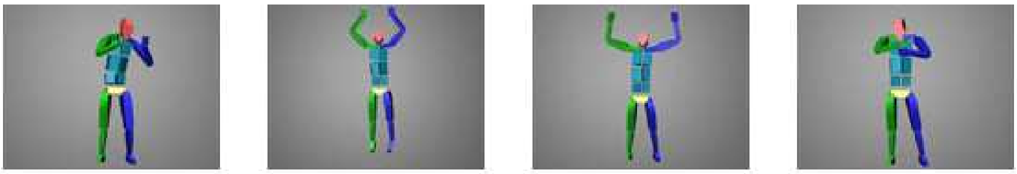
12. (Kleinsmith, et al., 2003): Kleinsmith, A., Fushimi, T., Takenaka, H., and Bianchi-Berthouze, N. (2003). *Towards bi-directional affective human-machine interaction.*

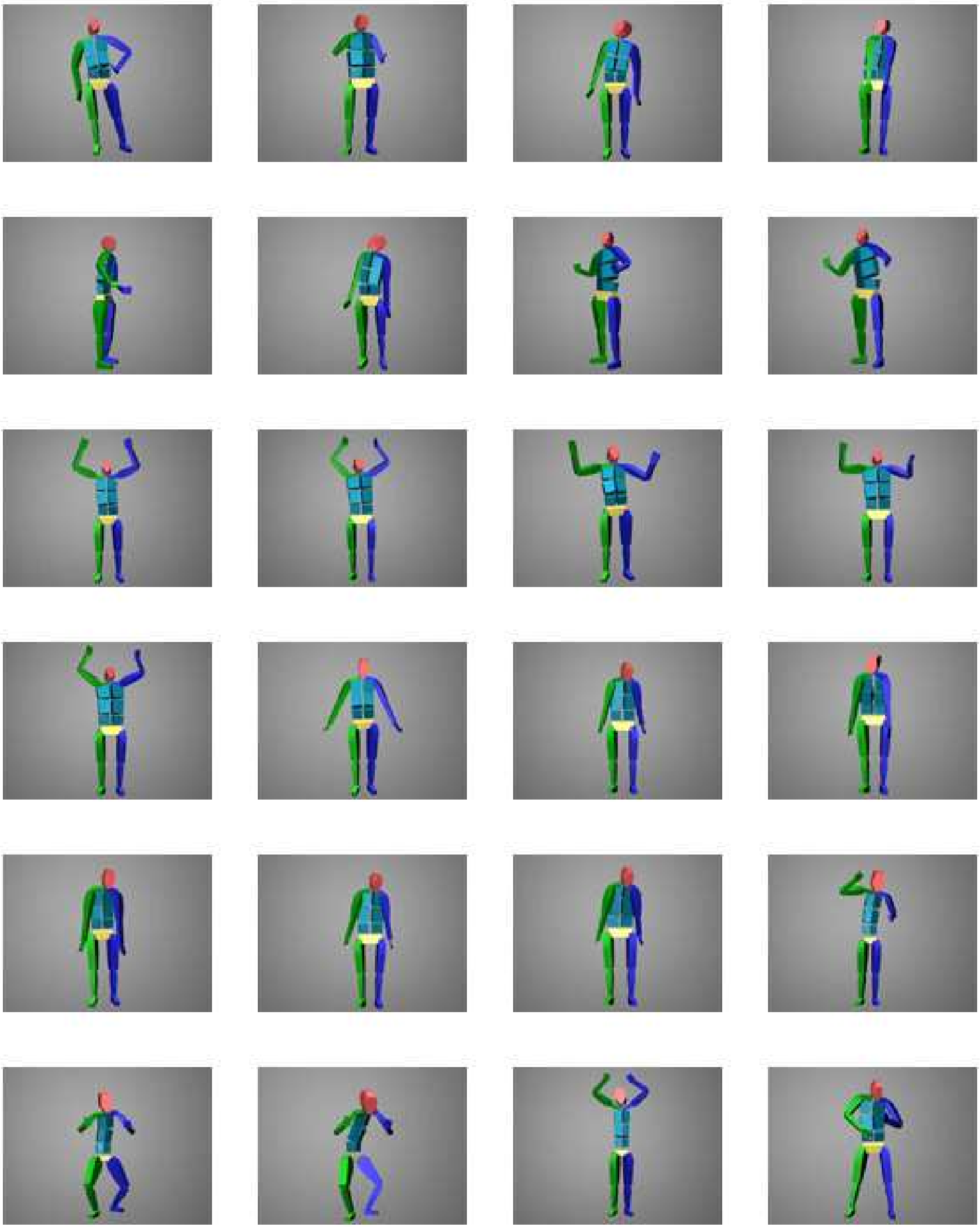
Special issue of the Journal of 3D-Forum Society. vol. 17, pp. 61-66.

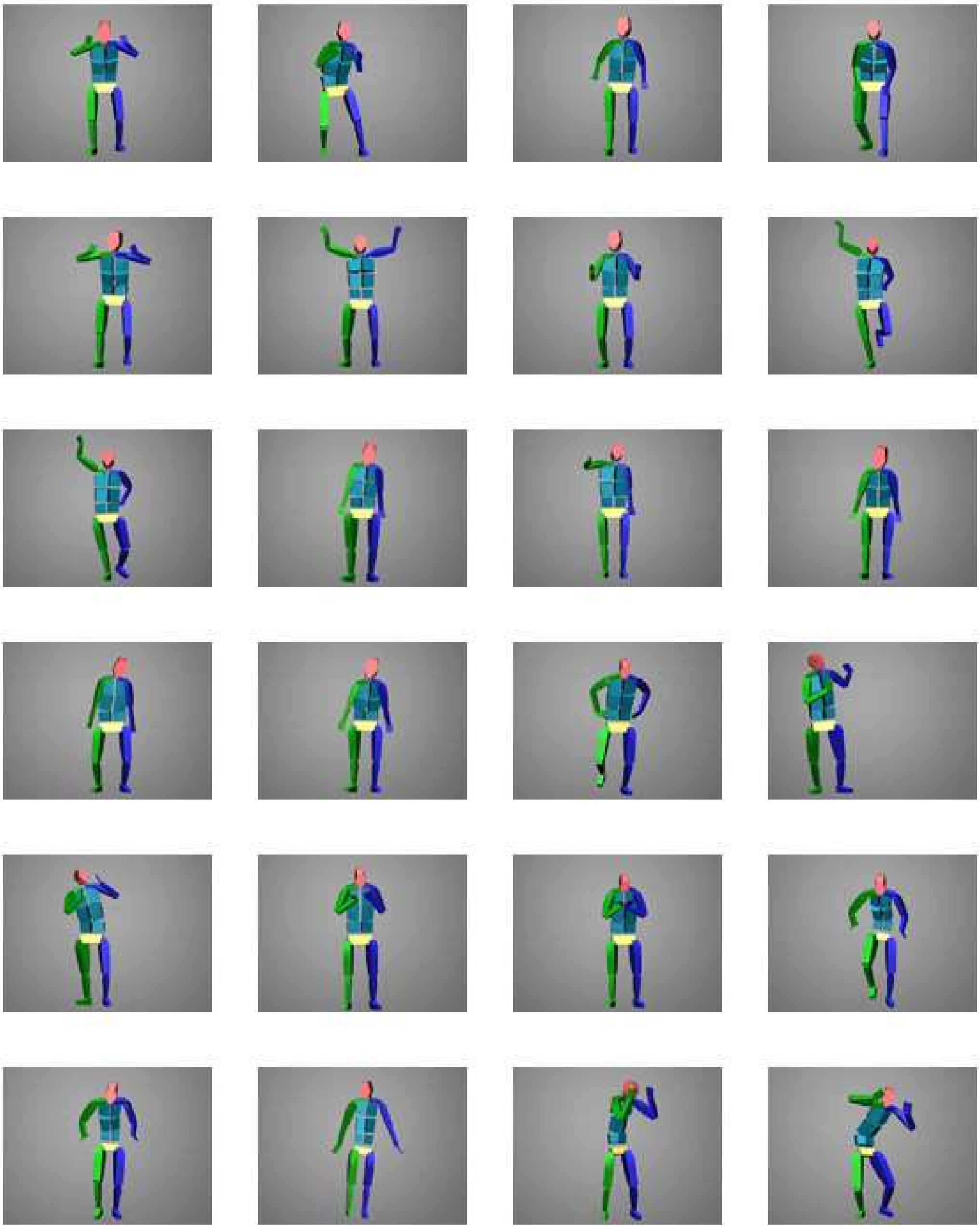
13. (Bianchi-Berthouze, et al., 2003): Bianchi-Berthouze, N., Fushimi, T., Kleinsmith, A., Hasegawa, M., Takenaka, H., and Berthouze, L. 2003). *Learning to recognize affective body postures.* In Proceedings of the IEEE International Symposium on Computational Intelligence for Measurement Systems and Applications. pp. 193-198.

