Perception of Emotional Body Expressions in Narrative Scenarios and across Cultures

Dissertation

zur Erlangung des Grades eines
Doktors der Naturwissenschaften

der Mathematisch-Naturwissenschaftlichen Fakultät
und
der Medizinischen Fakultät
der Eberhard-Karls-Universität Tübingen

vorgelegt
von

Ekaterina Volkova

aus Tver

July 2014

Tag der mündlichen Prüfung:	09.10.2014
Dekan der Math.-Nat. Fakultät:	Prof. Dr. W. Rosenstiel
Dekan der Medizinischen Fakultät:	Prof. Dr. I. B. Autenrieth
1. Berichterstatter:	Prof. Dr. Heinrich H. Bülthoff
2. Berichterstatter:	Prof. Dr. Dirk Wildgurber
Prüfungskommission:	Prof. Dr. Heinrich H. Bülthoff Prof. Dr. Dirk Wildgurber Prof. Dr. Detmar W. Meurers Dr. Betty J. Mohler

I hereby declare that I have produced the work entitled: "Perception of Emotional Body Expressions in Narrative Scenarios and across Cultures", submitted for the award of a doctorate, on my own (without external help), have used only the sources and aids indicated and have marked passages included from other works, whether verbatim or in content, as such. I swear upon oath that these statements are true and that I have not concealed anything. I am aware that making a false declaration under oath is punishable by a term of imprisonment of up to three years or by a fine.

Ekaterina Volkova

Tübingen, _____ _____

Date Signature

Abstract

Emotions are ubiquitous in human lives and their importance is now well-recognised in science. Non-verbal emotional expression in human communication is realised via various means: facial expressions, emotional prosody and body motions. In real life these means complement and amplify each other, making emotion expression maximally specific, unambiguous and fitting the context. However, such multimodal expression is challenging to study due to natural innumerability of possible combinations and variations. Hence, experiments on emotion perception are often performed with well-controlled uni-modal stimuli, with the exception of studies on multi-sensory integration. In this thesis we examine perception of emotional body expressions, since especially this channel of information has been largely under-investigated despite it being important for emotion expression.

Some aspects of core affective phenomena, such as fearful reaction to threat, aggression, bonding, and drive for procreation are universal for people and several other animal species. Nevertheless, the way people express and perceive emotions via body motion greatly depends on multiple factors, such as immediate context, social and cultural background. In this thesis we present several studies that aim to gain deeper understanding of perception of emotional body expressions. We focus on body motion generated by humans during meaningful speech in naturalistic narrative scenarios and investigate its perception by using categorisation and rating tasks. Across our experiments we use a rich set of emotion categories that aims to cover many aspects of human emotional experience. Our list includes five positive categories (*amusement, joy, pride, relief, surprise*), five negative categories(*anger, disgust, fear, sadness, shame*) and the *neutral* category.

We report on the acquisition process and final form for a large database of motion capture data we collected for our research. The actors who took part in the motion capture experiment were narrating stories, expressing emotions in a natural way using all available media in a normal way without unnecessary exaggeration of motion patterns. The database is in open access and contains not only the collected motion sequences but also various meta-information, such as intended emotion of the actor, physical properties of the motion sequences, and, importantly, categories assigned to the motion sequences by naive observers, the latter being collected during the second study.

In the study that followed the motion capture experiment we

investigated the perception of collected motion samples using stick-figure displays of the upper body alone. Our results show that despite the fact that motion samples came from naturalistic scenarios and are thus expressed emotions in a subtle, non-stereotypical manner, most of the emotions are recognised at above chance level. The agreement among observers is high as most of the motion sequences have unique emotion labels that were assigned to them by most participants. We also found a significant bias towards anger, as observers were accurate at recognising anger when it was intended as such by the actors, additionally observers tended to assign the category of anger to many other motion sequences that were intended as other emotions, hence, the false alarm rate for anger was high as well. Our analysis shows a connection between properties of motion (such as average speech, span and number of peaks) with what emotion categories the observers assign to motion sequences.

In the third study we used a subset of the previously collected motion sequences and ran an exhaustive cross-cultural study among participants from Germany, England and South Korea. In this experiment we used motion sequences representing ten emotion categories (all categories used previously except for the *neutral* category). We asked our participants to rate motion sequences along three affect scales (valence, arousal and dominance) and three motion scales (speed, span and brokenness). Additionally, we recorded participants' display rules for each of the emotion categories. Our results show multiple cross-cultural differences in the perception of motion sequences, these differences tracing back to the cross-cultural differences in display rules. One of the fundamental and previously unreported differences we found is that perceived increase in brokenness of motion patterns was associated with an increase in arousal levels in German and English cultures, in contrast to Korean culture, where an increase in brokenness was associated with a decrease in valence.

Research presented in this thesis expands current knowledge on the perception of emotional body expressions and fits within the existing context of related research. Our collected data and the results of the experiments can be useful to researchers from various scientific fields: neuroscientists, cognitive psychologists, computer scientists and linguists. On the application side, our findings can inform automatic emotion recognition and motion synthesis for animated virtual characters.

"I may have allowed myself some flicker of emotion in the recent past," said Death, "but I can give it up any time I like."

Terry Pratchett
Soul Music

Acknowledgments

There are many people I am very grateful to, but first and foremost I would like to thank my advisor Betty Mohler who supervised my work through all these years. She gave me the freedom to develop my ideas and at the same time wisely guided my enthusiasm into the time-frame of the doctoral project. The amount of knowledge I gained from her about academic writing and research in general cannot be compared to any course or book. Under her supervision I never felt either exploited or forgotten, since Betty was always supportive and helpful even before my doctoral studies began.

I want to thank my supervisor Professor Bülthoff because without his approval and funding my doctoral project would not be possible. Despite him being the director of the institute and an incredibly busy person, Professor BülthoFF was always giving me his valuable feedback on my projects and manuscripts within very short time. I am especially grateful for his generous support of my lengthy visits to Seoul, Korea and Oxford, UK, where I collected data for my cross-cultural study — an experiment that would have been impossible to accomplish in Tübingen alone.

I would also like to thank my advisors and co-authors Stephan de la Rosa and Trevor Dodds, to whom I could always turn for advice concerning statistical analysis and experiment setups. My experiments would also be unrealistic to accomplish without technical support of Joachim Tesch.

Many thanks go to my Advisory Board members Professor Detmar Meurers and Professor Dirk Wildgruber, who were always very positive and supportive about my research, yet, like Betty and Professor Bülthoff, helped me to define the reasonable frame of my project and gave valuable suggestions concerning the best directions to take. I am grateful to Professor Brian Parkinson in University of Oxford who

allowed me to be part of his group for a few months and do research under his supervision. I would also like to thank Professor Martin Giese for his expert advice at the initial stages of my research.

My special thanks to our brilliant secretary Dagmar Maier, without her help my paperwork would have been utter chaos. I thank the Graduate School, its Dean Professor Herbert Horst, and its lecturers for the amazing courses in neuroscience that opened a whole new world for me and for the financial support in my research-related travels. I am very grateful to the administrators Tina Lampe and Katja Thieltges for their patience and support from my first day on in the doctoral program to the last.

I would of course like to thank all the numerous participants who took part in my experiments and diligently accomplished long experiments. I am also hugely indebted to my faithful actors without whom this thesis would not even begin. I am very grateful to everybody who was helping me to collect data, recruit participants and run the studies across different countries: Anna Wellerdiek, Steffi Davey, Maria Niedernhuber, Junsuk Kim, and Marc Briel.

The years of work at the MPI for Biological Cybernetics would not have been so much fun if not the unique atmosphere of collaboration and creativity that I have had the privilege to experience. While it would be hard to enumerate all the people responsible for my happy years at this research institute in addition to those already named, I would like to thank my colleagues and peers Markus Leyrer, Iva Pyrianokova, Aurelie Saulton, Stephan Streuber, Florian Soyka, Francois Du Toit, Sally Linkenauger, Jeanine Stefanucci, Martin Dobricki, Christian Wallraven, and many others.

I want to thank my friends Anna, Alex, Anne, and Abhilash for reading through my texts, correcting my English, and giving valuable feedback — all that allowed me to improve my manuscripts greatly. Similar thanks go to my beloved husband Norbert, but to him, like to the rest of my family, I owe so much more. They supported me in the dark hours of desperation and disbelief in myself. They cheered for my every step forward toward my goal, they helped me to go on and to never give up. It were my parents to start with who helped me to develop the natural curiosity that is so crucial in a scientist. Their praise is worth many awards for me, without them I would have never made it to where I am now.

Contents

List of Figures

List of Tables

1 Synopsis

1.1 Introduction

In the early years of cognitive psychology, emotions had been largely ignored, treated in cognitive science as undesirable noise that interfered with reasonable thinking. By now the importance of emotions is well established and its various aspects, from subconscious affective reaction towards threatening stimuli (Meeren et al., 2005) to aesthetic experience of art (Robinson, 2007), from the effect of emotions on memory (LaBar and Cabeza, 2006) to their influence on general health (Salovey et al., 2000; Gallo and Metthews, 2003) are under intense scrutiny in many fields of fundamental and applied research. Emotion perception and expression are an integral part of human-human interaction. During communication, we receive and transmit emotional information through many channels: prosody, facial expressions, word choice, posture, and body motion.

Although emotions are indispensable in our lives, it does not mean that we experience them in the same way. Naturally, core components of emotion perception are common across most people and other animals, e.g., basic affective systems (Panksepp, 2005). There is also ample evidence that there may be a few basic, universal categories (Ekman, 1971, 1992) shared between multiple populations. However, cultural background (Mesquita and Frijda, 1992) and social context (Keltner and Haidt, 1999) influence emotion perception and expression as well. Culture-specific display rules, usually acquired in childhood during the socialisation process (Harris, 1995), form expectations and interpretations of observed emotional behaviour (Tsai et al., 2006) as well as modulate emotional expression of a person. Thus, it is important to keep in mind that physiological responses, many elements of expressive behaviour, response to sudden threat, etc. are influenced by the biology of human beings and are common to all people and several other species, while culture influences emotional experience at higher levels of cognition (Matsumoto and Hwang, 2012).

Thus there is much variability in **what** emotional aspects to study, and there are also several distinct approaches on **how** to study emotion. One school describes emotions in terms of values along several dimensions, typically three: *valence*, *arousal* and *dominance*, although the terminology and number of scales vary(Russell and Mehrabian, 1977; Fontaine et al., 2007). The categorical approach (Ekman, 1971, 1992; Ekman and Cordaro, 2011), where emotions are labelled as specific concepts with distant meaning and attributed ways of expression and experience, is also very popular. Another way to study emotions — the appraisal theory — was suggested by Arnold (1960) and further developed by Lazarus et al. (1970). The appraisal theory argues that our emotions are grounded in the evaluation of current events and situation. Scherer (2001) developed the "multi-level sequential check model", which elegantly accounts for complex, context-dependent emotional experiences in human life.

Research on emotion has been flourishing in the past decades, but some aspects, especially emotional body expressions have been largely ignored. Instead, researchers of emotional expression mostly focused on studying facial expression. The common opinion was that most of non-verbal expression takes place via facial expressions while the human body was often perceived as a tool for actions (e.g., walking, grasping, and carrying), or gesticulation accompanying speech process. However, the body is also an important medium for emotional expression (De Meijer, 1989; de Gelder et al., 2010). During communication, body motion can highlight and intensify emotional information conveyed by face and voice, add extra nuances of meaning to emotional expressions, or contrast emotional information coming from other channels. For instance, head position significantly affects the perception of facial expressions (Lyons et al., 2000). Compare a normal smiling face expressing joy and a smiling face with the head tilted backwards, expressing pride, add arms crossed in front of the chest and the expression of arrogant pride is complete being very different from a typical expression of happiness, despite the fact that the facial expressions are very similar. Thus body motion should be considered as a separated channel of emotion expression, along with facial expressions and prosody.

At the early stages of research on emotional body language, it was suggested that body posture and motion can only express general levels of activation (arousal) and are unable to communicate specific emotions. This observation was made on the basis of experiment re-

sults showing that expression different emotion categories were often confused with each other when they shared a similar level of movement activation (Ekman, 1965). By now several studies (including our research, see Chapters 4 and 5), have established that the high levels of accuracy and consistency with which observers label body motion for emotion categories cannot be attributed exclusively to movement activity (Wallbott, 1998).

In the following chapters we investigate motion that accompanies fluent speech in narrative scenarios and is thus communication bound. In human-human communication, body motion is usually referred to as *gestures* which accompany verbal communication and are typically subdivided into the following groups (Ekman, 1965):

Regulators' main function is to coordinate the conversational flow. By pointing or stretching one's arm towards an interlocutor a person can invite them to speak, or by shaking a hand at them in a prohibiting gesture can ask them to keep silent. Orientation of the head, torso or the whole body can perform a similar function.

Illustrators accompany speech and illustrate the verbal message. These gestures can take the form of pictographs, ideographs, or deictic (defined by context) movements. An example may be stretching one's arms wide apart while taking about the size of a fish caught, or turning an imaginary key when speaking about unlocking a door. Illustrators make speech more vivid and effective (Dodds et al., 2011).

Emblems are gestures that have a specific symbolic meaning and can be directly translated into words, e.g. the thumbs-up sign or sticking your tongue out. Emblem gestures are often strongly culture-dependent and infrequently have emotionally coloured meaning. Sometimes, elements of same emblems are shared as part of the full body motion patterns and take on a different meaning. Thus, shaking a fist in front of somebody's face means aggression, while putting the fist in the air and stretching the arm upwards means pride and joy of accomplishment.

Self-adaptors refer to behaviours performed without obvious intention and usually involve self-touching. Scratching oneself, twisting a wisp of hair, rubbing one's forehead are just a few examples of self-adaptors. Especially in emotional context it is sometimes

hard to tell emblem gestures and self-adaptors apart. For instance, putting one's palm to the face and/or covering your eyes with your palm is often interpreted as sign of being ashamed or embarrassed. In Korean culture an embarrassed person would often rub the back of the neck with their palm. Both gestures involve self-touching, yet they have an emblematic meaning.

All gestures can be modulated by various emotions, their motion style altered to express a particular emotional state more clearly. Turning away from a person more abruptly that usual means not only that you wish to end the conversation, but also communicates that you are angry with them. Clapping your hands very slowly in an exaggerated manner tells the interlocutor your ironic display of, in fact, not being pleased with them.

However, it is important to remember that a body motion or posture does not necessarily have to be a canonical gesture to express emotion, as style variations of one and the same general movement (walking, throwing, knocking) can be sufficient to support the percept of emotional states (Pollick et al., 2001b; Troje, 2002; Roether et al., 2010). While the face can express the emotional state of a person, their head and body orientation can show the location of the stimuli causing this emotion or the direction of the necessary course of actions (de Gelder et al., 2004). Emotions expressed by the body (e.g. anger or fear) are easier to recognise from the distance than those expressed by the face (de Gelder et al., 2004).

Emotional body expressions, both static postures and dynamic motion, are rather accurately recognised by human observers (Wallbott, 1998; Dittrich et al., 1996; Pollick et al., 2001b; Atkinson et al., 2004, 2007). However, one of the most pressing questions in this field of research still remains open, since the precise relationship between the style of motion and perceived emotions is still largely unclear. When studying emotion perception from body motion, most stimuli are collected with the help of motion capture or video recording. Alternatively, motion stimuli can be synthesised using systematic manipulation of chosen motion features. Synthesis of believable motion trajectories however is far from a trivial task and, to the best of our knowledge, has been attempted only for walking motion (Troje, 2002; Roether et al., 2010), and for accent gesticulation in speech production for virtual avatars (Hartmann et al., 2006; Levine et al., 2009).

Depending on the level at which emotions are studied, more atten-

tion has to be paid to the naturalness of stimuli to ensure ecological validity of results. For instance, to trigger fear in participants, a subliminal exposure to a photograph of a spider is enough (Tamietto and De Gelder, 2010), whereas if researchers want to study emotion in context of everyday life, e.g. emotion transfer during communication (Parkinson et al., 2012), a real dialogue situation with genuine discussion is needed (Oertel et al., 2013).

In the three presented studies we collected a large database of motion sequences in naturalistic settings, then investigated perception of the acquired motion samples using stick figure displays. We used categorisation and rating tasks and performed general and cross-cultural studies. Our work fits within the existing context of related research, detailed in the next section, yet it also expands the current knowledge on emotion perception of emotional body expressions produced in narrative scenarios.

1.2 Related Research on Emotional Body Expressions

1.2.1 Databases of Emotional Stimuli

To study body emotion expressions, data sets of good quality, preferably collected from various actors and spanning multiple emotion categories or other properties of affect under investigation, are often needed. Over the course of the last few decades multiple datasets have been accumulated especially in the last ten years, as the interest in emotional body language has increased (Johansson, 1973; Ekman and Friesen, 1976; Kudoh and Matsumoto, 1985; Kamachi et al., 1998; Gross and Shi, 2001; Burkhardt et al., 2005; Cowie et al., 2005; Pantic et al., 2005; Bänziger et al., 2006; Clavel et al., 2006; Gunes and Piccardi, 2006; Hwang et al., 2006; Ma et al., 2006; Kleinsmith et al., 2006; Bänziger and Scherer, 2007; Zara et al., 2007; Busso et al., 2008; Busso and Narayanan, 2008; Yin et al., 2008; Metallinou et al., 2010; Kaulard et al., 2012; Koelstra et al., 2012; Sneddon et al., 2012; Aubrey et al., 2013).

To the best of our knowledge, at the beginning phase of the research presented in this thesis there was no database available that satisfied all the requirements important for our research aims. We needed a dataset of motion sequences in motion capture format (as opposed to video recordings), collected during acts of coherent and

natural speech, and spanning across a rich set of emotion categories. The new database is described in Chapter 3, here we will give a short account of the most relevant databases that have inspired our database design and formed our choices of technology and emotion induction techniques.

One of the best known motion capture datasets is the CMU Graphics Lab database[1], which contains over 2600 full body motion sequences in different formats of motion capture. Its motion capture data represents various actions, such as walking and dancing. Unfortunately, relatively few of the motion sequences in the CMU database have emotional content. In contrast, the database developed by Ma et al. (2006) is focused on how emotion influences motion style and contains over four thousand motion capture sequences of actions like waving or throwing that display four categories — *neutral, angry, happy, sad*. This database is a valuable resource, but the motion patterns were captured in purely non-verbal scenarios, thus excluding the context of emotion expression during speech production. Moreover, the actors were fully aware that only their body motion is important for the recording and thus could have exaggerated their expressive body motions.

While our aim was to create a database based on motion capture, much can be learned from video-recoding based databases. Not all of these datasets are freely accessible and have often been collected as a necessary part of further emotion research. A few of such studies have used video recording or motion capture for further feature extraction and automatic emotion recognition from dynamic biological motion or static poses (Kleinsmith et al., 2006; Castellano et al., 2007). Among the video-based datasets made available for researchers is the GEneva Multimodal Emotion Portrayals (GEMEP) corpus (Bänziger et al., 2006), which used a wide range of emotion categories and a refined emotion induction technique involving pseudo-linguistic sentences. Another resource is the Interactive Emotional Dyadic Motion Capture (IEMOCAP) database (Busso et al., 2008), a database of video-recorded and motion captured scripted and spontaneous interactions between dyads of actors.

When designing a motion capture database, it is also important to consider its context (de Gelder, 2009; Parkinson, 2013) and naturalness. These factors ensure the recorded motion sequences will have higher ecological validity and will thus allow future research to gain deeper in-

[1] http://mocap.cs.cmu.edu

sight into the way people express emotions in real life. Moreover, such motion patterns could also provide a useful source of data for motion synthesis, e.g., for virtual character or robot animation - when applied to a virtual avatar together with facial animation, the resulting multimodal emotion expression will look more natural and not uncanny. In this regard, Oertel et al. (2013) are a very good example of giving their actors the freedom of naturalistic conversation in a domestic setting. The authors have non-intrusively gathered social interaction data by means of audio and video recording and motion capture, creating a valuable corpus for studying naturalistic conversational interaction.

1.2.2 Emotion Expression through Body Motion

In the past years, a considerable number of studies have researched the association between body movements and emotional states (Sogon and Masutani, 1989; De Meijer, 1989; Boone and Cunningham, 1998; Wallbott, 1998; Pollick et al., 2001b; Clarke et al., 2005; Atkinson et al., 2007; de Gelder, 2009). Various paradigms and methods have been employed, e.g. brain imaging (Pichon et al., 2008), psychophysical studies where a range of morphs between two motion styles is used to find the perception threshold between two categories (Roether et al., 2010), emotion recognition (Atkinson et al., 2004) and emotion rating. These studies are usually based on recording emotionally expressive behaviour, specifically body movements, by means of photo, video or motion capture. The emotional body expressions are often performed by professional, amateur, or lay actors following or imagining affect-inducing scenarios. In general the results of previous studies show that human observers are very good at recognising emotions from body motion, especially if the actors are expressive. When the stimuli used for emotion recognition are more subtle however, the accuracy rate can significantly decrease. The level of agreement among participants is also highly dependant not only on the origin of the stimuli but also on the task itself — choosing from a predefined set of categories yields very different results from open set labelling (Winters, 2009).

Research on emotional expression through body motion in humans is complex for several additional reasons. One complication is that the human body has hundreds of degrees of freedom, making it difficult to record and analyse its motion with precision and efficiency. Additionally, the body is used for action and emotion expression simultaneously, e.g., when a person is throwing a piece of paper into a paper basket in

frustration or rage, the motion contains both kinds of information — the action (*throwing*) and the emotion (*anger*). Moreover it is necessary to separate motion trajectories from the shape of the body and other attributes of appearance, such as age, gender, and clothing.

While general discussion of functions of emotional posture and motion has been postulated in 19th century by Darwin (1872), it was not until 1973, when Johansson developed the now widely used technique of biological motion representation — point light displays. This technique retains motion information but eliminates form information by marking the moving figure is marked with a few illuminated points positioned at the main body parts and joints (Johansson, 1973). In the years that followed, the point-light technique was frequently used in research on perception of biological motion in general and emotion perception studies in particular. Some research focussed on gait styles and their modulation by emotional state (Troje, 2002; Roether et al., 2010), while other studies used actions (Pollick et al., 2001b) or non-verbal portrayals of emotions (Atkinson et al., 2004; Clarke et al., 2005; Beck et al., 2012). Importantly, these studies showed that body motion conveys enough information for the observers to accurately recognise the emotion category of the stimuli. However, one feature of all these studies was the awareness of the actors that only their body language was of primary interest to the researchers. This could potentially change the way the actors moved, making the emotion expression for stereotypical and exaggerated. In three studies presented in this thesis we made sure that the collected motion sequences are maximally natural and close to daily emotional expression.

Related research on perception of emotional body motion followed several approaches discussed earlier. Some studies asked participants to rate stimuli along affect scales (Dael et al., 2013), but mostly the categorical approach on a predefined set of labels was used. The most frequently used emotion are the *basic emotions* as postulated by Ekman (1971): *anger, disgust, fear, joy, sadness, surprise*. The advantage of using the six emotion categories is their relative universality, despite the fact that this claim is often challenged (Elfenbein and Ambady, 2002). The list of six emotions has been used in multiple studies across various modalities and populations, which allows for comparisons across studies. Fewer studies have expanded the list of emotions in an attempt to cover more of human emotional experience (De Meijer, 1989; Wallbott, 1998; Pollick et al., 2001b; Bänziger et al., 2006; McDonnell et al., 2009; Beck et al., 2012). In our studies we too used a rich set of

emotion categories: five positive (*amusement, joy, pride, relief, surprise*), and five negative (*anger, disgust, fear, sadness, shame*) emotion categories as well as the *neutral* category.

1.2.3 Cross-cultural Differences in Emotion Perception

According to Matsumoto (2006), culture is "a shared system of socially transmitted behaviour that describes, defines, and guides people's ways of life". This definition is embracing more than just the emotional side of human life, but multiple studies have shown how cultural background influences the way people express and perceive emotions via attitudes, beliefs, and values about emotions, their meaning and appropriate display (Eid and Diener, 2001; Tsai et al., 2006; Matsumoto and Hwang, 2012). Since culture and language are often linked, cross-cultural differences have been found at the level of conceptualisation of emotional experience itself. For example, shame is categorised very differently in Chinese and English languages and cultures (Bedford, 2004).

Related research traditionally distinguished between two broad classes of cultures: *independent*, often also referred to as *individualistic*, typical for European and North American cultures; and *interdependent*, or *collectivistic*, which many cultures in Asia are often associated with. Persons from independent cultures tend to focus on individual aspects of the self, their feelings, emotions, thoughts and the self-related consequences thereof. People from interdependent cultures pay more attention to relational aspects of the self, being attentive and acutely aware of how their actions and emotions affect their surrounding (Chentsova-Dutton and Tsai, 2010). Naturally, such strict division of all cultures into two classes is bound to over-generalise and more flexible approaches to describe and classify cultures have been undertaken (Green, 2005; Tracy and Matsumoto, 2008). Moreover, cultures are changing structures and can be largely influenced by shifts in economical and political organisation (Lu and Yang, 2006). Nevertheless, different types of cultures are important to take into account since they influence the expression and perception of emotion.

Cross-cultural differences on the level of perception of facial expressions (Friesen, 1973), have been studied as well, but few researchers examined cross-cultural differences of emotion perception in whole

body posture (Matsumoto and Kudoh, 1987; Kleinsmith and Silva, 2005; Kleinsmith et al., 2006), and even fewer used dynamic stimuli despite the obvious importance in emotion expression in the body (Scherer et al., 1988; den Stock and Righart, 2007; de Gelder, 2009; de Gelder et al., 2010). Our third study addresses the question of cultural influence on the perception of the emotional body expressions as opposed to the physical properties of the motion itself.

1.3 Aim and Structure of the Thesis

This thesis aims at deepening the existing understanding of perception of emotional body expressions. Previous research, as discussed before, was based on few emotion categories and/or used exaggerated non-verbal displays of emotion categories. The studies, presented in the following chapters have investigated perception of dynamic body expressions using classification and rating tasks. The major focus across all the studies is kept on the naturalness of stimuli that we have collected using motion capture technology for this research under controlled yet ecologically valid conditions. Because ecological validity was particularly important throughout all studies presented in this thesis, we aimed to cover a broad spectrum of human emotional experience and hence used a set of eleven emotion categories: *amusement*, *joy*, *pride*, *relief*, *surprise*, *anger*, *disgust*, *fear*, *sadness*, *shame*, and *neutral*.

The motivation for the motion sequences collection study, its methods and the resulting database that is now available to the community are discussed in detail in Chapter 3. The resulting database contains a large set of natural emotional body expressions typical of monologues. In our motion capture setup, amateur actors were narrating coherent stories while their body movements were recorded in three dimensions at a high frame rate, thus providing detailed information about body movements. The collected motion capture data is well-suited for further perceptual studies and allows the manipulation of movement on a body joint basis.

We further investigated perception of emotion using the motion sequences from the new database. During the motion capture sequences the actors were employing all their expressive means, such as facial expressions, emotional prosody and body motion. Moreover, the actors were unaware that only their body motion will be used for further investigation of emotion perception. Hence, one of the next logical

questions was to find out whether emotion categories displayed via isolated dynamic human upper body stick figures can can be recognised at above chance level. To investigate this we ran a large-scale study where a set of motion sequences was categorised by multiple participants. The details of these studies and the results are presented in Chapter 4.

While categorisation is an important and well established tool in emotion research, we wanted to obtain a more detailed picture of perception of emotional body expressions using rating scales. Since previous research has established influence of the cultural background on emotion perception, we recruited participants from three different cultures - German, English, and Korean - in order to investigate previously under-studied cross-cultural differences in perception of dynamic emotional body expressions. This study was conducted in three corresponding countries and is reported in full detail in Chapter 5.

1.4 General Discussion

1.4.1 Chapter 3: The MPI Emotional Body Expressions Database for Narrative Scenarios

The aim of this study was to collect material for further use in perception studies. The resulting database of motion capture is unique, open for access, and can be useful to scientists from different research fields. The motion sequences included in the database have been collected in a naturalistic yet well controlled environment of emotional narration. The actors were narrating an unabridged story, their facial expressions, speech and body motion were recorded with the help of video camera, microphone and motion capture equipment. Thus, the actors were not aware that only their body motion was of primary importance to the researchers. This ensured more natural, non-exaggerated emotional expression via body motion, which is a crucial feature of our database and the following research.

Another valuable feature of the database is the fact that we used the rich set of emotion categories, a relatively rare property for databases of emotional stimuli, with some notable exceptions. We used not only the six basic emotions (*anger, disgust, fear, joy, sadness, surprise*) but expanded the list with four extra emotional categories: *amusement,*

pride, relief and shame and added the category of *neutral*.

The data format of motion capture is of great value itself for the purposes of motion pattern analysis, as the human motion was recoded in 3D space and at high frame rate. This format is different from video recordings and has several advantages for researchers interested in motion analysis and motion modification. The motion capture format allows to display biological motion and change many of its properties, e.g. speed of selected trajectories, magnitude, and the position of joints in space. The visual representation, for instance, point lights, full or parts of skeleton or even virtual character can be applied to the motion capture files.

Researchers from several branches of science may find our database useful, but its primarily aimed at computer scientists, psychologists, neuroscientists, and linguists. The motion sequences could be used in emotion perception research, motion analysis and synthesis, in behavioural and brain imaging research. The motion can be used for building models of emotional behaviour to synthesise expressive natural looking motion for virtual character animation. Here it is important to point out again that our motion was acquired in naturalistic settings, where actors were using all their expressive means in synchrony to narrate a story. Thus, if resulting motion patterns are used in model building and motion synthesis for virtual character animation, the resulting motion is less likely to look uncanny or exaggerated, especially when combined with possible facial animation and emotional prosody.

In addition to the motion sequences, our database provides meta-data for each motion sample that can inform researchers' choice when selecting stimuli for their experiment. Moreover, this meta-data can be of value to researchers in psychology, as it includes actors' intended emotion displays, the original annotations of full narration texts, and also the perceived emotions obtained during an emotion recognition study (Volkova et al., 2014b) in Chapter 4. The resulting categorisation responses from the observers are included in the database and are available as metadata with each motion sequence file. As each motion sequence was categorised by eleven observers, the most frequently chosen emotion category (the perceived emotion) is included into the database as a separate value. The proportion of responses the perceived category takes (consistency) and its correspondence to the emotion intended by the actor (accuracy) are provided in the database as well.

1.4.2 Chapter 4: Emotion Categorisation of Body Expressions in Narrative Scenarios

The experiments presented in this chapter show that people recognise naturally expressed emotions at above chance level, which is impressive and important, since the motion sequences come from a naturalistic narration setting and the visual stimuli only provide information about the upper body motion.

Experiment 1 uses 80 motion sequences obtained during actors performing non-verbal solitary short scenarios and is thus most similar to more classical emotion recognition studies — the body motion of the actor was the primary source of emotional information, although the face and voice (screams, grunts, sighs, laughter) were also used. The emotional scenarios were arranged in a random sequences and were very short. The results of the perceptual study show that all emotions were recognised at rates far higher than chance level. Another important note is that we did not use the category of *neutral* in this experiment, thus employing ten emotion categories.

In experiment 2 eight out of eleven emotions are recognised at above chance level, including the *neutral* category. The average recognition level of 19% is surprisingly high for the number of emotion categories, taking into consideration that most of the motion sequences used in this experiment came from narration tasks where all expressive channels were used by the actors and the motion was placed in the continuum of story-related context. Additionally, the motion sequences were categorised with high consistency between participants. Namely, for 85% out of 1700 motion sequences, response distribution has a unique modal value.

