AALTO UNIVERSITY
SCHOOL OF SCIENCE AND TECHNOLOGY
Faculty of Electronics, Communication and Automation
Department of Media Technology

Timo Idänheimo

Recognizing emotions from biological motion in a point-light display

Master's Thesis

In Espoo, May 10th 2010.

Supervisor: Professor Tapio Takala
Instructor: Professor Tapio Takala

AALTO UNIVERSITY
School of Science and Technology **ABSTRACT OF MASTER'S THESIS**

Author and Name of the Thesis:
Timo Idänheimo
Recognizing emotions from biological motion in a point-light display

Date: **10.5.2010** Number of pages: **72+9**

Faculty, department and professorship:
Faculty of Electronics, Communications and Automation
Department of Media Technology
AS-75 Media Technology

Supervisor and Instructor:
Professor Tapio Takala

The objective of this master's thesis was to examine the recognition of emotion from motion captured movement. Our goal was to research how accurately humans can recognize different emotions from biological motion. We also wanted to know how much the rate is decreased when the motion is turned upside down.

We used optical motion capture system to record movements from our actors and to separate the pure motion data from image data. With this method we got so called point-light display stimuli, where white dots are moving against dark background. The white dots were attached to our actor's main joints. For creating the movements we used two amateur actors, male and female, who acted a simple scene of knocking on the door with ten different emotions. These emotions were afraid, angry, excited, happy, neutral, relaxed, sad, strong, tired and weak. After the capture, we showed these displays as a stimulus to 35 test subjects and asked them to describe the emotions or the overall feeling of the actor with one of the ten pre-given adjectives. Purpose was to test how accurately the acted emotions could be recognized in the point-light display.

The results indicated that our test subjects could recognize emotions with varying rate from a point-light display. The recognition rate was average 20 % for normal oriented displays and on average 19 % for upside down displays. For a single emotion, we got 11 % to 44 % recognition rate (tired 44 %, afraid 11 %). We also discovered that the emotions were confused to another which was expected according to the theory. Our results also indicated that our test subjects experienced a slight learning during the experiment. The recognition rate was better in the latter half of our experiment.

Keywords: Emotions, motion capture, biological motion, pattern recognition, point-light display

AALTO YLIOPISTO

Teknillinen korkeakoulu **DIPLOMITYÖN TIIVISTELMÄ**

Tekijä ja työn nimi:

Timo Idänheimo

Emootioiden tunnistaminen biologisesta liikkeestä valopiste-esityksessä

Päivämäärä: **10.5.2010** Sivumäärä: **72+9**

Tiedekunta, osasto, professuuri:

Elektroniikan, tietoliikenteen ja automaation tiedekunta

Mediatekniikan laitos

AS-75 Viestintätekniikka

Työn ohjaaja ja valvoja:

Professori Tapio Takala

Tämä diplomityö käsittelee emootioiden tunnistamista tallennetusta liikkeestä. Tarkoituksenamme on tutkia kuinka hyvin ihmiset pystyvät tunnistamaan eri emootiot ainoastaan biologisen liikkeen perusteella. Halusimme tutkia myös kuinka paljon tunnistustarkkuus muuttuu, oletetusti pienenee, jos liikkeen kääntää ylösalaisin.

Tutkimuksessamme käytimme optista liikkeenkaappausjärjestelmää, minkä avulla pystyimme erottelemaan näyttelijöiden liikedatan kuvadatasta. Näin saadaan aikaiseksi ns. valopiste-esityksiä missä valkoiset pisteet liikkuvat mustalla taustalla. Valkoiset pisteet olivat sijoiteltu tallennuksen aikana näyttelijämme nivelten kohdalle. Esitysten luomiseen käytimme kahta harrastelija näyttelijää, miestä ja naista, jotka esittivät oveen koputus –liikkeen kymmenellä eri emootioilla. Nämä emootiot olivat pelokas, vihainen, innostunut, iloinen, neutraali, rentoutunut, surullinen, vahva, väsynyt ja heikko. Tämän jälkeen näytimme näitä valopiste-esityksiä 35 koehenkilölle joiden tarkoituksena oli yrittää tunnistaa näyttelijän emootio tai kuvailla sitä yhdellä valmiiksi annetuista adjektiiveista.