Appendix B

Posture Images Used in the Acted Study

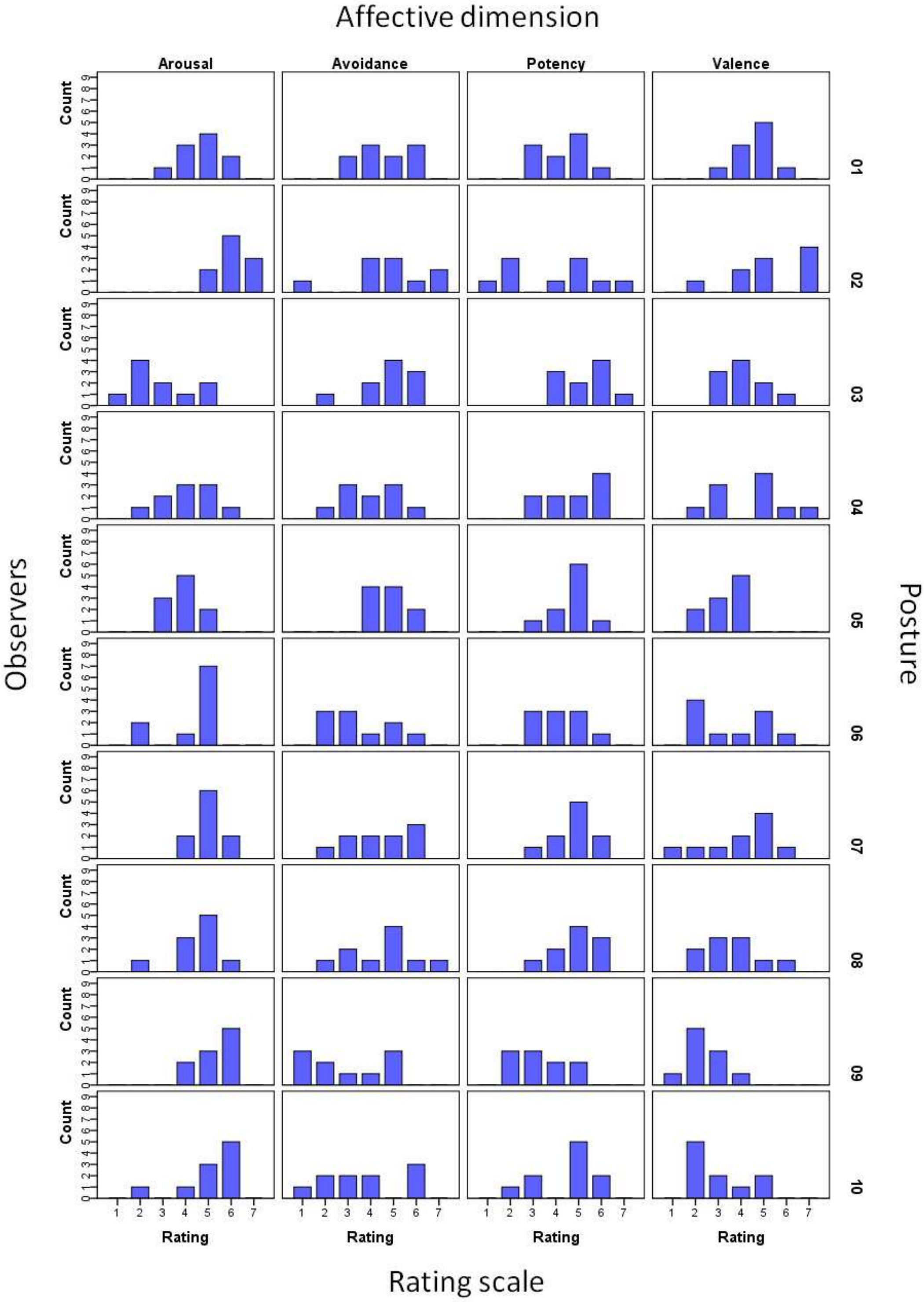


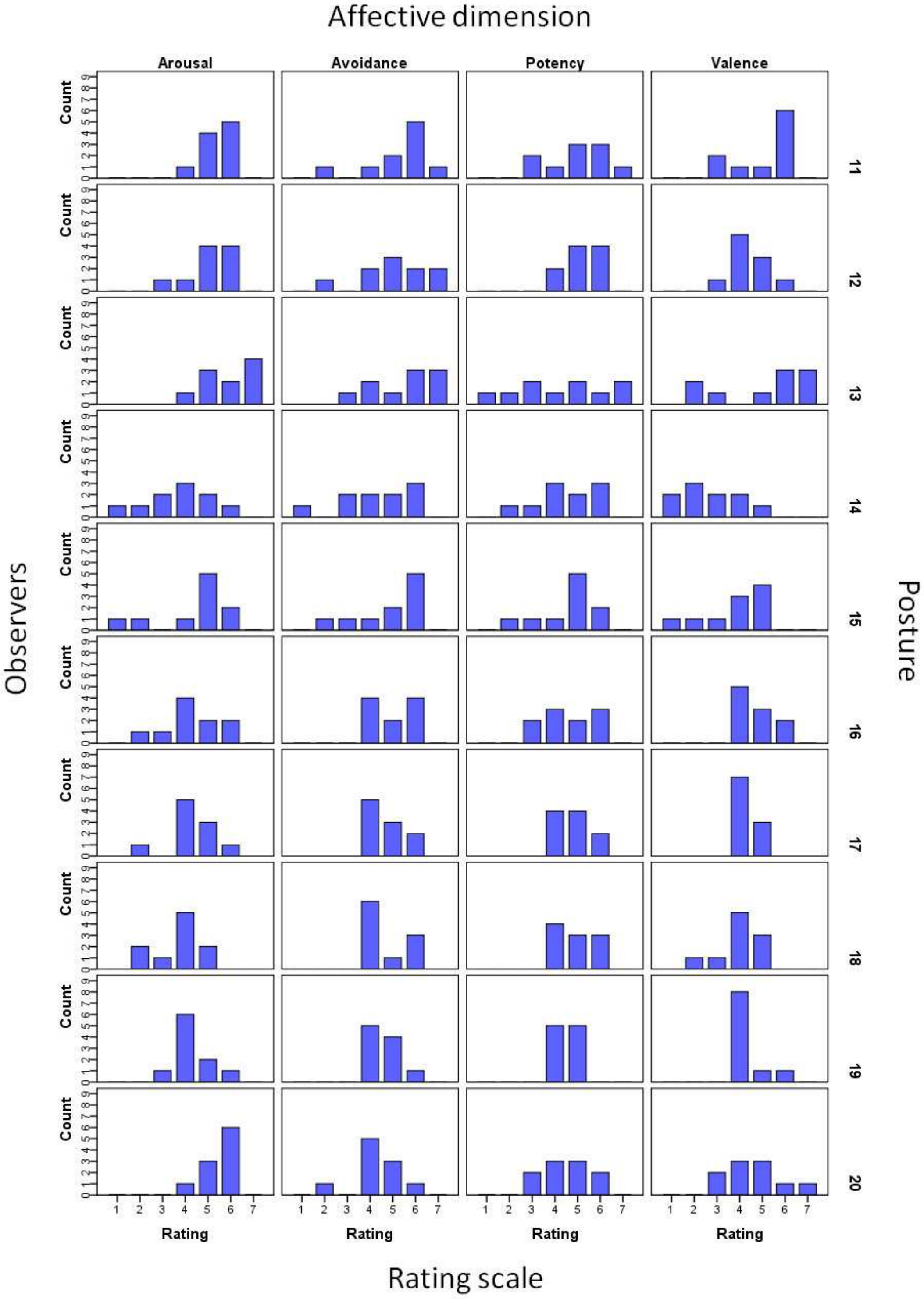


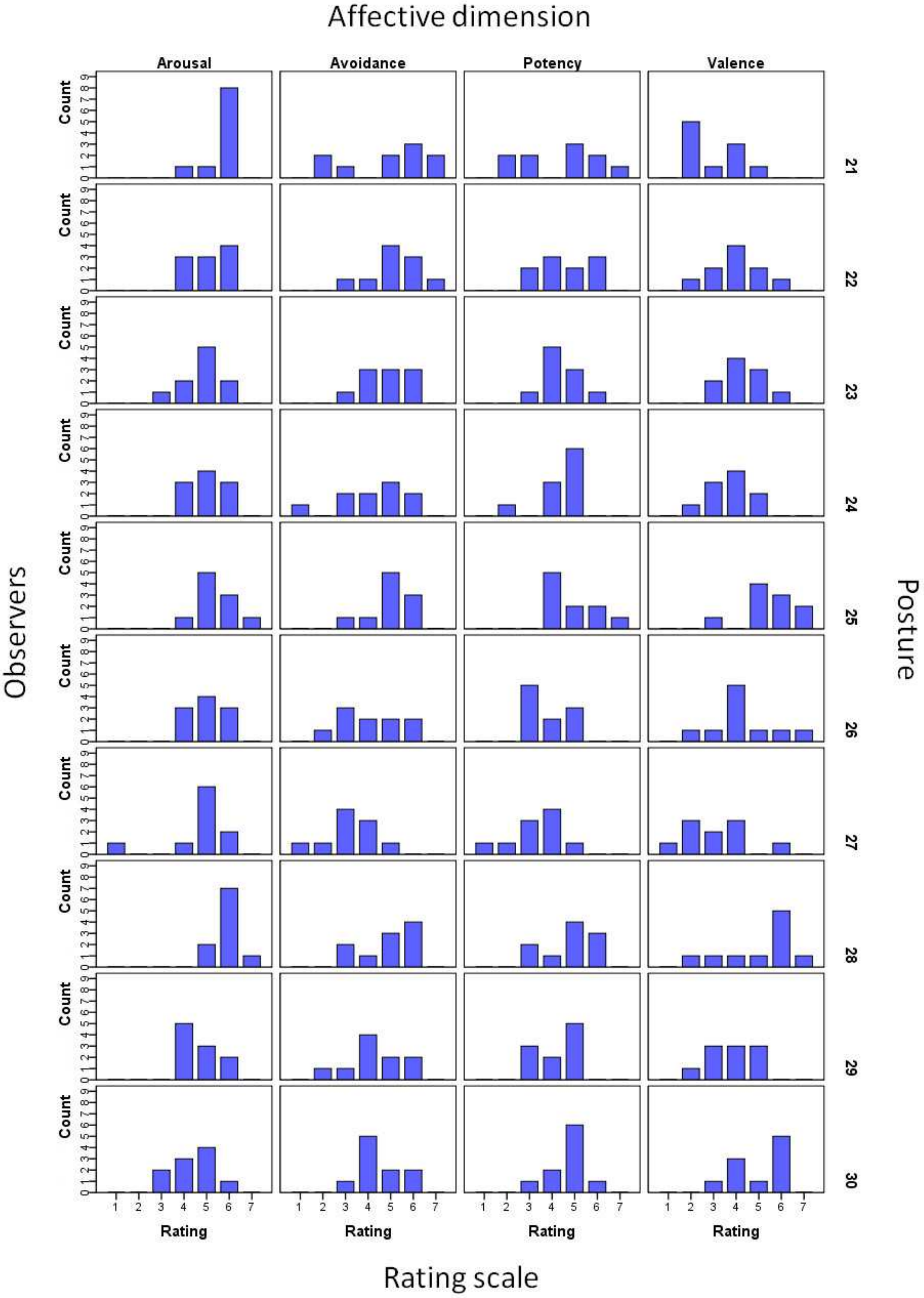


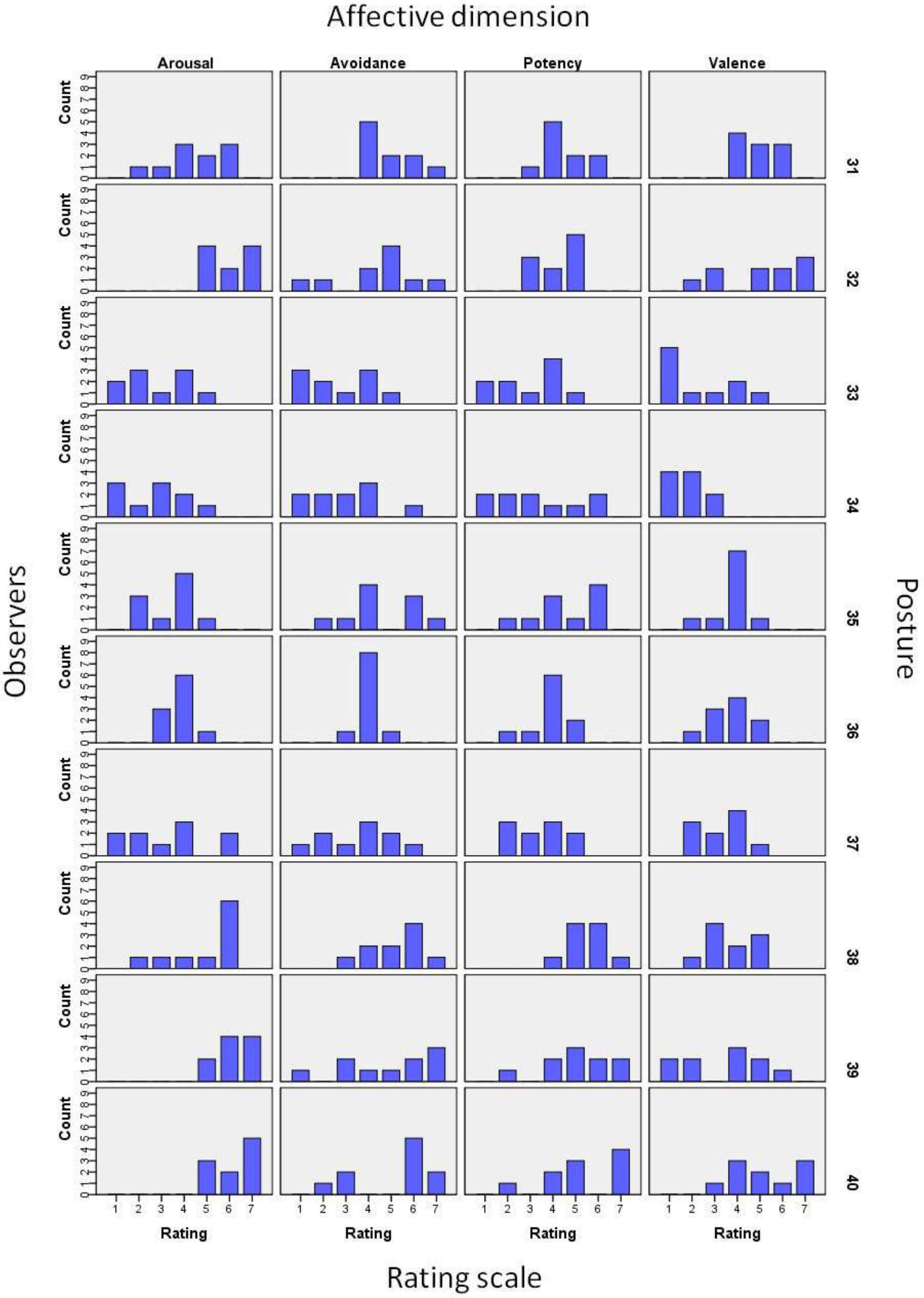
Appendix C

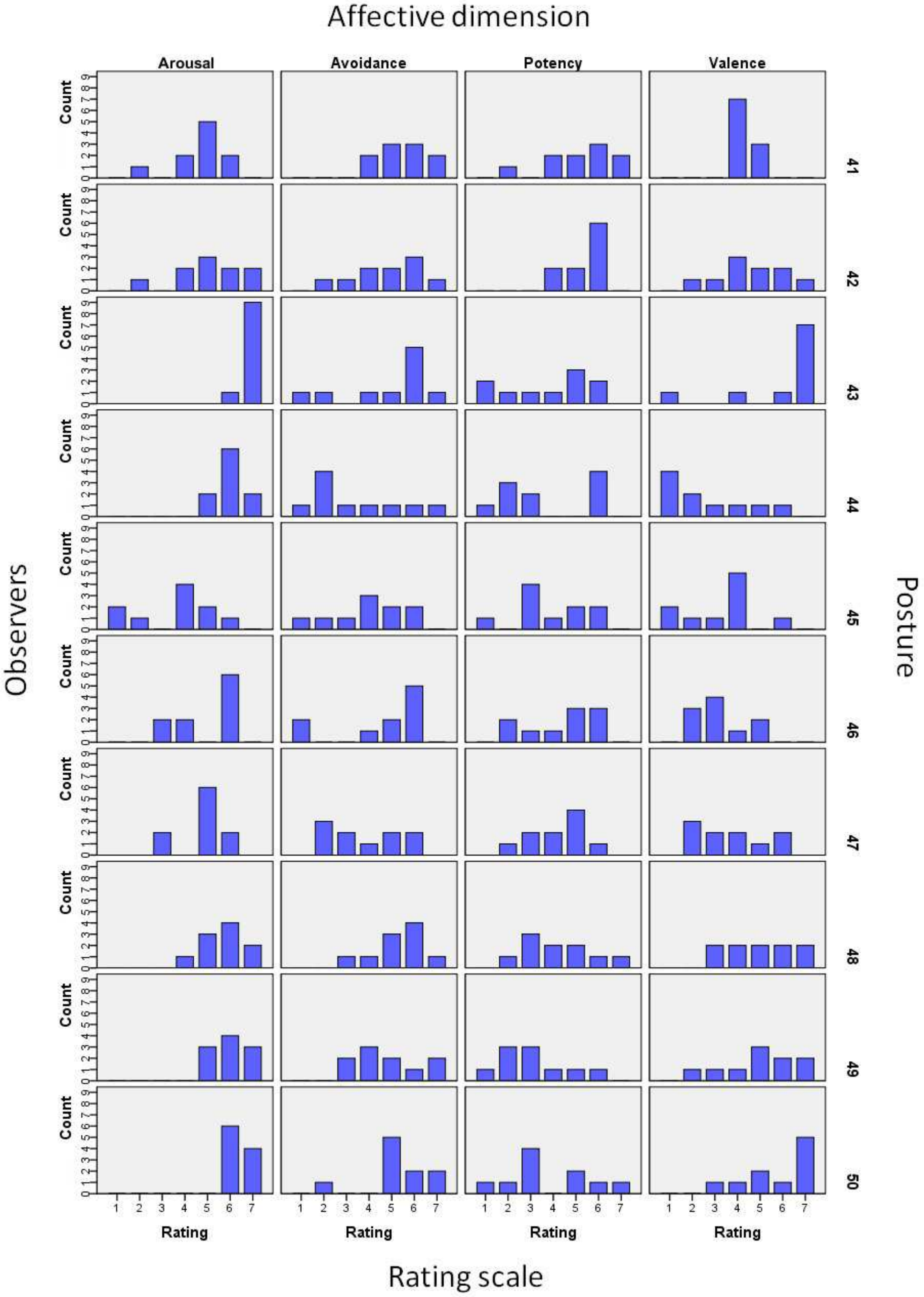
Affective Dimensions Overview of the Acted Postures Survey

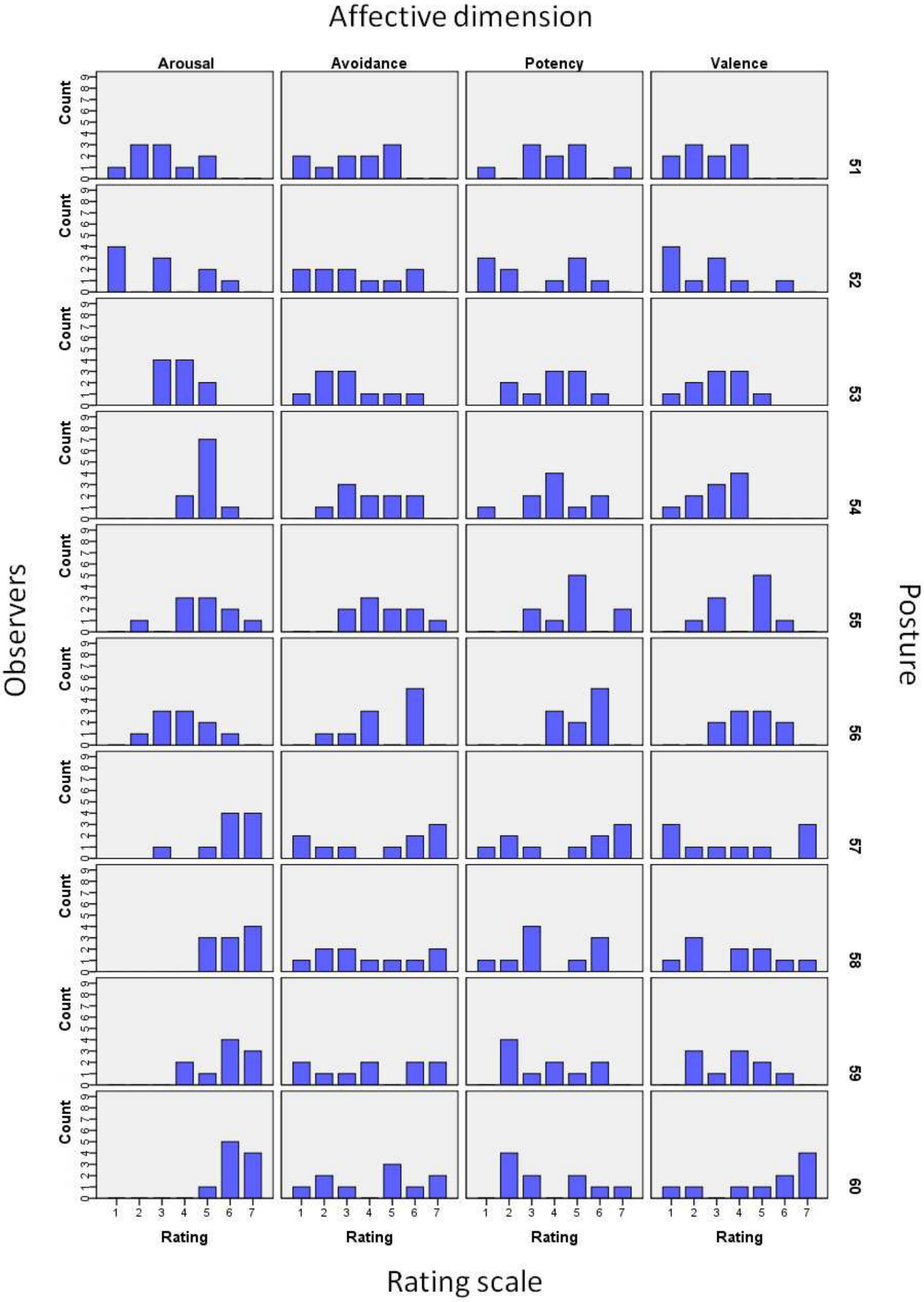

Affective dimension
Arousal
Avoidance
Potency
Valence
Count
0
1
2
3
4
5
6
7
8
9
01
02
03
04
05
06
07
08
09
10
Observers
Posture
1
2
3
4
5
6
7
Rating
Rating scale

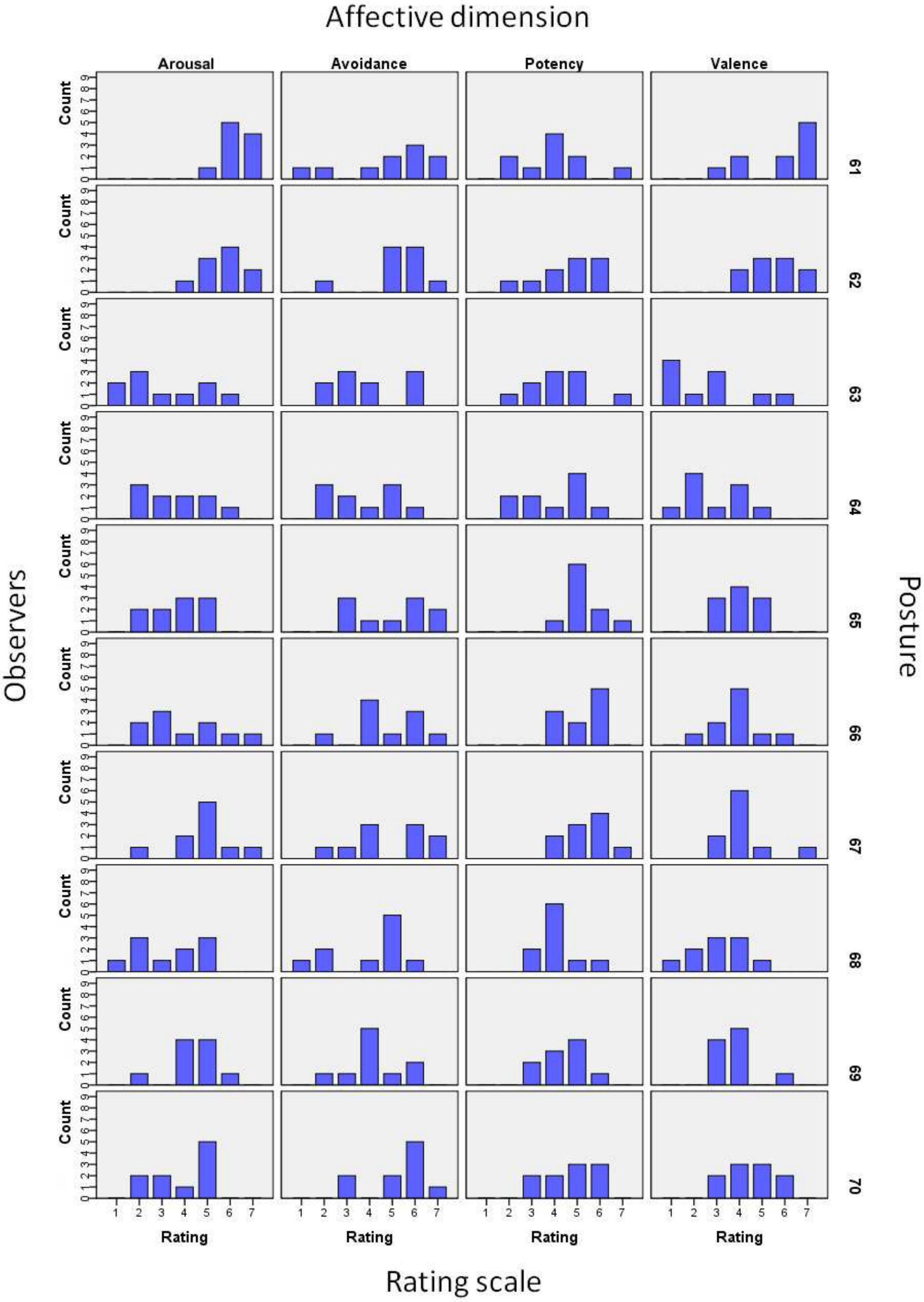

Affective dimension
Arousal
Avoidance
Potency
Valence
Count
0 1 2 3 4 5 6 7 8 9
Observers
Posture
11
12
13
14
15
16
17
18
19
20
1 2 3 4 5 6 7
Rating
Rating scale

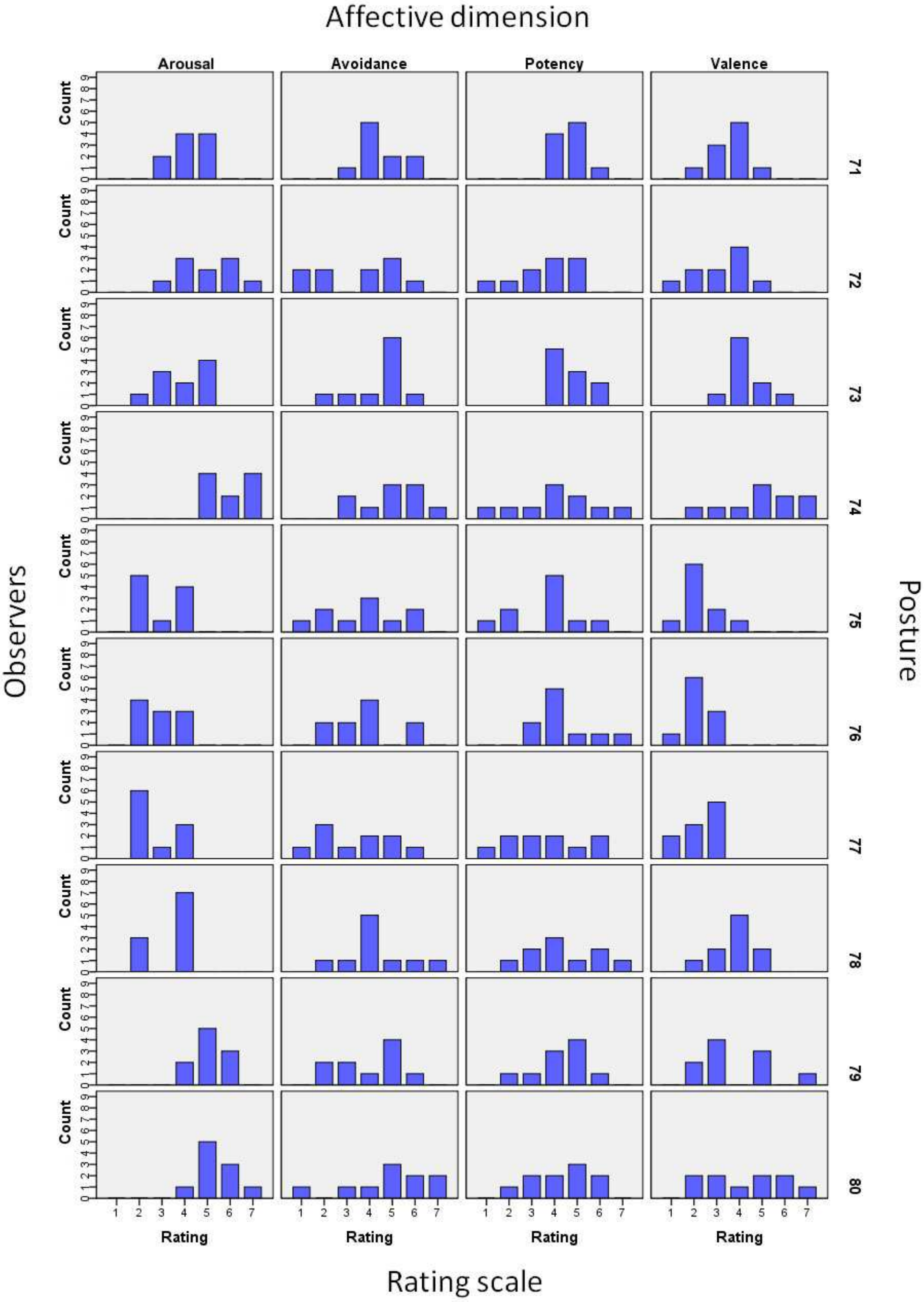

Affective dimension
Arousal
Avoidance
Potency
Valence
Count
0 1 2 3 4 5 6 7 8 9
21
22
23
24
25
26
27
28
29
30
Observers
Posture
1 2 3 4 5 6 7
Rating
Rating scale