Recognition accuracy and consistency levels differ among emotion categories. In both experiments, but most notably in experiment 2, observers have a considerable bias for categorising emotions as *anger* which is supported not only by recognition rates for this category but also by false alarm rates and post-hoc analyses that compared recognition rates across all emotion categories. The bias towards *anger* could be explained by the possible evolutionary importance of detecting anger as a potential threat. A mistake of failing to recognise a threat is more costly than mistaking harmless stimuli as dangerous. Previous research supports the importance of anger expressed via body motion: Pichon and colleagues have shown that a response to anger expression results in even more activation of defence mechanisms than a response

to fear expressions (Pichon et al., 2009). An observation similar to ours was recently made by Visch et al. (2014), where recognition of anger expressed in the body motion was the most robust under various stimuli degradation conditions. In addition to bias towards *anger*, participants had a bias towards the *neutral* category in experiment 2, as the *neutral* category also has high accuracy rate, false alarm rate and consistency levels. A likely reason for this bias is that many motion sequences did not possess properties that could communicate any particular emotion to the observer (see Chapter 4 for more detail).

By analysing the relationship between the response patterns and motion properties of the motion sequences, we found that distances in mean motion speed, number of peaks and mean span to some extent predict distances between response categories. These findings are encouraging for future work in automatic emotion recognition. In general, the patterns in which intended and observed emotional categories are organised are not random: while *anger* is well recognised, not infrequently it is taken for *joy* and *pride*, as well as motion sequences intended as *pride* or *joy* are often categorised as *anger*. This and several other examples are detailed in Chapter 4 but the general conclusion can be be made that confusions between emotion categories are not random but originate from certain properties of motion patterns that are shared among these emotions, may it be the general properties we have partially investigated or more specific features like elbow flexion, head tilt or orientation of limbs in the space relative to the observer.

Importantly, our results do not always support the theory of *basic* emotions (Ekman, 1992). In our experiments we used a rich set of emotions that goes beyond the basic ones. One of the major arguments for basic emotions is that they are universally recognised in most populations and are independent of expression medium. All emotions are recognised well above chance in experiment 1, where motion sequences were obtained from non-verbal short scenarios. However, the recognition rates in experiment 2 seem to suggest that for emotional body expressions in a natural setting, the basic emotions are perhaps not fitting. Namely, two out of three categories recognised below chance are basic: *disgust* and *surprise*, while non-basic emotions of *amusement*, *pride* and *shame* are recognised above chance. This allows us to conclude that the *distinctive universal signals* proposed by Ekman as one of the characteristics for basic emotions are not always present in our upper body motion patterns captured during natural expression.

1.4.3 Chapter 5: Cross-cultural Differences in Perception of Dynamic Emotional Body Expressions

Although previous research has investigated various aspects of cross-cultural differences in emotion perception, there are still many open questions left, since the phenomenon of culture is as complex as the one of emotion and its manifestations are versatile and prone to change (Parkinson et al., 2004). We used dynamic upper body expressions of ten emotion categories (*amusement, joy, pride, relief, surprise, anger, disgust, fear, sadness, shame*) and investigated the connection between perception of emotional body expressions and cultural display rules across three cultures (English, German and Korean). We used rating tasks of motion sequences and imagined emotional motion along six five-point scales — three affect scales (valence, arousal, dominance) (Lang, 1980; Bradley and Lang, 1994) and three motion scales (span, speed, brokenness). The scales were presented as rows of simple pictures, thus avoiding the problem of translation of each value along each scale into the three corresponding languages.

In one of the two experimental conditions the participants we asked to categorise the motion sequences in addition to rating them. The categorisation task allowed us to measure recognition accuracy across cultures and investigate if in-group advantage, previously shown only for emotional prosody (Scherer et al., 2001) and facial expressions (Elfenbein, 2013), could be found, since most motion sequences originated from German amateur actors. Indeed, The recognition accuracy was higher for participants coming from Germany, which supports the in-group advantage. While in general all emotion categories were recognised at above chance level, some of the *basic* emotions (*disgust, surprise*) were recognised worse than other emotions. Additionally, the clustering patterns of mis-categorisations can give valuable insight into common and culture-specific confusions between emotion categories.

Among all emotions, the ratings of the emotion of *shame* are very different between Korean and the Western cultures — in Korean culture, its dominance (level of control over the situation) is higher, and the motion span is wider. Cross-cultural differences of conceptualisation and perception of shame have been studied before (Fontaine, 2006; Liem, 1997; Stipek, 1998; Tracy and Matsumoto, 2008). One reason for this could be the fact that in Asian countries, where the characteristic of a person being humble and respectful is very important, shame is

shown more often and serves also as a social signal for appeasement at an early stage of an unpleasant situation.

The results of the factor analysis show that one of the main observed differences between Western (English and German) and Asian (Korean) cultures is how the scale of motion brokenness is used in the rating tasks. In Western cultures this scale is associated with arousal and similar results were found by Dael et al. (2013) and previous research. The correlation between perceived brokenness and arousal was direct, hence, the higher perceived brokenness of the motion, the higher the arousal. In Korean culture however, higher brokenness is associated with lower valence. In Western cultures higher valence was on in its turn associated with higher span of motion, this motion scale shared with the one of dominance. In Korean culture the scale of dominance does not seem to be associated with any of the motion scales. Similar results for individualistic, most often Western cultures have been found in previous research. For instance motion span or openness of body posture are associated with higher valence (Wallbott, 1985), while Pollick et al. (2001b) found strong positive correlations between perceived intensity of anger expressions and movement velocity.

Our results show that same motion sequences are perceived somewhat differently across the three cultures, mainly between the Korean and the two Western cultures. The dominance scale is the one where two culture types differ in motion sequence ratings and display rules. In most cases of statistically significant cross-cultural differences for a particular scale$\times$ emotion combination, the mean rating obtained from Korean participants is closer to the middle of the scale. With respect to the display rule ratings, this observation confirms findings from previous research that people in Asian cultures express emotions in a more subtle manner than people from individualistic cultures (Scherer et al., 1988).

By implementing a condition that involved only rating tasks without explicit categorisation, we specifically tested whether people can access affective properties of motion directly, but as our post-questionnaire results show, for most trials participants still assigned some category implicitly. While such labelling undoubtedly takes place due to the integration of perceived motion features, it seems that most of the time participants first assign a label to the motion sequence and then adapt their ratings to this particular instance of emotional expression. This statement is also supported by the fact that across all cultures ratings of display rules were more extreme (further from the

middle of the scale) than the ratings of motion sequences.

In summary, our findings support the hypothesis that cultural background influences immediate perception and interpretation of dynamic emotional body motion. This factor is important to take into consideration in scenarios where culture is an important factor in human-human or human-computer communication. Such basic properties of motion such as overall speed, span and brokenness have an important impact on the perceived affective state of a speaker. These findings can be useful for various culture-specific and multicultural scenarios, such as development of a virtual museum guide with multilingual interface or video material production for online education courses aimed at the general public.

1.5 Future Work

Our work has laid solid foundation for further research in several directions. The motion stimuli database we have developed and presented in this thesis provides not only the motion capture sequences but additional useful information about the actor, the intended emotion, the perceived emotion labels and a few properties such as motion duration, motion span, speed and brokenness. This database can prove to be a valuable resource of stimuli for future investigations of emotion perception in the fields of cognitive psychology and neuroscience. However, it would be interesting to analyse the available motion patterns to attain even deeper understanding as to what are the crucial properties of motion patterns that allow an observer to recognise a particular trajectory or a combination of such as a specific emotion category. In the studies presented in this thesis we have so far only developed the initial foundation for such analysis and its link to perceived properties of emotional expressions. In such analysis however, it would be important to differentiate between meaningful or emblematic motion patterns such as head shakes or hand claps and more general features such as increased joint flexion, lower motion energy, etc., since these two aspects are likely to exist on different levels of perception of biological motion. The former related to action recognition and the latter to more general motion style processing.

A detailed motion analysis with an elaborate feature extraction procedure can be useful not only for perceptual studies — it can be further used for automatic emotion recognition based on supervised

machine learning, when training is performed on sequences that are labelled with a predefined set of emotion categories. Unsupervised machine learning methods, when only existing physical properties of motion are used for clustering of available sequences into groups, can prove to be of great empirical value as well. This machine learning approach can reveal several distinct patterns for a group of motions usually labelled by human observers as one emotion category. In contrast, it can also show that some emotion categories share very similar emotion patterns and need further information from other channels like facial expression, social or language context, for successful disambiguation.

Yet another fascinating direction of further research would be motion synthesis or motion modulation. Informed by previous research on perceived emotion properties of body expression and their connection to perceived and physical properties of motion sequences, it would be interesting to develop motion models that could successfully convey various emotion categories and their subtle variations. Such models could be used in elaborate psychophysical studies, both of behavioural and brain imaging varieties. Importantly, models of emotional body expressions could be used for the animation of virtual characters in video games, including serious and educational games, which could significantly improve human-computer interaction.

2 Declaration of Contribution of the Candidate

This thesis is presented in the form of a collection of manuscripts that are at the time of thesis submission either published or prepared for publication. The bibliographic details for each manuscript and their order of appearance in the thesis are set out below. The description of the contributions of each author is included for each manuscript.

Chapter 3

Volkova, E., Mohler, B. J., de la Rosa, S., and Bülthoff, H. H. (2014a). The MPI Database of Emotional Body Expressions Common for Narrative Scenarios. *submitted*

The idea for this study was proposed by the candidate. Design, stimuli generation, experimental work, and analysis of the study have been predominantly developed and implemented by the candidate. The work has been presented at several scientific conferences and workshops by the candidate. The co-authors supervised the work of the candidate by giving advice, offering knowledge and criticism at every stage of the study, including the analysis of the results and the revising the manuscript.

Chapter 4

Volkova, E., Mohler, B. J., Dodds, T. J., Tesch, J., and Bülthoff, H. H. (2014b). Emotion Categorisation of Body Expressions in Narrative Scenarios. *Frontiers in Psychology*, 5(623)

The idea for this study was proposed by the candidate. Design, stimuli generation, experimental work, and analysis of the study have been predominantly developed and implemented by the candidate. The work has been presented at several scientific conferences and workshops by the candidate. The co-authors supervised the work of

the candidate by giving advice, offering knowledge and criticism at every stage of the study, including the analysis of the results and the revising the manuscript.

Chapter 5

Volkova, E., Mohler, B. J., Parkinson, B., Wildgruber, D., Bülthoff, H. H., and de la Rosa, S. (2014c). Cross-cultural Differences in Perception of Dynamic Emotional Body Expressions. *in preparation*

The idea for this study was proposed by the candidate. Design, stimuli generation, experimental work, and analysis of the study have been predominantly developed and implemented by the candidate. The work has been presented at several scientific conferences and workshops by the candidate. The co-authors supervised the work of the candidate by giving advice, offering knowledge and criticism at every stage of the study, including the analysis of the results and the revising the manuscript.

3 The MPI Emotional Body Expressions Database for Narrative Scenarios

Volkova, E., Mohler, B. J., de la Rosa, S., and Bülthoff, H. H. (2014a). The MPI Database of Emotional Body Expressions Common for Narrative Scenarios. *submitted*

3.1 Abstract

Emotion expression in human-human interaction takes place via various types of perceptual information, including body motion. Research on the perceptual-cognitive mechanisms underlying the processing of natural emotional body language can greatly benefit from datasets of natural emotional body expressions that facilitate stimulus manipulation and analysis. The existing databases have so far focused on few emotion categories which display predominantly prototypical, exaggerated emotion expressions. Moreover, many of these databases consist of video recordings which limit the ability to manipulate and analyse the physical properties of these stimuli.

We present a new database consisting of a large set (over 1400) of natural emotional body expressions typical for monologues. To achieve close-to-natural body emotional expressions, amateur actors were narrating coherent stories while their body movements were recorded with motion capture technology. The resulting 3-dimensional motion data recorded at a high frame rate (120 frames per second) provides fine grained information about body movements and allows the manipulation of movement on a body joint basis. For each expression it gives the positions and orientations in space of 23 body joints for every frame. We report the results of physical motion properties analysis and of a recognition study. The reactions of observers from the emotion recognition study are included into the database. Moreover, we recorded the intended emotion expression for each expression from the actor

to allow for investigations regarding the link between intended and perceived emotions. The motion sequences along with the accompanying information are made available in a searchable MPI Emotional Body Expression Database. We hope that this database will enable researchers to study expression and perception of naturally occurring emotional body expressions in greater depth.

3.2 Introduction

Emotions shape human communication and have been a long-standing subject of research in many fields of science including anthropology, psychology, and neuroscience (Darwin, 1872; Ekman, 1971; de Gelder et al., 2010). Previous research has largely focused on the examination of perceptual and cognitive processes underlying the recognition of facial emotional expressions. Surprisingly, body movements have been largely neglected in emotion research although they make an important contribution to emotion recognition and even modulate the interpretation of facial emotion (de Gelder, 2009; de Gelder et al., 2010).

To study body emotion recognition, data sets that allow control over the quality of body movements and the visual representation of data are an important requirement. Over the course of the last few decades multiple datasets, often collected as part of research reflecting on various sides of affect in humans, have been accumulated, some dating as much as 40 years back (Johansson, 1973; Ekman and Friesen, 1976; Kudoh and Matsumoto, 1985; Kamachi et al., 1998). Due to the progress in recording technology and data storage most of these corpora and databases have been collected during the last ten years (Kamachi et al., 1998; Gross and Shi, 2001; Burkhardt et al., 2005; Cowie et al., 2005; Pantic et al., 2005; Bänziger et al., 2006; Clavel et al., 2006; Gunes and Piccardi, 2006; Hwang et al., 2006; Ma et al., 2006; Bänziger and Scherer, 2007; Zara et al., 2007; Busso et al., 2008; Busso and Narayanan, 2008; Yin et al., 2008; Kleinsmith et al., 2006; Metallinou et al., 2010; Kaulard et al., 2012; Koelstra et al., 2012; Sneddon et al., 2012; Aubrey et al., 2013).

Many high quality datasets of human body motion are now available, but we will focus on those that deal with emotional body language. From the perspective of data format, the databases can be primarily collections of motion capture data or video recordings. One of the best known motion capture datasets is from the CMU Graphics Lab

(see Gross and Shi (2001) for one of the initial reports or access the database at http://mocap.cs.cmu.edu). It contains over 2600 full body motion sequences available in different formats. However, a very small proportion of motion sequences in the CMU database has explicit emotional content. Another noteworthy dataset has been developed by Hwang et al. (2006) and presents a systematic and well controlled motion capture set of non-emotional gestures and simple actions from many actors. Ma et al. (2006) have collected an extensive database from 30 actors, where 4080 motion capture sequences encompass waving and other non-verbal actions displayed by four emotion categories — *neutral, angry, happy, sad*. This database is a valuable resource, but the motion patterns were captured in purely non-verbal scenarios, thus excluding context of emotion expression during speech production.

The USC CreativeIT database (Metallinou et al., 2010) consists of short, scripted actions performed by pairs of actors. The interactions were recorded with full-body motion capture technology, video and audio. However, the database's primary focus is on different acting properties (interest, naturalness, creativity) rather than on the emotional content of the interactions. The display of actions was kept as natural and intuitive as possible and the setup itself does not influence the actors' expressive behaviour. Actors did not receive specific instructions as to which emotions to express, neither did they write an acting script for themselves. The database also provides observers' annotations for each recording with regards to the emotion dimensions (valence, activation, and dominance) and performance properties. Specific emotional expression labels are not provided.

Several studies have created video recording datasets, sometimes combined with audio recordings, to capture nonverbal emotional expressions typical for communication or free expressions of emotions. Not all of these datasets are available as databases but they make an important contribution to the best practices of data acquisition for emotional body expressions. The aim of most of the studies is to gain a deeper understanding of what properties of motion facilitate attribution of specific emotion categories or general emotional dimensions (De Meijer, 1989; Wallbott, 1998; Atkinson et al., 2004). A few studies have used video recording or motion capture for further feature extraction and automatic emotion recognition from dynamic biological motion or static poses (Kleinsmith et al., 2006; Castellano et al., 2007). Among the video-based datasets made available for researchers are the GEneva Multimodal Emotion Portrayals (GEMEP) corpus (Bänziger

et al., 2006) and the Interactive Emotional Dyadic Motion Capture (IEMOCAP) database (Busso et al., 2008). GEMEP uses a wide range of emotion categories and a refined emotion induction technique involving pseudo-linguistic sentences. The resulting corpus of about 2000 video sequences contains acoustic information, facial expressions and body motion. IEMOCAP is a database of video-recorded and motion captured scripted and spontaneous interactions between dyads of actors. Only facial expressions and general hand movements of one actor in each dyad were motion captured.

Databases of video recordings are restricted to the recorded 2-dimensional data. Additionally, video recordings make the examination of contributions of shape (e.g. height), texture (e.g. colour of clothes and skin), and motion cues to the recognition of body expressions difficult. Even with several cameras positioned at different view points, the retrieval of 3-dimensional body motion is a complicated, noise prone procedure. State-of-the-art motion capture systems that provide data for body motion in 3-dimensional space and at high frame rates on the other hand allow: (1) extraction of motion trajectories for further analysis and (2) alterations of the position and movement of individual body joints. Shape and texture cues can easily be manipulated during the animation process using standard software packages.

Context is another important factor for emotion production and recognition (de Gelder, 2009; Parkinson, 2013). Not only do context-set emotion expression allow future research to gain deeper insight into the way people normally express and perceive emotions, but could also make a valuable source of raw data for human motion modelling, e.g., for virtual character or robot animation. Oertel et al. (2013) have used an intricate setup to record naturalistic conversation in a domestic setting. Having given their participants no restriction over the conversation and interaction flow, the authors have non-intrusively gathered social interaction data by means of audio and video recording and motion capture, creating an ideal corpus for studying naturalistic conversational interaction.Thus, a database of motion expression is needed where the actors work with a rich set of emotion categories and their emotion expression is induced without strong conscious effort on their part and without unnecessary exaggeration of the expressions. The motion capture setup should not restrict their motion or influence their emotion expression in any way. While purely non-verbal emotion expression has its valid place in human lives, most emotion instances occur during communication and are accompanied by speech produc-

tion. Thus, for the reasons of ecological validity, it is advantageous to capture motion that occurs during emotional narration. However, the naturalness of emotion expression should ideally be combined with control over data stream by the experimenter. For example, it is useful for actors and experimenters to agree upon an acting script before the motion capture session.

Using state-of-the-art motion capture technology (Roetenberg et al., 2009), we have initially recorded a total of 5.4 hours of motion captured narrations performed by amateur actors, from which a total of 86 minutes, split into 1447 motion sequences were selected and included into the new database. We captured narrations of whole stories and not separate sentences or actions taken out of context and randomised in order. Special effort has been made to collect motion patterns typical for free emotional expression in real life. We used no specific emotion induction technique apart from letting the actors be immersed in the story. The actors were to recount the emotions of the narrator and the characters of the story, imagining they are telling the story to a child or several children. The actors annotated the texts for emotions prior to the motion capture, effectively creating personalised acting scripts.

Most related research has used fewer emotion categories, e.g., the basic emotions (Ekman, 1992). We argue that research on human emotional experience does not have to be constrained to universal emotions. We thus developed a list of eleven emotions that were operated with all stages of our research. The list was built by analysing emotion labels from several previous works that used more categories than only the basic emotions (De Meijer, 1989; Wallbott, 1998; Pollick et al., 2001b; Bänziger et al., 2006; McDonnell et al., 2009; Beck et al., 2012). The following criteria were taken into account during the final emotion categories list compilation: 1) broader span than the basic emotions, 2) manageable size, 3) balance between negative and positive emotions, and 4) categories frequently used in related research. Our final list of emotion categories used five positive (*amusement, joy, pride, relief, surprise*) and five negative (*anger, disgust, fear, sadness, shame*) emotion categories as well as the *neutral* category.

During the motion capture sessions, the text of the story, the emotion category the actor intended to express at every phrase, and the motion capture data were automatically synchronised during the narration process. The performance was also recorded on video and audio, leaving our actors unaware that only their body motion was of primary interest to the researchers and thus they did not exaggerate

the emotion expression through this particular channel. The fact that the motion capture took place in the context of narration, the resulting connections between the original text, its emotional annotation by the actor, their motion pattern and the perception thereof can be valuable for linguists as well. It is also important that our database format is not video but motion capture, which allows one to easily process the data, perform qualitative and quantitative analysis, change the presentation display method by altering underlying body proportions and appearance, modify various motion properties of the whole body or some of its parts in terms of timing and positioning and have full control of the motion stimuli.

Prior to database creation we conducted an extensive emotion recognition study using the collected motion sequences. We used upper human body display to present the motion sequences to multiple observers and let them categorise the stimuli using the same set of emotion categories as had been used by the actors. While a detailed account of the results of the recognition study can be found in Volkova et al. (2014b), a few aspects that specifically fit the scope of this article are reported here for the first time. Importantly, all observers' responses from the recognition study are included in the database, allowing for comparison between actors' intended emotions and the perceived emotion categories and making it a valuable feature for the users of the database.

In the following sections we report detailed information about the acquired motion data and its properties across individual actors and acting tasks. We also include the results of the emotion recognition experiment, where each motion sequence was evaluated by 11 observers. The richness of available information — the extended set of emotion categories, the intended and perceived emotion labels provided for each motion sequence, the corresponding text and physical properties, is likely to prove useful for researchers in the fields of motion modelling, human emotional perception, and linguistics.

3.3 Materials and Methods

3.3.1 Ethics Statement

The database and the emotion recognition experiment described later in this manuscript use human volunteers. Informed written consent was

obtained prior to any experiment or recording from all participants and actors. Participants and data from participants were treated according to the Declaration of Helsinki. The recording methods of the database and the subsequent validation experiment were approved by the local ethics committee of the University of Tübingen.

3.3.2 Texts

Most emotion induction techniques have used short situation descriptions or vignettes to aid enactment of various emotions during the recording sessions. One drawback of this approach is that actors often have to randomly switch between emotions. In our study we asked our actors to perform narrations of original fairy tale stories. Upon initial introduction to the motion capture scene, each actor was asked about their present and past acting experience. We preferred actors that were able to express their emotions freely and yet were careful to avoid recruiting professional theatre actors with many years of experience as they are prone to exaggerate their emotion displays and use stereotypical motion patterns. Various issues of acting *vs.* emotion elicitation, as well as lay *vs.* professional acting are discussed in detail by Bänziger and Scherer (2007).

Each recruited actor was asked to choose three texts out of nine pre-selected fairy tales. The texts of the stories were taken from a large collection of fairy tales recorded by Andrew Lang in a collection of twelve books, published between 1889 and 1910. All nine stories are thus written in unabridged English in consistent style. Andrew Lang's language, being eloquent and poetic, its rich vocabulary likely to trigger various emotions. The actors familiarised themselves with the texts and then read them out loud. The actors' speech was recorded for further processing. At this point it was important that the actor read the texts with appropriate speed. The texts were first split into sentences, then each sentence was split into utterances according to the acoustic pauses in actors' recorded speech. The duration of the pauses used for text splitting was set between 150 ms and 250 ms as these durations usually correspond to brief pauses between clauses but are too long for inter-lexical pauses. For more detail on prosodic pauses duration see Goldman-Eisler (1972); Zellner (1994); Campione and Véronis (2002).

The resulting split narrations were on average almost three hundred utterances long (M=298.5, SD=36.09), each utterance contain-

ing a few word tokens (M=6.84, SD=2.59). The split narrations were then provided as input to our custom-made online annotation tool (www.epetals.org). The actors were asked to use the annotation system and assign one of the eleven emotional labels to each utterance, thus effectively creating personalised acting scripts for themselves. The order of the utterances during the annotation process corresponded to the natural flow of the text. The annotations typically employed the full range of available emotion categories for each annotation (a minimum of eight categories occurred in only one of the annotations). The frequency between categories varied greatly, *neutral* naturally being the most frequent emotion and *shame* the least frequent. These acting scripts were then presented to the actor during the motion capture sessions with the help of PsychoPy software (Pierce, 2007).

During all stages of the motion data collection process, starting with the audio recordings, the actors were asked to imagine that they were narrating the stories to a child or children, which placed them into a socialisation scenario, where they could reflect on which emotions should be expressed at any moment of the story and in which way. Indeed, the actor was never alone in the room during the audio recording or the motion capture and was in fact telling the stories to one or two experimenters.

3.3.3 Apparatus

The motions of the actors were recorded with the Xsens MVN suit (Roetenberg et al., 2009) in a large, quiet room. Xsens MVN is a full body suit made of lycra, it includes 17 sensors aligned with anatomical landmarks of the body (see Figure 3.1). Each sensor is composed of an accelerometer, a gyrometer and a magnetometer. The placement of the sensors and corresponding cables in the suit enables the user to perform unrestricted actions. Two master units in the back of the suit transmit the data from the sensors wirelessly to a computer, where the software MVN Studio (Roetenberg et al., 2009) maps the captured data onto a actor-scaled human skeleton-like 3D model and records the resulting motion output in real-time, at 120 FPS. Although inertial motion capture systems can introduce errors in the absolute position of the suit wearer in space, the data for relative orientation, acceleration, and velocity they provide is reliable. In our motion capture setup the actor was seated on a stool, thus the possible drift was of no consequence since the centre of the skeleton was stationary in real

world.

The motion capture sessions were also video recorded with the permission of the actors. Due to privacy protection of the actors, the video and audio recordings were used only for reference. Another important function of the the video camera and the microphone in the setup was to create an impression that every aspect of emotion expression, e.g., prosody and facial expression, were important for the research purposes of the study. This prompted the actors to express the emotions in a more natural way, without exaggerating or suppressing any of the emotion expression channels. The open source presentation software used for acting scripts display (PsychoPy, Pierce (2007)) and the motion capture software (MVN Studio) ran on a Dell Precision M6400 laptop (Intel Core2 Duo 2.8 GHz, 4GB RAM, nVidia Quadro FX 3700M graphics card with 1024MB VRAM). The laptop was attached to a 20″ Dell external monitor with a resolution of 1280x1024 pixels and used to show to the actors the current utterance or acting motivation and the corresponding emotion category label. An external keyboard and pedals, designed for the experiment, were connected to the laptop via USB (Figure 3.1).

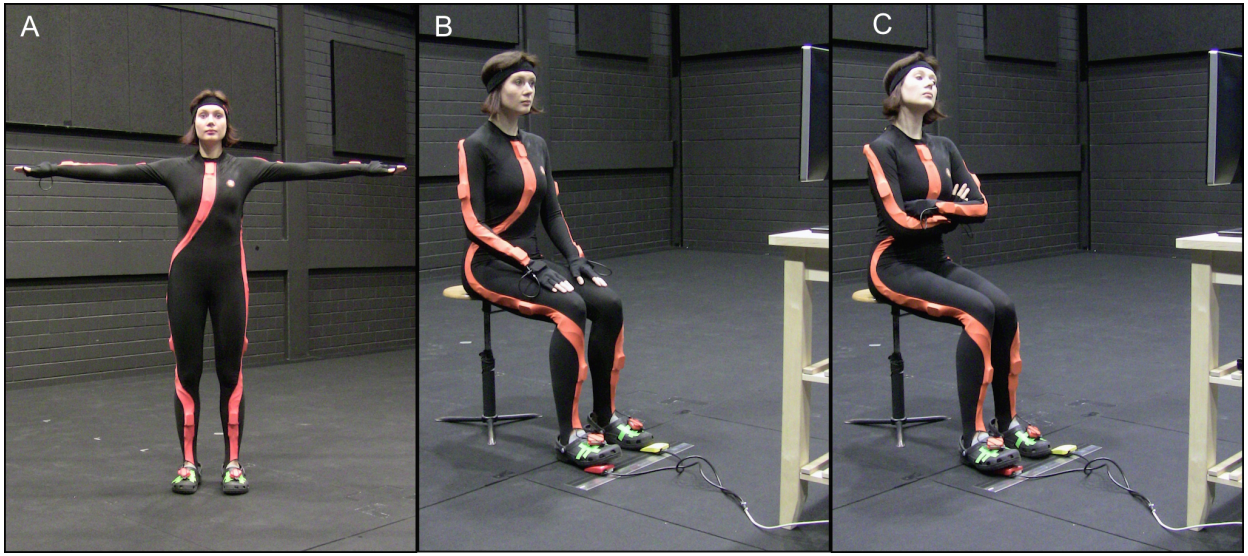

Figure 3.1: Setup for motion capture sessions. (A) An actor in motion capture Moven Xsens suit, t-pose (B) Acting setup for motion capture of narrative scenarios: actor in neutral pose, stool, pedals and display. (C) Actor expressing pride.

3.3.4 Motion Capture Procedure

Eight actors (4 female), 20 to 33 years old (M = 25.6, SD = 4.24) participated in the motion capture sessions. Informed written consent was obtained from each actor regarding their audio, video, text annotation, and motion capture data. Only the latter two types of data, due to privacy reasons, are available to the community. All actors received monetary compensation for their participation and none were aware of the end purpose of the motion capture until all the data was collected. After that the actors were offered debriefing concerning the goal of the study. All actors had normal or corrected to normal visual acuity, the male actors originated from Germany, one female actor came from India, two - from England, and one from Ireland. All participants' command of English was either at the native speaker's level or high enough to act the English texts smoothly without difficulties.

After having annotated the three chosen texts with the available emotion categories, each actor came for the first motion capture session, where they were given four types of short scenarios to act out. Each scenario type encompassed ten emotional categories (all the categories listed in the previous section with the exception of *neutral*). The emotion categories were randomised for each scenario and each actor. In each scenario instruction, the actor would see the goal emotion and the description of the situation, so the actors did not have to interpret and label the emotional content of the short scenarios themselves. The short scenarios largely resemble classical emotion induction vignettes used in related literature (Pollick et al., 2001b). These scenarios are later compared to the coherent narrations. The scenario types increase in their verbal content and subtlety of emotions and were performed in this order:

1. Solitary **non-verbal** emotional scenarios (**NS**): the instructions indicated that the actor was to imagine that he or she was alone. Example: (*Pride*) "You are sitting alone in your room at your desk and have just finished doing an online IQ test. According to the results, you are among the 5% smartest people in the world!"

2. Communicative **non-verbal** emotional scenarios (**NC**): the instructions indicated that the actor was in company with a friend / friends / a colleague / colleagues. Example: (*Surprise*) "You are at a cafe catching up with a friend you haven't seen for a long time. She tells you that she is now doing skydiving and ice

climbing."

3. Short **sentences** without direct speech (**SN**): The actor was asked to act out a short preselected sentence from a fairy-tale, the sentence contained no direct speech, the person acted on behalf of the narrator of the story. Example: (*Joy*) "One of Rapunzel's tears fell on prince's face and he could see again."

4. Short **sentences** with direct speech (**SD**): The actor was asked to act out a short preselected sentence from a fairy tale, the sentence contained direct speech, the person acted on the behalf of a character and, if needed, the narrator of the story as well. Example: (*Fear*) " "Please don't kill me!" – cried the rabbit – "I might be useful to you in the future!" "

In the three following motion capture sessions each actor worked on one story at a time. The text of each story was shown phrase by phrase in its natural order as was determined by the previous delimitation and annotation processes. Each phrase was shown together with the emotional label that had been assigned to it by the actor during the fairy tale text annotation process, e.g., Joy*: "...you have a kind heart...".*

In all recording sessions, each scenario type and story narration were rehearsed and recorded two to three times, in order to allow actors to get used to the text of the story and the emotion flow, but only the last recording was later used for the recognition study and the motion capture dataset compilation. During the recording sessions the actors were seated on a stool 1.5 meters away from the Dell monitor (Figure 3.1). The size of the text on the screen was large enough for the actors to read it effortlessly. They progressed through short scenarios or coherent stories by pressing the right foot pedal. This allowed them to maintain their own performance speed and also kept their upper body free for the full range of emotional expressions. The timing of pedal presses was recorded for synchronisation of acting script presentation and the motion capture stream. Especially in narration tasks, in order to minimise the risk of asynchrony between the actor's progress through the story and the recorded pedal presses, we instructed the actors to finish acting out each phrase before proceeding to the next one. It is important that, although the actors felt involved in the narration process, they were not required to recall personal emotional memories — a technique often used for emotion induction, which could affect

actor's overall emotional state if some particularly strongly unpleasant emotional memories were recalled.