Tuloksista pystyimme toteamaan, että koehenkilömme tunnistivat vaihtelevalla tarkkuudella emootiot näyttelijöidemme valopiste-esityksistä. Tunnistustarkkuus oli keskimäärin 20 % oikeinpäin olleille esityksille ja noin 19 % väärinpäin olleille esityksille. Yksittäisille emootioille saimme noin 11–44 %:n tunnistustarkkuuden (väsynyt 44 %, pelokas 11 %). Lisäksi totesimme useiden emootioiden sekoittuneen keskenään, mikä oli odotettua teorian perusteella. Tuloksemme osoittivat myös koehenkilöille pientä oppimista emootioiden tunnistamisessa kokeen aikana, esitysten tunnistamistarkkuuden ollessa heikompaa alkupään herätteissä kuin loppupään.

Avainsanat: Emootiot, liikekaappaus, biologinen liike, hahmon tunnistus, valopiste-esitys

TABLE OF CONTENT

FOREWORD

Finally it is done!

This thesis took a lot of commitment and effort. It was funded by Aalto University's key projects, aivoAALTO and made possible by Department of Media Technology in Aalto University. Therefore I would like to thank the whole aivoAALTO team for very mind-opening meetings and support to this thesis. Especially professor Riitta Hari, Niklas Ravaja and Elina Pihko.

My thesis would not have been possible without the staff of the Department of Media Technology. I also want to thank our wonderful research team, especially Klaus Lehtonen without whom this thesis would have been totally impossible. And also a big thanks to all who participated the experiment.

I can't imagine better instructor and supervisor to my thesis than professor Tapio "Tassu" Takala. I am very grateful to Tassu for giving me an opportunity to end my studies even with tight schedule, from challenging topic and without coding anything special. Without Tassu's inspiration I would still be writing.

Finally I would like to thank my family for all the support along my nearly twenty years of studies. It has been priceless. And Nina, thank you for believing me and supporting me even in rough times. You are the best!

And now, the last teekkariwappu! Skumppaa!

In Ullanlinnanmäki 1st of May 2010

ABBREVIATIONS

AC	Alternating Current
ANS	Autonomic Nervous System
CCD	Charge-Coupled Device
CGI	Computer-Generated Imagery
CNS	Central Nervous System
DC	Direct Current
EEG	Electroencephalography
EDA	Electrodermal Activity
ECG	Electrocardiography
EMG	Electromyography
fMRI	functional Magnetic Resonance Imaging
FPS	Frames Per Second
GSR	Galvanic Skin Responce
LED	Light-Emitting Diode
PL	Point-Light
PNS	Peripheral Nervous System

“Refuse to express a passion, and it dies”

- William James, 1890

1 INTRODUCTION

One of the most important skills that we humans have, is the ability to perceive other people's actions and intentions. When walking on street and confronted with someone you know, you can nearly immediately recognize the mood or the emotion that your friend has. You might not even have to confront the person. Humans can recognize fairly accurately others feeling from long distant by just looking at their gait or posture.

The old saying goes like "a picture is worth a thousand words". When we are trying to understand feeling, we could say that actions speak louder than pictures. Our vision provides us a very rich source of information for different social events.

The research of emotions is one of the most complex research areas needing expertise from psychology, physiology, sociology, biology, brain imagine, philosophies, computer science etc. If mathematics is one of the purest areas of science, emotions research is at the other end of that scale. And the complexity merely increases when we examine the social situation between two humans.

Emotions are not the easiest phenomenon to research. They are often disregarded because they are not considered as scientific or logical. In Star Trek, Mr. Spock often pointed out to Captain Kirk that logical thinking in more productive than wasting energy to reacting emotional to events. Mr. Spock also considered emotions to be illogical.

In over 100 years of emotional research, we have learnt a lot of our emotions. We know fairly well were emotions are created and how they are affecting our body. Emotions can be measured with different physiological and psychological measurement tools. Heart rate, skin conductance, muscle activation, respiration rate and brains electrical activity measures the changes in our nervous system.