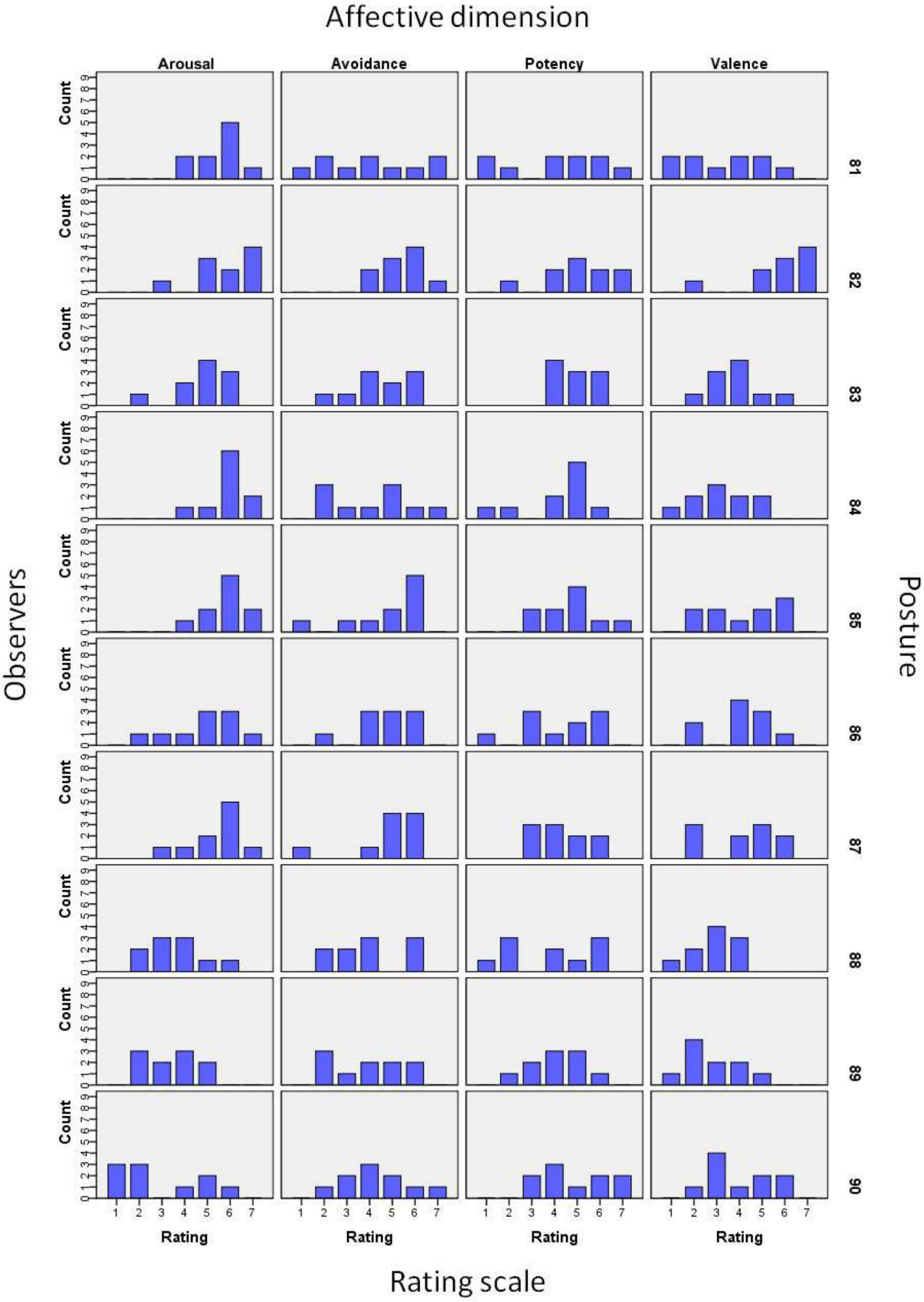

Affective dimension
Arousal
Avoidance
Potency
Valence
Count
0 1 2 3 4 5 6 7 8 9
31
32
33
34
35
36
37
38
39
40
Observers
Posture
1 2 3 4 5 6 7
Rating
Rating scale

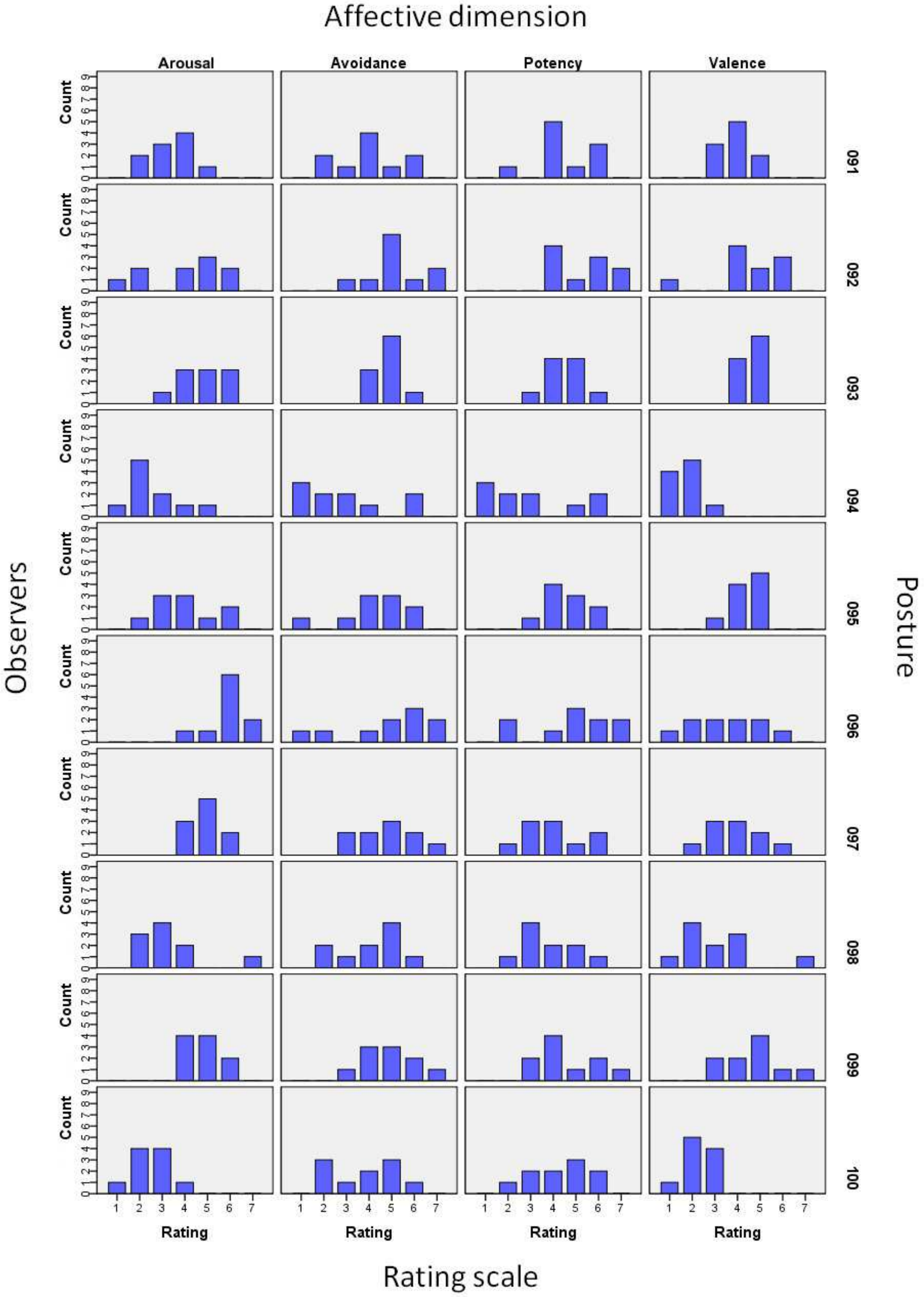

Affective dimension
Arousal
Avoidance
Potency
Valence
Count
0 1 2 3 4 5 6 7 8 9
41
42
43
44
45
46
47
48
49
50
1 2 3 4 5 6 7
Rating
Rating scale
Observers
Posture

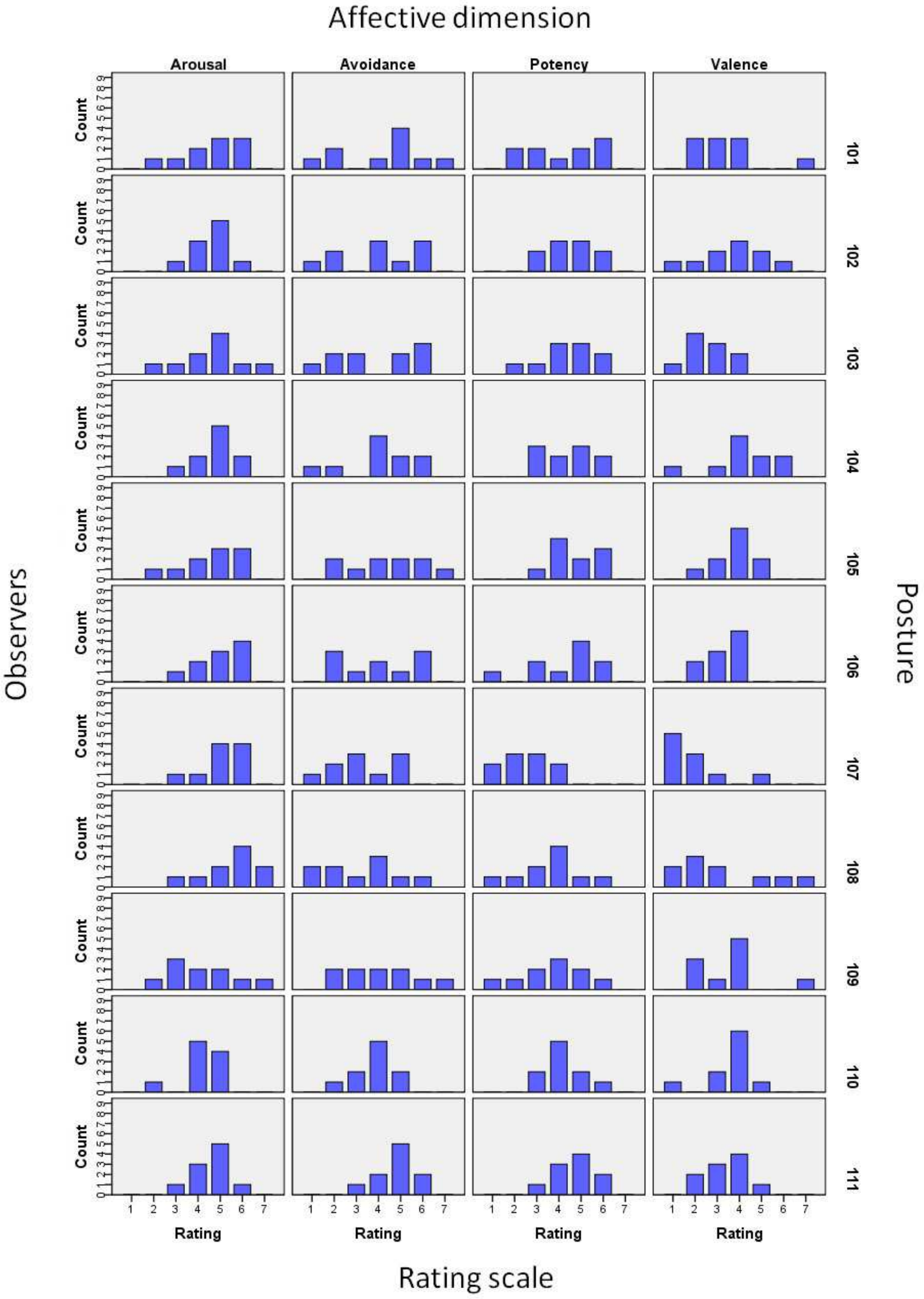

Affective dimension
Arousal
Avoidance
Potency
Valence
Observers
Count
0 1 2 3 4 5 6 7 8 9
Posture
51
52
53
54
55
56
57
58
59
60
1 2 3 4 5 6 7
Rating
Rating scale


Affective dimension
Arousal
Avoidance
Potency
Valence
Count
0 1 2 3 4 5 6 7 8 9
61
62
63
64
65
66
67
68
69
70
Observers
Posture
1 2 3 4 5 6 7
Rating
Rating scale


Affective dimension
Arousal
Avoidance
Potency
Valence
Count
0 1 2 3 4 5 6 7 8 9
71
72
73
74
75
76
77
78
79
80
1 2 3 4 5 6 7
Rating
Rating scale
Observers
Posture


Affective dimension
Arousal
Avoidance
Potency
Valence
Observers
Count
0 1 2 3 4 5 6 7 8 9
Posture
81
82
83
84
85
86
87
88
89
90
1 2 3 4 5 6 7
Rating
Rating scale


Affective dimension
Arousal
Avoidance
Potency
Valence
Observers
Count
0 1 2 3 4 5 6 7 8 9
Posture
091
092
093
094
095
096
097
098
099
100
1 2 3 4 5 6 7
Rating
Rating scale


Affective dimension
Arousal
Avoidance
Potency
Valence
Observers
Count
0 1 2 3 4 5 6 7 8 9
Posture
101
102
103
104
105
106
107
108
109
110
111
1 2 3 4 5 6 7
Rating
Rating scale

Appendix D

Motion Capture with the Nintendo Wii

D.1 Information Sheet for Participants

Thank you for participating in our study. This is one of a series of studies aimed at understanding people's experiences with video games. This study has been approved by University College London's Committee on the Ethics of Non-NHS Human Research. Please read through this information sheet and feel free to ask any questions. The experimenters will answer any general questions; however the specific aspects regarding this study cannot be discussed with you until the end of the session. The whole study will take approximately two hours.

You will be asked to play 2-4 sports video games using the Nintendo Wii$^{\mathrm{TM}}$.

Information that we collect will never be reported in a way that specific individuals can be identified. Information will be reported in a statistical and aggregated manner, and any verbal comments that you make, if written about in subsequent papers, will be presented anonymously.

IMPORTANT

However, if you are pregnant, suffer from heart, respiratory, back, joint or orthopedic problems, have high blood pressure, or if your doctor has instructed you to restrict your physical activity or if you have any other medical condition that may be aggravated by physical activity, or you are receiving treatment for any injury or disorder involving the fingers, hands or arms, I kindly suggest you not to participate.

For further information, please read Wii™(Nintendo©2007) Health and Safety precautions attached to these sheets.

PROCEDURES

- You will be asked to read, understand and sign a Consent Form. If you sign it the study will continue with your participation. Note that you can withdraw at any time without giving any reasons.
- Two photos will be taken for system calibration purposes only.
- You will then be fitted with a motion capture suit to track your body movement during game play.
- After the tasks you will be asked to complete a questionnaire about your experience.
- Thank you for your participation. Please do not discuss this study with others for about three months, as the study is ongoing.
- Any other questions? Please ask any questions that come to mind at this point. After this read and sign the Consent Form.

In case you have any enquiries regarding this study in the future, please contact:

Andrea Kleinsmith, UCL Interaction Centre, University College London, Remax House, 31-32 Alfred Place, London WC1E 7DP, Tel: +44 (0)20 7679 5242, Fax: +44 (0)20 7679 5295, A.Kleinsmith@cs.ucl.ac.uk, http://www.uclic.ucl.ac.uk/people/

D.2 Consent Form A

INVESTIGATORS: ANDREA KLEINSMITH AND NADIA BIANCHI-BERTHOUZE

To be completed by volunteers:

We would like you to read and answer the following questions carefully.

Age		
Nationality		
Gender	F	M
Have you read the information sheet about this study?	YES	NO
Have you had an opportunity to ask questions and discuss this study?	YES	NO
Have you received satisfactory answers to all your questions?	YES	NO
Have you received enough information about this study?	YES	NO
Which investigator have you spoken to about this study?	YES	NO
Do you understand that you are free to withdraw from this study?		
At any time	YES	NO
Without giving a reason for withdrawing	YES	NO
Do you understand and accept the risks associated with the health and safety precautions?	YES	NO
YES / NO Do you agree to take part in this study?	YES	NO
Do you agree to be video taped?	YES	NO
Do you agree to be audio taped?	YES	NO
Do you agree to have your body motions recorded with a motion system?	YES	NO

I certify that I do not have epilepsy.

Signed: Date:

Name in block letters:

Investigator:

In case you have any enquiries regarding this study in the future, please contact:

> Andrea Kleinsmith UCL Interaction Centre Remax House, 31-32 Alfred Pl. London WC1E 7DP Tel +44 (0)20 7679 5242 Fax +44 (0)20 7679 5295 A.Kleinsmith@cs.ucl.ac.uk http://www.uclic.ucl.ac.uk/people/

I, Andrea Kleinsmith, confirm that I have carefully explained the purpose of the study to the participant and outlined any reasonably foreseeable risks or benefits (where applicable).

Signed: Date:

D.3 Consent Form B

Name of the researcher: Andrea Kleinsmith

Please read and tick the following boxes:

- I agree for the researcher to use pictures and video of me during the experiment to be included in the thesis.

  YES

  NO

- I agree for the researcher to use pictures and video of me during the experiment to be included in articles to be published in academic journals, conference proceeding and other equivalent articles.

  YES

  NO

- I agree for the researcher to use pictures and video of me during the experiment to be presented in conferences and other equivalent academic presentations.

  YES

  NO

Would you like to receive through email a brief report on the findings of the study? If yes, please write your e-mail address.