3.3.5 Motion Capture Files Format

In our database the motion sequences are available in the following formats:

BVH (Biovision Hierarchy) was developed by a motion capture company called Biovision. BVH is one of the most popular motion data formats and is mainly used as a standard representation of movements in the animation of humanoid structures. The specification of a typical BVH file is relatively straightforward. In the first part of each file, that can be viewed in any text editor, one finds the skeletal hierarchy. It begins with the "root" node and continues to nest child "joint" bones. The number of joints and structure of the hierarchy solely depend on the previous motion capture recording and/or export setup. "Offset" describes the offset of each joint from its parent. "channels", typically six for the root (position and orientation) and three for the rest of the joints (orientation only) describe the motion data. Note that BVH rotations are recorded in Euler angles instead of quaternion terms. The keyword "motion" marks the beginning of the motion data. Number of frames and the frame time (sampling rate) are also given. Our files have 0.008333 frame rate, or 120 frames per second. The rest of the file contains the actual motion data. Each line is one sample of motion data in one frame. The numbers appear in the order of the channel specifications as the skeleton hierarchy was parsed. The hierarchy in our BVH files has 22 joints with 3 channels each and the root with 6 channels. Thus, every motion data line has 72 numbers.

MVNX (Moven Open XML format) contains position, orientation, acceleration, velocity, angular rate, and angular acceleration of each joint (also called segment) in 3D. The files have XML structure with elements such as "" or "<frames>" and attributes to the elements, e.g., "". The MVNX files consist of several major sections. The "mvnxInfo" section contains overview information on the number of frames, frame rate and other descriptive information. The "meshScale"

section is used for visualisation of the character in Moven Studio only, while "segments" defines all positions of joints. Unlike the BVH format, the "segments" section is not hierarchical, however, each segment has several points with their individual offset positions with respect to the origin of that segment. The "frames" section contains the actual motion data, where each frame is represented by one row containing 23 segments, 7 channels in each segment: 4 for quaternion orientation, 3 for the position. The last two values are the time of the frame in milliseconds and the timecode. Both time values rarely start from 0 in our data since motion sequences originate from longer motion caption sessions and the original timing values were retained. All values are set in the global coordinate system. Additionally, kinematic information is available in 1×3 vector form for each segment for velocity (m/sec), acceleration(m/sec^2), angular velocity (rad/sec) and angular acceleration (rad/sec^2) in corresponding file sections.

3.4 Results

3.4.1 Emotion Recognition Study and Motion Sequences Selection

We obtained seven motion capture files from each actor — four sets of short scenarios and three coherent stories. The files were then split into smaller sequences according to pedal press timestamps. The short scenario motion sequences were trimmed to cut away the time stretches before and after the motion. In sum, 320 motion sequences were obtained (10 emotions $\times$ 4 types of situations $\times$ 8 actors). The 24 coherent narrations (3 stories $\times$ 8 actors) were split into shorter motion sequences based on timestamps obtained from foot pedal presses. In total 1380 motion sequences were selected from narration motion capture data. The resulting 1700 motion sequences were further used for the emotion recognition experiment which was part of a larger perceptual study. While the results of the emotion recognition experiment in relation to emotion categories are given in full detail in Volkova et al. (2014b), here we will briefly report on its structure. We also report some results that were out of scope of the previous study but are highly relevant here.

The emotion recognition study used the full dataset of 1700 mo-

tion sequences. The motion sequences were displayed as dynamic stick-figure representations of human upper body. Fifty five participants (28 female) took part and each participant categorised a unique combination of 340 randomly chosen motion sequences. The stimuli blocks were organised in such a way that by the end of the experiment each of the 1700 motion sequences had been categorised by eleven participants. In every trial the participant was to choose one of the eleven emotion categories in a two-step response procedure, by first deciding whether the presented motion was *emotional* or *neutral*, then, if they had chosen the former, categorising the motion with one of the ten remaining emotion categories. The response could always be changed until the participant was satisfied and proceeded to the next trial.

As many as 85% of motion sequences have a unique modal value in the distribution of observers' categorisation. The category of the unique modal value is henceforth referred to as *perceived emotion*, while the emotion category obtained from the actor's annotation corresponding to the sequence is referred to as *intended emotion*. The frequency of the intended emotion categories as indicated by the actors for every selected motion sequence during the annotation phase (intended emotion frequency) is shown in Figure 3.2 (A). The frequency of the most given emotion response (i.e. perceived emotion frequency) for each motion sequence across emotion categories in the database is given in Figure 3.2 (B) and (C).

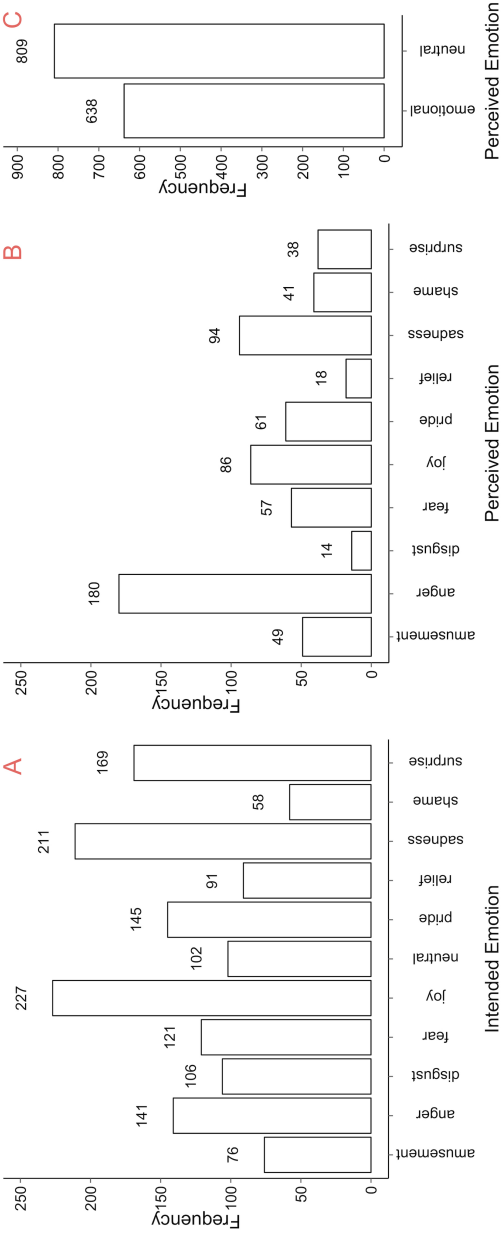

Figure 3.2: Motion sequence frequencies across intended (A) and perceived (B, C) emotions. Intended emotions originate from actors' text annotations while perceived emotions come from the categories forming unique modal value in observers' response distribution for every motion sequence. The perceived emotion frequencies are split into two graphs to allow same y-axis scale for (A) and (B) graphs. The *emotional* category in plot C is the sum of all frequencies in plot B.

The distribution of the perceived emotions across categories is very different from the distribution of intended emotions as indicated by the actors. The observed agreement between the intended and perceived emotion categories is 19%. This result suggests that the emotion intended by the actor is often perceived as a different category by the observer. It further highlights the importance of recording the intended emotions in the context of situational emotional expressions that are not based on predefined or exaggerated emotions, along with the perceived emotions based on the same motion sequences. The current database provides this information.

In our final database we included those motion sequences that have a unique modal value in the observers' response distribution, the total number amounting to 1447 motion sequences. While 85% of motion sequences have a unique perceived emotion category, the number of observer responses falling into the modal category varies. Thus we define the measure of *consistency* as the proportion of observers' responses that fall into the most frequently chosen emotion category. For example, if for a motion sequence six out of the eleven participants categorised the motion sequence in the same way, consistency equals 6/11=0.545. Note that, in cases when the response distribution is bimodal or multimodal (i.e. more than one category received the same number of responses), consistency is not defined.

The consistency is useful if the researcher wants to investigate the perception of emotions from body motion, for instance, to identify motion sequences or emotion categories that are perceived similarly across different observers. Our database organisation makes it easy to extract motion sequences with user-selected perceived categories and/or motion sequences with defined consistency rates. Figure 3.3 (A) shows the distribution of consistency rates across motion sequences perceived as *non-neutral*, showing that 0.3 is the most frequent consistency rate. Applying three cut-off levels (0.3, 0.5 and 0.7) we show the frequency of ten *non-neutral* emotion categories that include motion sequences with the cut-off consistency rate or higher (Figure 3.3 B, C, D). Motion sequences included in the graph (C) are a subset of motion sequences from the graph (B) and motion sequences from the graph (D) are a subset of motion sequences from the graphs (B) and (C). *Anger* and *sadness* are always the most frequent perceived emotion categories, indicating that the observes of the corresponding motion sequences were in high agreement with each other. On the contrary, *surprise, relief* and *disgust* are the least frequent and are not even present in the graph

(D), meaning that though these categories formed the modal value in the observers' responses for certain motion sequences, the latter were not numerous and the number of responses falling into the categories was never more than half of the total responses available.

The perceptual study that focused on the recognition of emotion categories, its results and implications are described in detail in Volkova et al. (2014b). In the rest of this section we analyse two fundamental sources of variability: the acting tasks they originated from and the actors they were performed by. For each of those sources we provide the following data: observed agreement between perceived and indented emotions (referred to as *recognition accuracy*), consistency rates, and physical motion properties. The latter are represented by four features: duration (in seconds), peaks (mean number of peaks and valleys across the x, y, z trajectories for each joint in question), speed (m/sec), and span (average distance between the joints, m). For the purposes of simplicity, here we used only the left and the right wrist joints of the underlying joint configuration for the motion properties extraction, as these were the most active joins in the motion production and are thus most suitable for motion characterisation.

3.4.2 Motion Sequences Distribution

Tables 3.2-3.4 show the final number of motion sequences included into the new database. Table 3.2 shows frequency of motion sequences across emotion categories and acting tasks, while Table 3.3 shows each actor's final contribution to the database across acting tasks and specific stories. The column name abbreviations stand for acting tasks described earlier in section *Motion Capture* and Table 3.1. Table 3.4 shows how many motion sequences were contributed by each actor into the final database across the emotion categories they intended to express. The next section gives detailed information on the results of the emotion recognition study, focusing on perceived emotion categories across acting tasks and individual actors.

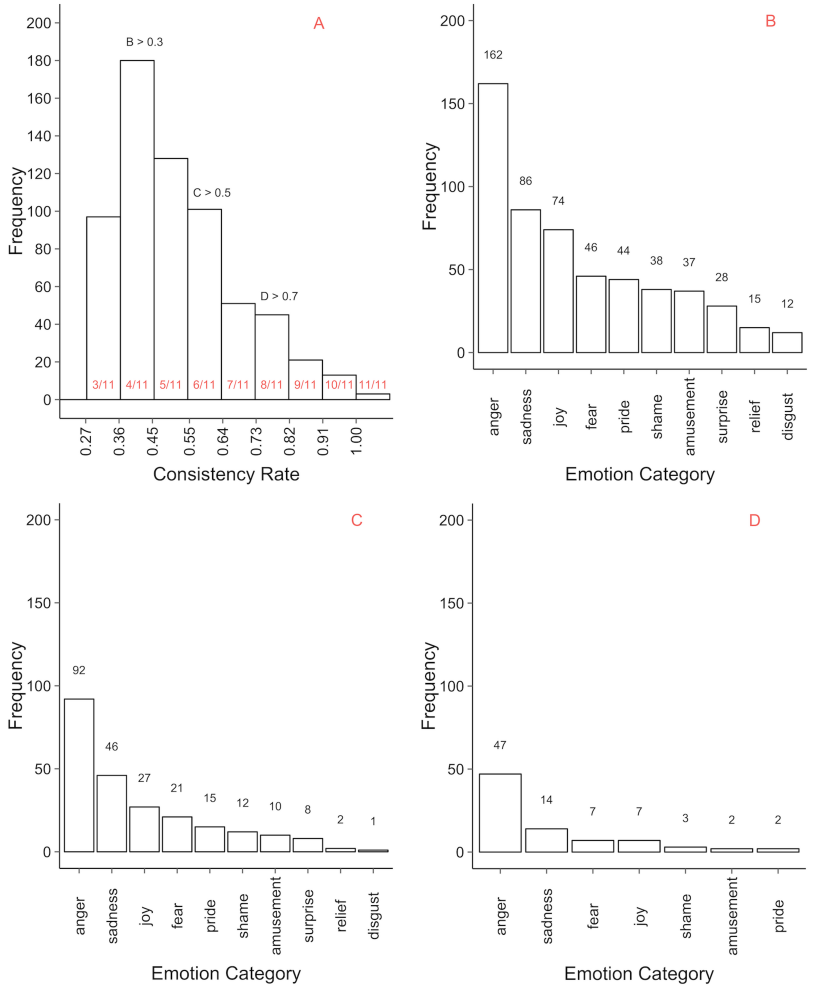

Figure 3.3: Emotion frequency distribution across consistency levels. A: histogram of consistency rates across motion sequences. The minimally possible consistency is always equal to one divided by the number of observations for the given stimulus and multiplied by two because there have to be at least least two observers assigning the same category to the stimuli to form a modal value. B: Distribution of perceived emotions across categories with consistency rate 0.3 or more. C: Distribution of perceived emotions across categories with consistency rate 0.5 or more. D: Distribution of perceived emotions across categories with consistency rate 0.7 or more.

Story Title	AnBh	DiMi	HeGa	LeSt	MaMa	NoVo	PaPi	SiGl	Count
Blue Beard (TBB)	✓				✓	✓	✓		4
Flower Princess (TFP)	✓	✓	✓				✓		4
Golden Goose (TGG)		✓	✓	✓					3
Hoodie Crow (THC)				✓	✓				2
Jack My Hedgehog (TJH)						✓	✓	✓	3
Owl And Eagle (TOE)					✓				1
Six Swans (TSS)								✓	1
Swineherd (TSH)			✓					✓	2
White Duck (TWD)	✓	✓		✓		✓			4
Sum	3	3	3	3	3	3	3	3	24

Table 3.1: Stories narrated by actors during motion capture sessions. Each actor performed three stories in total, one story per session. To increase actor's motivation and comfort, the choice of three stories out of available 9 was upon the actor and not the researcher, which resulted in the fact that some stories were acted by several actors (e.g., TBB, TWD) and some by only one actor (e.g., TSS).

3.4.3 Emotion Recognition and Consistency across Acting Tasks and Actors

Motion sequences included in the database originate not only from different actors but also from different acting tasks. In this section we organise motion sequences into three groups according to the tasks they were produced in: non-verbal, short sentences and narrations (see also Table 3.2). Figure 3.4 shows that non-verbal motion sequences have received higher recognition accuracy (0.35, with 1.0 being the absolute possible maximum) than their counterparts that included speech (0.25 and 0.21 for short sentences and narrations respectively). This suggests that it was easier for observers to recognise the actor's intended emotion when it was expressed non-verbally. The consistency however is on average the same across tasks: 0.49, 0.49 and 0.51 respectively (Figure 3.4), showing that observers had relatively high agreement rates among each other for all motion sequences even when they disagreed with the actors.

The actors who took part in this research were young adults with varying acting experience (currently active amateur to amateur actors in the recent past). As one can see, recognition accuracy rates vary greatly among individual actors, ranging from 0.07 to 0.4 (also see Table 3.5), but the consistency is comparable across all actors (Figure 3.5). Nevertheless, the number of motion sequences coming from each actor ranges from almost three hundred (292 from "AnBh") to under one hundred (77 from "MaMa").

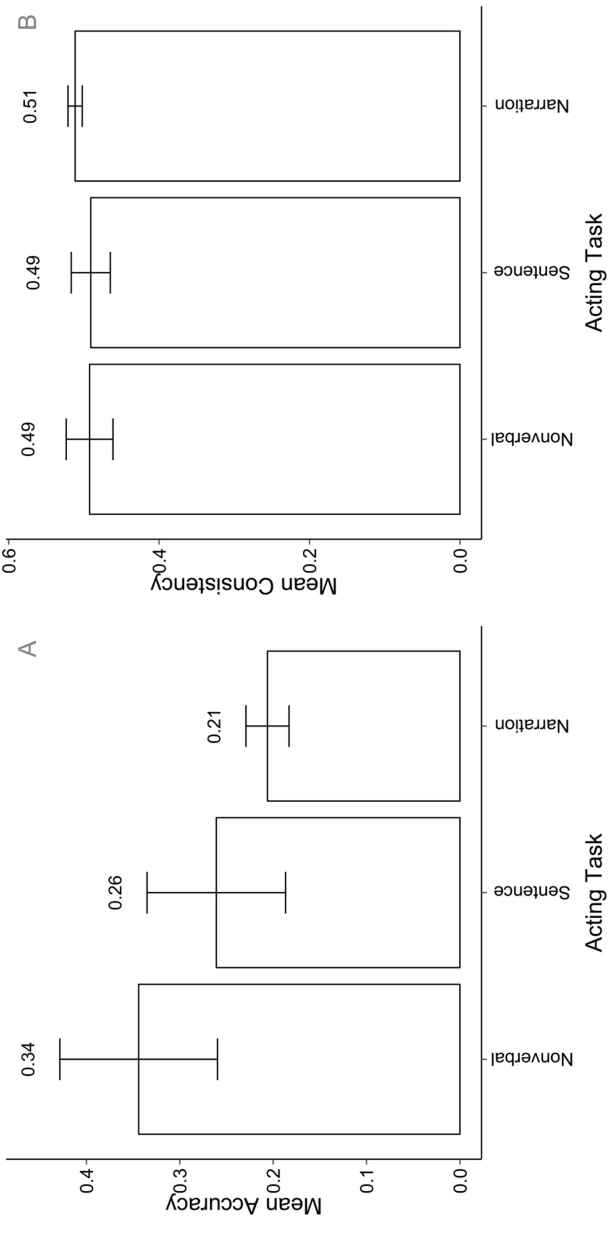

Figure 3.4: Emotion recognition accuracy across acting tasks for intended emotions and consistency rates for perceived emotions across acting tasks. All error bars represent 95% CI.

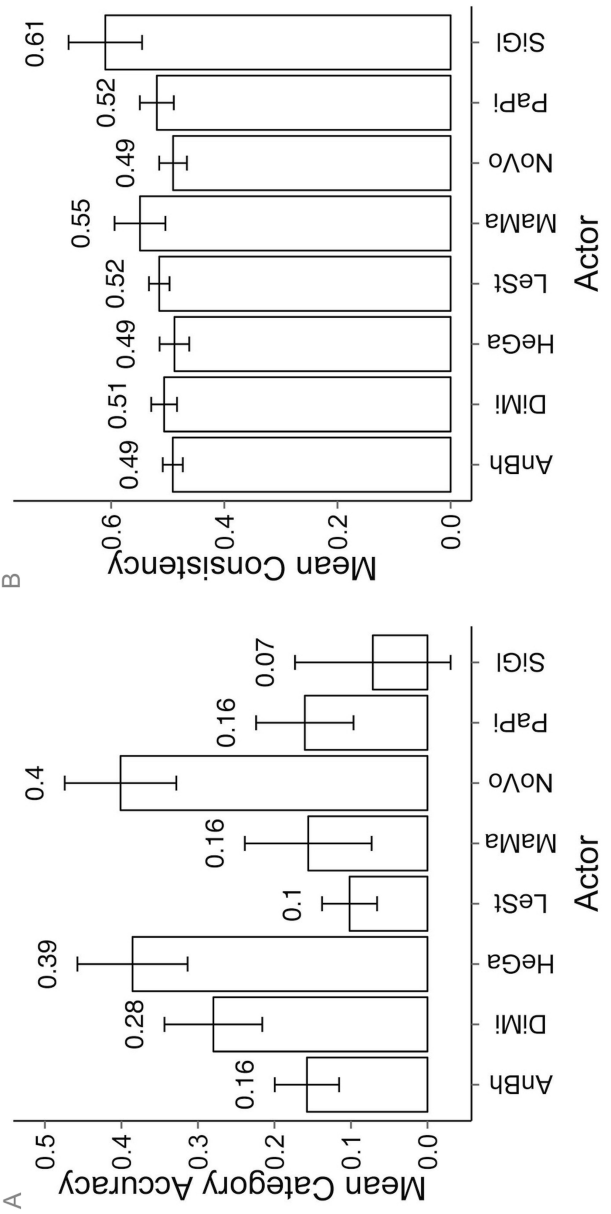

Figure 3.5: Average recognition accuracy across actors and observers' response consistency across actors. All error bars represent 95% CI.

Emotion	Short scenarios				Narrations								
	Non-verbal		Sentences										
	NS	NC	SD	SN	TBB	TFP	TGG	THC	TJH	TOE	TSS	TSH	TWD
amusement	7	6	8	6	4	31			8				6
anger	5	8	8	8	24	9	17	4	17		5	13	23
disgust	4	6	5	6	8	9	18	3	12	1		11	23
fear	5	7	7	8	29	12	4	4	6	2	6	3	28
joy	7	5	8	8	13	52	29	16	15	23	3	21	27
neutral					6	25	11	5	10	2		5	38
pride	8	5	8	8	18	39	13	1	17	6	2	12	8
relief	6	7	4	4	9	18	4	10	6	3	3	1	16
sadness	7	6	8	7	19	14	34	12	12	3	9	6	74
shame	6	7	7	8	3	8	8	2	4			2	3
surprise	7	6	5	7	20	18	26	21	12	4		11	32

Table 3.2: Frequencies in the final set of motion sequences across intended emotion categories and acting tasks. The abbreviations stand for: NS — non-verbal solitary, NC — non-verbal communicative, SD — sentences with direct speech, SN — sentences without direct speech, TBB — Blue Beard, TFP — Flower Princess, TGG — Golden Goose, THC — Hoodie Crow, TJH — Jack My Hedgehog, TOE — Owl and Eagle, TSS — Six Swans, TSH — Swineherd, TWD — White Duck.

Actor	Short scenarios				Narrations								
	Non-verbal		Sentences										
	NS	NC	SD	SN	TBB	TFP	TGG	THC	TJH	TOE	TSS	TSH	TWD
AnBh	7	8	9	10	98	67							93
DiMi	9	9	9	9		57	48						52
HeGa	9	7	9	8		57	35					54	
LeSt	7	9	8	8			81	78					84
MaMa	9	7	8	9						44			
NoVo	8	8	8	8	55				41				49
PaPi	6	7	7	10		54			47				
SiGl	7	8	10	8					31		28	31	

Table 3.3: Frequencies in the final set of motion sequences across actors and acting tasks. The abbreviations stand for: NS — non-verbal solitary, NC — non-verbal communicative, SD — sentences with direct speech, SN — sentences without direct speech, TBB — Blue Beard, TFP — Flower Princess, TGG — Golden Goose, THC — Hoodie Crow, TJH — Jack My Hedgehog, TOE — Owl and Eagle, TSS — Six Swans, TSH — Swineherd, TWD — White Duck.

	AnBh	DiMi	HeGa	LeSt	MaMa	NoVo	PaPi	SiGl	total
amusement	7	3	4	4	4	16	35	3	76
anger	30	22	20	15	4	24	6	20	141
disgust	14	11	13	30	3	8	12	15	106
fear	40	14	9	17	5	16	8	12	121
joy	52	27	33	46	26	10	17	16	227
neutral	25	10	18	20	2	12	10	5	102
pride	27	21	23	10	10	19	23	12	145
relief	25	14	4	22	6	5	6	9	91
sadness	41	31	15	61	6	30	8	19	211
shame	7	13	5	13	4	3	5	8	58
surprise	24	27	35	37	7	34	1	4	169
total	292	193	179	275	77	177	131	123	1447

Table 3.4: Frequencies in the final set of motion sequences across actors (rows) and intended emotion categories (columns).

	NoVo	HeGa	DiMi	AnBh	LeSt	SiGl	PaPi	MaMa	total
neutral	9	18	9	18	20	3	9	2	88 (.86)
anger	12	17	9	9	5	7	2	4	65 (.46)
sadness	15	1	16	8	1	2	2	2	47 (.22)
fear	14	3	10	3		2			32 (.26)
joy	7	14	2	2	1	2		1	29 (.12)
pride	4	10	1	3	1	1	3	1	24 (.16)
surprise	9	3	1					2	15 (.08)
amusement		2		2		1	3		8 (.10)
shame			5	1		1	1		8 (.14)
disgust	1		1			1	1		4 (.03)
relief		1				2			3 (.03)
total	71 (.40)	69 (.38)	54 (.28)	46 (.15)	28 (.10)	22 (.16)	21 (.16)	12 (.15)	323 (.22)

Table 3.5: Frequencies of motion sequences across actors where intended and perceived emotion categories coincide, sorted by frequency within each emotion category (rows) and actor (columns). Values in round brackets represent the proportions of the frequencies in relation to the whole database (see Table 3.4).

Physical Properties of Motion Sequencesacross Acting Tasks and Actors

Figure 3.6 shows motion properties across the three acting tasks. On average, motion sequence have longer duration in the narration task than in short scenarios, and also lower speed and number of trajectory peaks. Such slower and smoother profile in narration scenarios can be probably explained by the fact that it was accompanied by coherent speech plus the fact that narration task is the only of the three that contained *neutral* motion accruing to actors' intentions. Note however, that there is almost no difference in motion span across the acting tasks, showing that the actors used approximately the same space during all the motion capture sessions.

Actors vary not only across recognition rate by the observers, but also by their ways to produce body motion, which is reflected in Figure 3.6. While the duration of motion sequence is mostly uniform among the actors (M=3.42, SD=0.55, min=2.47, max=4.21), motion speed, peaks in motion trajectories and especially motion span vary greatly among individual actors (speed, m/sec: M=0.14, SD=0.03, min=1.18, max=1.96; span, m: M=0.26, SD=0.09, min=0.14, max=0.38; number of peaks, *count*: M=4.62, SD=0.99, min=3.58, max=6.65). The variation in span among actors can be partially due to the differences in individual motion style, but also due to different body configurations, e.g., maximum arm span.

3.4.4 Final Database Format

All motion sequences are available to the community, each motion sequence coming with additional meta-information that describes its various aspects, e.g., the physical properties of the motion, the intended emotion according to the actor and the response categories from the participants of the recognition study. Information about the source of the motion sequence is also presented — the acting task the sequence comes from (short scenario or a full narration with the story title) and for most motion sequences what exactly the task was, namely, the motivation sentence for short scenarios and the specific phrase that was acted out during the production of that motion sequence. Some information about the actors is available as well, such as the actor's code name, their gender, age, and mother tongue. Table 3.6 gives an overview of avail-

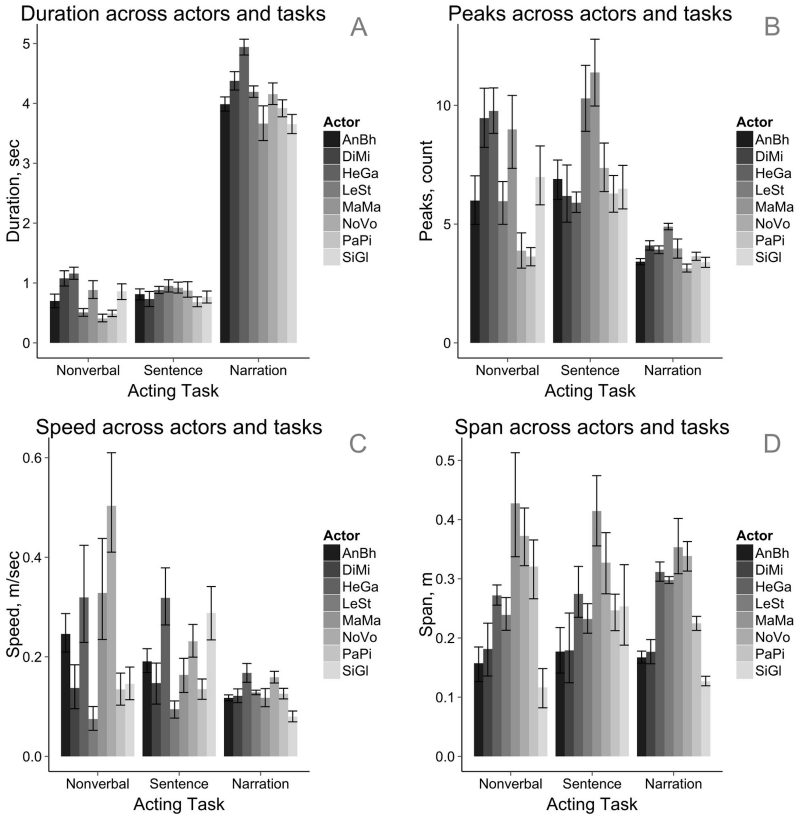

Figure 3.6: Physical properties of motion sequences across acting tasks and individual actors. All error bars represent 95% CI. The panels show: (A) Duration, sec; (B) Peaks in motion trajectories of right and left wrists across acting tasks; (C) Average motion speed; (D) Average motion span. As the bar plots show, physical properties of the motion sequence depend both on the acting task and on the individual the actor.

able information for each motion sequence. The database is free for access after a short registration procedure. The user can filter through and search for the information in the fields and select a subset of the motion sequences. For each selected motion sequence various formats as well as the descriptive meta-data information are available for download[1].

[1]http://ebmdb.tuebingen.mpg.de

Column Name	Description
	Actor-dependent Motion Properties
Motion Id	Unique motion sequence file name
Intended Emotion	One of the eleven emotion categories as intended by the actor
Intended Polarity	*Positive, negative* or *neutral* polarity of the motion sequence as intended by the actor
Duration	Duration of the motion in seconds
Peaks	Number of peaks and valleys in motion trajectory along x, y, z axes for the left and the right wrist joints
Speed	Average speed in m/sec for the left and the right wrist joints
Span	Average span in meters between the left and the right wrist joints
Acting Task	Nonverbal, Sentences, Narration
Acting Sub-task	Specific sub-task or story title (see Table 3.2)
Actor	The id name for one for the eight actors who performed the motion sequence
Gender	Actor's gender ("f" – female, "m" – male)
Age	Actor's age ($min = 22, max = 30$)
Handedness	Actor's handedness ("r" – righthanded, "l" – lefthanded
Native Tongue	Actor's mother language (German, English, Hindi)
	Observer-dependent Motion Properties
Perceived Category	One of the eleven emotion categories as perceived by majority of the observers
Perceived Polarity	*Positive, negative* or *neutral* polarity of the motion sequence as perceived by majority of the observers
Accurate Category	"1" when intended and perceived emotions coincide, "0" otherwise
Accurate Polarity	"1" when intended and perceived polarity coincide, "0" otherwise
Responses	The list of eleven responses to the motion sequence from all the observers
Consistency	The proportion of responses taken by the unique modal value, which is also recorded in "Perceived Category"
Text	The text that served as acting motivation (not spoken out lout in non-verbal tasks)

Table 3.6: Online database table overview

3.5 Discussion

Our new database gives researchers access to human body motion patterns produced during emotional narrations. This is unique and valuable for several reasons. First, the narration performance was kept as close to natural as possible. The actors were not aware that only their body motion was of primary importance to the researchers. Their facial expressions and voice were recorded along with the body motion and the actors were narrating an unabridged story. Second, the rich set of emotion categories is another valuable feature of the database - we used not only the six basic emotions (*anger, disgust, fear, joy, sadness, surprise*) but expanded the list with four extra emotional categories: *amusement, pride, relief and shame* and added the category of *neutral*. Finally, the motion capture format itself is a big advantage for motion pattern analysis, as it depicts human motion in 3D space and with a high time resolution.

Researchers from several branches of science may find our database especially useful, i.e. psychologists, computer scientists, neuroscientists, and linguists. The motion sequences could be used in emotion perception research, motion analysis and synthesis, in behavioural and brain imaging studies. The motion capture took place in a naturalistic setting, ensuring that the resulting motion sequences represent expression of emotions via body language without unnecessary exaggeration. Thus, these patterns have a potential for model building in order to synthesise expressive natural looking motion for virtual character animation. Unlike video recordings, the motion capture format in our database allows one to display biological motion and change many of its properties, e.g. speed of selected trajectories, magnitude, scale and position in space relative to the rest of the body, and visual representation (point lights, full skeleton or parts of skeleton). Segments of the body structure can be concealed altogether. Individual properties of the actors, e.g., their age, gender, body shape and other factors potentially confounding results when pure human motion is in question, are also not available to the observer but using animation tools could be added in a controlled experiment. Since every motion sequence comes with the information about the actor who performed it, and with the information available about the actor (gender, age, native tongue), our data could be used for studying individual motion styles as well as differences and similarities of emotion expression across genders and cultures. The 24 annotations of texts can be used as a separate corpus for research in sentiment

analysis or used together with the motion capture files to study human gesticulation when it accompanies meaningful speech, including the semantic aspect of gesture production.