Emotions have elicited many theories about how they are emerging. The oldest theories date back to the late 1800's. Regardless the numerous amounts of theories, the question of how emotions are really created, still remain unsolved. But first we will explain the motivation for this thesis.

1.1 Why is this research needed?

This master thesis is a part of aivoAALTO -research project which is one of the key projects in Aalto University. AivoAALTO project is a multidisciplinary project that has expertise from areas such as brain research, media technology, psychology, neuroeconomics etc. The goal is to find new ways to study human social interaction, decision-making and the effect of cinema on human mind.

This thesis describes a method, which can be used later in brain research of human behavior or in enactive media projects in Aalto University. Department of Media Technology is participating in aivoAALTO project with brand new motion capture studio. This thesis was also a good way to test this new studio properly and find out if it is suitable for scientific research.

To understand more thoroughly human mind, the brain must be stimulated with controlled stimuli. The point-light displays have been used for few decades to insulate the human motion information from the image information. The point-light displays use white dots against black background to present the biological motion. These white dots are placed into different joints on actor's body.

The possible further uses of this research data and point-light displays are presented at the end of this thesis in chapter 10.

1.2 The research problem

This master thesis has two main research problems:

1. *How accurately can humans recognize different emotions from biological motion which contains no social interaction?*
2. *Do the following attributes affect to the recognition rate?*
    a. *Orientation of the point-light display*
    b. *Gender of the actor*
    c. *Gender of the participant*
    d. *Learning through repetition*

We also address the question of which emotions can be recognized from these point-light displays. Our hypothesis and more precise handling of the research problems are presented more thoroughly in chapter 6.

1.3 The research method

This research combines many technical and psychological areas into one whole. Our goal is to get accurate measurement throughout the research.

We used optical motion capture system to capture male and female actor's movement and transform the data into point-light displays. The actors performed the same movement with 10 different emotions. The action recorded was knocking on the door –motion. These actions were then converted into point-light displays.

After that we conducted an interview with PHP-questionnaire form online, where we showed these point-light emotion displays to our test subjects. Half of the displays were upside down oriented, which is supposed to decrease the rate of recognition. With every point-light display stimulus, the test subject had to choose one emotion from a pre-given list which would reflect the emotion of the actor best.

Finally we analyzed the data that we had from the interviews and try to find answers to our research questions.

1.4 The structure of this thesis

This thesis describes how to study human emotions and mind.

First we start with different motion capture systems and their history in chapter 2. We discuss about how pure biological motion stimuli is created for emotional research experiments. We go through different applications were motion capture can be used. In chapter 2 we discuss briefly about facial recognition which is not the core of this thesis but important when it comes to human emotions.

After this we take a look of the history of emotional research and psychology of emotions. We present two of the most dominant theories that describe emotions and briefly discuss about the controversy of these theories in chapter 3. Chapter 4 discusses more about human physiology and how different parts of our body affect our emotions. In this chapter we discuss about different psychophysiological ways to measure emotional responses from human body. Chapter 4 may not be the core knowledge for this thesis but it gives important point of view of our behavior in various situations and gives little background of what actually effects to our movements.

Chapter 5 handles the actual recognition of emotions from different situations. Our research is largely based on these researches presented in this chapter. Chapters 6 and 7 describe our test setup and hypothesis of the outcome of this research. After this we present the results and analysis and reflect these results to previous studies in chapter 5. Finally, chapters 9 and 10, we make conclusions and discuss about the possible future studies.

1.5 Terminology

There are many terms that describe human feelings, emotions and moods. All these adjectives have a slight different interpretation when it comes to scientific research. Most of the past researches use term emotion or affect. Emotion refers more often to one specific feeling whereas affect means a change. This is why we will be using the term *emotion* throughout this thesis.

2 MOTION CAPTURE

Motion capture is a process in which a live motion is recorded and transformed onto a digital model. Motion capture uses many mathematical terms to track a number of key points in predetermined space (Menache, 2000). From this tracking data, we can obtain a single seamless, three-dimensional representation of the motion that can be further transferred for example to an animated character.