E-mail

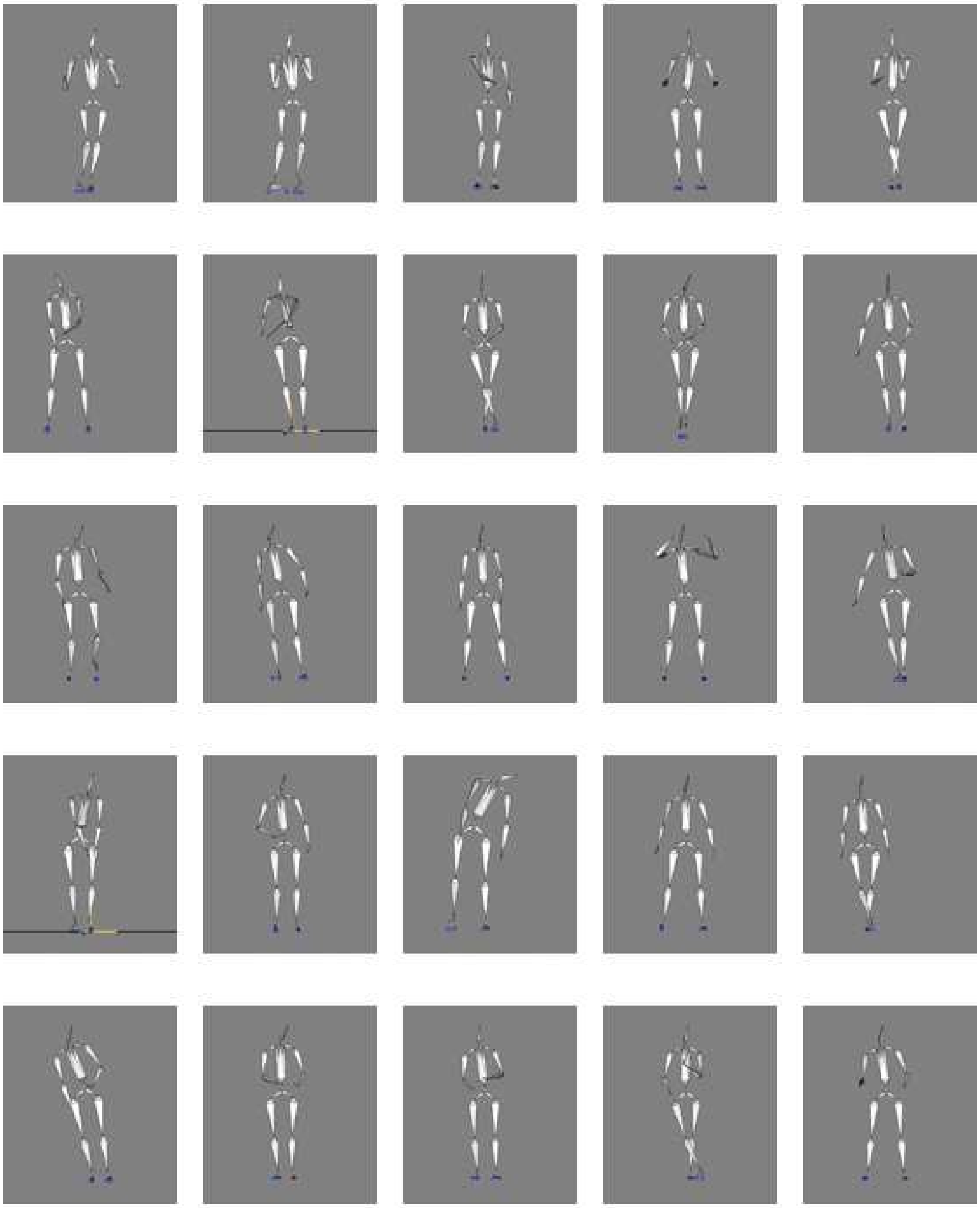
Participant's name Signature Date

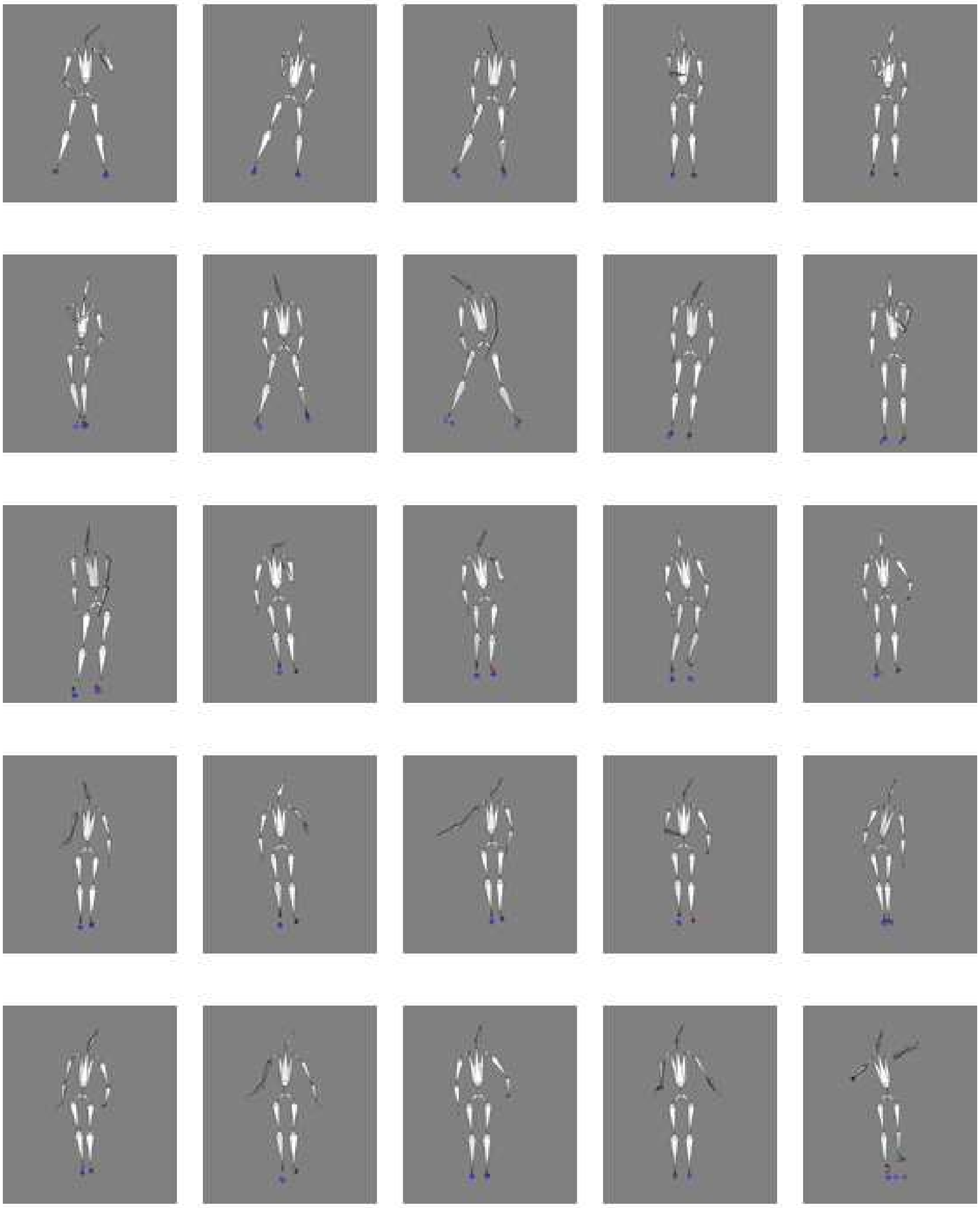
Andrea Kleinsmith

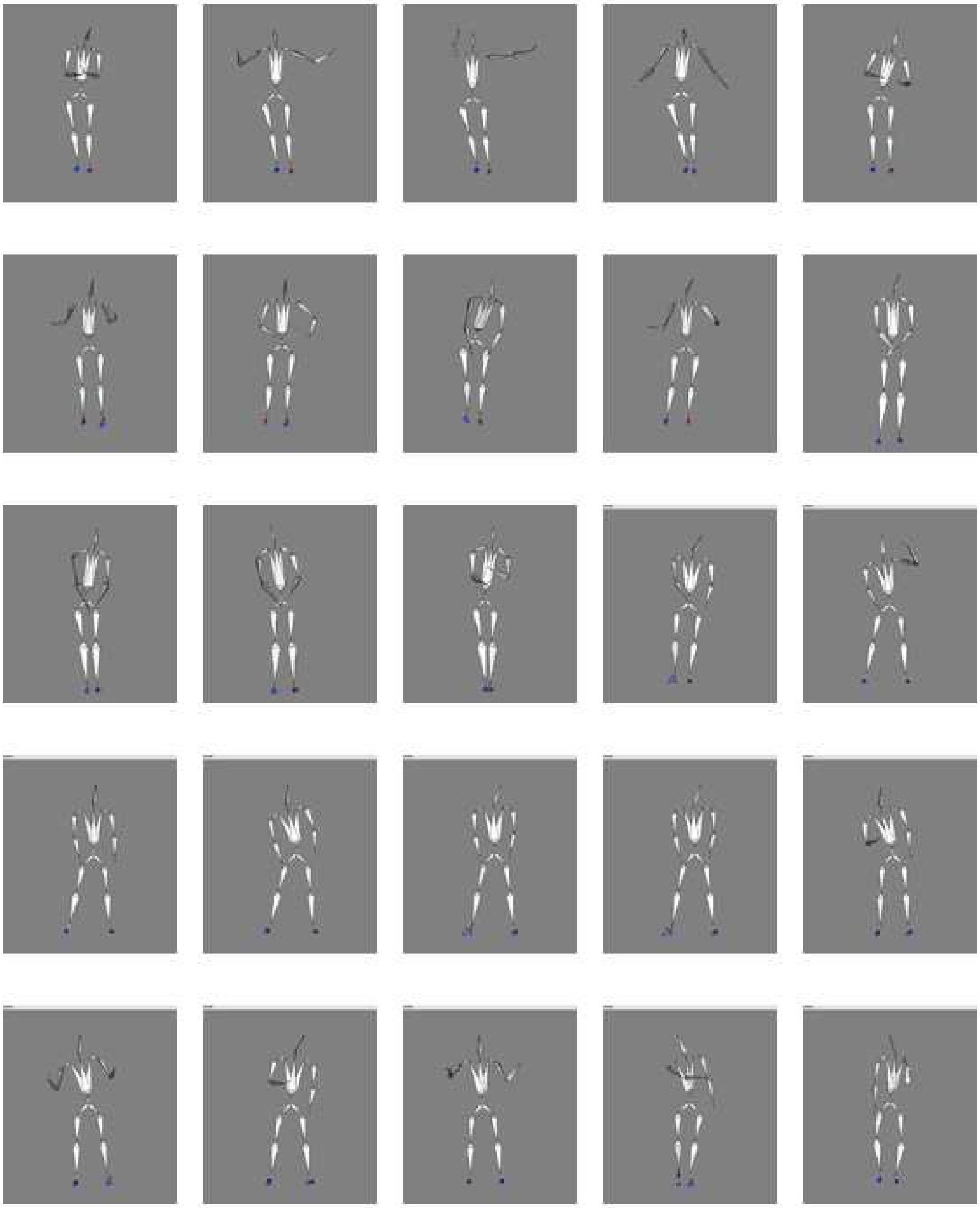
Researcher's name Signature Date

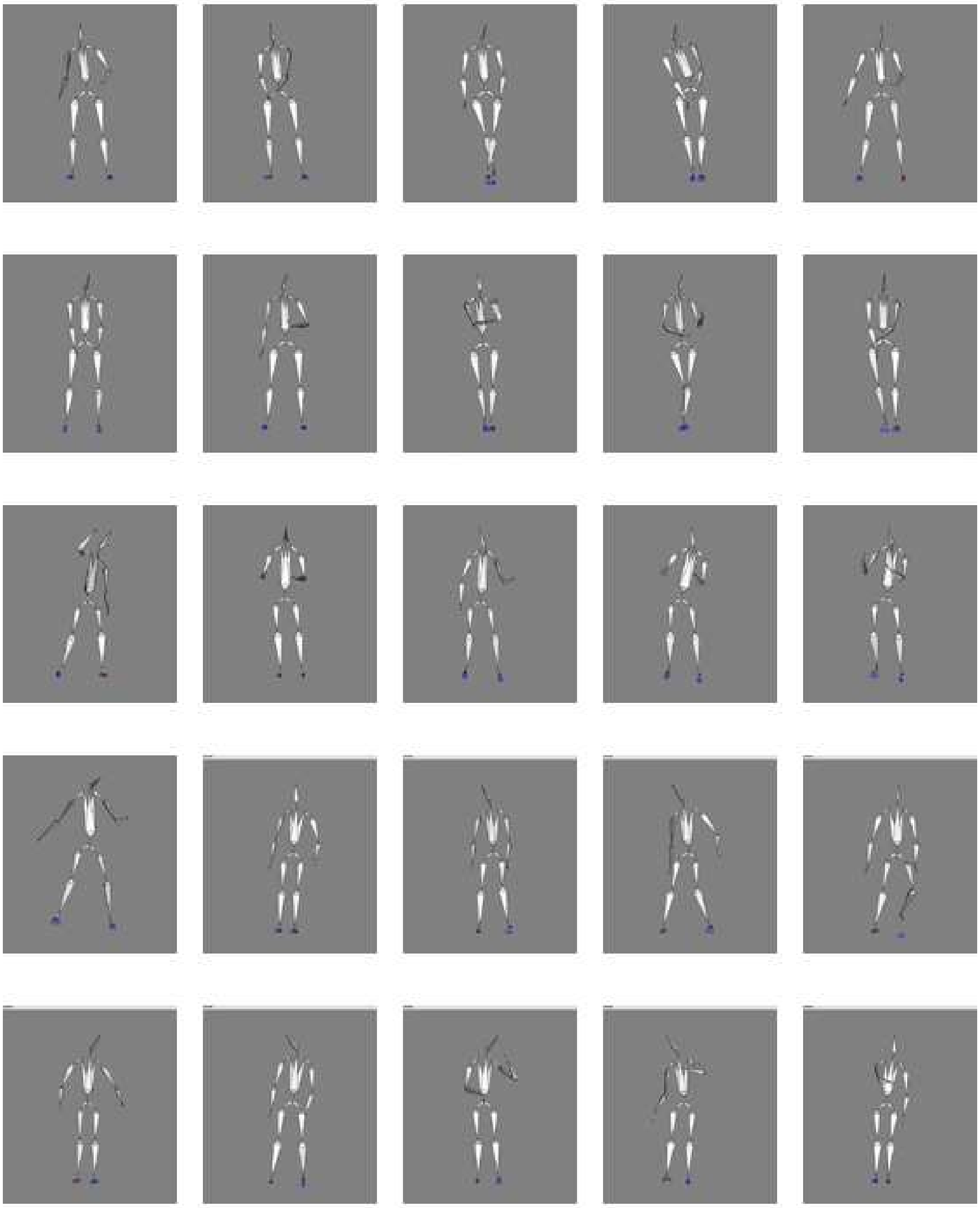
Appendix E

Posture Images Used in the Non-acted Study



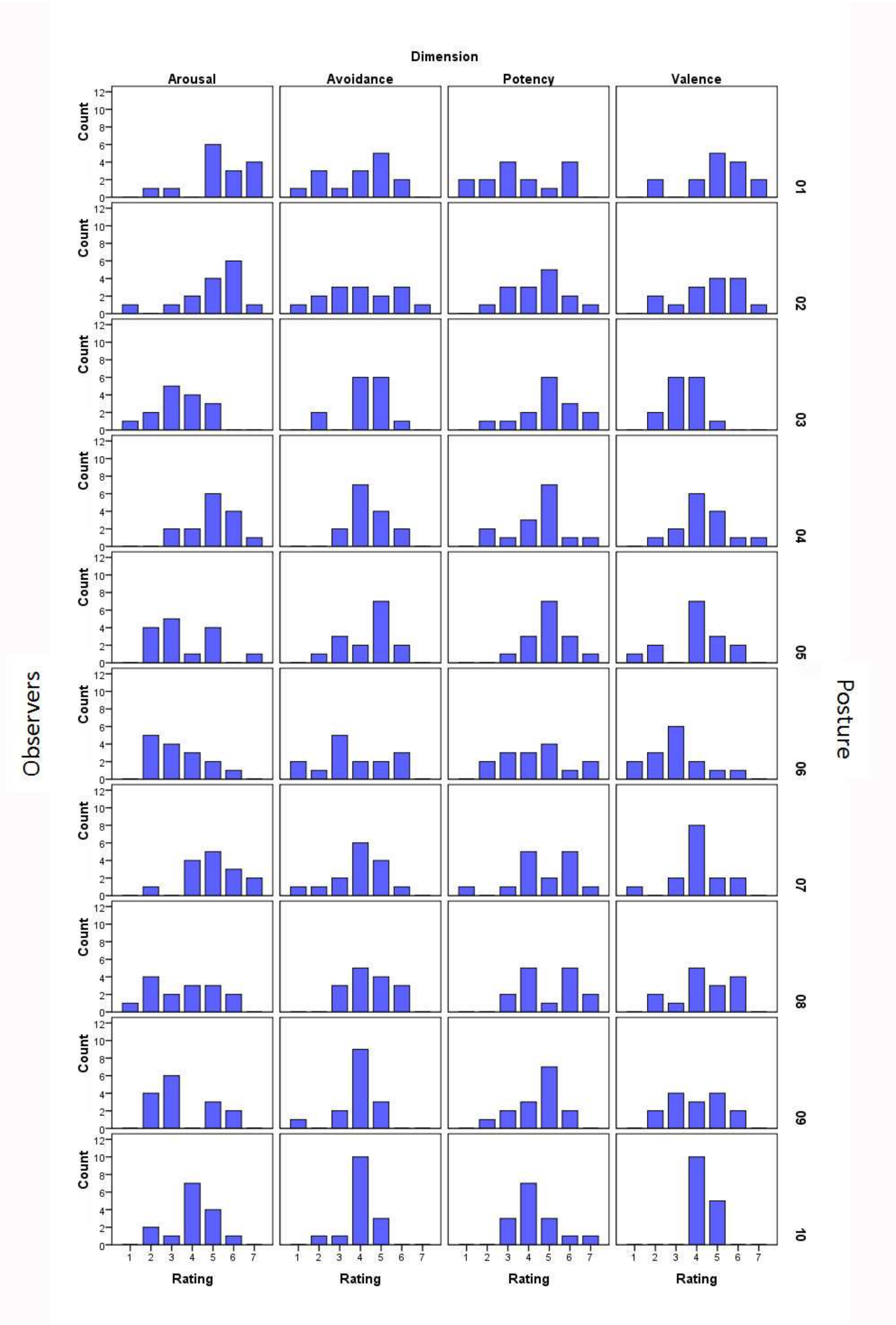


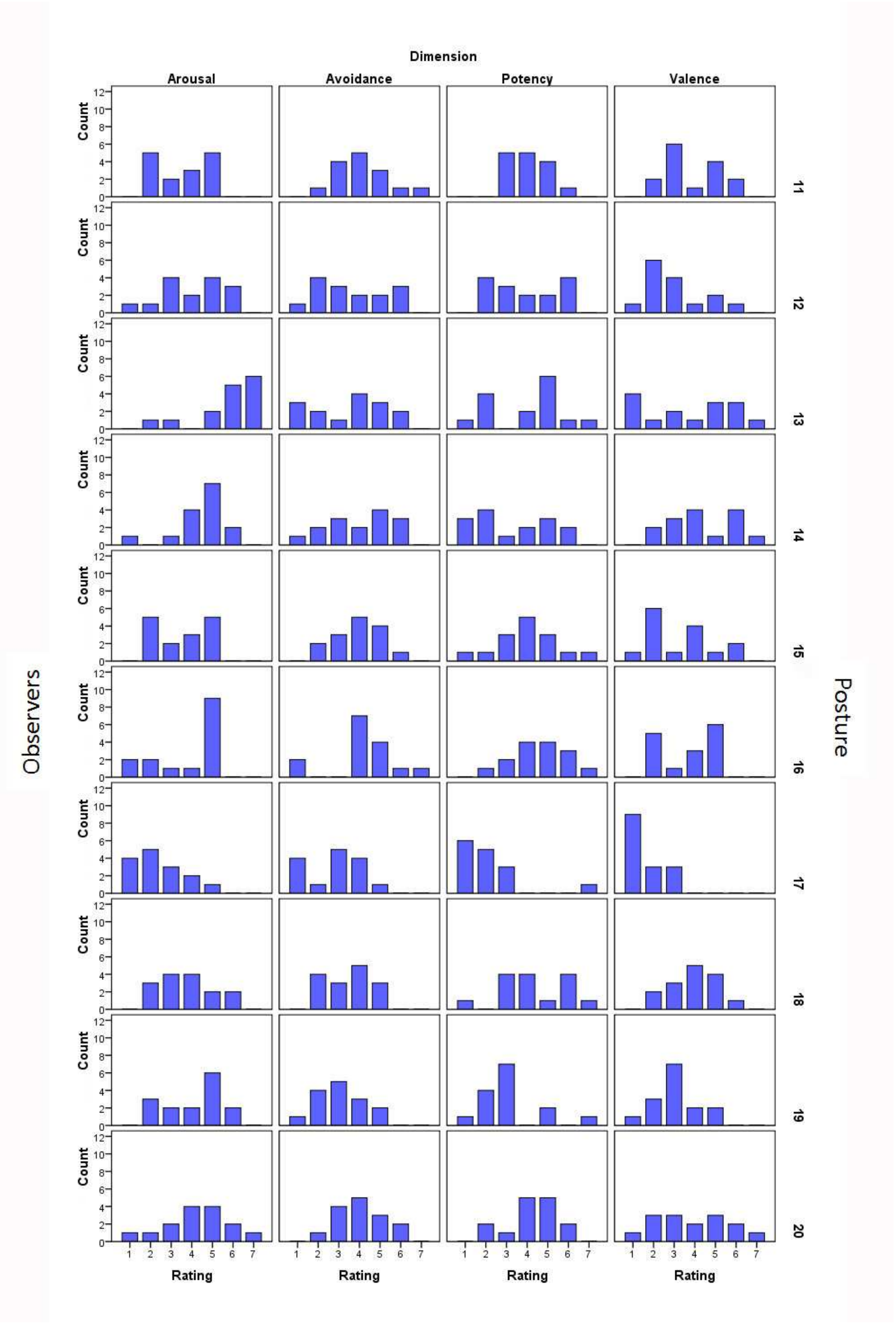