Our database provides not only the motion sequences labeled for actors' intended emotion displays and the original annotations of full narration texts, but also the perceived emotions obtained during an emotion recognition study (Volkova et al., 2014b). The participants of our recognition study observed motion patterns of upper body of the actors applied to a stick-figure the display of standardised size. Additionally, the observers did not watch the full narrations but were presented with short motion sequences in randomised order. The simplification of the stimuli was important as we wanted to investigate the amount of information the body motion itself brings into the emotion expression in narrative scenarios. As we found out, the participants were able to recognise the emotions at the above chance level even in the impoverished stimuli. The resulting categorisation responses from the observers are included in the database and are available as metadata with each motion sequence file. As each motion sequence was categorised by eleven observers, the most frequently chosen emotion category (the perceived emotion) is included into the database as a value separate from the full list of responses. The proportion of responses the perceived category (consistency) takes and its correspondence to the emotion intended by the actor (accuracy) are provided in the database as well.

3.6 Conclusions

In this study we first collected a large database of motion sequences from multiple actors by designing a way to elicit and record motion in narrative scenarios. Related research (Pollick et al., 2001b; Atkinson et al., 2004; Roether et al., 2010) gained insight into how emotional body expressions are produced and perceived, but the used datasets of emotional portrayals were, sometimes deliberately, exaggerated and thus unlikely to occur in typical day-to-day life experience. We aimed to collect natural emotional body expressions to deepen the understanding of the role of emotional body expressions in natural human communication. We thus chose coherent texts as basic material for our study. The resulting emotional monologues are a close approximation to a bed-time story or an anecdote told to friends at a party, and not an

attempt to reproduce stage acting settings.

We used eleven emotional categories, a richer set than in most related research, in order to allow actors to encode and express their perception of story texts in full detail. During the motion capture sessions the actors were free to express emotions through their speech, face and body. This freedom of emotion expression was an important aspect of our research, since we were interested in keeping the amount of information expressed via body motion at natural level of a typical narration scenario. We make a database of emotion body expressions available to the scientific community that includes the motion capture files, videos, the context by way of the language used and the intended emotions expressed by the actor as well as the perceived emotions determined by a emotion recognition study of upper body language expression.

4 Emotion Categorisation of Body Expressions in Narrative Scenarios

Volkova, E., Mohler, B. J., Dodds, T. J., Tesch, J., and Bülthoff, H. H. (2014b). Emotion Categorisation of Body Expressions in Narrative Scenarios. *Frontiers in Psychology*, 5(623)

4.1 Abstract

Humans can recognise emotions expressed through body motion with high accuracy even when the stimuli are impoverished. However, most of the research on body motion has relied on exaggerated displays of emotions. In this paper we present two experiments where we investigated whether emotional body expressions could be recognised when they were recorded during natural narration. Our actors were free to use their entire body, face and voice to express emotions, but our resulting visual stimuli used only the upper body motion trajectories in the form of animated stick figures. Observers were asked to perform an emotion recognition task on short motion sequences using a large and balanced set of emotions (amusement, joy, pride, relief, surprise, anger, disgust, fear, sadness, shame and neutral). Even with only upper body motion available, our results show recognition accuracy significantly above chance level and high consistency rates among observers. In our first experiment, that used more classic emotion induction setup, all emotions were well recognised. In the second study that employed narrations, four basic emotion categories (joy, anger, fear and sadness), three non-basic emotion categories (amusement, pride and shame) and the "neutral" category were recognised above chance. Interestingly, especially in the second experiment, observers showed a bias towards anger when recognising the motion sequences for emotions. We discovered that similarities between motion sequences across the emotions along such properties as mean motion speed,

number of peaks in the motion trajectory and mean motion span can explain a large percent of the variation in observers' responses. Overall, our results show that upper body motion is informative for emotion recognition in narrative scenarios.

4.2 Introduction and Related Research

Emotion is an integral part of human-human interaction. During communication, we receive and transmit emotional information through many channels: prosody, facial expressions, word choice, posture, and body motion. The human body is often perceived as a tool for actions (e.g., walking, grasping, and carrying), but it is also an important medium for emotional expression (De Meijer, 1989; de Gelder et al., 2010). During communication, body motion can highlight and intensify emotional information conveyed by other channels (e.g., hitting the table with a clenched fist while expressing anger with the voice and the face), add extra nuances of meaning to emotional expressions (e.g., bowing slightly while greeting someone to show respect), or contrast emotional information coming from other channels (e.g., crossing your arms while saying "This is just great." implies that you are actually displeased).

The research on emotional body language is particularly challenging because of the complexity of biological motion, since the human body has hundreds of degrees of freedom and can be used for action and emotion expression simultaneously. Here we will briefly mention the research most relevant to our work, for a more detailed and comprehensive survey please see the survey by Kleinsmith and Bianchi-Berthouze (2013). Earlier studies of biological motion mostly relied on still frame or video recordings for stimulus generation. Johansson (1973) developed now widely used technique of biological motion representation that retains motion information but eliminates form information. The moving figure is marked by a small number of illuminated points or stripes, that are positioned at the main body parts and joints. In the resulting point-light stimuli only these bright marks are visible to the observer. Such stimuli are strongly degraded, and so the identity of initial actors, as well as their age, gender, and body shape are hidden from the observer. The following years have seen point-light technique frequently applied in research on perception of biological motion, including emotion recognition studies. Some of

the earlier studies concentrated on emotion perception from dance (Walk and Homan, 1984; Dittrich et al., 1996; Brownlow et al., 1997). Many studies have investigated the recognition of human actions (Pollick et al., 2001a) and intentions (Manera et al., 2010), identity (Loula et al., 2005), gender (Troje, 2002; Pollick et al., 2005; Brooks et al., 2008) and emotion (Pollick et al., 2001b; Atkinson et al., 2004; Clarke et al., 2005; Beck et al., 2012; Ennis and Egges, 2012) from biological motion using point-light displays. These studies showed that this degraded representation of the body motion still conveys enough information for the observers to accurately recognise the stimuli.

Not only is biological motion itself a complex phenomenon, the factors that influence emotional perception and expression in the body motion are numerous and often interact with each other. For instance, gender of the observer has an effect on emotion recognition accuracy, as well as the gender of the performer of the motion. In several studies it has been shown that female participants are better at recognising neutral or negatively coloured actions (Sokolov et al., 2011), especially if the actor is male, while male participants recognise positive emotions in body language expressed by female actors with high accuracy (Krüger et al., 2013).

According to Giese and Poggio (2003) there are two distinct neural mechanisms in the brain that facilitate recognition of biological information: one for motion, another for form. Atkinson et al. (2007) determined that both form and motion signals are important for the perception of affect from the body motion. Using point-light biological motion stimuli Heberlein et al. (2004) have further investigated the neural systems involved in emotion recognition in normal population and subjects with brain damage. While biological motion per se and emotional body expressions are understandably not one and the same thing, correlation has been found between a subject's ability to discern emotional cues from point-light displays and the subject's ability to discriminate biological from non-biological motion. This observation that differences in emotion recognition may be related to more basic differences in processing biological motion per se are supported by studies for typically developed adults (Alaerts et al., 2011) and for participants with Aspergers Syndrome Condition related atypicalities (Nackaerts et al., 2012). A detailed review on the tight connection between the processing of biological motion and social cognition, and hence, disturbances in both these aspects of human mind due to atypical development (Aspergers Syndrome Condition, Down Syndrome,

pre-term birth) can be found in Pavlova (2012).

Studies on emotional body language have also investigated various aspects of emotion expression through body motion, such as how emotions modulate various actions, like walking (Roether et al., 2010) or knocking (Pollick et al., 2001b). Other research used general non-verbal portrayals of emotions (Atkinson et al., 2004; McDonnell et al., 2009; Kleinsmith et al., 2011; Beck et al., 2012), but the actors were still well aware that their body motion was of primary interest to the researchers since the tracking technology was focused on the body by, e.g. using full body suits, covering the face by a mask, restricting finger movement (Atkinson et al., 2004; McDonnell et al., 2009). Even though the used setups are completely justified by the research questions pursued in the related studies, such restrictions are very likely to prevent actors from expressing emotions in a natural way that would be typical of normal human-human interactions.

Our research aim was to investigate human perception of emotional body expressions that were captured in narrative settings, naturalistic yet well-controlled. For this we gathered a large dataset of motion patterns of the upper body using a non-restrictive inertial body capture suit (Volkova et al., 2014a). The motion patterns served further as stimuli in emotion recognition experiments. We also argue that it is valuable to use a rich set of emotion categories for the categorisation process. We conducted two perceptual experiments that evaluate the emotion recognition accuracy and consistency based exclusively on the upper body motions. Motivation behind focusing on upper body motion came partially from previous research by Glowinski et al. (2011), who successfully used videos of upper body emotional expression from the Geneva Multimodal Emotion Portrayal Corpus (Banziger and Scherer, 2010) to cluster recordings across the valence-arousal dimensions. Additionally, focusing on upper body motion allowed us to let the actors be seated during the motion capture sessions, a pose more common for narration situations in daily life, which in turn benefited our data recording and post-processing setups.

According to our null-hypothesis, the amount of information expressed through the body alone during narration should not be sufficient for an observer to recognise the emotion, since most of the information is expressed through the facial expressions, the speech prosody and, importantly, verbal content. However, our results show that even using stick figure stimuli and a large number of emotion categories, the recognition accuracy was above chance level, suggest-

ing that upper body motion produced during narrative scenarios is informative for emotional categorisation. Moreover, the responses from the observers are highly consistent and the agreement between participants was rather high according to Kendall's coefficient of concordance. Finally, we evaluated how much variance in the categorisation performance could be explained by motion statistics of the stimuli.

4.3 Materials and Methods

4.3.1 Motion Sequences Acquisition

Eight amateur actors were asked to perform a variety of natural emotionally expressive tasks. The motion sequences are distributed across the following eleven emotion categories: five positive (*amusement, joy, pride, relief, surprise*), five negative (*anger, disgust, fear, sadness, shame*) and *neutral*. Table 4.1 shows the number of motion sequences in descending order, representing intended emotion categories and acting tasks with the total number of motion sequences amounting to 1700. The motion was captured with the help of an Moven Xsens suit (Roetenberg et al., 2009) at the rate of 120 frames per second.

Each actor came separately for four motion capture sessions, thus amounting to 32 motion capture sessions in total. In the first session each actor received four blocks of short scenarios to act out: *solitary non-verbal scenarios*, where the actor was instructed that they were to imagine they were alone; *communicative non-verbal scenarios*, where the actor was instructed to imagine they were in company of one or more people they knew; *short sentences without direct speech*, meaning only narrator's text was present; and finally *short sentences with direct speech*, where narrator's as well as a story character's text were present. In each block all emotion categories except for *neutral* were used for emotion induction. The motivation text to act out (and in the case of short sentences also to speak out) was shown on a computer screen along with the emotion labels the actors were instructed to portray. The actors went through the blocks in the described order, the order of emotion categories within blocks was randomised for each actor and block.

In the next three motion capture sessions each actor worked at one story a time. Each actor chose three stories out of the available ten, according to their own preference. Before the motion capture

sessions each story was first split into utterances and annotated by the actors for eleven emotion categories. During the motion capture session the narration was shown utterance by utterance in its natural order and the emotion labels assigned by the actors were shown above each utterance. In both short scenarios and in full narrations, the actors were seated on a backless stool and progressed through short scenarios or full narrations by pressing a foot pedal, which allowed them to maintain their own speed and keep the upper body free for the expression of emotions. The timing of pedal presses was recorded for synchronisation of acting script presentation and the motion capture data. The narrations were on average three hundred utterances long (M=298.5, SD=36.09), each utterance containing a few word tokens. Each story annotation typically encompassed the full range of available emotion categories, yet the frequency between categories varied greatly, *neutral* naturally being the most frequent emotion and *shame* the least frequent.

The short scenarios are similar to classic motivation vignettes used for emotion induction in actors (see Bänziger and Scherer (2007) for a review of emotion induction methods). The actor portrays an emotion for a few seconds and then returns to a neutral pose. In contrast, during the narration task, the actors were immersed into the story and were displaying emotions in a way that was maximally close to natural day-to-day emotion expression. The actors were always free to express their emotions via face, voice and body, their performance was also captured on video. Additionally, the plot and the word choice of the story contributed to the naturalness of emotion expression. Although fairy tales may seem a source of extreme and dramatic emotions, this impression mostly comes from the fact that the density (but not necessarily the intensity) of emotion instances in many fairy tales is indeed higher than in, e.g., novels (Mohammad, 2012) which still leaves them as a suitable textual material for our purposes because of their conciseness and clear identification of bad and good characters.

	Short scenarios				Narrations
	Non-verbal		Sentences		
Emotion	Solitary	Communicative	Without direct speech	With direct speech	
joy	8	8	8	8	237
sadness	8	8	8	8	200
surprise	8	8	8	8	167
pride	8	8	8	8	137
anger	8	8	8	8	133
neutral	0	0	0	0	116
fear	8	8	8	8	112
disgust	8	8	8	8	106
relief	8	8	8	8	80
amusement	8	8	8	8	59
shame	8	8	8	8	33
Subtotal	80	80	80	80	
Total			320		1380
Grand total			1700		

Table 4.1: Count of motion sequences across emotion categories and acting scenarios. Two major types of acting tasks were *short scenarios* and *narration*. The former were of two kinds: *nonverbal* (solitary or social) and *short sentences* (without direct speech or with).

4.3.2 Hardware and Software

The two experiments were programmed in the Unity 3D engine and ran on a MacBook Pro laptop. Two viewing conditions were used: a large screen with 102-inch in diagonal and a 17-inch laptop display. The participants were seated 1.5 m away from the large screen and 40 cm away from the laptop screen. In both experiments, the motion sequences were mapped onto a stick figure, which represented a human figure from the chest up, including the arms (Figure 4.1). The pelvis and legs were excluded from the stimulus display, since our main research question dealt with the upper body. The resulting stick figure displayed biological motion of the real actors, but its configuration came from the underlying general skeleton model and not from the actor. Thus the body size, the proportions and other form cues were kept the same for all stimuli. When using the large screen display the size of the stick figure was matched to the one of an average person and was adjusted to correspond to the height of the average person seated. The maximal horizontal visual angle for the stimuli on the large screen was 59°, corresponding to the maximum arm span of the stick figure (1.7m) viewed from 1.5m distance. In the laptop screen viewing condition the maximal horizontal visual angle was 19.3°.

4.3.3 General Procedure and Participants

In total, 87 volunteer participants were recruited: Table 4.2 gives the distribution of the participant numbers, gender and age across the experiments as well as viewing conditions. Informed written consent was obtained before every experiment session. Participants and the obtained data were treated strictly according to the Declaration of Helsinki. The experiments were approved by the local ethics committee of the University of Tübingen. All participants received monetary compensation for their participation, all had normal or corrected to normal visual acuity. None of the participants were aware of the purpose of the experiment. Due to the location of the experiment, many of the participants (58 out of 87) were German native speakers. All participants' command of English was sufficient to understand the instructions and the meaning of all the emotion categories used in the experiment.

General overview information for each experiment is presented in Table 4.3. Experiment 1 used only the 80 motion capture sequences

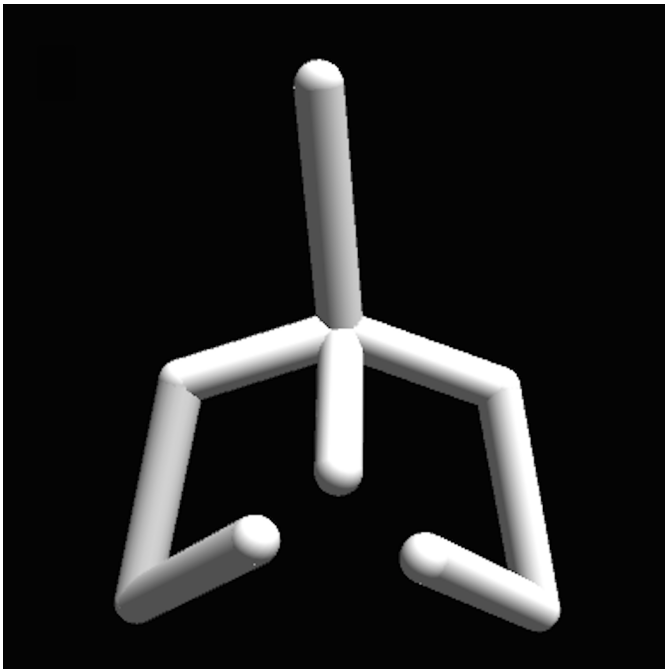

Figure 4.1: The stick-figure stimuli display. The pelvis and legs were excluded from the stimulus display, which includes only the upper body. The biological motion was displayed at 60 FPS and came from real actors. The configuration of joins came from an underlying general skeleton model, keeping the body size, the proportions and other form cues the same for all stimuli.

from non-verbal solitary emotional scenarios (Table 4.1, column 2). The task for each of the 32 participants was to choose between ten emotion categories, that is all categories mentioned in the beginning of Section 4.3 except for *neutral*, since no short scenarios included this category. Within each trial, the participant could always change their response before proceeding to the next trial. The motion sequence playback was set on the infinite playback loop, thus allowing the participants to watch the animation as many times as needed to perform the recognition task. Each of the animations was shown two times during the experiment. The trials were organised into two sessions and no animation occurred in one single session twice. The order of the

animations was randomised for each session. To avoid fatigues, the participants took short 5-minute breaks between the sessions.

The experiment 2 used the full dataset of 1700 motion sequences, where 81% of the motion sequences come from narrations. Because no single participant could possibly categorise 1700 stimuli in one experiment session, the full dataset of motion sequences was organised into five equal blocks, making sure that the proportions of emotion categories were equal in each block. For each five participants the blocks were generated anew. Each participant categorised a unique block of 340 randomised motion sequences. Each of the motion sequences was seen by eleven participants throughout the experiment. Thus, our participant pool for this experiment amounted to 55 individuals.

In each of the 340 trials, motion sequence playback was set on three iterations, then the participant completed a two-fold response task and proceeded to the next trial. In the two-fold response task the participants were first asked to decide whether the animation carried any emotional information or whether it was completely *neutral*. If it was the former, the participant was to choose a specific emotion from the ten remaining categories. The response could always be changed until the participant was satisfied and proceeded to the next trial. In order to avoid fatigue, short breaks of minimum 30 seconds were implemented after every 50 trials.

	Participants	Age	Viewing Condition
Exp. 1	32 (16 f.)	M=27.91, SD=7.66	laptop screen & large display
Exp. 2	55 (28 f.)	M=29.96, SD=9.96	large display

Table 4.2: Participants of the experiment. The columns show, from left to right: experiment number, number of participants and the gender distribution, participant age, and viewing conditions.

	Stimuli	Playback	Trials	Duration	Categories	Acting Tasks Used
Exp. 1	80	∞	160	$\approx$ 1.5h.	10 emotions (all except *neutral*)	nonverbal, solitary
Exp. 2	1700	3	340	$\approx$ 3h.	11 emotions	all, full dataset

Table 4.3: Experiment setups. The columns show, from left to right: experiment number, number of stimuli used in the experiment, animation playback (infinite or limited to 3 repetitions), number of trials in one experiment session, average session duration in hours, categories used, motions sequences source.

4.4 Results

4.4.1 Recognition Accuracy

In experiment 1, where only motion sequences from short non-verbal solitary scenarios were used, the average emotion recognition was 35.2% (Figure 4.2, A). All emotion categories were correctly identified by participants on above chance level (10%), all Holm-corrected p-values are below 0.001 for one-tailed t-tests (exact values for recognition accuracy, $t-$ and $p-$values for all experiments as well as an additional pilot study can be found in Supplementary materials, one-tailed t-tests are used since we are interested in recognition accuracy that is significantly higher than chance level). The between-participant factor of viewing condition had a non-significant effect on the recognition accuracy: 38% accuracy for large display condition vs. 31% accuracy in the desktop condition, ANOVA for viewing conditions: $F(1,30) = 2.78$, $p = .10$, $\eta_p^2 = .021$. The within-participants factor of emotion categories had a large effect on the recognition accuracy: $F(9,270) = 15.41$, $p < .001$, $\eta_p^2 = .28$. False alarm rates are at approximately 6% across all emotion categories with the exception of *shame* and *anger* (Figure 4.2, B). This experiment established the upper threshold for emotion recognition in our dataset. It was unlikely that participants would be able to recognise the more subtle emotional body expressions taken from the narrations better than those from non-verbal scenarios.

The overall recognition rate for experiment 2 was 18% (see Figure 4.3, A). The majority of the emotion categories were correctly identified by participants on above chance level (9%), most Holm-corrected p-values are below 0.001 for one-tailed t-tests. However, three emotion categories, *disgust*, *relief*, and *surprise* were recognised at below chance level. Recognition accuracy was affected by the within-participant factor of expressed emotion category, ANOVA $F(10,540) = 122$,$p < .001$, $\eta_p^2 = .68$. False alarm rates across most emotions are similar to those in experiment 1 (ca. 6%), with two important exceptions: *anger* (9.3%, almost the same rate as in experiment 1 - 9.7%) and most importantly *neutral* with exceptionally high false alarm rate of 30.8% (Figure 4.3, B).

In experiment 2 observers' responses can also be analysed according to the two-stage response structure they gave. In every trial the participant was first given the choice between *neutral* and *emotional*. If the participant considered the motion sequence to express an emotion other than *neutral*, the observer was to choose among the ten remaining

emotion categories. At any point of time within one trial the participant could change their response. In order to reflect the two-stage response structure, we ran two separate ANOVA's, one for the *neutral-emotional* level, the other for the ten non-neutral categories. Figure 4.4 (A), shows that participants recognised whether a motion sequence was emotional or neutral at above chance level. The analysis of this first stage was performed on just two categories — *emotional* vs. *neutral*. The results show that the within-participant factor of emotionality had a significant effect on response accuracy: ANOVA $F(10, 540) = 9.86$, $p < .001$, $\eta_p^2 = .08$. For the second stage of the response, the analysis of accuracy for ten emotion categories (all except for *neutral*) as a within-participant factor shows that emotion categories had a significant effect on response accuracy (Figure 4.4, B): ANOVA $F(9, 486) = 48.27$, $p < .001$, $\eta_p^2 = .44$. Note that according to the two-stage analysis, all emotion categories were recognised at above chance level (50% for the first step, 10% for the second step of analysis), and all Holm-corrected p-values are below 0.05.

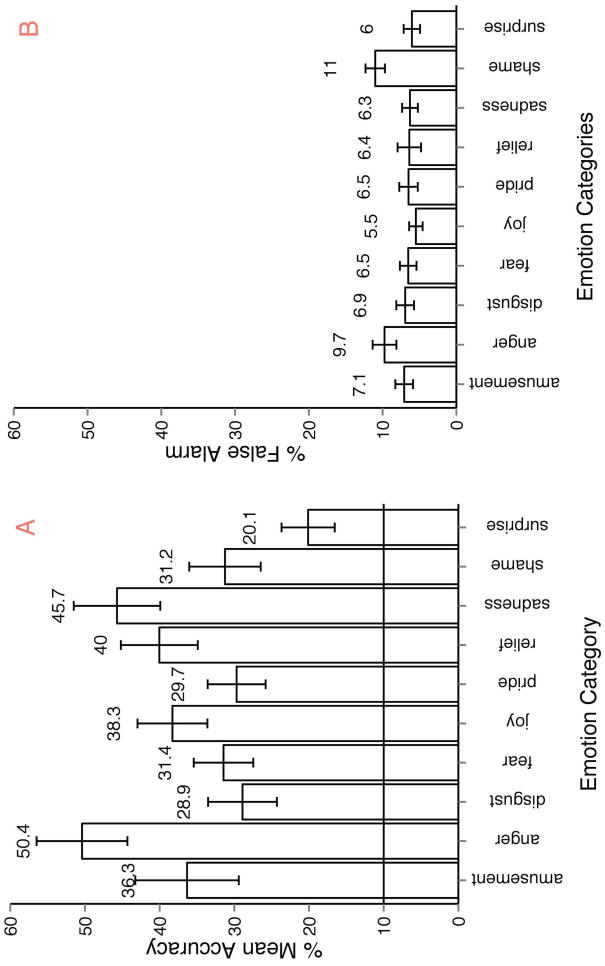

Figure 4.2: Emotion recognition in experiment 1. A: Accuracy across emotion categories. The horizontal line shows the chance level threshold of 10%. B: False alarm rate across emotion categories. All error bars show 95% confidence intervals. The participant observed a stick-figure representation of human upper body motion. The task was to recognise the emotion category expressed by the actor and respond by choosing one of the buttons with the corresponding emotion category. The motion sequence was set on infinite loop playback, the participant could always alter their choice before proceeding to the next trial.

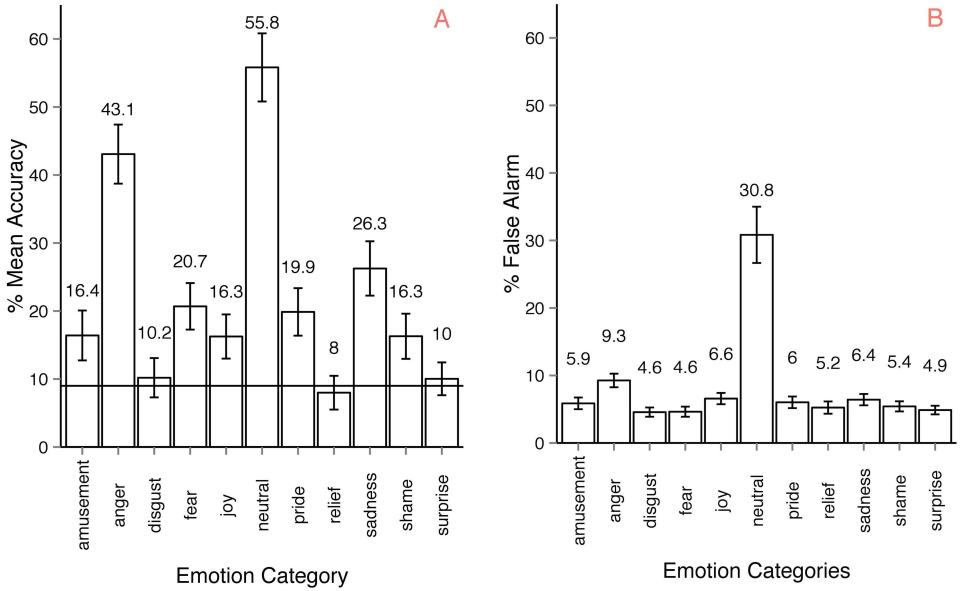

Figure 4.3: Emotion recognition in experiment 2. A: Accuracy across emotion categories. The horizontal line shows the chance level threshold of 9%. B: False alarm rate across emotion categories. All error bars show 95% confidence intervals. The participant observed a stick-figure representation of human upper body motion. The task was to recognise the emotion category expressed by the actor and respond by first choosing either *neutral* or *emotional* category, then, if the *emotional* option was chosen, select the oppropriate emotion category of out ten non-neutral categories available. Each motion sequence was shown three times after which the participant had unlimited time to respond. The participant could always alter their choice before proceeding to the next trial.

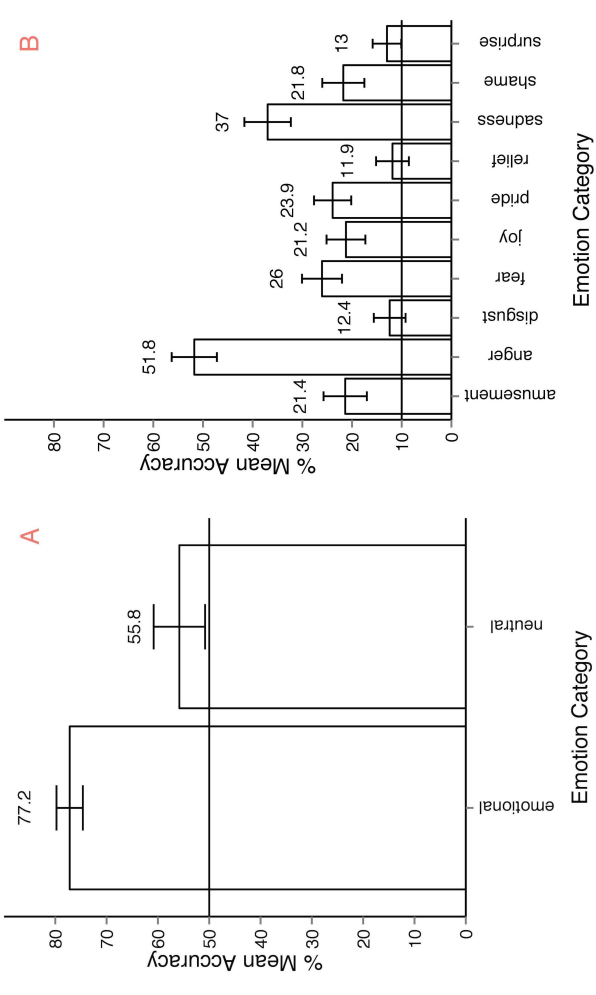

Figure 4.4: A: Accuracy across emotion categories for experiment 2, response stage 1 — *neutral* vs. *emotional*. The horizontal line shows the chance level threshold of 50%. B: Accuracy across emotion categories for experiment 2, response stage 2 — discrimination between all emotion categories except for *neutral*. The horizontal line shows the chance level threshold of 10%. All error bars show 95% confidence intervals.

	Raters	Categories	Items	IRA (W)
Exp. 1	32	10	80	0.26
Exp. 2	11	11	1700	0.24

Table 4.4: Inter-rater agreement in two experiments according to Kendall's coefficient of concordance W. The columns show, from left to right: number of raters (participants) per motion sequence, number of categories available for categorisation, number of items to categorise and the inter-rater agreement.

4.4.2 Inter-rater Agreement and Response Consistency

Inter-rater agreement (IRA) gives a score of how much consensus there is in the ratings given by observers. This measure can be calculated in various ways, depending on the number of raters. When more than two raters are available, Fleiss' *kappa* (Fleiss, 1971) is often used to measure IRA. However, since in our experiment the observers operated with non-parametric data ratings (emotion categories) we used Kendall's coefficient of concordance (Kendall and Smith, 1939) to obtain an estimate for IRA. Kendall's W ranges from 0.0 (no agreement) to 1.0 (complete agreement). In experiment 1 W is equal to 0.26 (averaged across the two experiment sessions) and in experiment 2 W is equal to 0.24 (see Table 4.4). IRA measures general consensus among the annotators and is not useful for calculating agreement for each motion sequence.

We thus developed an alternative measure for estimating agreement among observers and will henceforth refer to it as consistency (c). Consistency is the percentage of observers' responses falling into a particular emotion category that forms the modal value in the response distribution. For example, if for a specific motion sequence all responses fall into one category, $c = 100\%$. If three out of ten responses fall into one category and no other category received three responses or more, $c = 30\%$. The minimally possible c is always 100 divided by the number of observations for the given stimulus and multiplied by 2 because there have to be at least least two observers assigning the same category to the stimuli to form a modal value. In cases when the response distribution is bimodal or multimodal, c cannot be defined. In the rest of the section we will focus on the consistency results of experiment 2 since it encompassed all 1700 motion sequences. As Figure 4.5 shows, most responses to animations were distributed among five

or fewer emotion categories. Regardless of the intended emotion of the actor, for 1452 motion sequences (85% of 1700) observers preferred one emotion category for a given motion sequence over other categories. Moreover, in 40% of the cases (670 out of 1700 motion sequences) the modal value received more than half of the responses (at least six out of eleven responses). Figure 4.6 shows the levels of consistency across all emotion categories, most of which are above 40%.