The location in the space can be monitored with sensors or markers that are attached to actor's joints (Menache, 2000). The actor is placed in the middle of many cameras that usually work with infrared light. The movement of one or more actors is sampled many times in a second. Motion capture allows placing a performer to the scene that in other ways could not be possible or might be too dangerous to do in real life (Trager, 1999).

Motion capture has many useful applications for many kinds of users. In this chapter we shall explain about different types of motion capture systems. We also present a few applications in which motion capture can be used.

2.1 History of motion capture

History of motion capture can be placed in the late 1800s, when to scientist Étienne-Jules Marey and Eadweard Muybridge conducted independent studies of human and animal motion. They both used multiple photographs of moving objects to present the biological movement (Menache, 2000). These studies have had a major impact in biology, medicine, photography and nowadays also in animation (Menache, 2000).

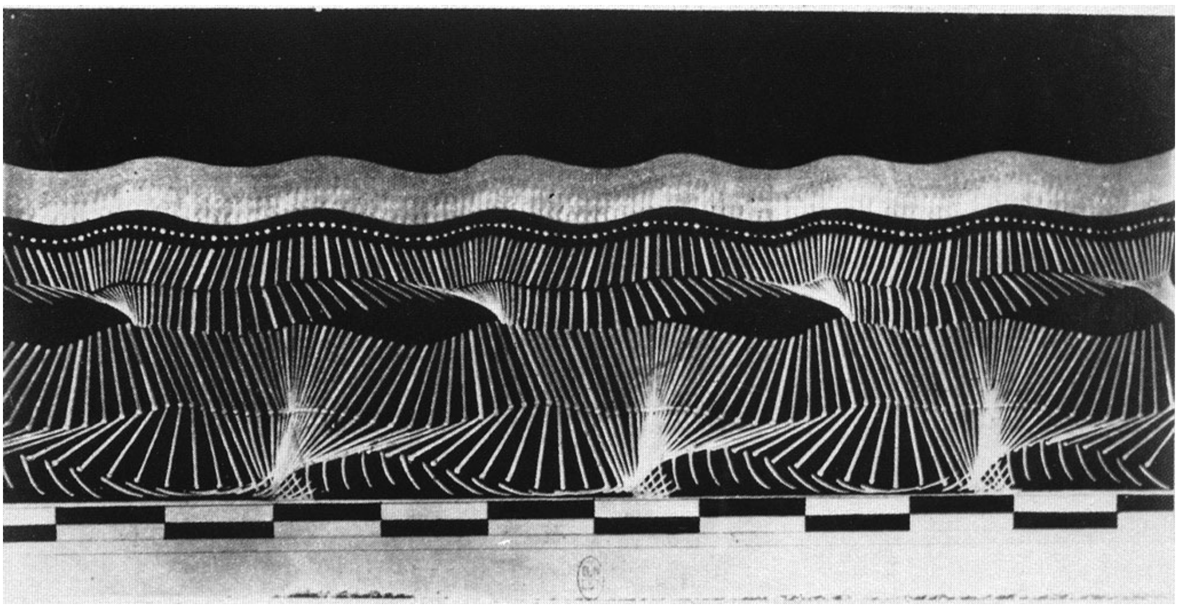


Figure 1. Marey's chronophotography with man walking

Marey was a French physiologist and physician, who had a passion for understanding biological movement (Blake & Shiffrar, 2007). This passion led to a development of "chronophotography" which is a high-speed technique for taking multiple photographs on a single photographic plate (Blake & Shiffrar, 2007). Marey made several little improvements to his methods that are still used today. Marey already used black suits and small markers on his actors to highlight joints and the kinematics of gait.

Muybridge was an English photographer, who can be considered as the inventor of motion pictures and a photographic wizard (Leslie, 2001). He was best known from his work with multiple cameras to capture motion, especially the galloping horse, and his zoopraxiscope. This spinning "film projector" used a large glass disk as a film, in which the figures were running around the edge. It required an artist to draw every picture on the glass (Leslie, 2001). With this technique, Muybridge showed the world's first movie in the fall of 1879 (Leslie, 2001). Afterwards Muybridge used the device to collect money by showing simple movies to paying customers.