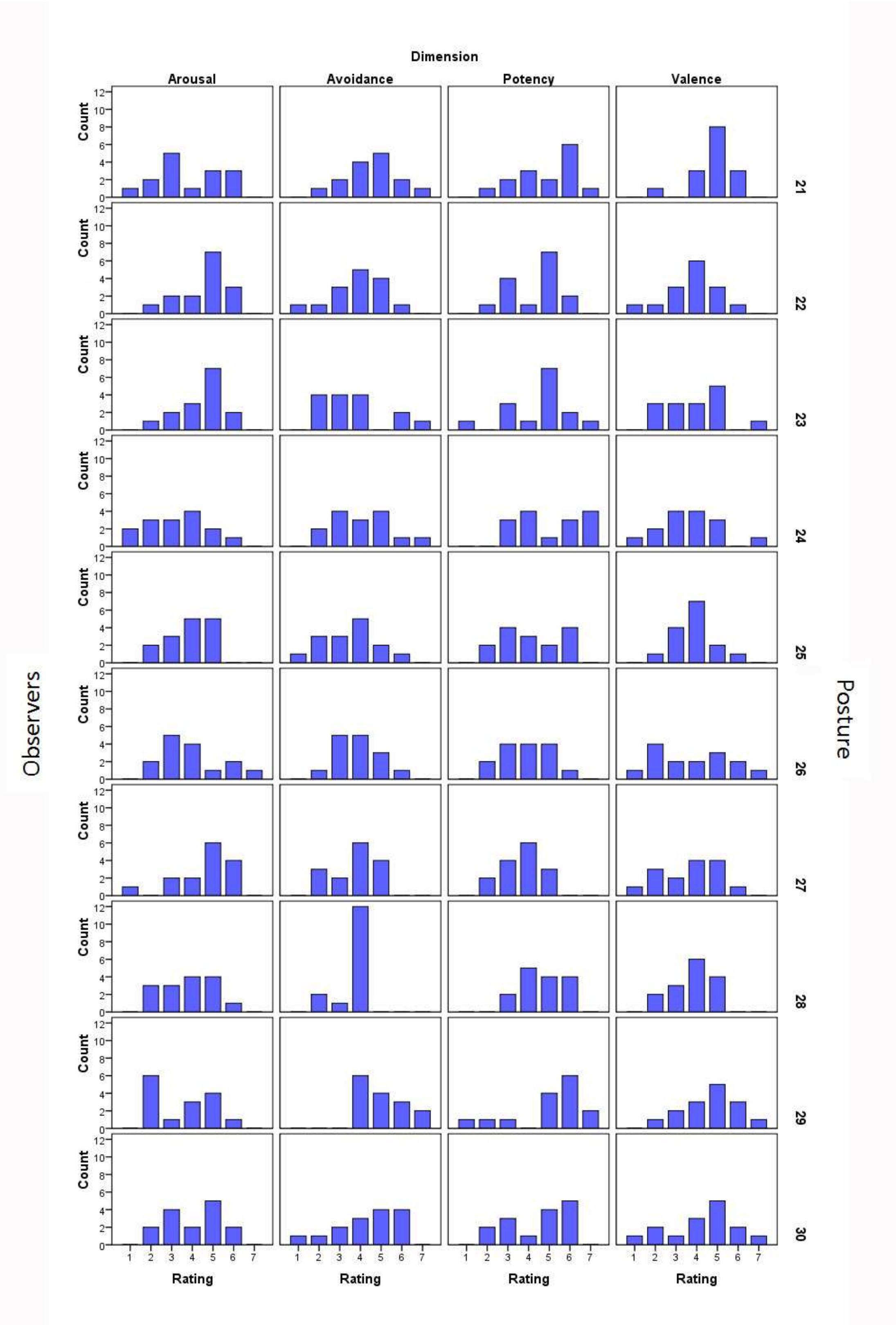


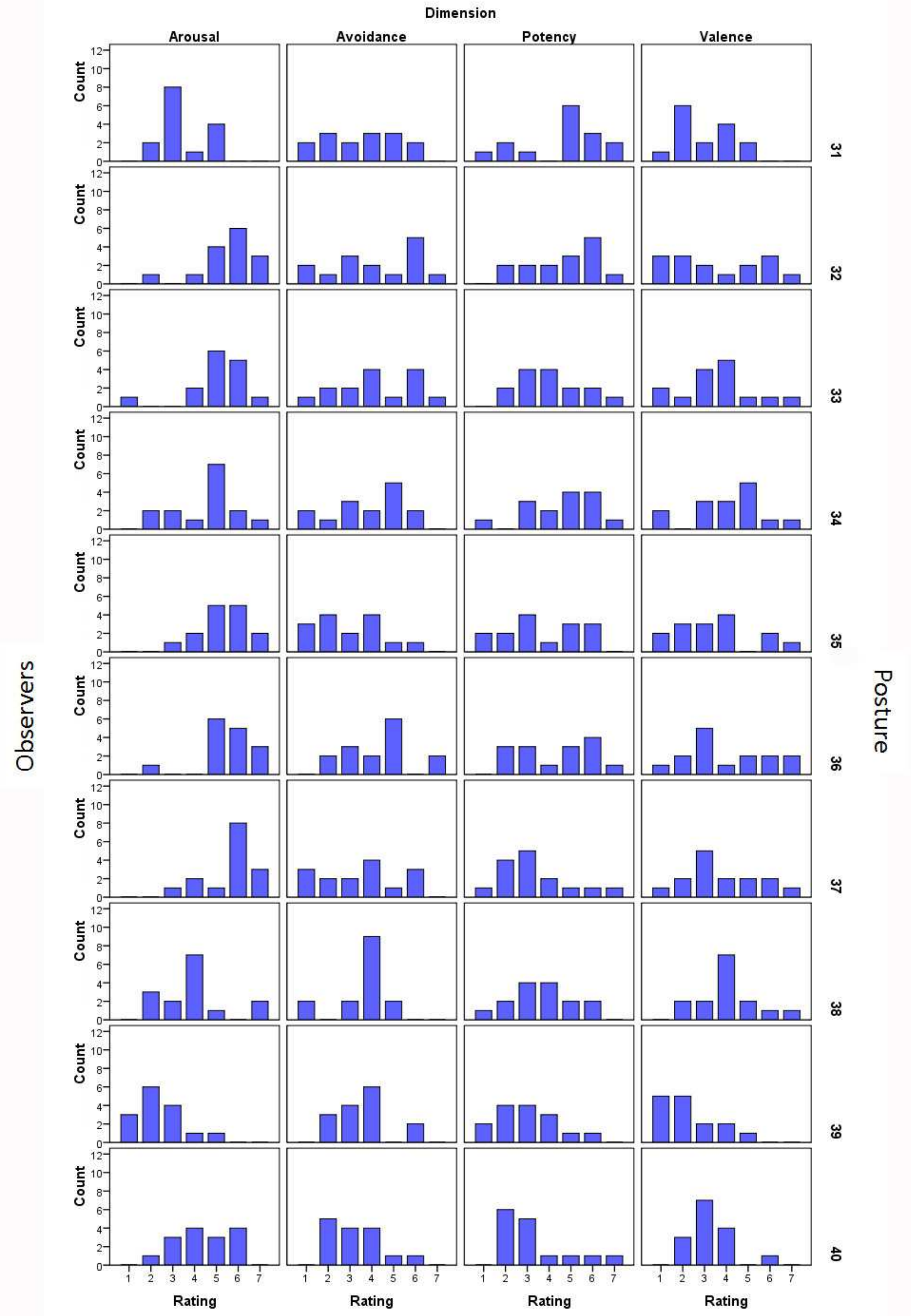
Appendix F

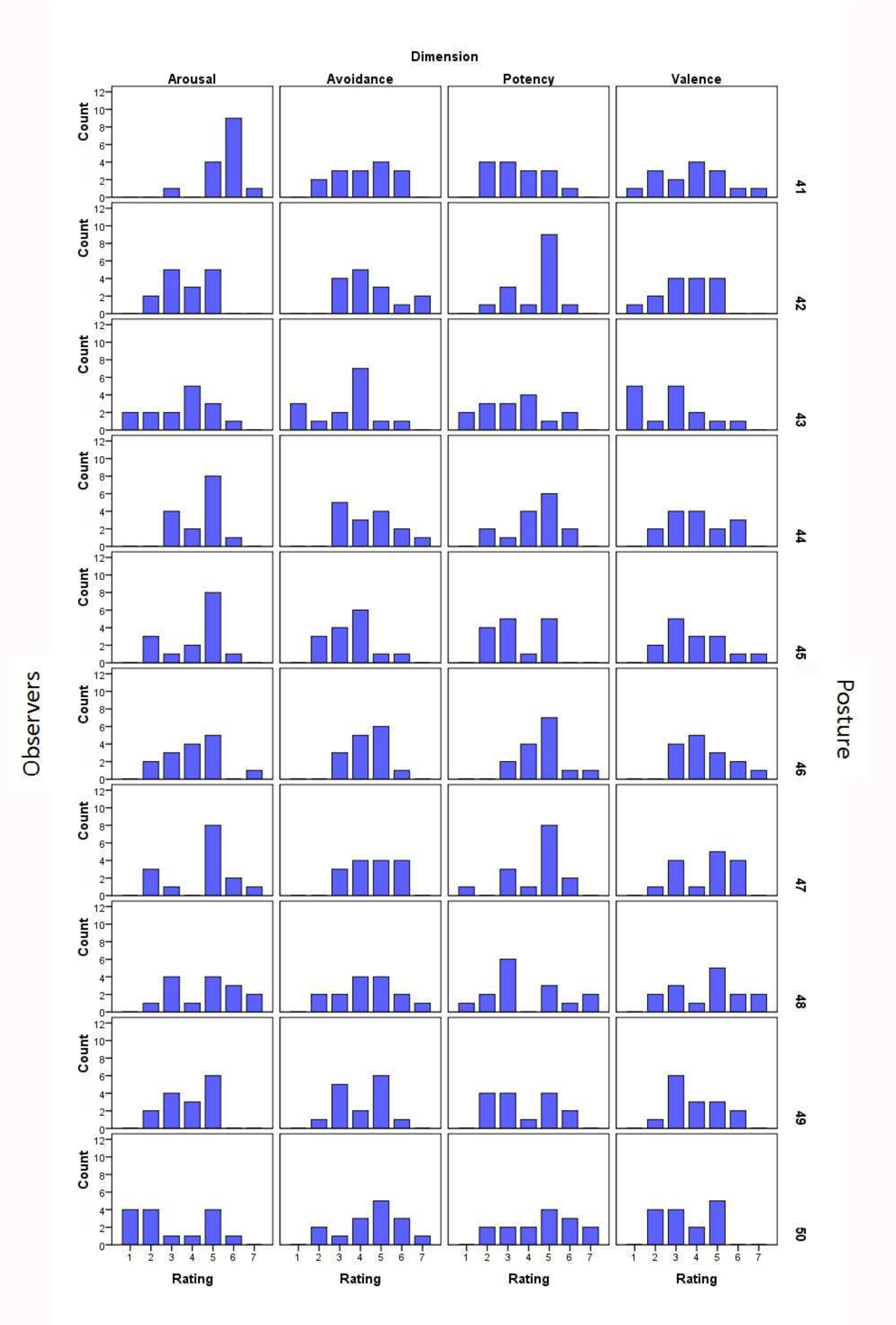
Affective Dimensions Overview of the Non-Acted Postures Survey

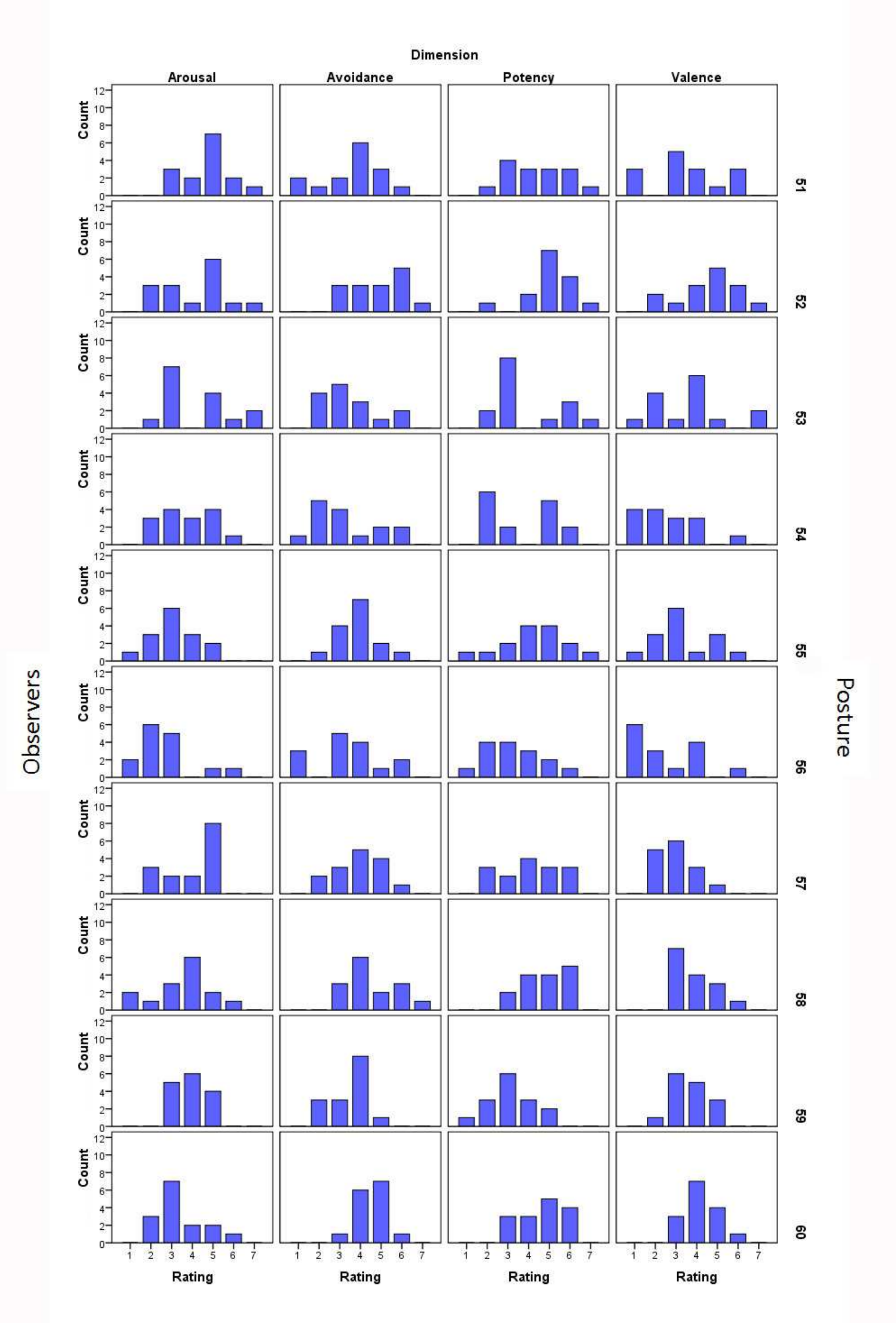

Dimension
Arousal
Avoidance
Potency
Valence
Count
0
2
4
6
8
10
12
Rating
1
2
3
4
5
6
7
01
02
03
04
05
06
07
08
09
10
Observers
Posture