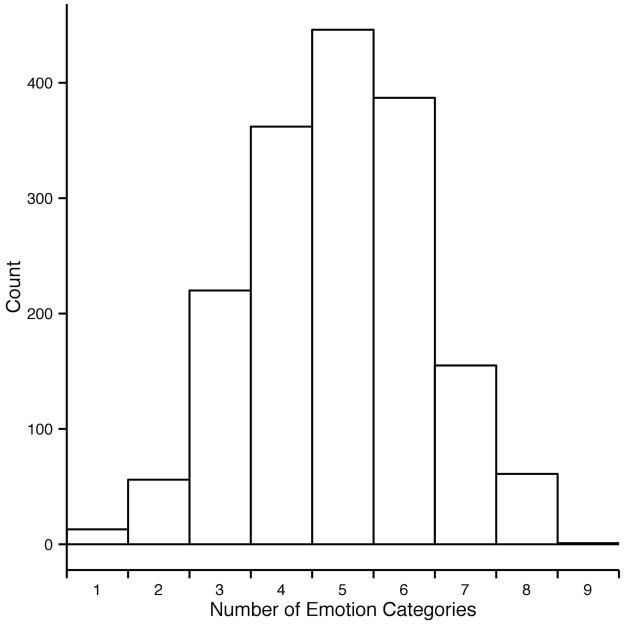

Figure 4.5: Distribution of number of distinct categories given to each motion sequence. Each motion sequence received in total eleven responses. For most motion sequences the responses fall into five or fewer categories.

4.4.3 Response Bias

Bias in the categorisation of motion sequences into two emotions (*anger* and *neutral*) was observed in experiment 2. In experiment 1 the *neutral* category was not used, but the recognition accuracy of *anger*, though

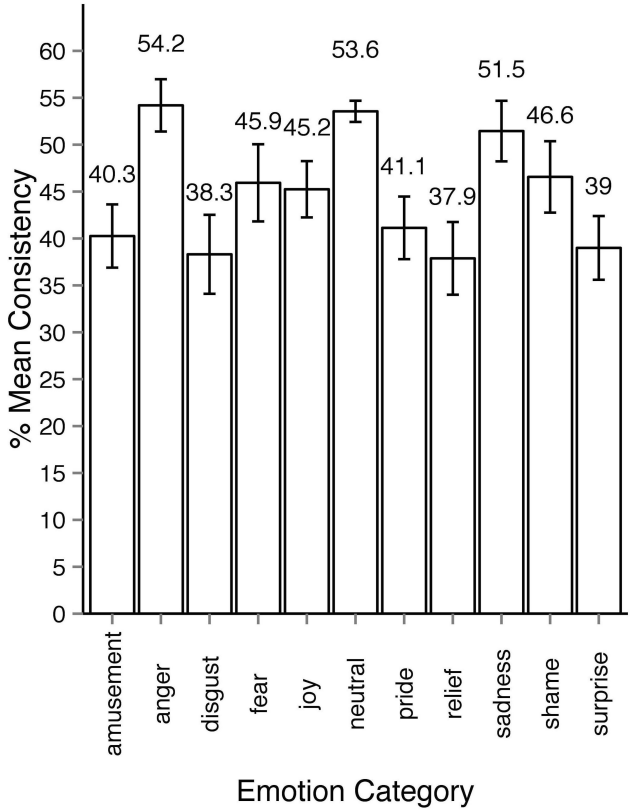

Figure 4.6: Average consistency levels across emotion categories in experiment 2. The sequences are assigned emotion labels according to the modal category in observers response distribution. All error bars show 95% confidence intervals.

highest among other categories, is not significantly different from *sadness* ($p = 1.0$) and relief ($p = 0.08$) according to Holm-corrected pairwise comparison. In experiment 2 both *neutral* and *anger* are significantly different from all other emotions — all p-values are below 0.001 with the exception of the comparison between *anger* and *neutral* themselves where $p = 0.004$, thus still significantly different between each other.

In experiment 2, the *neutral* category was chosen more frequently

than other categories. The frequency of *neutral* responses was 34% on average, several times more than the actual number of *neutral* animations in the dataset (7%). Indeed, the results of experiment 2 showed that the observers often made mistakes in recognising an emotion when deciding whether the motion sequence conveyed any emotion at all. Confusions between two *non-neutral* categories were less frequent and rather systematic (see Section 4.4.4 for more detail). By comparing Figures 4.3 and 4.4 one can see that once the *emotional* vs. *neutral* stage is separated from the second stage of response, all emotion categories were recognised above chance. False alarm rate contributed important information concerning the observers' response patterns. As the right side bar plots of Figures 4.2-4.3 show, false alarm rates for most categories lie under 10%. In experiment 2 however, the *neutral* category clearly received more false alarms than other emotion categories (30.8%). This means that in many cases when an emotion category other than *neutral* was intended by the actor, it was nevertheless perceived as *neutral* by the observers. The reasons behind the bias towards *neutral* and *anger* are discussed in Section 4.5.

4.4.4 Motion Properties

Recognition rates, inter-rater agreement and response consistency all clearly show that observers' choices of response categories are not random. When an observer failed to recognise an intended emotion from a motion sequence, there was a tendency for other categories to occur in the responses depending on the actor's intentions. For instance, in experiment 1, *sadness* was accurately recognised in 45.7% of the trials, and the 54.3% errors are distributed unevenly among the rest of emotion categories with 16% falling into the category of *shame*. Shame on the other hand was accurately recognised in 31.2% of the trials and in 25.8% it was categorised as *sadness*. Figure 4.7 shows more examples and full detail on distribution of response categories across intended emotions. Since upper body motion was all the information available to the observers, commonalities in motion patterns between different categories could possibly explain confusions between intended emotion categories and observers' responses.

Motion analysis should provide a way to compare motion sequences that have been stably labelled for certain emotion categories. Similarities and differences between motion sequences can be looked for at different levels. At a semantic level, meaningful patterns and

gestures like shaking fists in anger, crossing arms and tilting head when expressing pride and clapping hands for joy can be considered. On the other hand, each motion sequence can be described as a set of more descriptive statistics, e.g., speed, peaks in motion trajectory, span of motion, duration and so on. Compare clapping your hands in a fast energetic manner to express joy and clapping your hands slowly to express contempt, cold anger or disgust. The motion trajectories are the same yet motion properties like speed and span differ depending on what emotion is expressed. Extreme joy and hot anger are most likely recognised by their fast, broad motion, while the motion profile of sadness is notable for its low speed.

To test whether motion properties can predict response categories, we ran a multiple regression analysis using response categories as the dependent variable and several extracted motion properties as independent variables. For motion properties extraction we used only the left and the right wrist joints of the underlying body structure, because in our setup they were the most mobile joints for all emotions. For each motion sequence we calculated the mean motion speed, the average number of peaks in the trajectories for the x, y and z axes and the mean span of the motion, defined as average distance between the wrist joints during the motion sequence. For the purpose of this particular analysis the intended emotions from the actors could not be used, since we aimed to gain insight behind the observers' decision during the recognition tasks. Thus for each motion sequence we first established the distribution of response labels across emotion categories, similar to the response consistency analysis described in Section 4.4.2. Only motion sequences with a unique modal value were included in the analysis (78 out of 80 for experiment 1 and 1452 out of 1700 in experiment 2), where the emotion category representing the response modal value was used as the label for each motion sequence. Figure 4.8 gives an overview of the resulting values for mean motion speed (A), peak count (B) and mean span (C).

In order to perform the analysis we have calculated a distance matrix for response patterns for intended emotions and distance matrices for each of the motion properties. We tested if motion properties (speed, peaks, and span) significantly predicted response based distance between emotions for both experiments. The results of the regression analysis for experiment 1 indicated that the three predictors explained 48.3% (adjusted R^2) of the variance ($R^2 = .48, F(3,96) = 29.99, p < .001$). Mean motion speed significantly predicted emotion distance

($\beta = .35, p < .001$), as did number of peaks ($\beta = 0.32, p < .001$), and mean motion span ($\beta = 0.36, p < .001$). The results of the regression analysis for experiment 2 indicated the three predictors explained 34.9% (adjusted R^2) of the variance ($R^2 = .36, F(3, 117) = 22.53, p < .001$). The distance in mean motion speed across emotions significantly predicted emotion distance ($\beta = 0.20, p = .014$), as did number of peaks ($\beta = 0.31, p < .001$), and mean motion span ($\beta = 0.33, p < .001$). The assumptions of independence and multicollinearity were met for all predictors in both experiments.

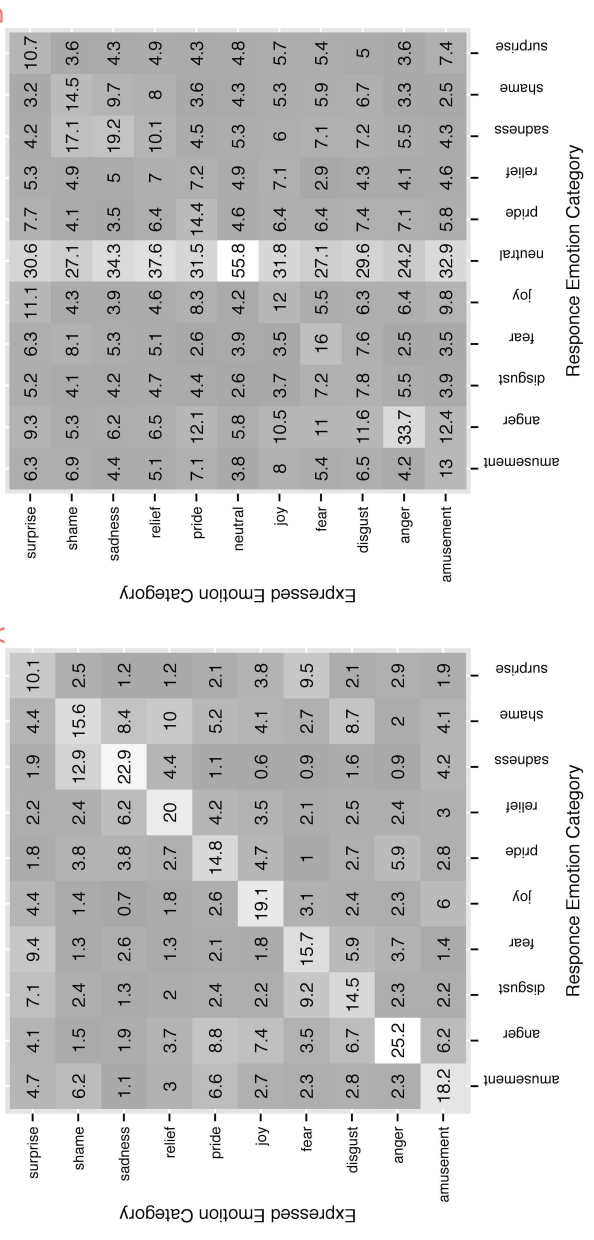

Figure 4.7: Detailed correspondence between intended emotion categories and observers' responses in experiment 1 (A) and experiment 2 (B). On y axis the intended categories are shown, on x axis - the response categories. The diagonal shows response accuracy. Each row sums up to 100. All values are in %.

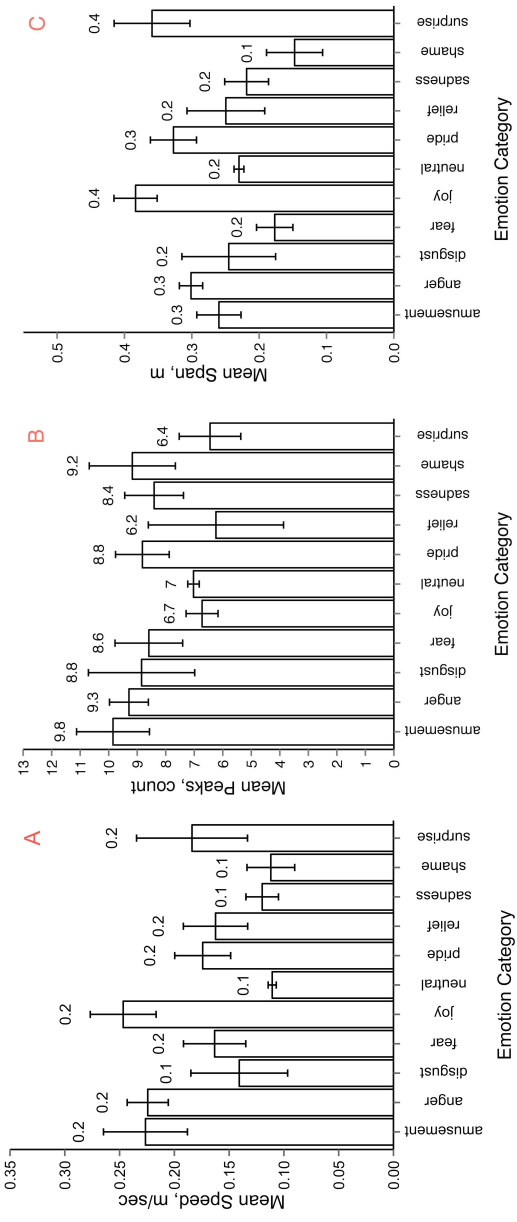

Figure 4.8: Average values for (A) motion speed (m/sec), (B) peaks (raw count) and (C) span (m) across emotion categories. Modal values of the response distributions were used for grouping motion sequences into emotion categories. All error bars show 95% confidence intervals.

4.5 Discussion

The experiments presented in this paper demonstrate that people recognise naturally expressed emotions with only upper body motion available. Since the motion sequences come from a naturalistic narration setting and the visual stimuli only provide information about the upper body motion, the recognition rate is impressive. In Experiment 1 the recognition is high across all ten emotion categories. This can be contributed to the fact that the motion sequences used in this experiment came from non-verbal solitary short scenarios, so more information was likely conveyed in the motion of the body since the auditory channel was not used. In experiment 2 seven out of ten emotions are recognised at above chance level despite the fact that most of the motion sequences used in this experiment came from narration tasks where all expressive channels were used by the actors. Not only is the recognition above chance but motion sequences were categorised with high consistency between participants. For 85% of the motion sequences response distribution has a unique modal value and for 40% of all motion sequences the majority (over 50%) of the observers' responses fall into one category.

Notably, recognition accuracy and consistency levels differ among emotion categories. Specifically, observers have a bias for categorising emotions as *anger* in both experiments, most prominently in experiment 2. Recognition and false alarm rates, as well as high consistency rates for *anger* show that people are prone to categorise motion sequences with *anger* more often than other categories, with the exception of *neutral* in experiment 2. A possible explanation for the bias towards *anger* categorisation could be the evolutionary importance of detecting anger as a potential threat regardless of which channel (language, prosody, face, or body) it is perceived through. Several previous works support the importance of anger expressed via body motion: Pichon and colleagues have shown that a response to anger expression results in even more activation of defence mechanisms ("fight or flight") than a response to fear expressions when the perceiver is the target of the anger (Pichon et al., 2009). A similar tendency is discussed in a recent study by Visch et al. (2014), where recognition of anger expressed in the body motion was the most robust under various stimuli degradation conditions (including a condition where only parts of the upper body were portrayed). Others have found that a bias towards anger is even more pronounced in violent offenders (Kret and de Gelder, 2013).

Moreover, participants have a bias towards the *neutral* category in experiment 2, as the *neutral* category also has high accuracy rate, false alarm rate and consistency levels. A likely reason for this bias is that many motion sequences did not possess properties that could communicate any particular emotion to the observer, as is supported by the high false alarm rate.

In order to gain insight into errors in emotion recognition of the intentions of the actor and factors behind them, we analysed the relationship between the intended emotions, the responses, and motion properties of the motion sequences. We found that distances in mean motion speed, number of peaks and mean span between motion sequences stably marked for certain emotions can to some extent predict distances between response categories. These findings do not belittle the significance of meaningful motion trajectories and gestures, e.g., fist shakes, hand claps and head nods. However, many motion properties, e.g., motion speed, are easier to extract from any biological motion than specific gestures. These findings are encouraging for future work in automatic emotion recognition and are in agreement with related research. Huis in 't Veld et al. (2014) found that both active expression and passive viewing of emotions via body motion activate similar muscle groups in the upper body. Interestingly, a study by Magnée et al. (2007) showed emotion specific facial muscle activity that was independent of stimuli type (facial expressions, bodily expressions or face-voice combinations). It would be intriguing to investigate whether spontaneous reaction to emotional stimuli in the body is as modality independent as the one observed in the face.

Our results also provide some evidence that using a rich set of emotions that goes beyond the basic Ekman emotions for body motion recognition is valuable. One of the major arguments for basic emotions (Ekman, 1992) is that they are saliently recognised in most populations regardless of age, gender or culture and are independent of expression medium (face, body, voice). However, the recognition rates in our experiments seem to suggest that for emotional body expressions in a natural setting the basic Ekman emotions are not sufficient. In experiment 2, two out of three categories recognised below chance are basic: *disgust* and *surprise*, while non-basic emotions of *amusement*, *pride* and *shame* are recognised above chance. However, all emotions, basic and non-basic, are recognised well above chance in experiment 1, where motion sequences were obtained from purely non-verbal short scenarios. This allows us to conclude that the *distinctive universal signals*

proposed by Ekman as one of the characteristics for basic emotions are not always present in our upper body motion patterns captured during natural expression.

4.6 Conclusion

Body motion is an important source of information in emotion expression. This research adds a novel approach by focusing on the perception of emotional body language occurring naturally in narrative scenarios. We used a rich set of eleven emotion categories in two perceptual experiments that investigated emotion recognition of upper body movements on stick figure stimuli. Almost all emotion categories achieved recognition accuracy that surpassed the chance level (ranging from 8% to 58%). Response consistency between the participants is strong, as for most motion sequences the distribution of response categories has a unique modal value, meaning that most observers chose one category as their response. Further, in 40% of the motion sequences more than half of the participants agreed on this modal value. In our experiments there is a strong bias for the *anger* category among observers' responses. This can be explained by the ecological importance of early anger detection in human environment as one of the survival strategies of fight or flight. There was additionally a strong bias towards *neutral*, which might be due to the low amount of movement in natural scenarios. In order to further consider what factors were driving the errors in recognition performance, we performed a multiple regression analysis using the descriptive statistics of motion, namely speed, peaks in motion trajectory and span. Our findings show that the investigated motion properties can serve as predictors for patterns in response categories. Overall, the results demonstrate that the information contained in upper body motion in natural scenarios is enough for people to recognise emotion.

4.7 Supplementary Material

4.7.1 Pilot Study

We conducted a pilot study on a subset of motion sequences. The pilot study employed motion sequences from the whole dataset (non-verbal expressions, short sentences and narrations), but for every emotion category only 20 motion sequences were randomly selected and used

throughout the study, amounting to 220 stimuli in total for each participant. The number of animations for each emotion was balanced in order to discover any possible response bias towards emotion categories. In each trial the participant could choose between eleven emotion categories (*amusement, joy, pride, relief, surprise, anger, disgust, fear, sadness, shame, neutral*) or the *cannot identify* response option in case they found themselves unable to categorise the stimuli. Within each trial, the participant could always change their response before proceeding to the next trial. Each motion sequence was used once during the experiment, in each trial the motion sequence was played back three times. The order of stimuli was randomise for each participant.

Twelve participants took part in the pilot study (7 female, age: M=26.16, SD=5.63) The stimuli were presented on a 17-inch laptop display, the participants viewed the stimuli from 40 cm distance. The recognition rate in the pilot study was 15% which was nevertheless above chance level of 9% (based on 11 possible categories). Table 4.9 reports recognition accuracy, as well as t- and p-values for the recognition above chance level for each emotion category. According to Holm-corrected p-values, only four emotion categories, namely *anger*, *fear*, *neutral*, and *sadness* were recognised at above chance level. The response option of *cannot identify* was used rarely, on average only in 3.6% of the cases, no single category of intended emotion received this label more often than other categories. Similarly to Experiments 1, emotion categories had a significant effect on the overall recognition rate, ANOVA $F(10, 110) = 8.81, p < .001, \eta_p^2 = 0.43$. Table 4.13 shows pairwise t-test comparisons with Holm-corrected p-values as post-hoc analysis.

Two major observation can be made on the basis of the pilot study results. First, despite the fact that each category had an equal number of representative motion sequences, there is obvious bias towards *anger* and *neutral* categories, which is reflected in the recognition accuracy rates and the false alarm rates (see Figure 4.9). Second, since the *neutral* category has a high false alarm rate, it could be interpreted as a default category the observers use when they are uncertain in what emotion is expressed in the stimuli. While the distinction between *neutral* and *absence of any emotion* is vague, the most likely explanation for high false alarm rates for neutral is the subtlety of emotional expression in those motion sequences. This is supported by the fact that in the pilot study the "cannot identify" option was available and yet barely used. Based on these results in Experiments 1 and 2 in the manuscript we did not use a *cannot identify* option.

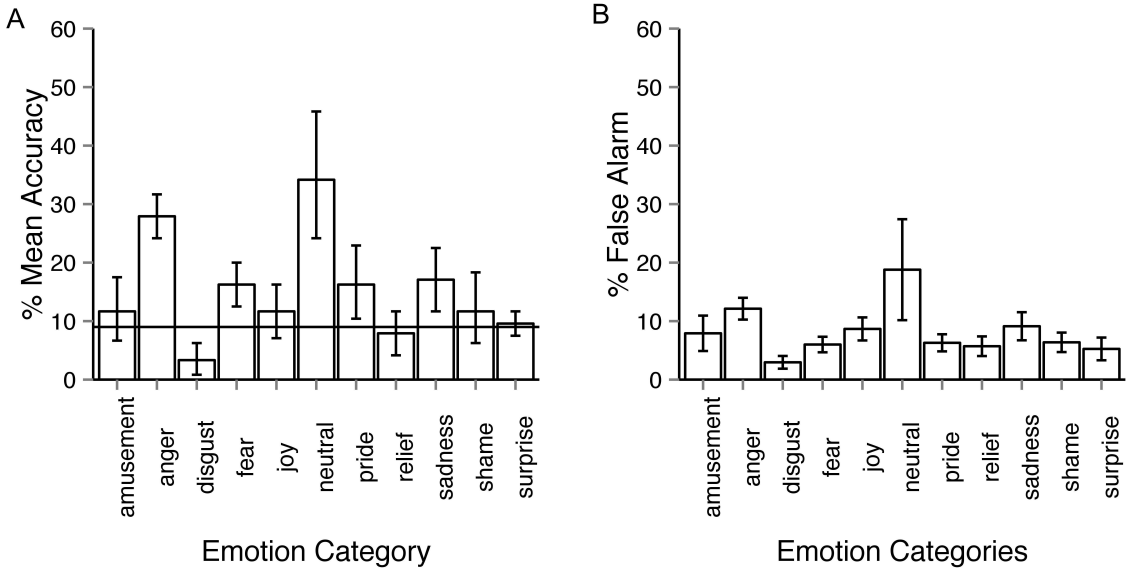

Figure 4.9: Emotion recognition in experiment 2. A: Accuracy across emotion categories. The horizontal line shows the chance level threshold of 9%. B: False alarm rate across emotion categories. All error bars show 95% confidence intervals.

4.7.2 Supplementary Tables and Figures

Emotion	Acc., %	$t, \mu = 0.1$	p_t
amusement	36	7.75	$< .001$
anger	50	13.54	$< .001$
disgust	28	8.37	$< .001$
fear	31	10.97	$< .001$
joy	38	12.33	$< .001$
pride	29	10.34	$< .001$
relief	40	11.90	$< .001$
sadness	45	12.57	$< .001$
shame	31	9.05	$< .001$
surprise	20	5.79	$< .001$

Table 4.5: Experiment 1, recognition accuracy for each emotion category with t and Holm-corrected p-values for performance above chance level (μ).

Emotion	Acc., %	$t, \mu = 0.09$	p_t
amusement	16	3.99	.001
anger	43	15.49	$< .001$
disgust	10	0.81	.59
fear	20	6.74	$< .001$
joy	16	4.41	$< .001$
neutral	55	18.74	$< .001$
pride	19	6.14	$< .001$
relief	7	-0.79	.78
sadness	26	8.53	$< .001$
shame	16	4.33	$< .001$
surprise	10	0.84	.59

Table 4.6: Experiment 2, recognition accuracy for each emotion category with t and Holm-corrected p-values for performance above chance level (μ)

Category	Acc., %	$t, \mu = 0.5$	p_t
non-neutral	77	20.89	$< .001$
neutral	55	2.32	.01

Table 4.7: Experiment 2, recognition accuracy for *neutral* category and the remaining ten categories combined with t and Holm-corrected p-values for performance above chance level (μ)

Category	Acc., %	$t, \mu = 0.1$	p_t
amusement	21	5.60	$< .001$
anger	51	18.54	$< .001$
disgust	12	2.11	.03
fear	26	8.36	$< .001$
joy	21	6.17	$< .001$
pride	23	7.86	$< .001$
relief	11	1.70	.04
sadness	36	11.82	$< .001$
shame	21	5.94	$< .001$
surprise	12	2.72	.01

Table 4.8: Experiment 2, recognition accuracy for all *emotional* categories except for *neutral* with t and Holm-corrected p-values for performance above chance level (μ)

Emotion	Acc., %	$t, \mu = 0.09$	p_t
amusement	11	0.89	0.97
anger	27	9.50	$< .001$
disgust	3	-3.98	1.00
fear	16	3.5	.02
joy	11	1.10	0.87
neutral	34	4.26	$< .006$
pride	16	2.18	.18
relief	7	-0.54	1.00
sadness	17	2.97	.05
shame	11	0.81	.97
surprise	9	0.50	.97

Table 4.9: Pilot study, recognition accuracy for all emotion categories with t and Holm-corrected p-values for performance above chance level (μ)

	amusement	anger	disgust	fear	joy	pride	relief	sadness	shame
anger	0.00235	-	-	-	-	-	-	-	-
disgust	0.62488	1.1e-07	-	-	-	-	-	-	-
fear	1.00000	4.8e-06	1.00000	-	-	-	-	-	-
joy	1.00000	0.01905	0.18660	0.77481	-	-	-	-	-
pride	0.82124	3.7e-07	1.00000	1.00000	0.29204	-	-	-	-
relief	1.00000	0.08997	0.04679	0.29204	1.00000	0.08997	-	-	-
sadness	0.18660	1.00000	9.1e-05	0.00193	0.62488	0.00024	1.00000	-	-
shame	1.00000	3.7e-06	1.00000	1.00000	0.72584	1.00000	0.27551	0.00159	-
surprise	0.00019	1.3e-14	0.27551	0.04012	1.4e-05	0.16494	1.2e-06	9.9e-11	0.04679

Table 4.10: Experiment 1 post-hoc analysis, pairwise t-tests with Holm-corrected p-values

	amusement	anger	disgust	fear	joy	neutral	pride	relief	sadness	shame
anger	<2e-16	-	-	-	-	-	-	-	-	-
disgust	0.18017	<2e-16	-	-	-	-	-	-	-	-
fear	0.84967	<2e-16	0.00045	-	-	-	-	-	-	-
joy	1.00000	<2e-16	0.18738	0.84967	-	-	-	-	-	-
neutral	<2e-16	0.00437	<2e-16	<2e-16	<2e-16	-	-	-	-	-
pride	1.00000	<2e-16	0.00164	1.00000	1.00000	<2e-16	-	-	-	-
relief	0.01086	<2e-16	1.00000	4.4e-06	0.01292	<2e-16	2.6e-05	-	-	-
sadness	0.00118	1.6e-10	1.2e-09	0.28530	0.00098	<2e-16	0.16453	1.4e-12	-	-
shame	1.00000	<2e-16	0.18738	0.84967	1.00000	<2e-16	1.00000	0.01276	0.00102	-
surprise	0.16453	<2e-16	1.00000	0.00032	0.18017	<2e-16	0.00122	1.00000	7.1e-10	0.18017

Table 4.11: Experiment 2 post-hoc analysis, pairwise t-tests with Holm-corrected p-values

	amusement	anger	disgust	fear	joy	pride	relief	sadness	shame
anger	<2e-16	-	-	-	-	-	-	-	-
disgust	0.00170	<2e-16	-	-	-	-	-	-	-
fear	1.00000	<2e-16	3.5e-05	-	-	-	-	-	-
joy	1.00000	<2e-16	0.05650	0.73816	-	-	-	-	-
pride	1.00000	<2e-16	0.00041	1.00000	1.00000	-	-	-	-
relief	0.00077	<2e-16	1.00000	1.3e-05	0.02897	0.00016	-	-	-
sadness	1.3e-05	7.0e-11	<2e-16	0.00077	5.5e-08	7.4e-05	<2e-16	-	-
shame	1.00000	<2e-16	0.00124	1.00000	1.00000	1.00000	0.00056	1.9e-05	-
surprise	0.33939	<2e-16	1.00000	0.02685	1.00000	0.13302	0.74656	1.6e-11	0.28664

Table 4.12: Experiment 2 post-hoc analysis for all *emotional* categories except for *neutral*, pairwise t-tests with Holm-corrected p-values

	amusement	anger	disgust	fear	joy	neutral	pride	relief	sadness	shame
anger	0.00610	-	-	-	-	-	-	-	-	-
disgust	1.00000	1.5e-06	-	-	-	-	-	-	-	-
fear	1.00000	0.20677	0.08714	-	-	-	-	-	-	-
joy	1.00000	0.00610	1.00000	1.00000	-	-	-	-	-	-
neutral	1.4e-05	1.00000	8.1e-10	0.00141	1.4e-05	-	-	-	-	-
pride	1.00000	0.20677	0.08714	1.00000	1.00000	0.00141	-	-	-	-
relief	1.00000	0.00019	1.00000	1.00000	1.00000	2.1e-07	1.00000	-	-	-
sadness	1.00000	0.34718	0.04677	1.00000	1.00000	0.00293	1.00000	0.96890	-	-
shame	1.00000	0.00610	1.00000	1.00000	1.00000	1.4e-05	1.00000	1.00000	1.00000	-
surprise	1.00000	0.00096	1.00000	1.00000	1.00000	1.5e-06	1.00000	1.00000	1.00000	1.00000

Table 4.13: Pilot study post-hoc analysis, pairwise t-tests with Holm-corrected p-values

5 Cross-Cultural Differences in Perception of Dynamic Emotional Body Expressions

Volkova, E., Mohler, B. J., Parkinson, B., Wildgruber, D., Bülthoff, H. H., and de la Rosa, S. (2014c). Cross-cultural Differences in Perception of Dynamic Emotional Body Expressions. *in preparation*

5.1 Abstract

Cultural background is one of the factors underlying differences in perception and expression of emotions. While in many aspects emotions are universal across most human populations, cultural differences are especially prominent when the objects of scientific scrutiny are attitudes, values and beliefs about emotions, cultural norms, display rules, and other social aspects of human interaction. Most work on cross-cultural differences in emotion perception concentrates on facial expressions or on differences in systems of emotional concepts. Cross-cultural perception of emotional postures has also been studied to some extent, but predominantly on static stimuli.

In this experiment we combine emotion recognition tasks with scale-based responses in order to investigate how cultural background, along with other factors, influences perception of dynamic emotional body expressions. We find that same motion patterns observed by participants from various cultures (English, German and Korean) receive different ratings depending on the observer's culture and their display rules. Using factor analysis we found that higher brokenness of motion trajectory is associated in English and German cultures with higher arousal, while in Korean culture it is associated with higher negativity. This research demonstrates that emotional body expressions are interpreted differently across cultures and can inform further investigations in cognitive, social and cross-cultural psychology and

neuroscience. Moreover differences in perception of emotional body language are important to consider in our globalised world, e.g. in business management psychology dealing with multicultural teams, development of interactive social avatars and in e-learning applications.