Figure 2. The zoopraxiscope by Muybridge

The galloping horse was a medical study and a bet to prove that horse's all four hooves left ground simultaneously (Menache, 2000). Muybridge used 24 cameras to photograph the horse in a gallop. According to Menache (2000), Muybridge and Marey knew each other and others work but they followed a different path. Where Muybridge used multiple cameras to capture the movement, Marey used only one camera. Muybridge's work led to more entertaining fields of business, whereas Marey's work had a bigger effect in medicine and later in sports (Menache, 2000).

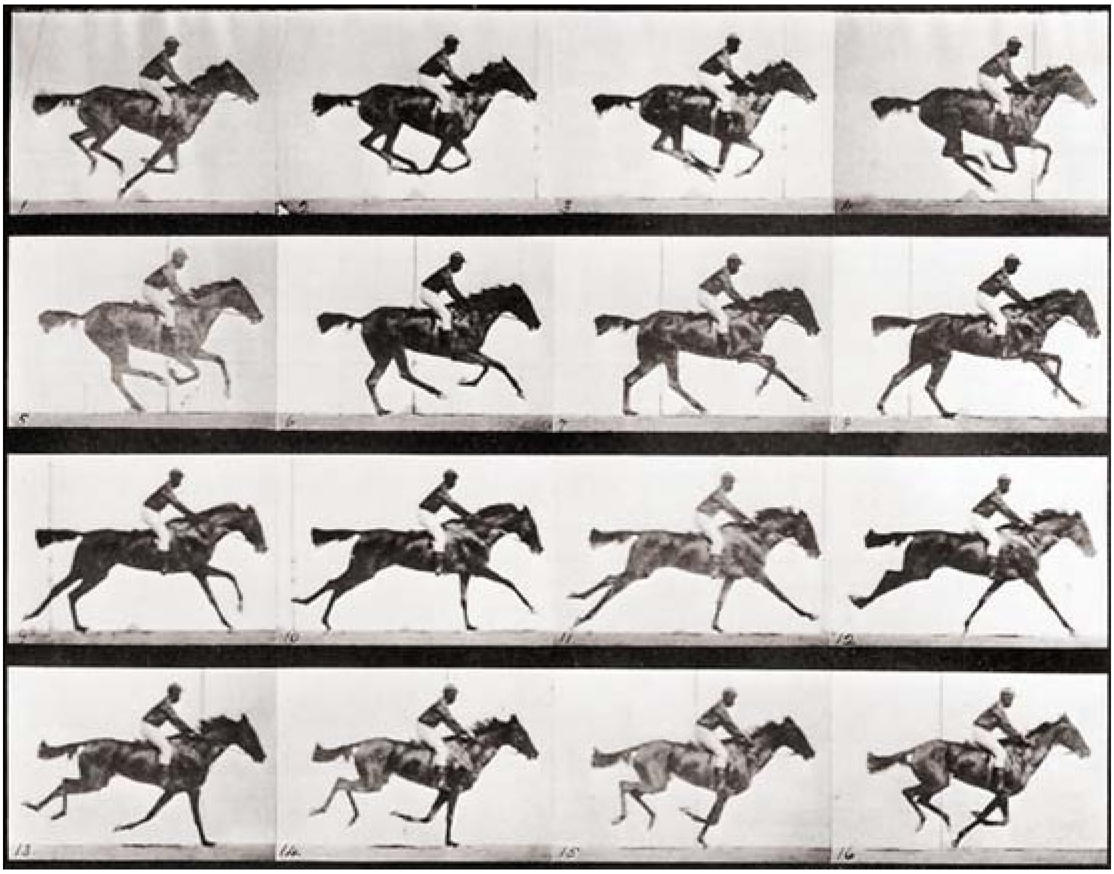


Figure 3. Galloping horse

For nearly a century, this field of research yielded almost no benefits, until in the 1970's Swedish psychologist named Gunnar Johansson (1973) introduced an optical motion capture system (Menache, 2000; Blake & Shiffrar, 2007). The technique was better known as point-light animation of biological motion (Johansson, 1973). According to Johansson (1973), to analyze a motion in physics, it is well known that you will need the concept of particle. The use of bright spots moving against a homogeneous contrasting background allowed Johansson to conduct analysis of human motion patterns without the interference of pictorial information. This method was also used by Marey about a century earlier.