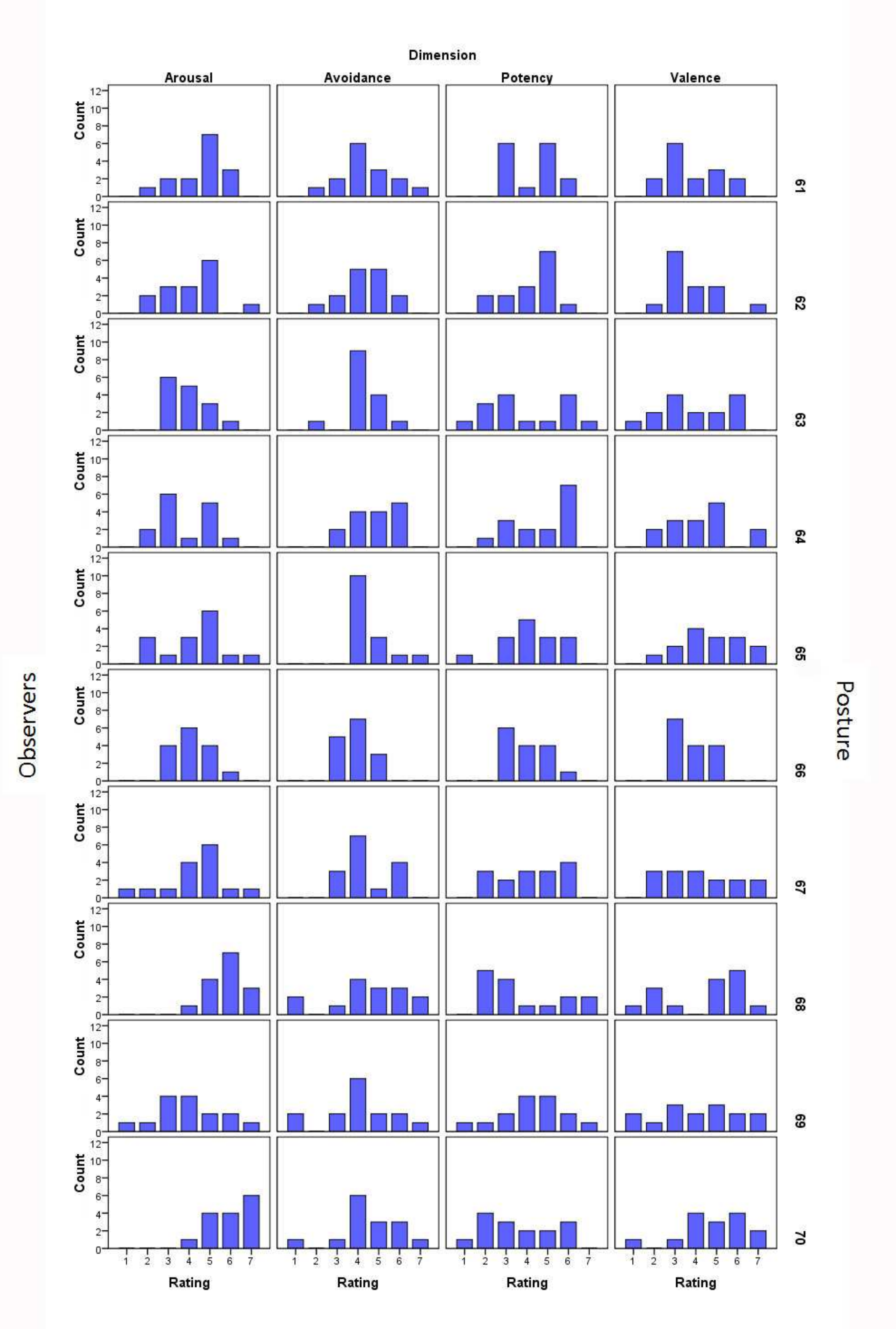

Dimension
Arousal
Avoidance
Potency
Valence
Count
0
2
4
6
8
10
12
Rating
1
2
3
4
5
6
7
11
12
13
14
15
16
17
18
19
20
Observers
Posture

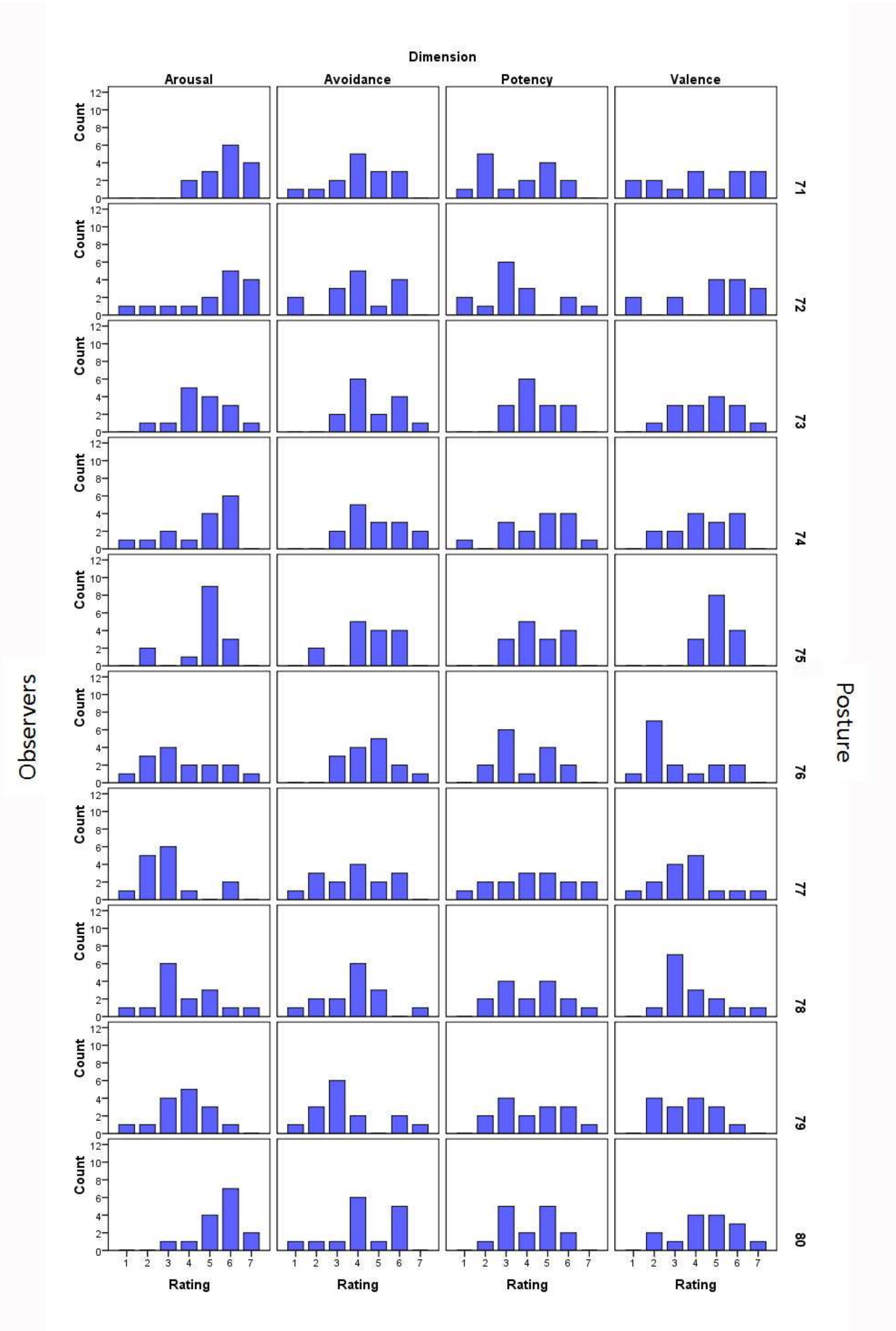

Dimension
Arousal
Avoidance
Potency
Valence
Count
0
2
4
6
8
10
12
Rating
1
2
3
4
5
6
7
21
22
23
24
25
26
27
28
29
30
Observers
Posture

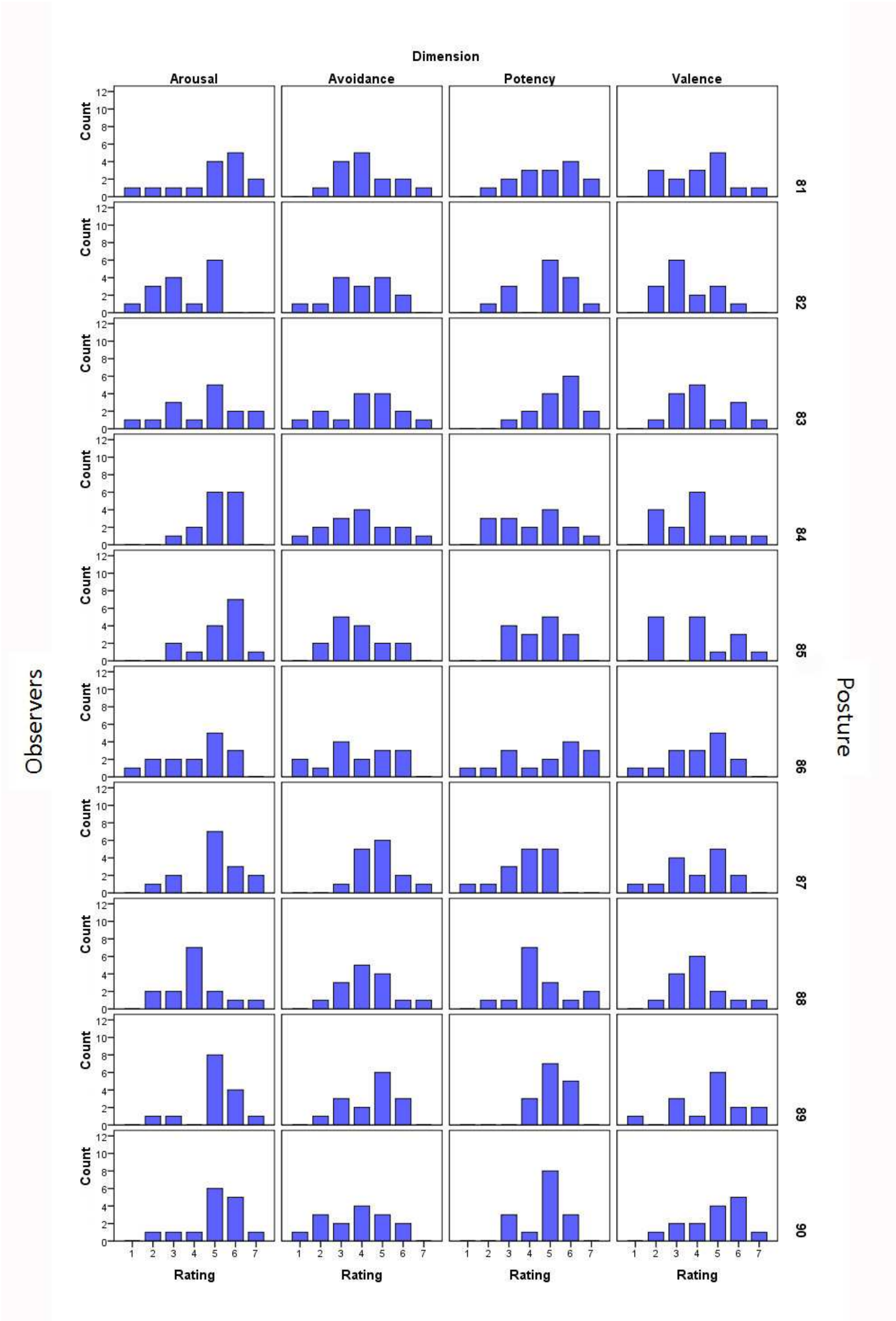

Dimension
Arousal
Avoidance
Potency
Valence
Count
0
2
4
6
8
10
12
31
32
33
34
35
36
37
38
39
40
1 2 3 4 5 6 7
Rating
Observers
Posture

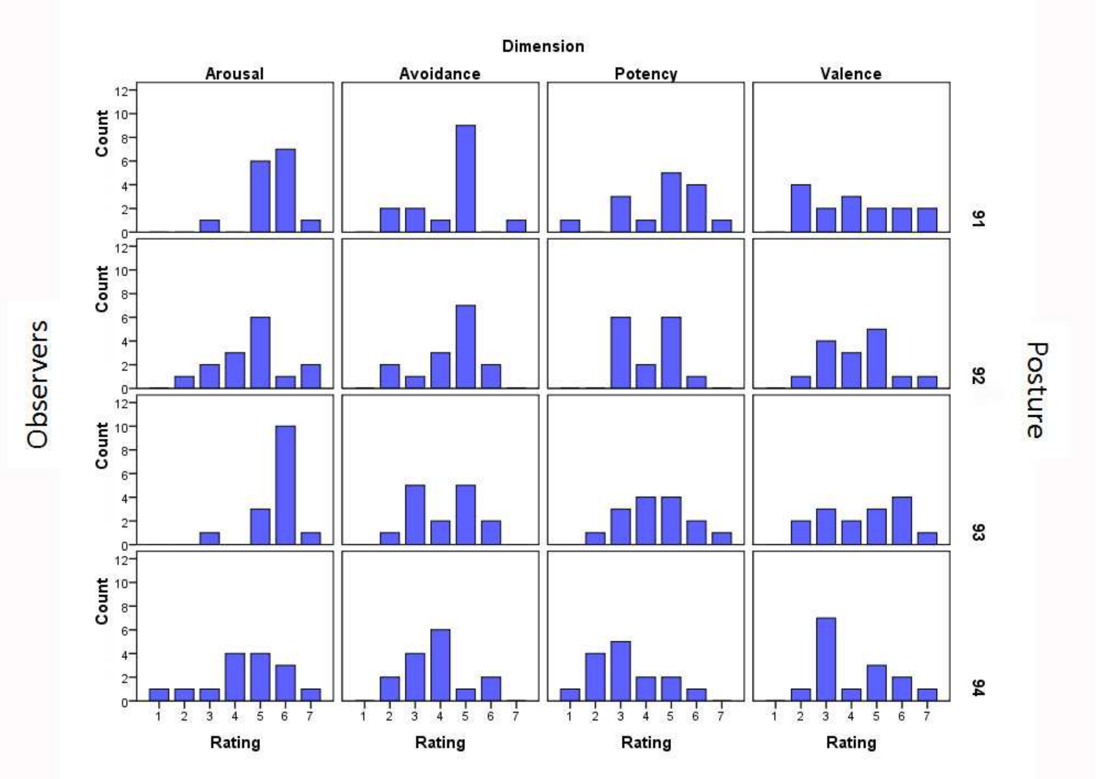

Dimension
Arousal
Avoidance
Potency
Valence
Count
0
2
4
6
8
10
12
Rating
1
2
3
4
5
6
7
41
42
43
44
45
46
47
48
49
50
Observers
Posture


Dimension
Arousal
Avoidance
Potency
Valence
Count
Rating
Observers
Posture
51
52
53
54
55
56
57
58
59
60
0
2
4
6
8
10
12
1
2
3
4
5
6
7


Dimension
Arousal
Avoidance
Potency
Valence
Count
0
2
4
6
8
10
12
Rating
1
2
3
4
5
6
7
61
62
63
64
65
66
67
68
69
70
Observers
Posture


Dimension
Arousal
Avoidance
Potency
Valence
Count
0
2
4
6
8
10
12
71
72
73
74
75
76
77
78
79
80
1
2
3
4
5
6
7
Rating
Observers
Posture


Dimension
Arousal
Avoidance
Potency
Valence
Count
0
2
4
6
8
10
12
Rating
1
2
3
4
5
6
7
81
82
83
84
85
86
87
88
89
90
Observers
Posture


Dimension
Arousal
Avoidance
Potency
Valence
Count
0
2
4
6
8
10
12
Rating
1
2
3
4
5
6
7
91
92
93
94
Observers
Posture

Appendix G

Glossary of Terms

Actors: The motion capture participants recruited for the acted study. They are not professional actors. The term is used to refer to the fact that the participants were explicitly asked to act out specific emotions through whole body postures.

Affect: *"Behavior that expresses a subjectively experienced feeling state (emotion); affect is responsive to changing emotional states, whereas mood refers to a pervasive and sustained emotion. A subjective feeling or emotional tone often accompanied by bodily expressions noticeable to others"* [psy08]. Cognitive states are considered under the umbrella of affect and are deemed possible for the automatic recognition of affect.

Affective computing: A multidisciplinary field of research concerned with *"computing that relates to, arises from, or deliberately influences emotions"* [Pic97].

Affective posture: A bodily configuration through which an affective state is displayed.

Agreement: Pertains to the human recognition of affect. The level to which the judgments made by one subset of observers matches the judgments made by a second subset of observers for the set of postures.

Apex: The most expressive, static instant of the postures.

Benchmark: Defined in this research, it is the average agreement across a number of trials between any two sets of observers that can be created using the entire pool of observers.

Configuration model: Used in the motion capture process, the actor and player configuration models define the arrangement and size of the individual's body. Its purpose is to fit each of the motion captures to the size and form of the actor's body.

Emotion: *"A conscious mental reaction (as anger or fear) subjectively experienced as strong feeling usually directed toward a specific object and typically accompanied by physiological and behavioral changes in the body"* [Mer07].

Ground truth label: An affective label assigned to a posture by a single observer or a group of observers. For a group of observers, the ground truth label is taken as the most frequent label in the discrete affective categories cases, and the median rating in the affective dimensions cases.

Observers: The participants of the posture judgment surveys for the acted, basic emotions study and the non-acted, non-basic affective states study.

Players: The motion capture participants recruited for the non-acted studies involving video game play.

Posture: *"The relative disposition of the various parts of something; esp. the position and carriage of the limbs or the body as a whole, often as indicating a particular quality, feeling, etc.; an attitude, a pose"* [oed08].

Recognition: Pertains to the automatic recognition of affect. The level at which an automatic model classifies the set of postures.

Replay window: The period of time of motion captured video game play in which the player views a replay of the point just played. It is during these sections that postural displays of affect are thought to most likely to occur.

Static posture: A single frame of motion capture data.

Bibliography

[ADBM05] S. Abrilian, L. Devillers, S. Buisine, and J-C Martin. EmoTV1: Annotation of real-life emotions for the specification of multimodal affective interfaces. In *HCI International*, New Orleans, 2005.

[ADGY07] A.P. Atkinson, W.H. Dittrich, A.J. Gemmell, and A.W. Young. Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures. *Cognition*, 104:59–72, 2007.

[ADK$^+$02] J. Ang, R. Dhillon, A. Krupski, E. Shriberg, and A. Stolcke. Prosody-based automatic detection of annoyance and frustration in human-computer dialog. In *Proceedings ICSLP*, Denver, Colorado, 2002.

[Ado02] R. Adolphs. Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, 12(2):169–177, 2002.

[ALC$^+$09] A. Ashraf, S. Lucey, J. Cohn, T. Chen, K. Prkachin, and P. Solomon. The painful face: Pain expression recognition using active appearance models. *Image and Vision Computing*, 27:1788–1796, 2009.

[AM94] P.A. Alexander and P.K. Murphy. The research based for APA's learner-centered psychological principles. In *Paper presented at the American Educational Research Association*, New Orleans, 1994.

[Ani07] Animazoo. *Gypsy 5 Motion Capture System.* http://www.animazoo.com/, Retrieved November 2007.

[AR09] S. Afzal and P. Robinson. Natural affect data – collection and annotation in a learning context. In *Proceedings of the Third International Conference on Affective Computing and Intelligent Interaction*, pages 22–28, 2009.

[Arg88] M. Argyle. *Bodily Communication.* Methuen & Co. Ltd, London, 1988.

[Arn60] M.B. Arnold. *Emotion and Personality.* Columbia University Press, New York, 1960.

[ATD04] A.P. Atkinson, M.L. Tunstall, and W.H. Dittrich. Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*, 33:717–746, 2004.

[Aut08] Autodesk. *Autodesk 3ds Max.* http://usa.autodesk.com/, Retrieved October 2008.

[AWH92] J. Aronoff, B.A. Woike, and L.M. Hyman. Which are the stimuli in facial displays of anger and happiness? Configurational bases of emotion recognition. *Journal of Personality and Social Psychology*, 62:1050–1066, 1992.

[AWH09] N. Amir, A. Weiss, and R. Hadad. Is there a dominant channel in perception of emotions? In *Proceedings of the Third International Conference on Affective Computing and Intelligent Interaction*, pages 241–246, 2009.