5.2 Introduction

Human emotions is a topic that has received much attention in the past few decades across such fields as cognitive psychology, neuroscience and affective computing. Factors that influence emotion expression and perception, both processes being largely interdependent, are currently under detailed investigation. While some aspects of human emotional experience are universal, spanning across genders, ages and countries, the factor of culture has been shown to influences the way emotions are displayed and perceived.

The definition of culture given by Matsumoto is inherently social: culture is "a shared system of socially transmitted behaviour that describes, defines, and guides people's ways of life" (Matsumoto, 2006). Traditionally, cultures have been put into two broad classes: independent, often also referred to as Western or individualistic; and interdependent, often referred to as Eastern or collectivistic. People from independent cultures tend to focus on individual aspects of the self while people from interdependent cultures pay more attention to relational aspects of the self (Chentsova-Dutton and Tsai, 2010). Such strict division of all cultures into two classes is of course rather general and more flexible approaches to describe and classify cultures have been undertaken (Green, 2005; Tracy and Matsumoto, 2008). Not all European cultures are individualistic and not all Asian cultures are collectivistic. It is also important to keep in mind that changes in economical and political organisation and cultural norms are interdependent and influence each other. One such example is Taiwan where relatively recent change in economical structure has led to considerable shift in cultural norms from a typical Chinese collectivistic culture to a more individualistic culture (Lu and Yang, 2006). Nevertheless, different types of cultures are important to take into account since they influence the expression and perception of emotion.

Before we report on previous findings in cross-cultural differences in emotion perception, it is necessary to note that such differences should not be expected in all aspects of human emotional experience

and that some affective processes are certainly universal. Mesquita and Frija have made a very justified remark that "Whether cross-cultural differences or similarities are found depends to an important degree on the level of description of the emotional phenomena" (Mesquita and Frijda, 1992). This observation is also confirmed more recently by Matsumoto and Hwang (2012). Thus, such aspects as physiological responses, some elements of expressive behaviour, response to sudden threat, etc. are influenced by the biology of human beings and a common for all people and several other species. Studies that focus on the universality of emotional expression and perception (Ekman, 1992) even highlight the necessity of common physiological response to the stimuli across all populations as one criterium for universality.

In contrast, work that emphasises influence of culture focuses on the social aspect of emotion and constructs that are not innate but are rather acquired during the first years of life during the socialisation process — attitudes, beliefs, values about emotion are culturally specific and are part of any child's explicit or implicit upbringing (Harris, 1995; Eid and Diener, 2001; Tsai et al., 2006; Matsumoto and Hwang, 2012). A good example is given by Zborowski, who shows that cultural background determines to a certain extent the expression of pain; although the sensation of pain likely feels very similarly for all people, and the display of pain via face, voice and body is also universally recognised, the social context and the extent in which pain is displayed varies greatly across various cultures Zborowski (1969).

The ways emotions are expressed in everyday life are extremely versatile and include facial expressions, vocal prosody, body posture and motion, and verbal means. Up to now few researchers examined cross-cultural differences of emotion perception in whole body posture (Matsumoto and Kudoh, 1987; Kleinsmith and Silva, 2005; Kleinsmith et al., 2006) despite its obvious importance in emotion expression (den Stock and Righart, 2007; de Gelder, 2009; de Gelder et al., 2010). Moreover, undoubtedly due to the abundance of various cultures and sometimes difficulties connected with conducting research in remote locations, language difficulties and other real world factors, a generous proportion of cross-cultural studies focused on the differences between American and Japanese cultures, which are regarded as typical representatives of Western and Asian cultures respectively. For example, Friesen (1973) compared facial expressions of Japanese and American subjects while viewing both neutral films, and films intended to cause stress. He showed that both groups expressed the same facial

expressions when watching the films alone, but differed when being in company with compatriots. Scherer et al. (1988) reported fewer hand, arm and whole body gestures produced in emotional situations by people from Japanese culture as compared to Americans.

Despite the seeming abundance of research on cross-cultural differences in emotion perception, some of which focus on postures and gestures, there are still many open questions in this field. In the context of biological motion it is not clear how much influence the cultural origin has on the perception of the dynamic emotional body expressions and how much the physical properties of the motion itself. However, studying this aspect of emotional behaviour is very important in our globalised world. Not only could deeper understanding of these processes inform human-human interaction in multi-cultural communication (Glikson and Erez, 2013) but also aid culture-aware human-computer interaction starting with automatic affect recognition to avatar animation in virtual environments (Qu et al., 2013; Quiros-Ramirez and Onisawa, 2013).

There are several distinct approaches in research of emotion perception — dimensional approach (Russell and Mehrabian, 1977; Fontaine et al., 2007), categorical approach (Ekman, 1971, 1992) and appraisal models (Arnold, 1960; Lazarus et al., 1970; Scherer, 2001). Most research on cross-cultural emotion perception follows one of the approaches when developing their hypotheses and methods: for instance, participants from various cultural backgrounds are asked to categorise emotional stimuli or to rate the stimuli along various scales, but rarely are two methods used together.

We conducted this study to investigate cross-cultural differences in perception of emotional body expression and the underlying factors by combining emotion recognition and rating tasks. We ran the experiment with participants from three cultures: English, German, and Korean and investigated recognition, perception and conceptualisation of the ten following emotional categories: amusement, joy, pride, relief, surprise, anger, disgust, fear, sadness, shame. The participants were asked to observe, categorise and rate stimuli, a dynamic human upper body stick figure, on the screen with the help of categorisation buttons and affect and motion scales.

The motion sequences used in the experiment are a subset of a large database of motion sequences recorded in emotional narrative scenarios with state-of-the-art motion capture technology (Volkova et al., 2014a). The perception of the motion sequences used in this experiment

was previously investigated in a large scale emotion recognition study (Volkova et al., 2014b). The selected 50 motion sequences, five for each emotion category, that had received highest inter-rater agreement ratings within their category and thus combine the naturalness of emotional expression due to the original motion capture setup and high quality as shown by the previous observers. In a separate task we also recorded the culture specific display rules by asking each participant to rate the emotional categories as concepts without stick figure display stimuli — participants were presented with an emotional category, e.g., "Relief", and were asked to imagine the most typical way it is usually expressed via body motion.

In the present study we define and test the following research questions. First, we aim to find out whether there are cultural differences in perception of dynamic emotional body language. Second, we want to measure the extent to which potential cross-cultural differences in perception of motion are facilitated by the cultural display rules. Third, we want to establish the degree to which perceived and physical properties of human motion influence the emotional ratings of the stimuli.

Our results show that differences between cultures exist in both perception of real human body motion and imagined ideal body motion for several emotion categories and across several rating scales. We show the connection between culture-specific display rules and perceived properties of motion sequences, as well as structure of ratings between affect and motion scales.

5.3 Materials and Methods

5.3.1 Motion sequences

We selected 50 motion sequences from a large dataset of motion capture sequences[1] (Volkova et al., 2014a). Ten emotion categories were used in the experiment: five positive (*amusement, joy, pride, relief, surprise*) and five negative (*anger, disgust, fear, sadness, shame*). Each emotion category was represented by motion sequences collected in naturalistic narrative environment and thus good examples of how these emotions are expressed in real life. Five motion sequences with highest agreement rates among the observers from a previous study (Volkova et al., 2014b)

[1] http://ebmdb.tuebingen.mpg.de

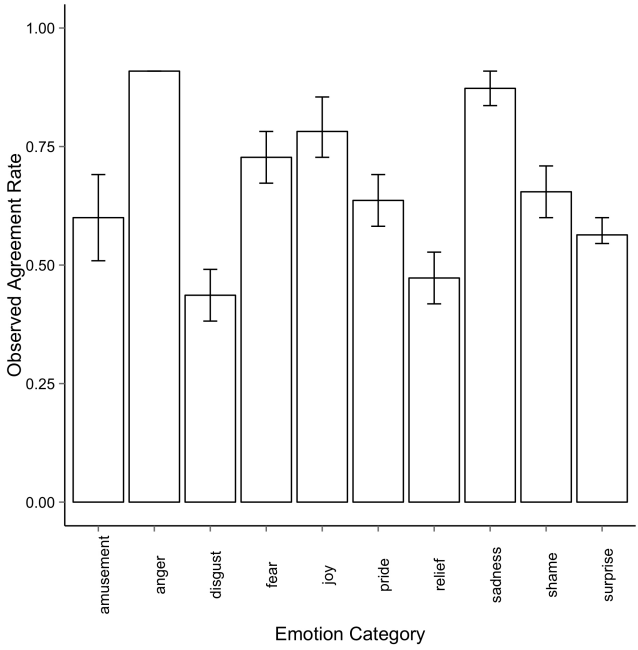

Figure 5.1: Motion stimuli quality. The observed agreement on the emotion labels of the motion sequences used in the study as obtained from previous research (Volkova et al., 2014b).

were chosen. The average observed agreement rates for each emotion category are shown in Figure 5.1.

The motion sequences originated from actors coming from four different countries: Germany (68%), Ireland (12%), England (10%) and India (10%). Each of the 50 stimuli was shown three times during each experiment session in random order, thus amounting to 150 trials. The stimuli were displayed in a form of a dynamic stick figure that represented human upper body (see Figure 5.2). During the instruction phase of the experiment the participants were informed that the motion came from a real person, however they were not told any details about the actors, e.g. their gender, cultural background, etc.

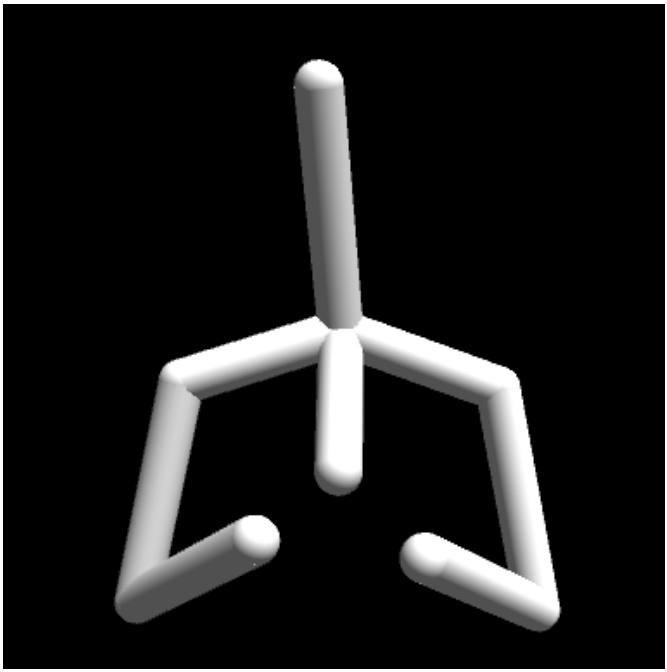

Figure 5.2: Stimuli display. The motion sequences were displayed in the form of a dynamic stick figure of human upper body.

5.3.2 Participants

The participants of this study came from three cultures: English, German and Korean. The 72 participants were between 18 and 30 years old (see Figure 5.3) and were distributed throughout the experiment in a balanced way: 36 per condition, 12 per culture in each condition, 6 per gender in each condition$\times$culture combination. All stages of the experiment were approved by local ethics committees and each participant signed a letter of informed consent. All participants received monetary compensation or course credit for their participation, all had normal or corrected to normal visual acuity. None of the participants were aware of the purpose of the experiment. The participants and the obtained data were treated strictly according to the Declaration of Helsinki.

We made sure that our participants were reliable representatives

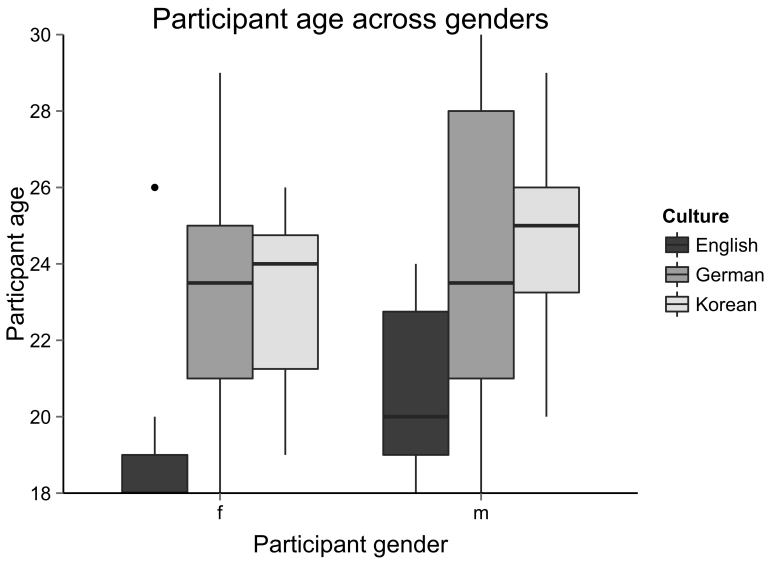

Figure 5.3: Participant age. Participant age across cultures and genders. In total, 72 participants took part in the experiment, 24 from each culture, out of which 12 in each culture were female. In each experimental condition there were 36 participants, 12 from each culture (6 female participants in each culture.

of each respective culture. The experiment was run in three different locations: Oxford, United Kingdom, for English culture, Tübingen, Germany for German culture and Seoul, South Korea for Korean culture. In each location we recruited only those participants who were native speakers of the language corresponding to the location, were born there or have been living in the country since early childhood for most of the time and have not been living abroad for extended periods of time in the course of the last few years.

5.3.3 General Experiment Flow and Main Experimental Task

The instructions for the experiment were originally written in English and then translated into German and Korean. The translations were

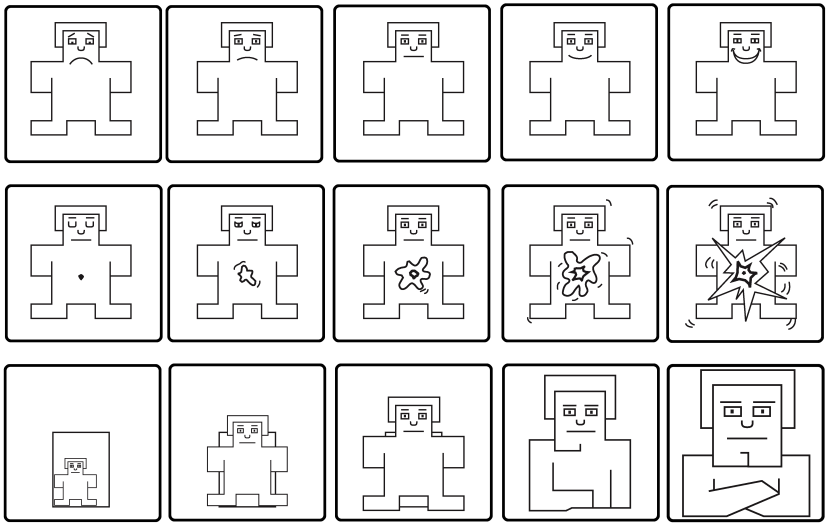

Figure 5.4: Affect scales. Three scales used for stimuli rating along valence, arousal and dominance dimensions. Adapted from Lang (1980). The affect scales were also used during the self-report of participant's emotional state. In the experiment software each scale was accompanied by a short description in corresponding language.

checked and back-translated several times by native speakers of all languages. Special effort was made to ensure closest correspondence of emotion category lists to their translated counterparts in other languages. In the beginning of each experiment session the participant was first given the experiment instructions. After that the participant signed the consent form and the experiment began. The first three trials were practice trials, in order for the participant to get used to the program interface and the main experiment task in experimenter's presence and ask questions if any should occur. The practice trials used motion sequences not included in the 50 selected stimuli. After the practice the participant completed the main experiment task alone.

Independent of the conditions described below, the main task of every trial was to rate the motion stimuli on three affect and three motion scales. The affect scales, shown in Figure 5.4 were adaptations of the well-known self-assessment manikin (SAM) task (Lang, 1980; Bradley and Lang, 1994), the motion scales (Figure 5.5) were developed

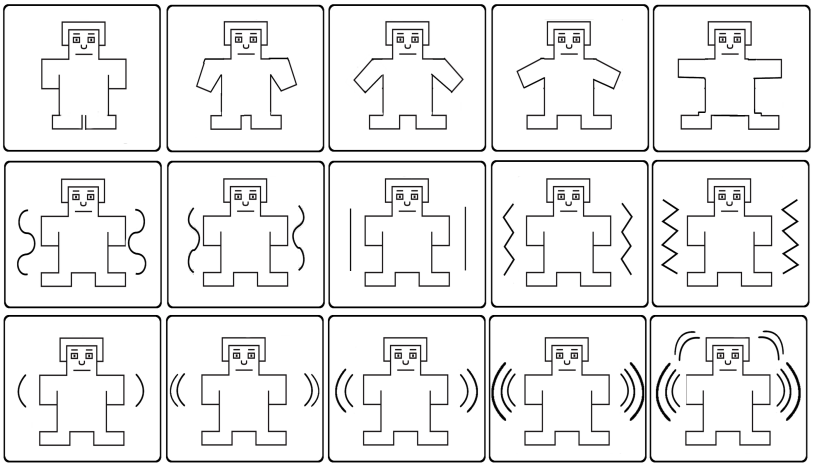

Figure 5.5: Motion scales. Three scales used for stimuli rating along span, brokenness and speed dimensions. Adapted from Lang (1980). In the experiment software each scale was accompanied by a short description in corresponding language.

specially for this experiment with the SAM task serving as a prototype. Before proceeding to the next trial, the participant could always change their ratings. Proceeding to the next trial was made impossible until six ratings were provided, each on five-point scale, values increasing from left to right :

Affect scales (see Figure 5.4):

1. Emotional valence — how negative or positive the actor represented by the stick figure feels at the moment.
2. Emotional arousal — how passive or active (calm or excited) the actor represented by the stick figure feels at the moment.
3. Emotional dominance — how little or much control the actor represented by the stick figure has over the situation.

Motion scales (see Figure 5.5):

1. Motion span — how narrow or open the motion is.
2. Motion brokenness — how smooth or jerky the motion is.

3. Motion speed — how slow or fast the motion is.

The participants were asked to report their own affective state with the help of the SAM scales (Figure 5.4) six times during the experiment, namely before the first non-practice trial, every 30 consequent trials, including one after the last trial. After filling out the scales to report their current emotional state, the participants were requested to have a short break to avoid fatigue. The reported values were used later to control for participants' emotional state influence on the ratings during the main experiment task. The overview of the experiment session flow is presented in Figure 5.6.

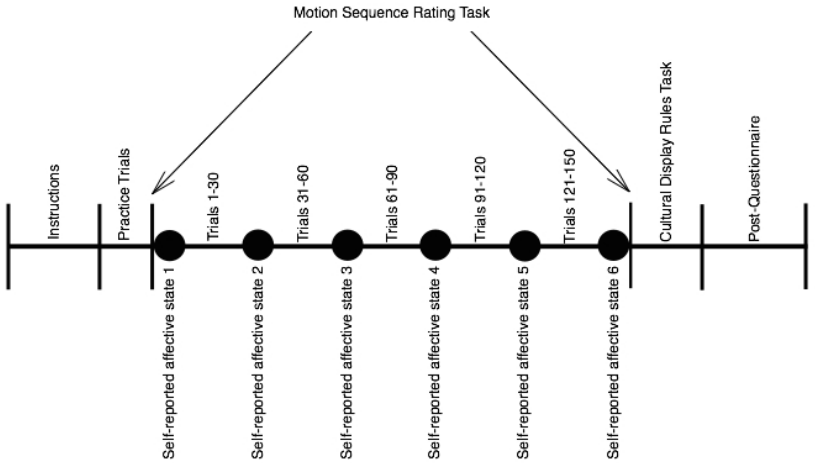

Figure 5.6: Experiment flow. Order of the experiment tasks during an experiment session. A typical session lasted 90 minutes. After the instructions and three practice trials the participants rated motion sequences (in CR Condition also classified them), reporting their own emotional state every 30 trials. They then completed the display rules rating task and other post-questionnaire questions.

5.3.4 Experimental Conditions

Each of the 72 participants took part in one of the two experimental conditions that we designed as a between-participant factor. The main task

(motion sequence rating along seven scales) and post-questionnaires were almost identical in all conditions.

In "only rating" condition (henceforth referred to as OR Condition) the participants were asked to rate motion sequences. The instructions and the main part of the experiment did not mention any of the emotion categories. The participants were asked to rate the emotional state of the actor, but special care was taken not to encourage them to label the motion sequence. For instance, the instructions were carefully written as not to mention any specific emotion category and always referred to more vague *emotional state* of the actor. This was done in order to avoid explicit reference to display rules before the first part of the post-questionnaire.

In "categorisation and rating" condition 2 (henceforth referred to as CR Condition) the participants were asked to first categorise the motion sequence using the list of emotion categories: *amusement, joy, pride, relief, surprise, anger, disgust, fear, sadness,* and *shame* (Table 5.1 gives emotion categories in English, German, and Korean). The order of emotion categories was randomised for each experiment session. The emotion and motion scales were hidden from view until the participant has chosen an emotion category. Here we expected to trigger participant's reference to display rules during the experiment. Participants of this condition also allowed us to measure emotion recognition accuracy and investigate potential cross-cultural differences based on direct labelling of motion sequences.

5.3.5 Post-questionnaire

In the end of each experiment session each participant filled out a post-questionnaire. First, the participants were presented with ten trials that dealt with the ten emotion categories used in this study. With the help of these trials we recorded each participant's display rules. The task was not anticipated by the participants as it was referred to as part of the post-questionnaire in the initial instructions. This was done so not to trigger participants display rules of emotion categories used in the experiment. For the participants of OR Condition this was also the first time during the experiment when they were asked to work with emotion categories directly. In each trial the participant saw an emotion category (e.g., "Joy") and was asked to imagine a person experiencing and actively expressing the most typical instance of this emotion. The participant was then asked to rate the imaginary person's emotional

English	German	Korean
Amusement	Belustigung	재미
Joy	Freude	기쁨
Pride	Stolz	자부심
Relief	Erleichterung	안도감
Surprise	Überraschung	놀람
Anger	Ärger	화남
Disgust	Ekel	혐오감
Fear	Angst	두려움
Sadness	Traurigkeit	슬픔
Shame	Scham	수치심

Table 5.1: Emotion categories used in CR Condition. The order of the categories, shown as buttons in experiment software, was randomised for each participant.

state along the affect scales and imaginary person's corresponding motion along the motion scales. During this task the participant was alone in the experiment room.

The rest of the post-questionnaire questions were asked verbally and filled in by the researcher into the experiment software in order to avoid any mistakes. The questions included the following items:

1. How difficult was the main experiment task on the scale from 1 (very easy) to 5 (very difficult)

2. (For OR Condition) While rating the motion of the stick-figure actor and their emotional state, did you think of an emotion category the movement represented? If yes, did you do it for every trial or only for a certain proportion of the trials?

3. While rating the motion sequences, did you try to make the same movement yourself? If yes, did you do it for every trial or only for a certain proportion of the trials?

4. (For the display rules task) When you were imagining a person expressing a certain emotion, did you try to make the imagined movement yourself? If yes, did you do it for every out of ten trials or only for a certain proportion of the trials?

Lastly, basic demographic information was collected from the participant, such as gender, age, handedness and acting experience.

5.4 Results

First, we consider whether participants' affective state had an effect on the ratings given to the motion stimuli. Second, we investigate cross-cultural differences in the two main tasks of the experiment: motion sequences and display rules, first each separately then in comparison to each other. Additionally, we perform factor analysis in order to gain insight into the influence of cultural background on emotional perception of human body motion. The effect of condition on participants' responses as well as the recognition accuracy rate of participants from CR Condition is reported, and several important observations from post-questionnaires.

5.4.1 Self-Reported Affect State

During the experiment the participants were requested to report their own affective state using the three emotional scales of valence, arousal, and dominance. For each participants six SAM measurements were collected: before the beginning of the motion sequence rating task of the experiment (after the instruction and the training trials phase) and after each 30 trials, the last measurement point being after the last, 150th, trial (see Figure 5.6).

The SAM measurements were used for control for fatigue effect, since a typical experiment session lasted more than one hour. A MANOVA analysis revealed that the reported ratings of valence, arousal and dominance were affected by the within-participant factor of measurement time: $F(15, 1278) = 3.917$, $p<.001$, Pillai's trace $V=0.131$. This result means that the participants' emotional state was changing while the experiment progressed. By running separate univariate ANOVAs on each SAM rating scale we found out that only the ratings along the valence scale were changing significantly during the experiment: $F(5, 426)=8.991$, $p<.001$, $ges=0.095$.

The important question to ask however is whether the significant change in the emotional state along the valence of the participants, notably only after the first set of trials (see Figure 5.7), influenced their response along the valence scale while rating the motion sequences.

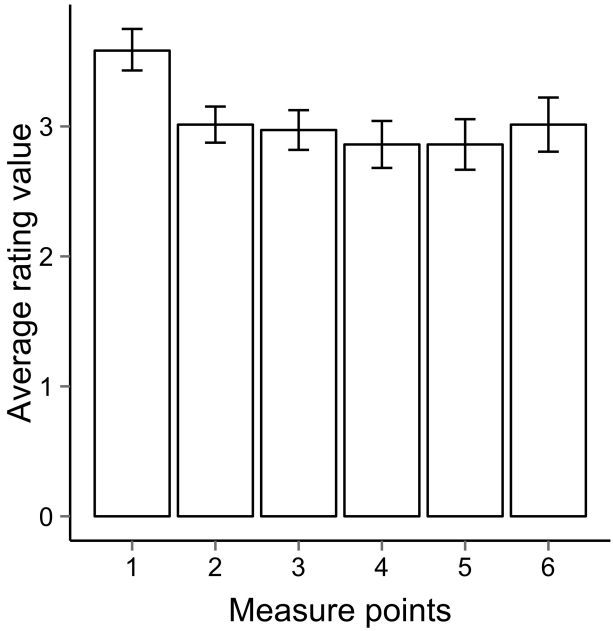

Figure 5.7: Self-reported valence across participants in the course of the experiment. Valence ratings and all participants. The six measured points are as follows: 1 – before the beginning of the first trial of the main task, 2-5 – during the main task after trials 30, 60, 90, 120 respectively, 6 – after the last trial of the main task. All error bars represent 95% CI, N in each bar is the total number of participants, 72. On average, participant's rating of valence decreed to neutral after the first block of trials and remained at this level for the rest of the experiment.

Linear regression analysis reveals that the effect of self-reported emotional state rating on the emotional scales ratings for motion sequences is non-significant: $F(1, 358)=2.735$, $p=.099$, adjusted R-squared = 0.004, beta = 0.087. In light of the results of the linear regression analysis no normalisation of the ratings of the motion sequences were necessary.

5.4.2 Motion Sequence Ratings

During each experiment session across all experimental conditions, the participants were asked to observe and rate the motion sequences along

the affect and motion scales. We analysed the effect of participant's culture (English, German, Korean) and experiment condition (OR, CR) as between-participant factors and emotion category of displayed motion as within-participant factor on the variation in ratings along six scales (valence, arousal, dominance, speed, span, brokenness).

Using Pillai's trace, we found a three-way interaction between the emotion of displayed motion, participant's culture and experimental condition in their effect on the affect and motion ratings ($V=0.290$, $F(108, 3564)=1.682$, $p<.001$). Additionally there is two-way interaction between emotion of displayed motion and experimental condition (Pillai's trace $V=0.209$, $F(54, 3564)=2.386$, $p<.001$) and a two-way interaction between emotion of displayed motion and participant's culture (Pillai's trace $V=0.709$, $F(108, 3564)=4.424$). However, no significant interaction between participant's culture and experimental condition was observed. Each independent factor taken separately had a main effect on the motion sequences ratings: participant's culture (Pillai's trace $V=0.322$, $F(12, 124)=1.98$, $p=.030$), experimental condition ($V=0.249$, $F(6, 61)=3.385$, $p=.006$) and emotion of displayed motion ($V=2.508$, $F(54, 3564)=47.419$, $p<.001$).

Table 5.2 gives an overview of the univariate ANOVA results for the independent factors and their combinations that had significant effect on each individual scale. The table shows that for univariate analyses there is no main effect of participant's culture, although it is still present in two- and three-way interactions.

Taken that two groups of our participants originated from Western cultures and one group from an Asian culture, we performed orthogonal contact analysis on each scale grouping English and German cultures together. The results show main effect of culture for two scales: speed ($F(2, 717)=2.234$, $p=0.040$) and brokenness ($F(2, 717)=17.23$, $p<.001$).

Figures 5.8 and 5.9 are a detailed depiction of differences among ratings, where the distinction is made of the level of scale type first, then on the emotion category and lastly on the factor of culture. As can be seen from the figure, there are cross-cultural differences on the level of certain scales and emotion categories. Table 5.3 gives a full account of these differences. The row numbers of Table 5.3 correspond to numbers above box plot groups in Figures 5.8 and 5.9. The scale$\times$emotion combinations in bold mean that a cross-cultural difference on the same scale for the same emotion was also found for display rules task.