Today, motion analysis and animation plays a significant role in advertising, entertainment, education and many other scientific researches (Amaya et al. 1996). New refined methods of point-light animation come out every year through the use of computer animation and combining motion capture to animation software (Blake and Shiffrar, 2007).

2.2 Requirements and types of motion capture

There are restrictions and requirements that apply to all motion capture systems. Trager (1999) notes that motion capture system must be easy to set up in every location and it must be quick to calibrate for different performers. It must also allow the performers to do the act as unrestricted as possible. No excess cabling is allowed from performer to a computer or screen. No restriction to extending and rotating joint is allowed during the performance. Motion capture must be as accurate as possible to represent the orientations of the joints and bone structure. It also has to provide good data for the gross body motion alongside with detailed motion capture data (Trager, 1999). Finally, it has to have an easy way to convert the motion capture data into a simple animated figure with simple skeleton system.

For tackling different problems of motion capture, there are many different ways to do motion capture. Menache (2000) categorizes motion capture systems into three groups based on where the capture sources and sensors are placed. The three groups are by Menache 2000:

- An outside-in system, which is the very basic human motion capture system that uses cameras as external sensors to collect data from the reflective markers (the source) placed on actor's body. Optical motion capture systems are in this category.
- Inside-out system in which the sensors are placed directly on actor's body. For example electromagnetic motion capture system in which sensors move in an externally generated electromagnetic field.
- Inside-in systems have sensors and sources placed on the actor's body. Electromechanical suits and exoskeleton systems have potentiometers or goniometers[1] as sensors and the actual joints inside the body as sources.

2.2.1 Optical motion capture system

Optical motion capture systems have become the most popular systems over the last decades. It is a very accurate method for capturing certain types of motions. The system offers the actor the most freedom during the act because it does not require any cables. On the other hand, optical motion capture system's data requires quite extensive post-processing (Menache, 2000; Trager, 1999).

Optical system requires at least three CCD (charge-couple device) cameras and a single computer that controls the inputs. CCD has an array of light-sensitive photoelectric cell, also known as a pixel, which captures light from different image sources (Menache, 2000). Cameras use typically infrared as a light source which illuminates the field of view. Infrared light is used because it creates less visual distortion for the performer (Menache, 2000).

[1] A goniometer in an instrument that either measures angle or allows an object to be rotated to a precise angular position.

Sampling rate is getting better and better when new equipment and faster data processing are being developed. Now a days, state-of-the-art motions capture camera can go up to 2000 frames per second with resolution of 0.3 megapixels. Cameras have also their own preprocessing module that deceases the post-processing time a lot (NaturalPoint, 2010).

Motion capture couldn't be possible if the performer wouldn't have any fixed points that the cameras could follow. The performer is usually equipped with black sticker suit which can be incorporate with directionally reflective balls, also known as markers (Trager, 1999; Vlasic, 2007). The size of the marker spheres can vary from a couple of millimeters to a few centimeters (Menache, 2000).

For reliable data acquisition, the optical motion capture system must have at least three cameras pointed to the performer and the markers. There can be up to 32 cameras in one motion capture system (Menache, 2000). Computer calculates from the camera data a 3D position for every marker (Trager, 1999; Vlasic, 2007). To be able to calculate the positions, at least two cameras must track a single point all the time. Extra cameras are needed for maintaining a direct line of sight (Menache, 2000). Adding more cameras makes tracking easier, but it also increases post-processing time when each marker becomes more complex.

This two camera tracking method is very vulnerable for errors in the data stream. If one or more markers are occluded or hidden in some point, the optical system can't follow the marker and it creates a hole into the stream (Trager, 1999). This can be avoided by adding more cameras. With optical systems, one must constantly balance with accuracy of data (increasing with increase of cameras) and the power available for the post-processing (Menache, 2000).