[BBK03] N. Bianchi-Berthouze and A. Kleinsmith. A categorical approach to affective gesture recognition. *Connection Science special issue on Epigenetic Robotics - Modeling Cognitive Development in Robotic Systems*, 15(4):259–269, 2003.

[BBKP07] N. Bianchi-Berthouze, W.W. Kim, and D. Patel. Does body movement engage you more in digital game play? and why? In *LNCS: Proceedings of the Second International Conference on Affective Computing and Intelligent Interaction*, pages 102–113, 2007.

[BCGWH08] S. Baron-Cohen, O. Golan, S. Wheelright, and J. Hill. *The Mindreading DVD.* http://www.jkp.com/mindreading/, Retrieved October 2008.

[BCMS99] M. Banerjee, M. Capozzoli, L. McSweeney, and D. Sinha. Beyond kappa: A review of interrater agreement measures. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 27(1):3–23, 1999.

[BFH$^+$03] A. Batliner, K. Fischer, R. Huber, J. Spilker, and E. Noth. How to find trouble in communication. *Speech in Communication*, 40:117–143, 2003.

[BHS$^+$04] A. Batliner, C. Hacker, S Steidl, E. Noth, and J. Haas. From emotion to interaction: Lessons learned from real human-machine dialogues. In *Proceedings of Workshop on Affective Dialogue Systems 2004*, pages 1–12. Springer-Verlag, 2004.

[BJM$^+$05] J.K. Burgoon, M.L. Jensen, T.O. Meservy, J. Kruse, and J.F. Nunamaker. Augmenting human identification of emotional states in video. In *Proceedings of the International Conference on Intelligent Data Analysis*, 2005.

[BL94] M.M. Bradley and P.J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49–59, 1994.

[BLF$^+$05] M.S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Recognizing facial expression: Machine learning and application to spontaneous behavior. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2*, pages 568–573. IEEE Computer Society, 2005.

[BLF$^+$06] M.S. Bartlett, G. Littlewort, M. Frank, C. Lainscseki, I. Fasel, and J. Movellan. Fully automatic facial action recognition in spontaneous behavior. In *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, pages 223–230, 2006.

[BNS75] M.H. Bond, H. Nakazato, and D. Shiraishi. Universality and distinctiveness in dimensions of Japanese person perception. *Journal of Cross-Cultural Psychology*, 6:346–357, 1975.

[BR07] D. Bernhardt and P. Robinson. Detecting affect from non-stylised body motions. In *LNCS: Proceedings of the Second International Conference on Affective Computing and Intelligent Interaction*, pages 59–70. Springer-Verlag, 2007.

[Bre03] C. Breazeal. Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, 59:119–155, 2003.

[Bul87] P.E. Bull. *Posture and Gesture.* Pergamon, Oxford, 1987.

[Cam02] N. Campbell. Recording and storing of speech data. In *Proceedings of the 4th Internatioanl Conference on Learning Resources and Evaluation*, 2002.

[CDCC05] R. Cowie, E. Douglas-Cowie, and C. Cox. Beyond emotion archetypes: Databases for emotion modelling using neural networks. *Neural Networks*, 18:371–388, 2005.

[Cen08] Center for the Study of Emotion and Attention. *The International Affective Picture System.* http://csea.phhp.ufl.edu/media/iapsmessage.html, Retrieved October 2008.

[CG07] E. Crane and M. Gross. Motion capture and emotion: Affect detection in whole body movement. In *LNCS: Proceedings of the Second International Conference on Affective Computing and Intelligent Interaction*, pages 95–101. Springer-Verlag, 2007.

[CK98] J.F. Cohn and G.S. Katz. Bimodal expression of emotion by face and voice. In *Proceedings of the 6th ACM Intemational Multimedia Conference on Face Gesture Recognition and Their Applications*, pages 41–44, 1998.

[CLV03] A. Camurri, I. Lagerlof, and G. Volpe. Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques. *International Journal of Human-Computer Studies*, 59(1-2):213–225, 2003.

[CMK$^+$06] G. Caridakis, L. Malatesta, L. Kessous, N. Amir, A. Raouzaiou, and K. Karpouzis. Modeling naturalistic affective states via facial and vocal expressions recognition. In *ICMI '06: Proceedings of the 8th International Conference on Multimodal Interfaces*, pages 146–154, New York, USA, 2006. ACM.

[CMR$^+$04] A. Camurri, B. Mazzarino, M. Ricchetti, R. Timmers, and G. Volpe. Multimodal analysis of expressive gesture in music and dance performances. In *Gesture-based Communication in Human-Computer Interaction*, pages 20–39, 2004.

[CNSD93] C. Cruz-Neira, D.J. Sandin, and T.A. DeFanti. Surround-screen projection-based virtual reality: The design and implementation of the CAVE. In *Proceedings of ACM SIGGRAPH*, pages 135–142, 1993.

[Coh60] J. Cohen. A coefficient of agreement for nominal scale. *Education and Psychological Measurement*, 20:3746, 1960.

[Cou04] M. Coulson. Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. *Journal of Nonverbal Behavior*, 28:117–139, 2004.

[Cov93] M.V. Covington. A motivational analysis of academic life in college. *Higher Education: Handbook of theory and research*, 9:50–93, 1993.

[CPM$^+$07] R. Colombo, F. Pisano, A. Mazzone, C. Delconte, S. Micera, M.C. Carrozza, P. Dario, and G. Minuco. Design strategies to improve patient motivation during robot-aided rehabilitation. *Journal of Neuroengineering Rehabilitation*, 4(3), 2007.

[CPM+09] C. Clavel, J. Plessier, J-C. Martin, L. Ach, and B. Morel. Combining facial and postural expressions of emotions in a virtual character. In *LNCS: Proceedings of the Ninth International Conference on Intelligent Virtual Agents*, pages 387–300. Springer-Verlag, 2009.

[Cro51] L.J. Cronbach. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3):297–334, 1951.

[CTV02] A. Camurri, R. Trocca, and G. Volpe. Interactive systems design: A KANSEI-based approach. In *NIME '02: Proceedings of the 2002 Conference on New Interfaces for Musical Expression*, pages 1–8, 2002.

[Dam94] A.R. Damasio. *Descartes' Error: Emotion, Reason and the Human Brain.* Avon Books, New York, 1994.

[Dar72] C. Darwin. *The expression of the emotions in man and animals.* Murray, London, 1872.

[Dav64] J.R. Davitz. Auditory correlates of vocal expression of emotional feeling. In J.R. Davitz, editor, *In The Communication of emotional Meaning*, pages 101–112, New York, 1964. McGraw-Hill.

[DBH+99] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, and T.J. Sejnowski. Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(10):974–989, 1999.

[DCCC03] E. Douglas-Cowie, N. Campbell, and R.P. Cowie. Emotional speech: Towards a new generation of databases. *Speech Communication*, 40(1-2):33–60, 2003.

[DCCS+07] E. Douglas-Cowie, R. Cowie, I. Sneddon, C. Cox, L. Lowry, M. McRorie, J.C. Martin, L. Devillers, S. Abrilian, A. Batliner, N. Amir, and K. Karpouzis. The HUMAINE database: Addressing the collection and annotation of naturalistic and induced emotional data. In *Proceedings of the Second In-*

ternational Conference on Affective Computing and Intelligent Interaction, pages 488–500. Springer-Verlag, 2007.

[DCSG06] K. D'Mello, S.D. Craig, J. Sullins, and A.C. Graesser. Predicting affective states expressed through an emote-aloud procedure from AutoTutor's mixed-initiative dialogue. *International Journal of Artificial Intelligence in Education*, 16:3–28, 2006.

[DG99] J. Decety and J. Grezes. Neural mechanisms subserving the perception of human actions. *Trends in Cognitive Sciences*, 3:172–178, 1999.

[dG06] B. de Gelder. Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience*, 7(3):242–249, 2006.

[dG09] B. de Gelder. Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society*, 364(3):3475–3484, 2009.

[dM89] M. de Meijer. The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior*, 13:247–268, 1989.

[DM06] P. Dunn and A. Marinetti. *Cultural adaptation: Necessity for eLearning.* http://www.linezine.com/7.2/articles/pdamca.htm, Retrieved January 2006.

[Doy04] J. Doyle. Prospects for preferences. *Computational Intelligence*, 20(2):111–136, 2004.

[DPW96] F. Dellaert, T. Polzin, and A. Waibel. Recognizing emotion in speech. In *Proceedings of ICSLP 1996*, pages 1970–1973, 1996.

[EA02] H. Elfenbein and N. Ambady. On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, 128:205–235, 2002.

[ECT98] G.J. Edwards, T.F. Cootes, and C.J. Taylor. Face recognition using active appearance models. In *Proceedings of European Conference on Computer Vision*, pages 581–695, 1998.

[EF67] P. Ekman and W. Friesen. Head and body cues in the judgment of emotion: A reformulation. *Perceptual and Motor Skills*, 24:711–724, 1967.

[EF69a] P. Ekman and W. Friesen. Nonverbal leakage and clues to deception. *Psychiatry*, 32:88–105, 1969.

[EF69b] P. Ekman and W. Friesen. The repertoire of non-verbal behavioral categories: Origins, usage and coding. *Semiotica*, 1:49–98, 1969.

[EF74] P. Ekman and W. Friesen. Detecting deception from the body or face. *Journal of Personality and Social Psychology*, 29(3):288–298, 1974.

[EF75] P. Ekman and W. Friesen. *Unmasking the Face: A Guide to Recognizing Emotions from Facial Expressions.* Prentice Hall, 1975.

[EF76] P. Ekman and W. Friesen. *Pictures of facial affect.* Consulting Psychologists Press, Palo Alto, CA, 1976.

[EF78] P. Ekman and W. Friesen. *Manual for the facial action coding system.* Consulting Psychology Press, Palo Alto, Ca, 1978.

[EF82] P. Ekman and W. Friesen. Felt, false and miserable smiles. *Journal of Nonverbal Behavior*, 6(4):238–252, 1982.

[EF08] P. Ekman and W.V. Friesen. *Pictures of Facial Affect.* http://www.paulekman.com/researchproducts.phpt, Retrieved October 2008.

[EFE82] P. Ekman, W.V. Friesen, and P. Ellsworth. What emotion categories or dimensions can observers judge from facial behavior? *Emotion in the Human Face, P. Ekman (Ed.)*, pages 39–55, 1982.

[eK05] R. el Kaliouby. Mind-reading machines: Automated inference of complex mental states. Tech report 636, University of Cambridge, Cambridge, 2005.

[Ekm94] P. Ekman. Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Bulletin*, 115:268–287, 1994.

[eKR04] R. el Kaliouby and P. Robinson. Real-time inference of complex mental states from facial expressions and head gestures. In *Proceedings of the IEEE International Workshop on Real Time Computer Vision for Human Computer Interaction at CVPR*, 2004.

[EMA$^+$02] H. A. Elfenbein, M. K. Mandal, N. Ambady, S. Harizuka, and S. Kumar. Cross-cultural patterns in emotion recognition: Highlighting design and analytical techniques. *Emotion*, 2(1):75–84, 2002.

[Fel04] L. Feldman Barrett. Feelings or words? Understanding the content in self-report ratings of emotional experience. *Journal of Personality and Social Psychology*, 87:266–281, 2004.

[Fel06] L. Feldman Barrett. Valence as a basic building block of emotional life. *Journal of Research in Personality*, 40:35–55, 2006.

[FH05] J. Fürnkranz and E. Hüllermeier. Preference learning. *Kunstliche Intelligenz*, 19(1):60–61, 2005.

[Fie05] A. Field. *Discovering Statistics Using SPSS.* Sage, London, 2005.

[Fin72] B.M. Finifter. The generation of confidence: Evaluating research findings by random subsample replication. *Sociological Methodology*, 4:112–175, 1972.

[Fle71] J.L. Fleiss. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 6(5):378–382, 1971.

[Fox08] E. Fox. *Emotion Science.* Palgrave Macmillan, Basingstoke, 2008.

[FR84] B. Fehr and J.A. Russell. Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General*, 113(3):464–486, 1984.

[Fre03] D. Freeman. *Creating Emotions in Games.* New Riders, San Diego, 2003.

[Fri53] N.H. Frijda. The understanding of facial expression of emotion. *Acta Psychologica*, 9:294–362, 1953.

[Fri72] W. Friesen. *Cultural Differences in Facial Expressions in a Social Situation: An Experimental Test of the Concept of Display Rules.* Doctoral dissertation, University of California, San Francisco, 1972.

[Fri88] N.H. Frijda. The laws of emotion. *American Psychologist*, 43(5):349–358, 1988.

[FSRE07] J.R.J. Fontaine, K.R. Scherer, E.B. Roesch, and P.C. Ellsworth. The world of emotions is not two-dimensional. *Psychological Science*, 18(12):1050–1057, 2007.

[FSS$^+$00] D. France, R. Shiavi, S. Silverman, M. Silverman, and D. Wilkes. Acoustical properties of speech as indicators of depression and suicidal risk. *IEEE Transactions on Biomedical Engineering*, 47(7):829–837, 2000.

[FT05] N. Fragopanagos and J.G. Taylor. Emotion recognition in human-computer interaction. *Neural Networks*, 18(4):389–405, 2005.

[GD04] K.M. Gilleade and A. Dix. Using frustration in the design of adaptive videogames. In *ACE '04: Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, pages 228–232, New York, 2004. ACM.

[GP03] M.A. Giese and T. Poggio. Neural mechanisms for the recognition of biological movements. *Neuroscience*, 4:179–191, 2003.

[GP06] H. Gunes and M. Piccardi. Observer annotation of affective display and evaluation of expressivity: Face vs. face-and-body. In *Proceedings of the HCSNet Workshop on Use of Vision in Human-Computer Interaction*, pages 35–42. Australian Computer Society, Inc., 2006.

[GP07] H. Gunes and M. Piccardi. Bi-modal emotion recognition from expressive face and body gestures. *Journal of Network and Computer Applications*, 30:1334–1345, 2007.

[Gra82] J.A. Gray. *The Neuropsychology of anxiety.* Oxford University Press, Oxford, 1982.

[Har60] H.H. Harman. *Modern Factor Analysis.* University of Chicago Press, Chicago, 1960.

[Hay99] S. Haykin. *Neural Networks: A Comprehensive Foundation.* Prentice Hall, New Jersey, 1999.

[HFH$^+$09] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten. The WEKA data mining software: An update. *SIGKDD Explorations*, 11(1), 2009.

[HH97] C.L. Huang and Y.M. Huang. Facial expression recognition using model-based feature extraction and action parameters classification. *Journal of Visual Communication and Image Representation*, 8(3):278–290, 1997.

[HH06] M. Hirai and K. Hiraki. The relative importance of spatial versus temporal structure in the perception of biological motion: An event-related potential study. *Cognition*, 99:B15–B29, 2006.

[HNvdM98] H. Hong, H. Neven, and C. von der Malsburg. Online facial expression recognition based on personalized galleries. In *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, pages 354–359, 1998.

[Hod08] J. Hodgins. *CMU graphics lab motion capture database.* http://mocap.cs.cmu.edu/, Retrieved December, 2008.

[Hof06] G. Hofstede. *Geert Hofstede Cultural Dimensions.* http://www.geert-hofstede.com/, Retrieved January 2006.

[HR83] J.A. Harrigan and R. Rosenthal. Physicians head and body positions as determinants of perceived rapport. *Journal of of Applied Social Psychology*, 13(6):496–509, 1983.

[Hum08] Humaine. *The Humaine Portal.* http://emotion-research.net/projects/humaine/toolbox/, Retrieved May 2008.

[Hut87] A. Hutchinson. *Labanotation.* Theatre Arts Book, 1987.

[IASF02] C.E. Izard, P.B. Ackerman, K.M. Schoff, and S.E. Fine. Self-organization of discrete emotions, emotion patterns, and emotion-cognition relations. In D.L. Marc and I. Granic, editors, *Emotion, Development, and Self-Organization: Dynamic Systems Approaches to Emotional Development*, pages 15–36, Cambridge, 2002. Cambridge University Press.

[Iza71] C.E. Izard. *The Face of Emotion.* Appleton-Century-Crofts, New York, 1971.

[Iza07] C.E. Izard. Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives in Psychological Science*, 2(3):260–280, 2007.

[Jam84] W. James. What is an emotion? *Mind*, 9:188–205, 1884.

[Jam32] W.T. James. A study of the expression of bodily posture. *Journal of General Psychology*, 7:405–437, 1932.

[JG93] D.H. Jonassen and B.L. Grabowski. *Handbook of individual differences, learning, and instruction.* Erlbaum, Hillsdale, N.J., 1993.

[JL02] P. Juslin and P. Laukka. Communication of emotions in vocal expression and music performance. *Psychological Bulletin*, 129(5):770–814, 2002.

[Kai60] H.F. Kaiser. The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20:141–151, 1960.

[KBP07] A. Kapoor, W. Burleson, and R.W. Picard. Automatic prediction of frustration. *International Journal of Human-Computer Studies*, 65(8):724–736, August 2007.

[KCT00] T. Kanade, J.F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 46–53, 2000.

[KdSBB06] A. Kleinsmith, P.R. de Silva, and N. Bianchi-Berthouze. Cross-cultural differences in recognizing affect from body posture. *Interacting with Computers*, 18:1371–1389, 2006.

[KEGB03] D. Keltner, P. Ekman, G. C. Gonzaga, and J. Beer. Facial expression of emotion. In R. Davidson, K. Scherer, and H. Goldsmith, editors, *Handbook of Affective Sciences*, New York, 2003. Oxford University Press.

[KF07] V. Kostov and S. Fukuda. Emotion in user interface, voice interaction system. In T.S. Huang, A. Nijholt, M. Pantic, and A. Pentland, editors, *IEEE International Conference on Systems, Man, and Cybernetics*, pages 798–803. IEEE, 2007.

[KFBB05] A. Kleinsmith, T. Fushimi, and N. Bianchi-Berthouze. An incremental and interactive affective posture recognition system. In *User Modeling 2005 Workshop: Adapting the Interaction Style to Affective Factors.* http://www.di.uniba.it/intint/UM05/list-ws-um05.html, 2005.

[KFTBB03] A. Kleinsmith, T. Fushimi, H. Takenaka, and N. Bianchi-Berthouze. Towards bi-directional affective human-machine interaction. *Special issue of the Journal of 3D-Forum Society*, 17(4):61–66, 2003.

[KGKW05] S. Kopp, L. Gesellensetter, N. Kramer, and I. Wachsmuth. A conversational agent as museum guide – design and evaluation of a real-world application. In J. Tao, T. Tan, and R. Picard, editors, *Intelligent Virtual Agents*, pages 329–343. Springer-Verlag, 2005.

[KJK81] P.R. Kleinginna Jr and A.M. Kleinginna. A categorised list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion*, 5(4):345–379, 1981.

[KKVB$^+$05] A. Kapur, A. Kapur, N. Virji-Babul, G. Tzanetakis, and P.F. Driessen. Gesture-based affective computing on motion capture data. In *Proceedings of the First International Conference on Affective Computing and Intelligent Interaction*, pages 1–7. Springer-Verlag, 2005.

[KLG08] M. Kamachi, M. Lyons, and J. Gyoba. *The Japanese Female Facial Expression Database.* http://www.kasrl.org/jaffe.html, Retrieved October 2008.