Scale	Effect	DFn	DFd	F	p	ges
valence	emotion	9	594	13.647	<.001	0.141
	culture × emotion	18	594	2.928	<.001	0.066
	condition × emotion	9	594	3.369	<.001	0.039
	culture × condition × emotion	18	594	2.170	.003	0.049
arousal	condition	1	66	8.525	.004	0.056
	emotion	9	594	45.261	<.001	0.268
	culture × emotion	18	594	4.522	<.001	0.068
	condition × emotion	9	594	6.543	<.001	0.050
	culture × condition × emotion	18	594	2.826	<.001	0.043
dominance	emotion	9	594	8.643	<.001	0.093
	culture:condition × emotion	18	594	2.041	.006	0.046
speed	emotion	9	594	21.812	<.001	0.104
	culture × emotion	18	594	2.467	<.001	0.025
	culture × condition × emotion	18	594	2.079	.005	0.021
span	emotion	9	594	60.570	<.001	0.290
	culture × emotion	18	594	2.080	.005	0.027
	condition × emotion	9	594	2.266	.016	0.015
brokenness	emotion	9	594	6.098	<.001	0.050
	culture × emotion	18	594	2.637	<.001	0.044

Table 5.2: Univariate analyses of effects on ratings along affect and motion scales for the motion sequence rating task

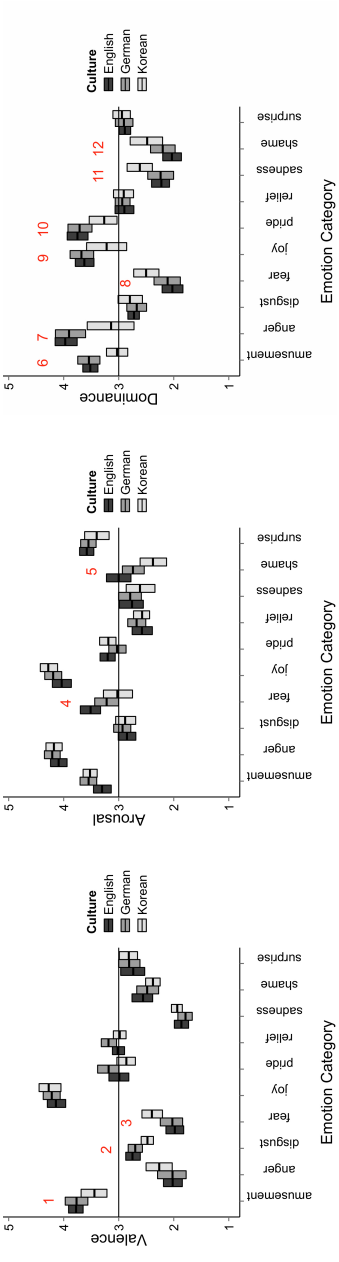

Figure 5.8: Ratings of motion sequences for each affect scale: valence, arousal, and dominance; for each displayed emotion across three cultures. The cross-boxes show the mean value and 95% CI. Numbers above box plot groups correspond to row numbers in Table 5.3

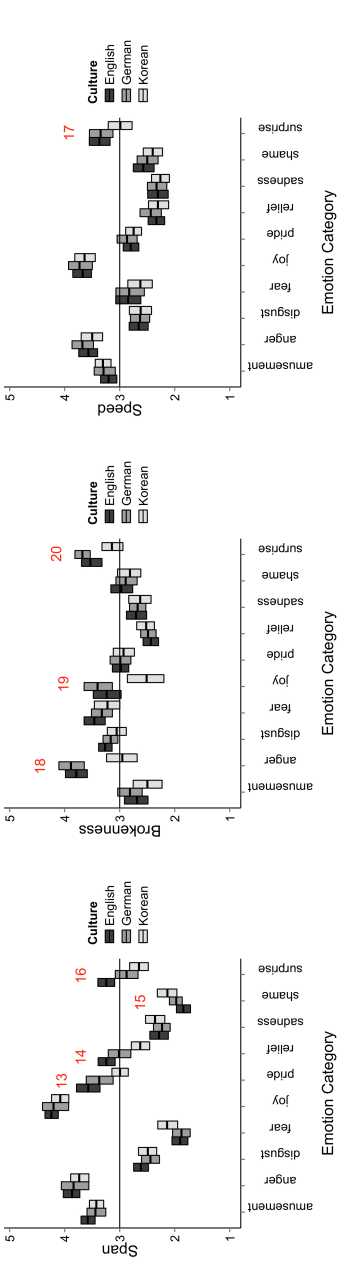

Figure 5.9: Ratings of motion sequences for each motion scale: span, speed, and brokenness; for each displayed emotion across three cultures. The cross-boxes show the mean value and 95% CI. Numbers above box plot groups correspond to row numbers in Table 5.3

			ANOVA			Pairwise Comparisons			Mean and SE values		
#	scale	emotion	F (2, 68)	p	es	E-G	E-K	G-K	English	German	Korean
1	valence	amusement	3.549	0.034	0.093	n.s.	n.s.	n.s.	3.775 (0.921)	3.767 (0.956)	3.442 (1.045)
2	valence	disgust	4.89	0.01	0.124	n.s.	0.013	0.038	2.75 (0.981)	2.7 (0.931)	2.478 (0.946)
3	valence	fear	6.029	0.004	0.149	n.s.	0.007	0.012	1.981 (0.906)	2.025 (0.856)	2.397 (1.001)
4	**arousal**	**fear**	4.158	0.02	0.108	n.s.	0.017	n.s.	3.508 (0.906)	3.219 (0.828)	3.025 (1.037)
5	arousal	shame	7.024	0.002	0.169	n.s.	0.001	n.s.	3 (0.795)	2.742 (0.939)	2.378 (0.908)
6	dominance	amusement	9.67	<.001	0.219	n.s.	0.001	0.001	3.517 (0.795)	3.542 (0.819)	3.031 (0.971)
7	**dominance**	**anger**	8.321	0.001	0.194	n.s.	0.001	0.002	3.972 (0.725)	3.903 (0.723)	3.139 (0.726)
8	**dominance**	**fear**	4.906	0.01	0.125	n.s.	0.013	0.037	2.025 (0.781)	2.111 (0.838)	2.503 (0.876)
9	**dominance**	**joy**	3.234	0.045	0.086	n.s.	n.s.	n.s.	3.625 (0.757)	3.678 (0.773)	3.219 (0.919)
10	**dominance**	**pride**	4.992	0.009	0.126	n.s.	0.018	0.022	3.75 (0.855)	3.711 (0.832)	3.264 (0.91)
11	dominance	sadness	3.949	0.024	0.103	n.s.	0.047	0.047	2.225 (0.807)	2.239 (0.82)	2.617 (0.855)
12	**dominance**	**shame**	3.541	0.034	0.093	n.s.	0.031	n.s.	2.033 (0.789)	2.2 (0.841)	2.483 (0.852)
13	span	pride	7.183	0.001	0.172	n.s.	0.001	0.035	3.572 (0.868)	3.372 (0.923)	2.994 (1.051)
14	span	relief	11.038	<.001	0.242	n.s.	<.001	0.01	3.244 (0.979)	3.014 (0.917)	2.628 (0.887)
15	**span**	**shame**	3.788	0.027	0.099	n.s.	0.023	n.s.	1.844 (0.794)	1.972 (0.965)	2.133 (1.062)
16	span	surprise	10.238	<.001	0.229	0.014	<.001	n.s.	3.244 (0.747)	2.875 (0.793)	2.65 (0.808)
17	speed	surprise	3.722	0.029	0.097	n.s.	n.s.	n.s.	3.369 (0.672)	3.347 (0.67)	2.989 (0.771)
18	brokenness	anger	16.292	<.001	0.321	n.s.	<.001	0	3.789 (0.772)	3.886 (0.671)	2.953 (0.896)
19	brokenness	joy	10.898	<.001	0.24	n.s.	0.001	<.001	3.236 (0.901)	3.403 (0.914)	2.511 (0.992)
20	brokenness	surprise	9.044	<.001	0.208	n.s.	0.007	<.001	3.533 (0.696)	3.678 (0.64)	3.142 (0.737)

Table 5.3: ANOVA, post-hoc analysis, mean and SE values of cross-cultural differences along scales and emotion categories in motion sequence ratings; row numbers refer to numbers above box plot groups in Figures 5.8 and **??**.

5.4.3 Cultural Display Rules

As the first part of post-questionnaire the participants were asked to rate ten emotion categories by imagining the most typical instance of each category and rating the imaginary person's emotional state and motion patterns. A MANOVA was performed with the parameters similar to the repeated measures MANOVA used for motion sequence rating analysis (Section 5.4.2). Thus, we analysed the effect of participant's culture and experiment condition as between-participant factors and displayed emotion category as a within-participant factor on the variation in ratings along six scales (valence, arousal, dominance, speed, span, brokenness) for imagined ideal motion for each emotion category.

The results show two-way interaction between participant's culture and displayed emotion category (Pillai's trace V=0.490, F(108, 3564)=2.937, $p<.001$) and main effects for displayed emotion category (Pillai's trace V=1.989, F(54, 3564)=32.746, $p<.001$) and culture (Pillai's trace V=0.448, F(12, 124)=2.985, p=0.001). Univariate analyses along each scale however shows only a main effect of displayed emotion category (see Table 5.4).

Scale	F(9, 594)	effect size
valence	56.952	0.435
arousal	9.497	0.100
dominance	6.671	0.073
speed	9.292	0.106
span	13.338	0.151
brokenness	4.976	0.060

Table 5.4: Univariate analyses of effect of expressed emotion on ratings of imagined motion in display rules rating task. All p-values are under .001

Similar to the motion sequence rating task analysis we performed orthogonal contact analysis on each scale grouping English and German cultures together. The results show main effect of culture contrast (Western *vs.* Asian) for brokenness (F(2, 717)=2.028, p=0.044). Note that a statistically significant difference for this scale was also found for the motion sequence rating task in the previous section.

Figures 5.10 and 5.11 is a detailed depiction of differences among ratings, where the distinction is made of the level of scale type first,

then on the emotion category and lastly on the factor of culture. As can be seen from the figure, here too there are cross-cultural differences on the level of certain scales and emotion categories. Table 5.5 gives a full account of these differences. The row numbers of Table 5.5 correspond to numbers above box plot groups in Figures 5.10 and 5.11. The scale×emotion combinations in bold mean that a cross-cultural difference on the same scale for the same emotion was also found for display rules task.

5.4.4 Ratings of Imagined and Observed Motion

In the display rules rating task participants were asked to imagine the most typical motion for each emotion category. As Figures 5.12 and 5.13 show, the ratings of typical imaginary motion sequences (blue box plots) are further away from the neutral line (3.0) than rating of motion sequences (black box plots). This indicates that either our motion capture sequences are examples of very subtle emotional expression and/or that the imagined typical examples of emotion categories are in fact stereotypical and extreme (see Figures 5.12 and 5.13).

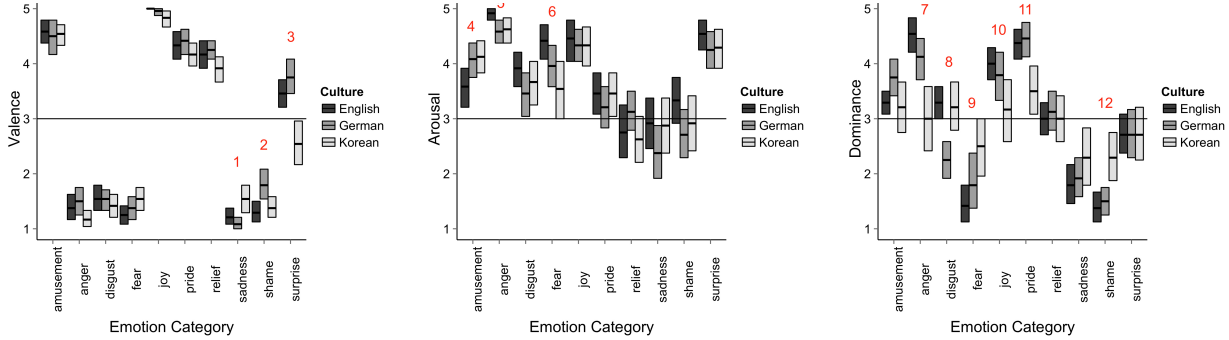

Figure 5.10: Ratings of display rules for each affect scale: valence, arousal, and dominance; for each displayed emotion across three cultures. The cross-boxes show the mean value and 95% CI. Numbers above box plot groups correspond to row numbers in Table 5.5

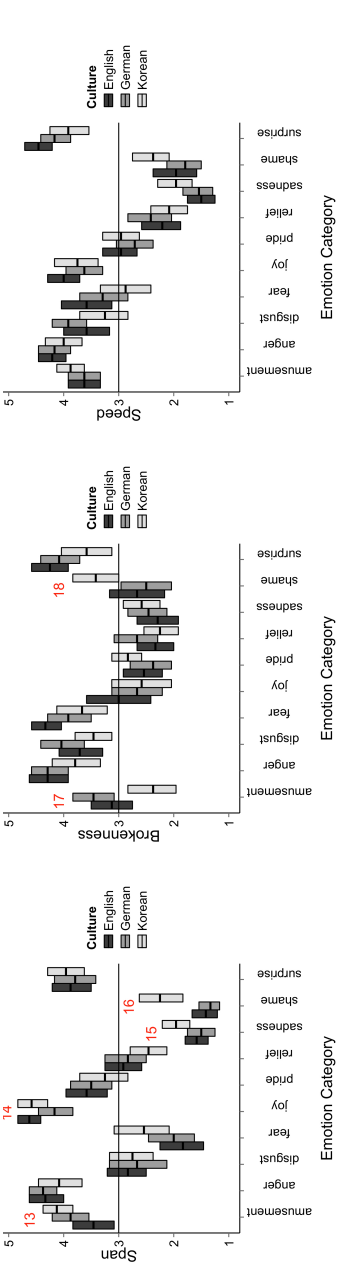

Figure 5.11: Ratings of display rules for each motion scale: span, speed, and brokenness; for each displayed emotion across three cultures. The cross-boxes show the mean value and 95% CI. Numbers above box plot groups correspond to row numbers in Table 5.5

			ANOVA			Pairwise Comparisons			Mean and SE values		
#	scale	emotion	F (2, 68)	p	es	E-G	E-K	G-K	English	German	Korean
1	valence	sadness	6.761	0.002	0.164	n.s.	0.024	0.002	1.208 (0.415)	1.083 (0.282)	1.542 (0.588)
2	valence	shame	5.272	0.007	0.133	0.01	n.s.	0.028	1.292 (0.464)	1.792 (0.721)	1.375 (0.495)
3	valence	surprise	13.059	<.001	0.275	n.s.	0.001	<.001	3.458 (0.658)	3.75 (0.847)	2.542 (1.021)
4	arousal	amusement	3.397	0.039	0.09	n.s.	n.s.	n.s.	3.583 (0.881)	4.083 (0.776)	4.125 (0.741)
5	arousal	anger	3.572	0.033	0.094	0.05	n.s.	n.s.	4.917 (0.282)	4.583 (0.504)	4.625 (0.576)
6	**arousal**	**fear**	3.833	0.026	0.1	n.s.	0.022	n.s.	4.417 (0.83)	3.958 (1.083)	3.542 (1.318)
7	**dominance**	**anger**	11.376	<.001	0.248	n.s.	<.001	0.003	4.542 (0.833)	4.125 (0.992)	3 (1.532)
8	dominance	disgust	8.231	0.001	0.193	0.002	n.s.	0.003	3.292 (0.806)	2.25 (0.897)	3.208 (1.215)
9	**dominance**	**fear**	5.125	0.008	0.129	n.s.	0.007	n.s.	1.417 (0.881)	1.792 (1.285)	2.5 (1.351)
10	**dominance**	**joy**	3.568	0.034	0.094	n.s.	0.037	n.s.	4 (0.722)	3.792 (1.103)	3.167 (1.435)
11	**dominance**	**pride**	8.397	0.001	0.196	n.s.	0.002	0.001	4.375 (0.711)	4.458 (0.779)	3.5 (1.142)
12	**dominance**	**shame**	8.09	0.001	0.19	n.s.	0.001	0.004	1.375 (0.711)	1.5 (0.659)	2.292 (1.122)
13	span	amusement	3.817	0.027	0.1	n.s.	0.024	n.s.	3.458 (0.977)	3.875 (0.797)	4.125 (0.741)
14	span	joy	3.455	0.037	0.091	n.s.	n.s.	n.s.	4.625 (0.576)	4.167 (0.761)	4.583 (0.654)
15	span	sadness	3.982	0.023	0.103	n.s.	n.s.	0.03	1.583 (0.584)	1.5 (0.59)	1.958 (0.624)
16	**span**	**shame**	11.296	<.001	0.247	n.s.	<.001	0	1.417 (0.654)	1.333 (0.482)	2.25 (0.989)
17	brokenness	amusement	7.475	0.001	0.178	n.s.	0.022	0.001	3.125 (0.9)	3.458 (0.932)	2.375 (1.135)
18	brokenness	shame	4.149	0.02	0.107	n.s.	n.s.	0.026	2.667 (1.239)	2.5 (1.142)	3.417 (1.139)

Table 5.5: ANOVA, post-hoc analysis, mean and SE values of cross-cultural differences along scales and emotion categories in display rules ratings; row numbers refer to numbers above box plot groups in Figures 5.10 and 5.11

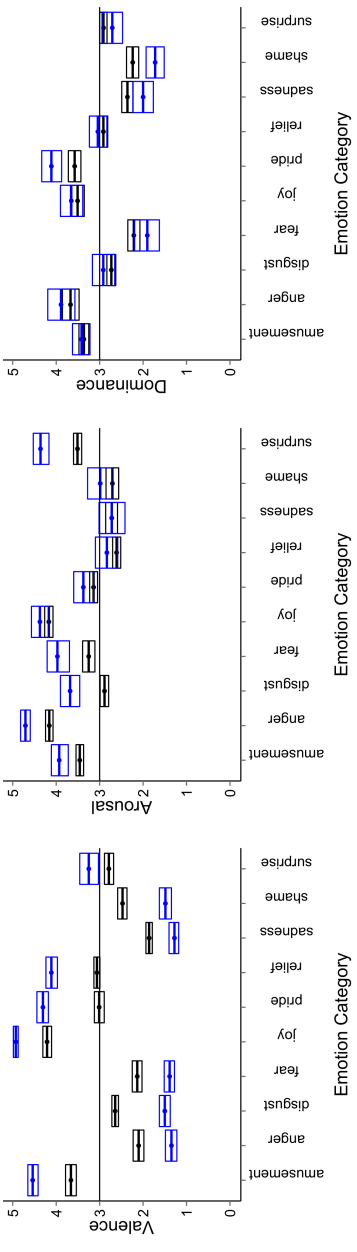

Figure 5.12: Relation between ratings of display rules and motion sequences for each affect scale: valence, arousal, and dominance. Box plots in blue show display rule ratings, black box plots show motion sequence ratings averaged for all cultures.

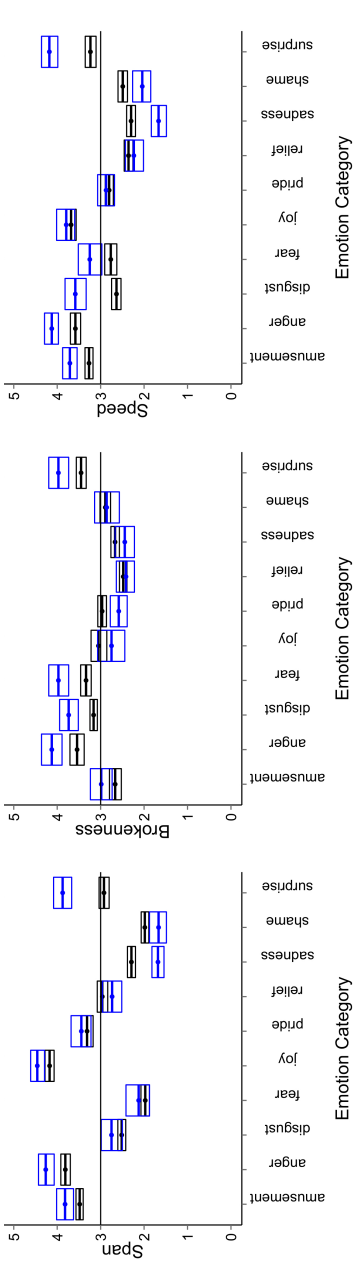

Figure 5.13: Relation between ratings of display rules and motion sequences for each motion scale: span, speed, and brokenness. Box plots in blue show display rule ratings, black box plots show motion sequence ratings averaged for all cultures.

5.4.5 Effect of Culture and Motion on Ratings

Several previous studies have observed stable connection between perceived emotional characteristics of biological motion and their physical properties — span or openness of body posture are associated with higher valence (Wallbott, 1985), while Pollick et al. (2001b) found strong positive correlations between perceived intensity of anger expressions and movement velocity.Generally, higher speed and span of motion and lower fluidity of motion have been associated with higher arousal, higher span of motion additionally associated with higher perceived dominance. We performed factor analysis on six rating scales to discover the relation between affect and motion scale rating separately for each culture. Within each culture we differentiated between motion sequence ratings and ratings of cultural display rules.

Six principal axis factor analyses with a varimax orthogonal rotation of 6 scales were conducted on N=240 samples each, motion sequence rating task and display rules rating task (24 participants $\times$ 10 emotion categories. An examination of the Kaiser-Meyer Olkin measure of sampling adequacy suggested that all the samples were factorable (see Table 5.6). We selected two factors for each analysis based on the number of first resulting eigenvalues whose values were greater than 1.0. Due to the fact that there were only two factors, loadings under .5 were discarded. The resulting underlying structure of factors and their loadings on each scale is shown in Table 5.6. Several observation can be made. First, groupings of scales into factors is always the same within each culture between the motion sequence ratings and cultural display rules. Second, the scale groupings into factors in English and German cultures are very similar to each other — in both Western cultures higher brokenness and higher speed profile of motion are associated with higher perceived levels of arousal, and wider span of motion means higher dominance and valence. Korean culture clearly stands out, because increased smoothness of motion (negative factor loadings of brokenness) is associated with higher positiveness of expressed emotion. Moreover, whereas higher dominance is European cultures is associated with higher span, in Korean model it is not present at all.

5.4.6 CR Condition and Recognition Accuracy

In CR Condition the participants were asked to categorise each emotion sequence before rating it. Thus, for these 36 participants (12 in each

		Motion Sequence		Display Rules	
Culture	scales	RC1	RC2	RC1	RC2
English	valence	.82			.82
	arousal		.83	.79	
	dominance	.83			.81
	span	.87			.82
	brokenness		.87	.85	
	speed		.80	.82	
	KMO	0.704		0.689	
	Eig. Val.	3.09	1.57	2.73	1.66
German	valence		.86		.85
	arousal	.79		.69	
	dominance		.78		.80
	span		.74		.73
	brokenness	.89		.85	
	speed	.85		.84	
	KMO	0.747		0.742	
	Eig. Val.	3.28	1.37	2.81	1.54
Korean	valence		.65		.83
	arousal	.89		.81	
	dominance				
	span	.84		.86	
	brokenness		-.91		-.79
	speed	.92		.85	
	KMO	0.750		0.689	
	Eig. Val.	2.98	1.18	2.37	1.46

Table 5.6: Factor analyses of motion sequence rating and culture display rules. RC1 and RC2 are the two principal components with Eigenvalues above 1.0. In each culture the distribution of scales across factors was consistent for motion sequence ratings and display rules ratings. In German and English cultures the composition of factors is the same but is rather different from the one in Korean culture. In Western cultures *valence*, *dominance* and *span* belong to one factor and *arousal, brokenness* and *speed* to the other, in Korean culture *arousal, span* and *speed* form one factor, *valence* and *brokenness* (with highly negative loading) the other and *dominance* does not figure in the resulting factors at all.

culture) we can measure the emotion recognition rate and investigate what factors have effect on it. In each trial in experiment sessions in CR Condition the recognition accuracy was recorded as "0" if the displayed and response emotion categories did not coincide and as "1" otherwise. The repeated measures ANOVA results show two-way interaction between displayed emotion category and participant culture: $F(18, 270)=2.15$, $p=.004$, $ges=0.110$. The main effects of culture ($F(2, 30)=13.27$, $p<.001$, $ges=0.10$) and displayed emotion ($F(9, 270)=34.7$, $p<.001$, $ges=0.498$) are also present (also see Figure 5.14, especially results for emotion categories marked with frames around corresponding box plot groups).

The effect of displayed emotion is not surprising since many studies have shown that the recognition rate varies across emotion categories, modalities and display type. Namely, dynamic stimuli have higher recognition rates than static stimuli, especially for subtle and social expressions (Kaulard et al., 2012), and observers seem to have a bias for certain categories, like *anger* (Pichon et al., 2009; Visch et al., 2014; Volkova et al., 2014b). The effect of culture however is notable and demands further investigation.

Mean accuracy rate varies among cultures (English: M=0.31, SD=0.46; German: M=0.43, SD=0.49; Korean: M=0.36, SD=0.48). Holm-corrected p-values for post-hoc pairwise t-test comparisons also show significant differences across all cultures (see Table 5.7, "overall" row). The fact that participants from English culture have lower accuracy than other cultures is somewhat surprising. Since our participants from English culture were usually younger that participants from other cultures (see Figure 5.3 for reference), we have conducted ANOVA with participant age as between participant predictor and emotion recognition accuracy as dependant variable. Participants' age has shown no significant effect of recognition accuracy ($F(1, 358)=0.13$, $p=.719$). We found significant differences across cultures for several emotion categories: *amusement, disgust, joy,* and *sadness*. Table 5.7 shows Holm-corrected p-values for post-hoc pairwise t-test comparisons for these four emotion categories for each culture combination. Taking into account that most of the animations used in this experiment originated from German actors, the higher accuracy among German participants is most likely an evidence for "in-group" advantage that was previously shown for facial expressions (Elfenbein, 2013).

Since recognition accuracy is not perfect, it is also important to investigate the errors participants make during categorisation task. In

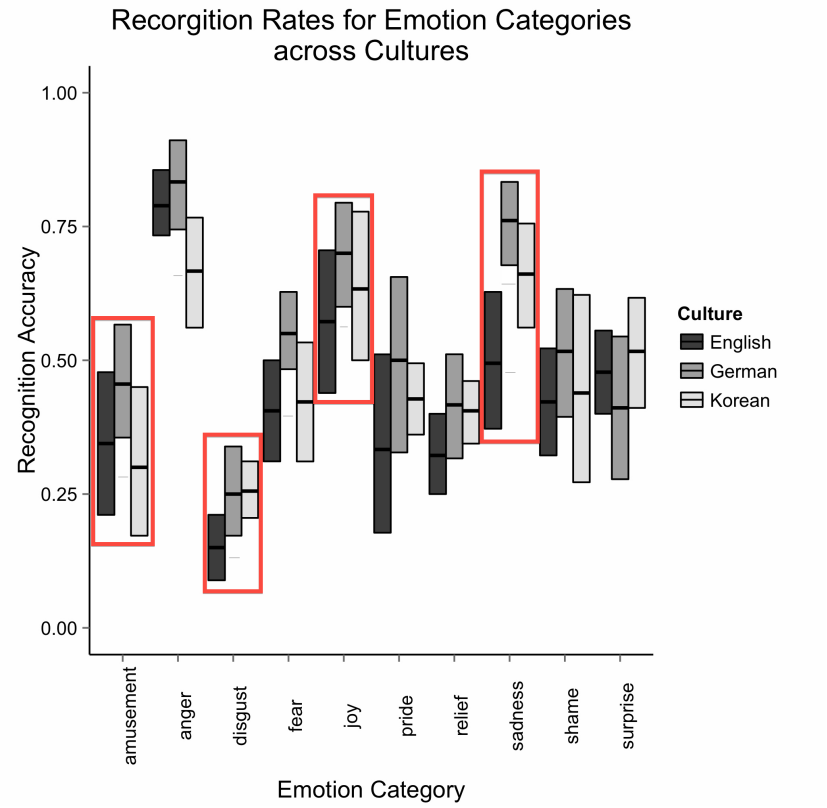

Figure 5.14: Recognition accuracy in CR Condition for each emotion category across cultures. Box plot groups in red frames mark emotion categories where statistically significant difference in recognition accuracy between cultured was found (see Table 5.7).

order to investigate the mis-categorisation patterns, confusion matrices are a useful tool. Figure 5.15 shows how participants' responses are distributed across displayed categories. The diagonals of the matrices represent the recognition accuracy rate presented in Figure 5.14, averaged across the three cultures in the top-left matrix and . From these matrices we can see that some emotions get confused with others. For the purpose of simplicity we use a 0.2 cut-off threshold, which

Emotion	English-German	English-Korean	German-Korean
Overall	<.001	.039	.017
amusement	.011	n.s.	.003
disgust	n.s.	.010	n.s.
joy	.038	n.s.	.038
sadness	.010	.018	n.s.

Table 5.7: Pair-wise post-hoc analysis for cross-cultural differences in emotion recognition rates, Holm-corrected p-values. "Overall" row shows cross cultural differences in emotion recognition for all ten emotions.

means that a mis-categorisation is taken into account when 20% or more responses form the displayed-response pair (marked with boxes in Figure 5.15). Across all cultures intended *amusement* is often interpreted as *joy* but a reversed situation is not as frequent. Moreover, *pride* is often perceived as *anger*. Confusion matrices for each individual culture reveal the following differences: only English participant often mis-interpreted the expression of *disgust* as *amusement* and display of *sadness* as *shame*; German participants frequently mis-categorised intended *surprise* as *anger* and Korean participants mis-took *shame* for *fear*.

5.4.7 Comparison between Intended and Observed Emotion Categories

One important distinction between OR and CR Conditions is that in the latter the ratings along the six scales can be traced back not only to the emotion categories that were expressed in the stimuli, but also to the emotion category given by the observer. As results reported in Section 5.4.2 demonstrate, the experimental condition is a main effect of the results. However, univariate analysis shows that only on the arousal scale is experimental condition a main effect with mean arousal higher in CR condition (M=3.365, SD=1.032) than in OC Condition (M=3.156, SD=1.168). One of processes a categorisation task can trigger is conscious matching of perceived properties of an observed motion sequence to the display rules of an emotion category one associated the motion sequence with. Even though the recognition rate is high considering the number of emotion categories and the mis-categorisations largely fall into clear patterns, it is important to

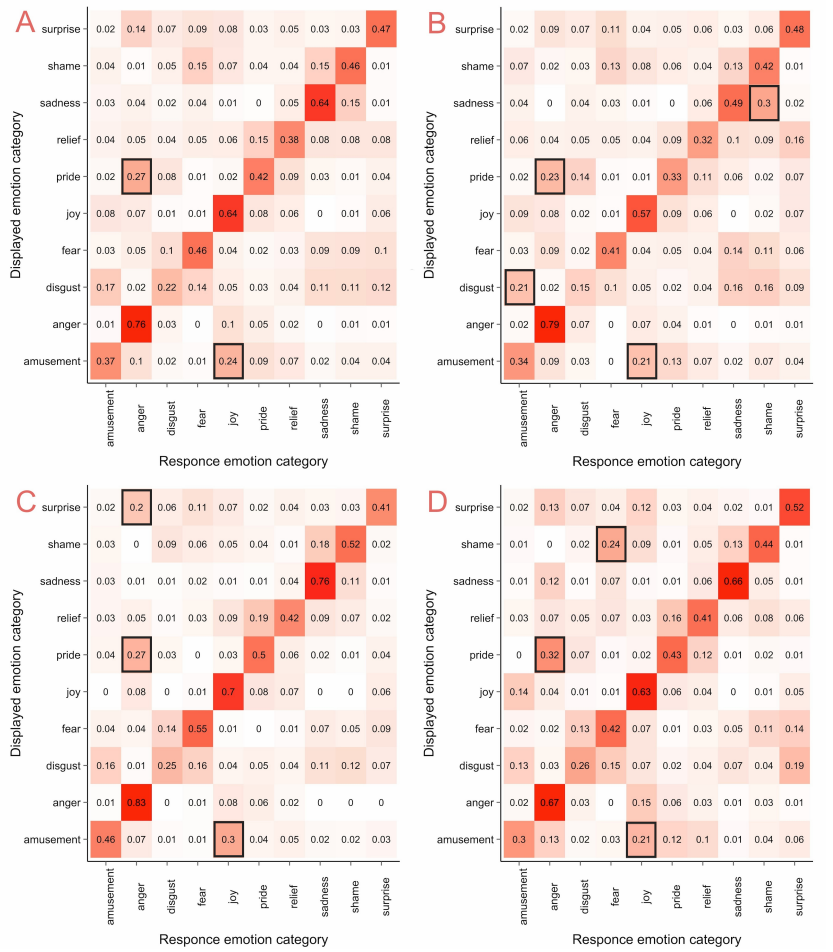

Figure 5.15: Confusion matrix for displayed and response emotion categories. (A) all cultures; (B) English participants; (C) German participants; (D) Korean participants. On y axis the intended categories are shown, on x axis - the response categories. The diagonal shows response accuracy. Each row sums up to 1.0.

control for possible differences between ratings of motion sequenced in reference to their original emotional levels and their new labels from

the categorisation task.

Hypothetically, if CR Condition triggers display rules and OR does not, ratings in CR conditions, bound to the response labels, should be closer to the ratings of the display rules. To test this hypothesis we obtained correlation values for the six ratings scales between motion sequence ratings scales and display rules of the category that was implied by the stimuli (Figure 5.16, *stimuli* emotion origin, blue colour) and between the motion sequence rating scales and the display rules of the category obtained during the categorisation task (Figure 5.16, *response* emotion origin, red colour). Especially for the affect scales the correlation for the response is higher, but the t-test shows no statistically significant difference between the two methods of motion sequence rating evaluation (t = 0.587, df = 8.813, p-value = 0.571, response factor M=0.682, stimuli factor M=0.639).

While CR Condition may indeed have triggered more conscious matching between presented stimuli and display rules, it is not clear whether participants in OR Condition did not categorise the stimuli mentally, without being prompted to do so. One of our post-questionnaire items aimed to answer this question.

5.4.8 Post-Questionnaire Responses

The responses on the following post-questionaire questions have been investigated: 1) perceived main task difficulty, 2) motion sequence repetition after the stick-figure, 3) imagined motion simulation during the emotion categories rating task (described in Section 5.3.5), and, for the participants of OR Condition, 4) implicit categorisation of the motion sequences in the main experimental task.

According to analysis of variance for the first three post-questionnaire questions, the between-participant factor of experimental condition had significant effect on reported difficulty of the task ($F(1, 70)=6.064$, $p=.016$, $ges=.079$). CR Condition was perceived as more difficult (M=3.472, SD=0.810) than the OR Condition (M=2.972, SD=0.909). Most likely, the extra task of choosing the emotion category from the list influenced the rating of difficulty. Moreover, participant gender had significant effect on the amount of motion sequence mimicking the participants used during the main experimental task ($F(1, 70)=6.936$, $p=.01$, $ges=.090$). Male participants reported higher amount of mimicking (M=17.97%, SD=21.58) than female participants (M=7.08%, SD=12.23). Note that the reported values represent the percentage of the 150 trials

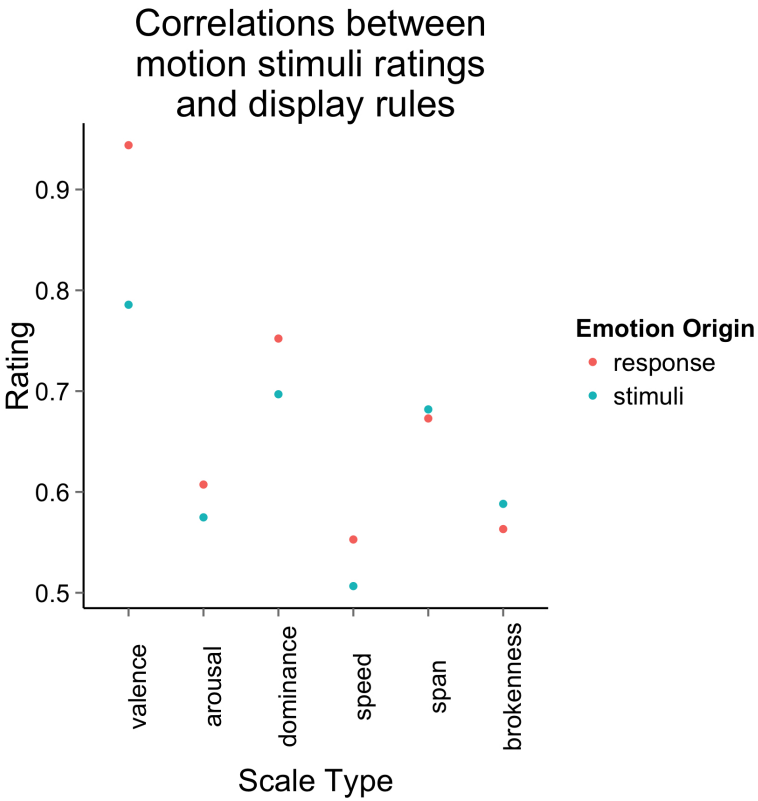

Figure 5.16: Correlation between motion stimuli rating and the display rules. For every trial in CR condition we performed correlation analysis between the values obtained in the motion sequence rating task and display rules ratings for a) original emotion category of the stimuli (blue), and b) response emotion category chosen by the participant (red)

and are thus rather low for both genders. Many participants reported that they repeated the motion after the stick figure whenever it was particularly hard for them to rate it. The histogram in Figure 5.17 (A) shows the distribution of the reported mimicked trials.

In the second part of the post-questionnaire the participants of OR Condition were asked if they still assigned emotional labels to

the motion sequences and if yes, how often (in percentage from the total amount of trials). Most of 36 participants reported to have given mental labels to motion sequences in 60% or more trials. The high rate of implicit labelling was unaffected by the between-participant factors of culture and gender. Figure 5.17 (B) shows that the majority of participants in the OR Condition have indeed categorised the stimuli.

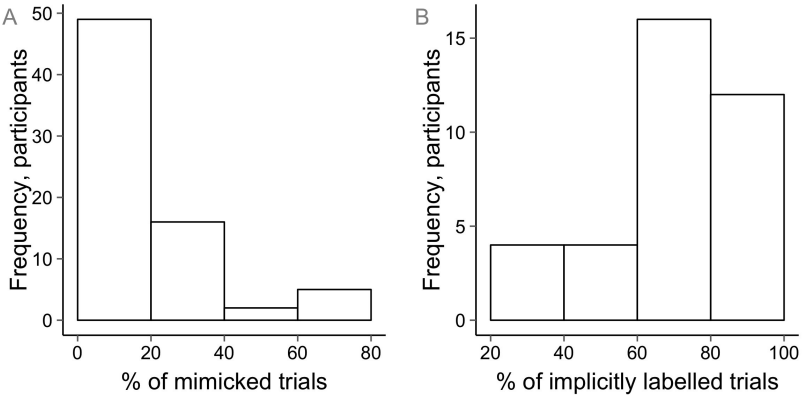

Figure 5.17: Post-questionnaire results: motion sequence mimicking and implicit labelling. (A) Distribution of reported proportion of trials where the participant mimicked the motion sequence shown in the trial (e.g., in order to assist the rating task). (B) Distribution of reported proportion of trials where the participant implicitly assigned an emotional label to the motion sequence.

5.5 Discussion

Previous research has investigated various aspects of cross-cultural differences in emotion perception. However, there are still many open questions, undoubtedly due to the complexity of the research topic in focus (Parkinson et al., 2004). Firstly, emotion perception is extremely versatile. It ranges from reflex reactions to threatening stimuli (Meeren et al., 2005) to aesthetic experience of art forms like music and literature (Robinson, 2007). It can focus on verbal aspects of human communication or nonverbal expression of emotion, the latter differentiating among facial expressions, emotional prosody and body language. Even

body language itself is a vast theme, encompassing such topics as iconic gestures (hand wave, fist clenching), postures (static full body expressions) and whole body motion or motion of its parts. In this paper we have used dynamic upper body expressions as the main study material. We investigated the connection between perception of emotional body expressions and cultural display rules across three cultures (English, German and Korean) using rating tasks of motion sequences and imagined emotional motion. Six five-point ratings scales were used - three affect scales (valence, arousal, dominance) and three motion scales, specially designed for this study (span, speed, brokenness). The scales were presented as rows of pictograms, giving the researches the advantage of avoiding excessive dependance on each culture's language.

The motion sequences stimuli were chosen from a big dataset and represented ten emotion categories: amusement, joy, pride, relief, surprise, anger, disgust, fear, sadness, shame. In one of the two experimental conditions the participants also had to categorise the motion sequences, which allowed us to measure recognition accuracy and category clustering patterns. Our participants were shown the same set of motion sequences, and the results show that they are perceived somewhat differently across different cultures. The differences can be traced back to the cultural display rules which we recoded in the end of each experiment session and which the participants were not aware of during the first part of the experiment. The fact that we still observe cultural differences in display rules diminishes the possibility of the motion sequences themselves having great impact on the display rules task, e.g., a situation where a person when asked to imagine a most typical instance of *joy* would imagine a motion sequence previously seen in the experiment instead of a motion pattern exemplary for their culture.

One of the major differences observed between Western (English and German) and Asian (Korean) cultures is how the scale of motion brokenness is utilised during the rating tasks. In English and German cultures this scale is associated with affect arousal, and similar results for single Western culture population were found by Dael et al. (2013). The higher perceived brokenness of the motion, the higher the arousal. In Korean culture lower brokenness, or higher fluidity of motion trajectory is associated with higher positiveness of the person who is expressing the emotion. For Western cultures higher valence is associated with higher span of motion, this motion scale shared with

the one of dominance. In Asian cultures the scale of dominance does not seem to be associated with any of the motion scales.

When analysing average scale ratings given to each emotion category, we observe statistically significant cross-cultural differences, mainly between the Korean and the two Western cultures. The dominance scale is the one where two culture types differ in both display rules and motion sequence ratings (seven out of ten emotions for motion sequence ratings and six emotions for display rules). Judging by the results of factor analysis, the scale of dominance is expressed, perceived and conceptualised very differently in Korean culture. Another important observation to make is that in most cases when a cross-cultural difference exists for a particular scale$\times$emotion combination, the mean rating obtained from Korean participants is closer to the neutral value of 3.0. When speaking about display rules, this observation confirms findings from previous research that people in Asian cultures like Korea, China, Japan, express emotions in a less extreme manner than people from individualistic cultures (Scherer et al., 1988).

According to our results, the emotion of *shame* is imagined and perceived differently between Korean and the Western cultures - in Korean culture its dominance (level of control over the situation) is higher, and the motion span is wider. Cross-cultural differences of conceptualisation and experience of shame have beed studied before (Fontaine, 2006; Liem, 1997; Stipek, 1998; Tracy and Matsumoto, 2008). The results show that in Asian countries, where the characteristic of a person being humble and respectful is very important, *shame* is shown more often and serves also as a social signal for appeasement of an unpleasant situation at an early stage. It is important to keep in mind however, that despite very close translations of emotion categories lists into German and Korean languages, the conceptualisation of emotional experience differs between languages and cultures (Kim, 1985; Bedford, 2004).

Interestingly, the valence of surprise is mildly negative in Korean culture, and mildly positive in English and German cultures, according to the display rules ratings. However, the motion ratings that expressed surprise were rated with mildly low valence in all cultures and in CR condition mis-categorised predominantly with other negative emotion categories like *anger*. Speaking about the overall recognition accuracy, we have found some evidence for in-group advantage for German culture specifically as the overall accuracy of German participants in CR Condition is higher by almost 10% on average than the accuracy of

other participants. These results are similar to in-group advantage that was previously shown only for emotional prosody (Scherer et al., 2001) and facial expressions (Elfenbein, 2013).

In the CR condition participants were asked to categorise the motion sequences along with rating them. The categorisation task was the sole difference between OR and CR conditions. One possibility is that categorisation task could urge participants to consciously match the perceived properties of the motion sequence with the display rules for the category they have chosen. In OR Condition we specifically avoided categorisation task to see if people can access affect properties of motion, but as the results of the post-questionnaire analysis show, for most trials participants still assigned internally some category. While such labelling undoubtedly takes place due to perceived motion features integration, it seems that participants first assign a label to the motion if they can find a suitable category and then analyse the motion sequence in relation to this category, adapting their ratings to this particular instance of emotional expression.

In summary, our findings support the hypothesis that cultural background influences immediate perception and interpretation of dynamic emotional body motion. This factor is important to take into consideration in scenarios where culture can play an important factor in human-human or human-computer interaction. Motion properties such as overall speed, span and brokenness are easy to access and obviously have important impact on perceived affective state of a speaker. These findings can prove useful when developing a virtual avatar for a video game, for successful collaboration in a multicultural environment and video material production for online education courses aimed at world-wide public.

Bibliography

Alaerts, K., Nackaerts, E., Meyns, P., Swinnen, S. P., and Wenderoth, N. (2011). Action and emotion recognition from point light displays: an investigation of gender differences. *PLoS One*, 6(6):e20989.

Arnold, M. B. (1960). *Emotion and personality.* Columbia University Press.

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

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., Young, A. W., and Others (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*, 33:717–746.

Aubrey, A. J., Marshall, D., Rosin, P. L., Vandeventer, J., Cunningham, D. W., and Wallraven, C. (2013). Cardiff Conversation Database (CCDb): A Database of Natural Dyadic Conversations. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*, pages 277–282.

Bänziger, T., Pirker, H., and Scherer, K. R. (2006). GEMEP-GEneva Multimodal Emotion Portrayals: A corpus for the study of multimodal emotional expressions. In *Proceedings of LREC*, pages 15–19.

Bänziger, T. and Scherer, K. (2007). Using Actor Portrayals to Systematically Study Multimodal Emotion Expression: The GEMEP Corpus. In Paiva, A. R., Prada, R., and Picard, R., editors, *Affective computing and intelligent interaction*, pages 476–487. Springer Berlin Heidelberg.

Banziger, T. and Scherer, K. R. (2010). Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) Corpus. In Scherer, K. R., Banziger, T., and Roesch, E. B., editors, *Blueprint for Affective Computing: A Sourcebook*, pages 271–294. Oxford, England: Oxford University Press.

Beck, A., Stevens, B., Bard, K. A., and Cañamero, L. (2012). Emotional body language displayed by artificial agents. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2(1):2–29.

Bedford, O. a. (2004). The Individual Experience of Guilt and Shame in Chinese Culture. *Culture & Psychology*, 10(1):29–52.

Boone, R. T. and Cunningham, J. G. (1998). Children's decoding of emotion in expressive body movement: the development of cue attunement. *Developmental psychology*, 34(5):1007–1016.

Bradley, M. M. and Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59.

Brooks, A., Schouten, B., Troje, N. F., Verfaillie, K., Blanke, O., and van der Zwan, R. (2008). Correlated changes in perceptions of the gender and orientation of ambiguous biological motion figures. *Current Biology*, 18(17):R728–R729.

Brownlow, S., Dixon, A. R., Egbert, C. A., and Radcliffe, R. D. (1997). Perception of movement and dancer characteristics from point-light displays of dance. *The Psychological Record*, (47):411–421.

Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., and Weiss, B. (2005). A database of German emotional speech. In *INTERSPEECH*, volume 5, pages 1517–1520.

Busso, C., Bulut, M., Lee, C.-C., Kazemzadeh, A., Mower, E., Kim, S., Chang, J. N., Lee, S., and Narayanan, S. S. (2008). IEMOCAP: Interactive emotional dyadic motion capture database. *Language Resources and Evaluation*, 42(4):335–359.

Busso, C. and Narayanan, S. S. (2008). Scripted dialogs versus improvisation: lessons learned about emotional elicitation techniques from the IEMOCAP database. In *INTERSPEECH*, pages 1670–1673.

Campione, E. and Véronis, J. (2002). A large-scale multilingual study of silent pause duration. *Speech Prosody*, pages 199–202.

Castellano, G., Villalba, S. D., and Camurri, A. (2007). Recognising human emotions from body movement and gesture dynamics. In *Affective computing and intelligent interaction*, pages 71–82. Springer.

Chentsova-Dutton, Y. E. and Tsai, J. L. (2010). Self-focused attention and emotional reactivity: the role of culture. *Journal of personality and social psychology*, 98(3):507–519.

Clarke, T. J., Bradshaw, M. F., Field, D. T., Hampson, S. E., and Rose, D. (2005). The perception of emotion from body movement in point-light displays of interpersonal dialogue. *Perception-London*, 34(10):1171–1180.

Clavel, C., Vasilescu, I., Devillers, L., Richard, G., Ehrette, T., and Sedogbo, C. (2006). The SAFE Corpus: illustrating extreme emotions in dynamic situations. In *The Workshop Programme Corpora for Research on Emotion and Affect Tuesday 23 rd May 2006*, pages 76–79.

Cowie, R., Douglas-Cowie, E., and Cox, C. (2005). Beyond emotion archetypes: Databases for emotion modelling using neural networks. *Neural networks*, 18(4):371–388.

Dael, N., Goudbeek, M., and Scherer, K. R. (2013). Perceived gesture dynamics in nonverbal expression of emotion. *Perception*, 42(6):642–657.

Darwin, C. (1872). *The expression of the emotions in man and animals*. London, UK: John Murray.

de Gelder, B. L. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3475–3484.

de Gelder, B. L., den Stock, J. V., Meeren, H. K., Sinke, C., Kret, M. E., and Tamietto, M. (2010). Standing up for the body. Recent progress in uncovering the networks involved in the perception of bodies and bodily expressions. *Neuroscience and Biobehavioral Reviews*, 34(4):513–527.

de Gelder, B. L., Snyder, J., Greve, D. N., Gerard, G., and Hadjikhani, N. (2004). Fear fosters flight: a mechanism for fear contagion when perceiving emotion expressed by a whole body. *Proceedings of the National Academy of Sciences of the United States of America*, 101(47):16701–16706.

De Meijer, M. (1989). The contribution of general features of body movement to the attribution of emotions. *Journal of nonverbal behavior*, 13(4):247–268.

den Stock, J. V. and Righart, R. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion*, 7(3):487–494.

Dittrich, W. H., Troscianko, T., Lea, S. E., and Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception-London*, 25(6):727–738.

Dodds, T. J., Mohler, B. J., and Bülthoff, H. H. (2011). Talk to the Virtual Hands: Self-Animated Avatars Improve Communication in Head-Mounted Display Virtual Environments. *PLoS One*, 6(10):e25759.

Eid, M. and Diener, E. (2001). Norms for experiencing emotions in different cultures: Inter- and intranational differences. *Journal of personality and social psychology*, 81(5):869–885.

Ekman, P. (1965). Differential communication of affect by head and body cues. *Journal of Personality and Social Psychology*, 2(5):726–735.

Ekman, P. (1971). Universals and cultural differences in facial expressions of emotion. In Cole, J., editor, *Nebraska symposium on motivation*, pages 207–283. Lincoln, NE: University of Nebraska Press.

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

Ekman, P. and Cordaro, D. (2011). What is Meant by Calling Emotions Basic. *Emotion Review*, 3(4):364–370.

Ekman, P. and Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists.

Elfenbein, H. A. (2013). Nonverbal Dialects and Accents in Facial Expressions of Emotion. *Emotion Review*, 5(1):90–96.

Elfenbein, H. A. and Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological bulletin*, 128(2):203–235.

Ennis, C. and Egges, A. (2012). Perception of Complex Emotional Body Language of a Virtual Character. In Kallmann, M. and Bekris, K., editors, *Motion in Games*, pages 112–121. Springer Berlin Heidelberg.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378–382.

Fontaine, J. R. J. (2006). Untying the Gordian Knot of Guilt and Shame: The Structure of Guilt and Shame Reactions Based on Situation and Person Variation in Belgium, Hungary, and Peru. *Journal of Cross-Cultural Psychology*, 37(3):273–292.

Fontaine, J. R. J., Scherer, K. R., and Roesch, E. B. (2007). The world of emotions is not two-dimensional. *Psychological*, 18(12):1050–1057.

Friesen, W. V. (1973). *Cultural differences in facial expressions in a social situation: An experimental test on the concept of display rules.* PhD thesis, ProQuest Information & Learning.

Gallo, L. C. and Metthews, K. A. (2003). Understanding the association between socioeconomic status and physical health: do negative emotions play a role? *Psychological bulletin*, 129(1):10–51.

Giese, M. A. and Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, 4(3):179–191.

Glikson, E. and Erez, M. (2013). Emotion Display Norms in Virtual Teams. *Journal of Personnel Psychology*, 12(1):22–32.

Glowinski, D., Dael, N., Camurri, A., Volpe, G., Mortillaro, M., and Scherer, K. R. (2011). Towards a Minimal Representation of Affective Gestures. *IEEE Transactions on Affective Computing*, 2(2):106–118.

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech*, 15(2):103–113.

Green, E. G. T. (2005). Variation of Individualism and Collectivism within and between 20 Countries: A Typological Analysis. *Journal of Cross-Cultural Psychology*, 36(3):321–339.

Gross, R. and Shi, J. (2001). The cmu motion of body (mobo) database. *Carnegie Mellon University*.

Gunes, H. and Piccardi, M. (2006). A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, pages 1148–1153. IEEE.

Harris, J. R. (1995). Where is the child's environment? a group socialization theory of development. *Psychological Review*, 102(3):458–489.

Hartmann, B., Mancini, M., and Pelachaud, C. (2006). Implementing expressive gesture synthesis for embodied conversational agents. In Gibet, S., Courty, N., and Kamp, J.-F., editors, *Gesture in Human-Computer Interaction and Simulation*, volume 3881 of *Lecture Notes in Computer Science*, pages 188–199. Springer Berlin Heidelberg.

Heberlein, A. S., Adolphs, R., Tranel, D., and Damasio, H. (2004). Cortical regions for judgments of emotions and personality traits from point-light walkers. *Journal of Cognitive Neuroscience*, 16(7):1143–1158.

Huis in 't Veld, E. M. J., Van Boxtel, G. J. M., and de Gelder, B. L. (2014). The Body Action Coding System I: Muscle activations during the perception and expression of emotion. *Social Neuroscience*, 9(3):249–264.

Hwang, B.-W., Kim, S., and Lee, S.-W. (2006). A full-body gesture database for automatic gesture recognition. *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14(2):201–211.

Kamachi, M., Lyons, M., and Gyoba, J. (1998). The japanese female facial expression (jaffe) database. *www.kasrl.org*.

Kaulard, K., Cunningham, D. W., Bülthoff, H. H., and Wallraven, C. (2012). The MPI Facial Expression Database — A Validated Database of Emotional and Conversational Facial Expressions. *PLoS One*, 7(3):e32321.

Keltner, D. and Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition & Emotion*, 13(1):505–521.

Kendall, M. G. and Smith, B. B. (1939). The Problem of m Rankings. *The Annals of Mathematical Statistics*, 10(3):275–287.

Kim, K.-h. (1985). Expression of Emotion by Americans and Koreans. *Korean Studies*, 9(1):38–56.

Kleinsmith, A. and Bianchi-Berthouze, N. (2013). Affective body expression perception and recognition: A survey. *Affective Computing, IEEE Transactions on*, 4(1):15–33.

Kleinsmith, A., Bianchi-Berthouze, N., and Steed, A. (2011). Automatic recognition of non-acted affective postures. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 41(4):1027–1038.

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

Kleinsmith, A. and Silva, P. D. (2005). Recognizing emotion from postures: Cross-cultural differences in user modeling. *User Modeling 2005*, pages 50–59.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., and Patras, I. (2012). DEAP: A Database for Emotion Analysis ;Using Physiological Signals. *Affective Computing, IEEE Transactions on*, 3(1):18–31.

Kret, M. E. and de Gelder, B. L. (2013). When a smile becomes a fist: the perception of facial and bodily expressions of emotion in violent offenders. *Experimental brain research*, 228(4):399–410.

Krüger, S., Sokolov, A. N., Enck, P., Krägeloh-Mann, I., and Pavlova, M. A. (2013). Emotion through Locomotion: Gender Impact. *PLoS One*, 8(11):e81716.

Kudoh, T. and Matsumoto, D. (1985). Cross-cultural examination of the semantic dimensions of body postures. *Journal of personality and social psychology*, 48(6):1440–1446.

LaBar, K. and Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, 7(1):54–64.

Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: computer applications. In Sidowski, J. B., Johnson, J. H., and Williams, T. A., editors, *Technology in mental health care delivery systems*, pages 119–l37. Ablex, Norwood, NJ.

Lazarus, R. S., Averill, J. R., and Opton, E. M. (1970). Towards a cognitive theory of emotion. *Feelings and emotions*, pages 207–232.

Levine, S., Theobalt, C., and Koltun, V. (2009). Real-time prosody-driven synthesis of body language. *ACM Trans. Graph.*, 28(5):172:1–172:10.

Liem, R. (1997). Shame and Guilt among First-And Second-Generation Asian Americans and European Americans. *Journal of Cross-Cultural Psychology*, 28(4):365–392.

Loula, F., Prasad, S., Harber, K., and Shiffrar, M. (2005). Recognizing people from their movement. *Journal of experimental psychology. Human perception and performance*, 31(1):210–220.

Lu, L. and Yang, K.-S. (2006). Emergence and composition of the traditional-modern bicultural self of people in contemporary Taiwanese societies. *Asian Journal of Social Psychology*, 9(3):167–175.

Lyons, M. J., Campbell, R., Plante, A., Coleman, M., Kamachi, M., and Akamatsu, S. (2000). The Noh mask effect: vertical viewpoint dependence of facial expression perception. *Proceedings of the Royal Society of London, Series B*, 267:2239–2245.

Ma, Y., Paterson, H. M., and Pollick, F. E. (2006). A motion capture library for the study of identity, gender, and emotion perception from biological motion. *Behavior research methods*, 38(1):134–141.

Magnée, M. J., Stekelenburg, J. J., Kemner, C., and de Gelder, B. L. (2007). Similar facial electromyographic responses to faces, voices, and body expressions. *Neuroreport*, 18(4):369–372.

Manera, V., Schouten, B., Becchio, C., Bara, B. G., and Verfaillie, K. (2010). Inferring intentions from biological motion: A stimulus set of point-light communicative interactions. *Behavior research methods*, 42(1):168–178.

Matsumoto, D. (2006). Culture and nonverbal behavior. *Handbook of nonverbal communication*, pages 219–235.

Matsumoto, D. and Hwang, H. S. (2012). Culture and Emotion The Integration of Biological and Cultural Contributions. *Journal of Cross-Cultural Psychology*, 43(1):91–118.

Matsumoto, D. and Kudoh, T. (1987). Cultural similarities and differences in the semantic dimensions of body postures. *Journal of nonverbal behavior*, 11(3):166–179.

McDonnell, R., Jörg, S., McHugh, J., Newell, F. N., and O'Sullivan, C. (2009). Investigating the role of body shape on the perception of emotion. *ACM Transactions on Applied Perception (TAP)*, 6(3):14:1–14:11.

Meeren, H. K. M., Van Heijnsbergen, C. C. R. J., and de Gelder, B. L. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences of the United States of America*, 102(45):16518.

Mesquita, B. and Frijda, N. H. (1992). Cultural variations in emotions: a review. *Psychological bulletin*, 112(2):179.

Metallinou, A., Lee, C.-C., Busso, C., Carnicke, S., and Narayanan, S. (2010). The USC CreativeIT database: a multimodal database of theatrical improvisation. In *Workshop on Multimodal Corpora, LREC*.

Mohammad, S. M. (2012). From once upon a time to happily ever after: Tracking emotions in mail and books. *Decision Support Systems*, 53(4):730–741.

Nackaerts, E., Wagemans, J., Helsen, W., Swinnen, S. P., Wenderoth, N., and Alaerts, K. (2012). Recognizing biological motion and emotions from point-light displays in autism spectrum disorders. *PLoS One*, 7(9):e44473.

Oertel, C., Cummins, F., Edlund, J., Wagner, P., and Campbell, N. (2013). D64: a corpus of richly recorded conversational interaction. *Journal on Multimodal User Interfaces*, 7(1-2):19–28.

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition*, 14(1):30 – 80. Neurobiology of Animal Consciousness.

Pantic, M., Valstar, M., Rademaker, R., and Maat, L. (2005). Web-based database for facial expression analysis. In *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pages 317–321.

Parkinson, B. (2013). Contextualizing Facial Activity. *Emotion Review*, 5(1):97–103.

Parkinson, B., Fischer, A. H., and Manstead, A. S. (2004). *Emotion in social relations: Cultural, group, and interpersonal processes*. Psychology Press.

Parkinson, B., Phiri, N., and Simons, G. (2012). Bursting with anxiety: Adult social referencing in an interpersonal Balloon Analogue Risk Task (BART). *Emotion*, 12(4):817–826.

Pavlova, M. A. (2012). Biological motion processing as a hallmark of social cognition. *Cerebral Cortex*, 22(5):981–995.

Pichon, S., de Gelder, B. L., and Grezes, J. (2008). Emotional modulation of visual and motor areas by dynamic body expressions of anger. *Social Neuroscience*.

Pichon, S., de Gelder, B. L., and Grezes, J. (2009). Two different faces of threat. Comparing the neural systems for recognizing fear and anger in dynamic body expressions. *Neuroimage*, 47(4):1873–1883.

Pierce, J. W. (2007). PsychoPy — Psychophysics software in Python. *Journal of Neuroscience Methods*, 163(1):8–13.

Pollick, F. E., Fidopiastisô, C., and Braden, V. (2001a). Recognising the style of spatially exaggerated tennis serves. *Perception*, 30:323–338.

Pollick, F. E., Kay, J. W., Heim, K., and Stringer, R. (2005). Gender recognition from point-light walkers. *Journal of experimental psychology. Human perception and performance*, 31(6):1247–1265.

Pollick, F. E., Paterson, H. M., Bruderlin, A., and Sanford, A. J. (2001b). Perceiving affect from arm movement. *Cognition*, 82(2):B51–B61.

Qu, C., Brinkman, W.-P. P., Ling, Y., Wiggers, P., and Heynderickx, I. (2013). Human perception of a conversational virtual human: an empirical study on the effect of emotion and culture. *Virtual Reality*, 17(4):307–321.

Quiros-Ramirez, M. A. and Onisawa, T. (2013). Considering cross-cultural context in the automatic recognition of emotions. *International Journal of Machine Learning and Cybernetics*.

Robinson, J. (2007). *Deeper Than Reason: Emotion and its Role in Literature, Music, and Art*. Clarendon Press.

Roetenberg, D., Luinge, H., and Slycke, P. (2009). Xsens MVN: full 6DOF human motion tracking using miniature inertial sensors. Technical report, Enschede, the Netherlands.

Roether, C., Omlor, L., and Giese, M. (2010). Features in the Recognition of Emotions from Dynamic Bodily Expression. In Ilg, U. J. and Masson, G. S., editors, *Dynamics of Visual Motion Processing*, pages 313–340. Springer US.

Russell, J. A. and Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3):273–294.

Salovey, P., Rothman, A., Detweiler, J., and Steward, W. (2000). Emotional states and physical health. *American psychologist*, 55(1):110–121.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. *Appraisal processes in emotion: Theory, methods, research*, 92:120.

Scherer, K. R., Banse, R., and Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, 32(1):76–92.

Scherer, K. R., Matsumoto, D., Wallbott, H. G., and Kudoh, T. (1988). Emotional experience in cultural context: a comparison between Europe, Japan, and the US. *Facets of emotions*, pages 5–30.

Sneddon, I., McRorie, M., McKeown, G., and Hanratty, J. (2012). The belfast induced natural emotion database. *Affective Computing, IEEE Transactions on*, 3(1):32–41.

Sogon, S. and Masutani, M. (1989). Identification of emotion from body movements: A cross-cultural study of americans and japanese. *Psychological Reports*, 65(1):35–46.

Sokolov, A. A., Krüger, S., Enck, P., Krägeloh-Mann, I., and Pavlova, M. A. (2011). Gender affects body language reading. *Front Psychol*.

Stipek, D. (1998). Differences between Americans and Chinese in the circumstances evoking pride, shame, and guilt. *Journal of Cross-Cultural Psychology*, 29(5):616–629.

Tamietto, M. and De Gelder, B. (2010). Neural bases of the non-conscious perception of emotional signals. *Nature Reviews Neuroscience*, 11(10):697–709.

Tracy, J. L. and Matsumoto, D. (2008). The spontaneous expression of pride and shame: evidence for biologically innate nonverbal displays. *Proceedings of the National Academy of Sciences of the United States of America*, 105(33):11655–11660.

Troje, N. F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of vision*, 2(5):371–387.

Tsai, J. L., Knutson, B., and Fung, H. H. (2006). Cultural variation in affect valuation. *Journal of personality and social psychology*, 90(2):288–307.

Visch, V. T., Goudbeek, M. B., and Mortillaro, M. (2014). Robust anger: Recognition of deteriorated dynamic bodily emotion expressions. *Cognition & Emotion*, 28(5):936–946.

Volkova, E., Mohler, B. J., de la Rosa, S., and Bülthoff, H. H. (2014a). The MPI Database of Emotional Body Expressions Common for Narrative Scenarios. *submitted*.

Volkova, E., Mohler, B. J., Dodds, T. J., Tesch, J., and Bülthoff, H. H. (2014b). Emotion Categorisation of Body Expressions in Narrative Scenarios. *Frontiers in Psychology*, 5(623).

Volkova, E., Mohler, B. J., Parkinson, B., Wildgruber, D., Bülthoff, H. H., and de la Rosa, S. (2014c). Cross-cultural Differences in Perception of Dynamic Emotional Body Expressions. *in preparation*.

Walk, R. D. and Homan, C. P. (1984). Emotion and dance in dynamic light displays. *Bulletin of the psychonomic society*, 22(5):437–440.

Wallbott, H. G. (1985). Hand movement quality: a neglected aspect of nonverbal behavior in clinical judgment and person perception. *Journal of Clinical Psychology*, 41(3):345–359.

Wallbott, H. G. (1998). Bodily expression of emotion. *European Journal of Social Psychology*, 28(6):879–896.

Winters, A. (2009). Perceptions of body posture and emotion: A question of methodology. *The New School Psychology Bulletin*.

Yin, L., Chen, X., Sun, Y., Worm, T., and Reale, M. (2008). A high-resolution 3D dynamic facial expression database. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pages 1–6. IEEE.

Zara, A., Maffiolo, V., Martin, J.-C., and Devillers, L. (2007). Collection and annotation of a corpus of human-human multimodal interactions: Emotion and others anthropomorphic characteristics. In *Affective computing and intelligent interaction*, pages 464–475. Springer.

Zborowski, M. (1969). *People in pain*. Jossey-Bass San Francisco.

Zellner, B. (1994). Pauses and the temporal structure of speech. *in E. Keller (Ed.) Fundamentals of speech synthesis and speech recognition*, pages 41–62.