[KM85] T. Kudoh and D. Matsumoto. Cross-cultural examination of the semantic dimensions of body postures. *Journal of Personality and Social Psychology*, 48(6):1440–1446, 1985.

[KMP01] A. Kapoor, S. Mota, and R.W. Picard. Towards a learning companion that recognizes affect. Tech report 543, MIT Media Laboratory, 2001.

[KOIH04] S. Kamisato, S. Odo, Y. Ishikawa, and K. Hoshino. Extraction of motion characteristics corresponding to sensitivity information using dance movement. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 8(2):167–178, 2004.

[KPI04] A. Kapoor, R.W. Picard, and Y. Ivanov. Probabilistic combination of multiple modalities to detect interest. *Proceedings of the 17th International Conference on Pattern Recognition*, 3:969–972, August 2004.

[KS00] M. Kienast and W.F. Sendlmeier. Acoustical analysis of spectral and temporal changes in emotional speech. In R. Cowie, E. Douglas, and M. Schroeder, editors, *Speech and emotion: Proceedings of the ISCA workshop*, pages 92–97, 2000.

[LAJ05] M.Y. Lim, R. Aylett, and C.M. Jones. Affective guide with attitude. In J. Tao, T. Tan, and R. Picard, editors, *First International Conference on Affective Computing and Intelligent Interaction*, pages 72–79. Springer-Verlag, 2005.

[Lan80] P.J. Lang. Behavioral treatment and bio-behavioral assessment: Computer applications. *Technology in mental health care delivery systems*, pages 119–l37, 1980.

[Laz82] R.S. Lazarus. Thoughts on the relations between emotion and cognition. *American Psychologist*, 37(2):1019–1024, 1982.

[Laz91] R.S. Lazarus. *Emotion and Adaptation.* Oxford University Press, New York, 1991.

[Laz04] N. Lazzaro. Why we play games: Four keys to more emotion without story. Technical report, XEODesign, Inc., March 2004.

[LBA99] M.J. Lyons, J. Budynek, and S. Akamatsu. Automatic classification of single facial images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(12):1357–1362, 1999.

[Lev94] R.W. Levenson. Human emotion: A functional view. *The Nature of Emotion, P. Ekman and R.J. Davidson (Eds.)*, 1994.

[LH97] L. Leinonen and T. Hiltunen. Expression of emotional-motivational connotations with a one-word utterance. *Journal of the Acoustical Society of America*, 102(3):1853–1863, 1997.

[LK77] J.R. Landis and G.G Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977.

[LLCBB08] S. Lindley, J. Le Couteur, and N. Bianchi-Berthouze. Stirring up experience through movement in game play: Effects on engagement and social behaviour. In *Proceedings of the Workshop on Exertion Interfaces at CHI 2008*. ACM, 2008.

[LNP02] C. Lee, S. Narayanan, and R. Pieraccini. Classifying emotions in human-machine spoken dialogs. In *Proceedings of the International Conference on Multimedia and Expo*, 2002.

[LR78] M. Lewis and L.A. (Eds.) Rosenblum. *The development of affect*. Plenum Press, New York, 1978.

[Mat05] D. Matsumoto. Culture and nonverbal behavior. In V. Manusov and M. Patterson, editors, *Handbook of Nonverbal Communication*, Thousand Oaks, CA, 2005. Sage.

[McD26] W. McDougall. *An Introduction to Social Psychology*. Luce, Boston, 1926.

[ME72] A. Mehrabian and N. Epstein. A measure of emotional empathy. *Journal of Personality*, 40:525–543, 1972.

[Meh68] A. Mehrabian. Inference of attitude from the posture, orientation, and distance of a communicator. *Journal of Consulting and Clinical Psychology*, 32:296–308, 1968.

[Mer07] Merriam-Webster Online Dictionary. http://www.m-w.com/dictionary/emotion, Retrieved December 2007.

[Mes03] B. Mesquita. Emotions as dynamic cultural phenomena. In R. Davidson, K. Scherer, and H. Goldsmith, editors, *Handbook of Affective Sciences*, New York, 2003. Oxford University Press.

[MF69] A. Mehrabian and J. Friar. Encoding of attitude by a seated communicator via posture and position cues. *Journal of Consulting and Clinical Psychology*, 33:330–336, 1969.

[MK83] D. Matsumoto and H. Kishimoto. Developmental characteristics in judgments of emotion from nonverbal vocal cues. *International Journal of Intercultural Relations*, 7:415–424, 1983.

[MK87] D. Matsumoto and T. Kudoh. Cultural similarities and differences in the semantic dimensions of body postures. *Journal of Nonverbal Behavior*, 11(3):166–179, 1987.

[MN89] L. McClenney and R. Neiss. Post-hypnotic suggestion: A method for the study of nonverbal communication. *Journal of Nonverbal Behavior*, 13:37–45, 1989.

[MPP06] Y. Ma, H.M. Paterson, and F.E. Pollick. A motion capture library for the study of identity, gender, and emotion perception from biological motion. *Behavior Research Methods*, 38(1):134–141, 2006.

[MR74] A. Mehrabian and J. Russell. *An Approach to Environmental Psychology.* MIT Press, Cambridge, 1974.

[MvHdG05] H. Meeren, C. van Heijnsbergen, and B. de Gelder. Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences of the USA*, 102(45):16518–16523, 2005.

[NC93] S. Neill and C. Caswell. *Body language for competent teachers.* Routledge, London, 1993.

[NNT99] R. Nakatsu, J. Nicholson, and N. Tosa. Emotion recognition and its application to computer agents with spontaneous interactive capabilities. In *Proceedings of the 3rd ACM Conference on Creativity Cognition*, pages 135–143, 1999.

[NR05] T. Nef and R. Riener. Armin - design of a novel arm rehabilitation robot. *Rehabilitation Robotics. 9th International Conference on*, pages 57–60, 2005.

[oed08] *Oxford English Dictionary.* http://dictionary.oed.com/entrance.dtl, Retrieved November, 2008.

[OJL87] K. Oatley and P.N. Johnson-Laird. Towards a cognitive theory of emotions. *Cognition & Emotion*, 1:29–50, 1987.

[OMM75] C.E. Osgood, W.H. May, and M.S. Miron. *Cross-cultural universals of affective meaning.* University of Illinois Press, Urbana, 1975.

[OST57] C.E. Osgood, G.J. Suci, and P.H. Tannenbaum. *The measurement of meaning.* University of Illinois Press, Chicago, 1957.

[OT90] A. Ortony and T.J. Turner. What's basic about basic emotions? *Psychological Review*, 97:315–331, 1990.

[Oud03] P.Y. Oudeyer. The production and recognition of emotions in speech: Features and algorithms. *International Journal of Human-Computer Studies*, 59:157–183, 2003.

[Pan82] J. Panksepp. Towards a general psychobiological theory of emotions. *The Behavioral and Brain Sciences*, 5:407–467, 1982.

[PBBvDN09] M. Pasch, N. Bianchi-Berthouze, B. van Dijk, and A. Nijholt. Movement-based sports video games: Investigating motivation and gaming experience. *Entertainment Computing*, 9(2):169–180, 2009.

[PC96] C. Padgett and G.W. Cottrell. Representing face images for emotion classification. In *Proceedings of Conference on Advances in Neural Information Processing Systems*, pages 894–900, 1996.

[Pet99] V. Petrushin. Emotion in speech: Recognition and application to call centers. In *Proceedings of Artificial Neural Networks in Engineering*, pages 7–10, 1999.

[Pic97] R. Picard. *Affective Computing.* MIT Press, Cambridge, 1997.

[Pic98] R. Picard. Toward agents that recognize emotion. In *Actes Proceedings of IMAGINA*, pages 153–165. Springer-Verlag, 1998.

[pic08] *The Psychological Image Collection at Stirling.* http://pics.psych.stir.ac.uk/, Retrieved October 2008.

[PLRC02] F.E. Pollick, V. Lestou, J. Ryu, and S-B. Cho. Estimating the efficiency of recognizing gender and affect from biological motion. *Vision Research*, 42:2345–2355, 2002.

[Plu80] R. Plutchik. A general psychoevolutionary theory of emotion. *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion, R. Plutchik and H. Kellerman (Eds.)*, pages 3–31, 1980.

[PP07] M. Pasch and R. Poppe. Person or puppet? The role of stimulus realism in attributing emotion to static body postures. In *LNCS: Proceedings of the Second International Conference on Affective Computing and Intelligent Interaction*, pages 83–94. Springer-Verlag, 2007.

[PPB+04] R.W. Picard, S. Papert, W. Bender, B. Blumberg, C. Breazeal, D. Cavallo, T. Machover, M. Resnick, D. Roy, and C. Strohecker. Affective learning - a manifesto. *BT Technology Journal*, 22(4):253–269, 2004.

[PPBS01] F.E. Pollick, H.M. Paterson, A. Bruderlin, and A.J. Sanford. Perceiving affect from arm movement. *Cognition*, 82:51–61, 2001.

[PPJ02] H.M. Paterson, F.E. Pollick, and E. Jackson. Movement and faces in the perception of emotion from motion. *Perception, ECVP Glasgow Supplemental*, 31(118):232–232, 2002.

[PPKW04] H. Park, J. Park, U. Kim, and W. Woo. Emotion recognition from dance image sequences using contour approximation. In *LNCS: Proceedings of the Joint IAPR International Workshops on Structural, Syntactic, and Statistical Pattern Recognition*, pages 547–555. Springer-Verlag, 2004.

[PPM04] F.E. Pollick, H. Paterson, and P. Mamassian. Combining faces and movements to recognize affect [Abstract]. *Journal of Vision*, 4(8):232–232, 2004.

[PPS01] H.M. Paterson, F.E. Pollick, and A.J. Sanford. The role of velocity in affect discrimination. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, pages 756–761. Lawrence Erlbaum Associates, 2001.

[PR00a] M. Pantic and L.J.M. Rothkrantz. Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445, 2000.

[PR00b] M. Pantic and L.J.M. Rothkrantz. Expert system for automatic analysis of facial expression. *Image and Vision Computing Journal*, 18(11):881–905, 2000.

[psy08] *Dictionary of Psychology.* http://dictionary-psychology.com/index.php?$a = term\&d = Dictionary + of + psychology\&t = Affect$, Retrieved November 2008.

[PVH01] R.W. Picard, E. Vyzas, and J. Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10):1175–1191, 2001.

[PVRM05] M. Pantic, M. Valstar, R. Rademaker, and L. Maat. Web-based database for facial expression analysis. *Multimedia and Expo, IEEE International Conference on*, 2005.

[PVRM08] M. Pantic, M. Valstar, R. Rademaker, and L. Maat. *The MMI Face Database.* http://www.mmifacedb.com/, Retrieved October 2008.

[PWD06] M.V. Peelen, A.J. Wiggett, and P.E. Downing. Patterns of fMRI activity dissociate overlapping functional brain areas that respond to biological motion. *Neuron*, 49:815822, 2006.

[RC03] P. Rozin and A.B. Cohen. High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of Americans. *Emotion*, 3(1):68–75, 2003.

[RF99] J.A. Russell and L. Feldman-Barrett. Core affect prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, 76:805–819, 1999.

[RFGF96] G. Rizzolatti, L. Fadiga, V. Gallese, and L. Fogassi. Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3:131–141, 1996.

[RFH82] R.H. Rozensky and L. Feldman-Honor. Notation systems for coding nonverbal behavior: A review. *Journal of Behavioral Assessment*, 4(2):119–132, 1982.

[RM97] G. Rugg and P. McGeorge. The sorting techniques: A tutorial paper on card sorts, picture sorts and item sorts. *Expert Systems*, 14(2):80–93, 1997.

[RST$^+$05] N. Ravaja, T. Saari, M. Turpeinen, J. Laarni, M. Salminen, and M. Kivikangas. Spatial presence and emotions during video game playing: Does it matter with whom you play? In *Proceedings of the 8th Annual International Workshop on Presence*, pages 327–333, University College London, 2005.

[Rus80] J.A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

[Rus94] J.A. Russell. Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. *Psychological Bulletin*, 115:102–141, 1994.

[Rus97] J.A. Russell. Reading emotions from and into faces: Resurrecting a dimensional-contextual perspective. *In Russell, J.A. and Fernandez-Dols, J. (Eds.), The Psychology of Facial Expression*, 1997.

[Rus03] J.A. Russell. Core affect and the psychological construction of emotion. *Psychological Review*, 110(1):145–172, 2003.

[RYD94] M. Rosenblum, Y. Yacoob, and L.S. Davis. Human emotion recognition from motion using a radial basis function network architecture. In *Proceedings of the Workshop on Motion of Non-rigid and Articulated Objects*, 1994.

[SB08] T.F. Shipley and J.S. Brumberg. *Markerless motion-capture for point-light displays.* http://astro.temple.edu/ tshipley/mocap/MarkerlessMoCap.pdf, Retrieved December, 2008.

[SCDC$^+$01] M. Schroeder, R. Cowie, E. Douglas-Cowie, M. Westerdijk, and S. Gielen. Acoustic correlates of emotion dimensions in view of speech synthesis. In *Proceedings of Eurospeech 2001*, pages 87–90, 2001.

[SCGH06] N. Sebe, I. Cohen, T. Gevers, and T.S. Huang. Emotion recognition based on joint visual and audio cues. In *Proceedings of the International Conference on Pattern Recognition*, pages 1136–1139, 2006.

[SH98] V. Surakka and J.K. Hietanen. Facial and emotional reactions to Duchenne and non-Duchenne smiles. *International Journal of Psychophysiology*, 29:23–33, 1998.

[SR06] A.S. Sharaani and D.M. Romano. Basic emotions from body movements. In *Proceedings of The First International Symposium on Culture, Creativity and Interaction Design. HCI 2006 Workshops*, London, 2006.

[SSB$^+$04] N. Sebe, Y. Sun, E. Bakker, M.S. Lew, I. Cohen, and T.S. Huang. Towards authentic emotion recognition. In *IEEE International Conference on Systems, Man and Cybernetics*, pages 623–628, 2004.

[Sta96] F.K. Stage. Setting the context: Psychological theories of learning. *Journal of College Student Development*, 37(2):227–235, 1996.

[TCDFBP02] J.L. Tsai, Y. Chentsova-Dutton, L. Freire-Bebeau, and D.E. Przymus. Emotional expression and physiology in European Americans and Hmong Americans. *Emotion*, 2(4):380–397, 2002.

[Tha89] R.E. Thayer. *The biopsychology of mood and arousal.* Oxford University Press, New York, 1989.

[Tom62] S.S. Tomkins. *Affect, Imagery, and Consciousness: Vol 1. The Positive Affects.* Springer, New York, 1962.

[Tom63] S.S. Tomkins. *Affect, Imagery, and Consciousness: Vol 2. The Negative Affects.* Springer, New York, 1963.

[Tom94] S.S. Tomkins. Affect theory. *Approaches to emotion, K.R. Scherer and P. Ekman (Eds.)*, 1994.

[VdSRdG07] J. Van den Stock, R. Righart, and B. de Gelder. Body expressions influence recognition of emotions in the face and voice. *Emotion*, 7(3):487–494, 2007.

[VG05] A.J. Viera and J.M. Garrett. Understanding interobserver agreement: The kappa statistic. *Family Medicine Journal*, 37(5):36–363, 2005.

[vHMGdG07] C.C.R.J. van Heijnsbergen, H.K.M. Meeren, J. Grezes, and B. de Gelder. Rapid detection of fear in body expressions, an ERP study. *Brain Research*, 1186:233–241, 2007.

[Vic07] Vicon. *Vicon Motion Capture Systems.* http://www.vicon.com/, Retrieved November, 2007.

[VK02] D. Vastfjall and M. Kleiner. Emotion in product sound design. In *Proceedings of Journees Design Sonore*, 2002.

[vL63] R. von Laban. *Modern educational dance.* MacDonald & Evans, Ltd., London, 1963.

[vL71] R. von Laban. *The mastery of movement.* MacDonald & Evans Ltd, London, 1971.

[VL89] D.L. Verbyla and J.A. Litvaitis. Resampling methods for evaluating classification accuracy of wildlife habitat models. *Environmental Management*, 13(6):783–787, 1989.

[VV04] J. Vanrie and K. Verfaillie. Perception of biological motion: A stimulus set of human point-light actions. *Behavior Research Methods, Instruments, & Computers*, 36(4):625–629, 2004.

[Wal98] H.G. Wallbott. Bodily expression of emotion. *European Journal of Social Psychology*, 28:879–896, 1998.

[Wat30] J.B. Watson. *Behaviorism.* University of Chicago Press, Chicago, 1930.

[WCT88] D. Watson, L.A. Clark, and A. Tellegen. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6):1063–70, 1988.

[Wie95] A. Wierzbicka. *Emotions across languages and cultures: Diversity and universals.* Cambridge University Press, Cambridge, 1995.

[WS86] H.G. Wallbott and K.R. Scherer. Cues and channels in emotion recognition. *Journal of Personality and Social Psychology*, 51(4):690–699, 1986.

[Wun07] W. Wundt. *Outlines of psychology.* Wilhelm Englemann, Leipzig, 1907.

[Wun73] W. Wundt. *The Language of Gestures.* Mouton & Company, The Hague, 1973.

[Yam08] J. Yamadera. *MOCAPDATA.com.* http://www.mocapdata.com/, Retrieved October, 2008.

[Yan09] G.N. Yannakakis. Preference learning for affective modeling. In *Proceedings of the Third International Conference on Affective Computing and Intelligent Interaction*, pages 126–131, 2009.

[You43] P.T. Young. *Emotion in man and animal: Its nature and relation to attitude and motive.* Wiley, New York, 1943.

[YSLB03] S. Yacoub, S. Simske, X. Lin, and J. Burns. Recognition of emotions in interactive voice response systems. In *Proceedings of Eurospeech*, Geneva, 2003.

[YWS$^+$06] L. Yin, X. Wei, Y. Sun, J. Wang, and M.J. Rosato. A 3D facial expression database for facial behavior research. In *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, pages 211–216, 2006.

[YWS$^+$08] L. Yin, X. Wei, Y. Sun, J. Wang, and M.J. Rosato. *The Binghamton University 3D Facial Expression Database.* http://www.cs.binghamton.edu/ lijun/Research/3DFE/3DFEAnalysis.html, Retrieved October 2008.

[Zaj80] R.B. Zajonc. Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35(2):151–175, 1980.

[ZL65] M. Zuckerman and B. Lubin. *Manual for the Multiple Affect Adjective Check List.* Educational and Industrial Testing Service, San Diego, 1965.

[ZLSA98] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu. Comparison between geometry-based and gabor wavelets-based facial expression recognition using multi-layer perceptron. In *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, pages 454–459, 1998.

[ZTL$^+$04] Z. Zeng, J. Tu, M. Liu, T. Zhang, N. Rizzolo, Z. Zhang, T.S. Huang, D. Roth, and S. Levinson. Bimodal HCI-related affect recognition. In *Proceedings of the 6th International Conference on Multimodal Interfaces*, pages 137–143, New York, NY, USA, 2004. ACM.

[ZTP$^+$05] Z. Zeng, J. Tu, P. Pianfetti, M. Liu, T. Zhang, Z. Zhang, T.S. Huang, and S. Levinson. Audio-visual affect recognition through multi-stream fused HMM for HCI. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, pages 967–972, 2005.