The biggest problem with the optical systems is the occlusion of the markers. If cameras cannot see the markers, the data stream will be corrupted. Second problem is the already mentioned extensive post-processing needed. Optical motion capture is also quite location sensitive meaning that the equipment cannot be transferred from place to place easily (Vlasic, 2007). Environment must be very controlled with no reflective noise around. Also the right lighting conditions are important for successful capture (Menache, 2000).

On the other hand, the data is most often extremely accurate (Vlasic, 2007). This is because the performer can be equipped with a large number of markers. It is also very easy and fast to change the configuration of the markers if for example only an arm movement is wanted to track. The optical system also allows the performer a large performance area with no cables attached directly to the performer (Menache, 2000).

The optical motion capture system is widely favored in the computer animation and film industry communities (Vlasic, 2007). In this thesis, we use optical motion capture system manufactured by OptiTrack. This system is equipped with 12 infrared cameras. More about this motion capture system's specifications and details in chapter 7.3.1.

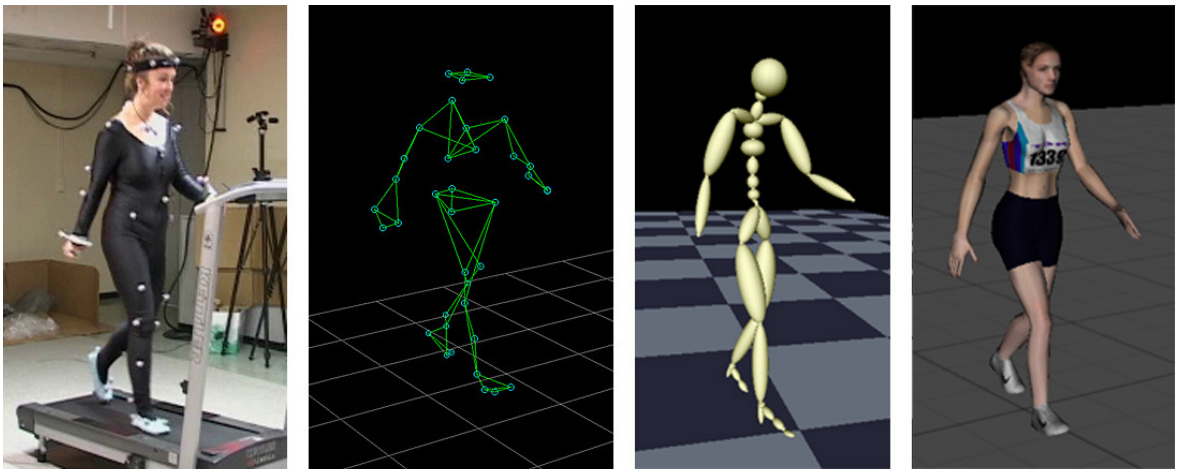


Figure 4. The optical motion capture system

2.2.2 Electromagnetic motion capture system

Electromagnetic motion capture systems were originally invented for military purposes. The system was first used in aircrafts for helmet-mounted displays in which the pilot could acquire a target by just locating the target visually (Menache, 2000). The system, also known as magnetic tracker, is very much used now a days for performance capture (Trager, 1999). It captures the motion from relative magnetic flux of transmitter's and receiver's coils.

Magnetic motion capture utilizes to "whole body" -capture typically from 11 to 18 sensors strapped on to various parts of performer's body. The number of sensors needed is dependent a lot of the usage of the motion capture. These sensors act as receivers for the capture and they are connected to an electronic control unit by individual cables. Transmitters are located nearby so that the receivers can measure their spatial relationship to the transmitters (Menache, 2000; Trager, 1999).

In practice the system works by the transmitters generating a low frequency electromagnetic field that the receivers detect. After the detection, the data is sent to electronic control unit for filtering and amplification (Menache, 2000). The final data stream holds the 3D positions and orientations for each receiver (Trager, 1999).

The electromagnetic motion capture system betters from the optical motion capture system in few ways. It doesn't have the occlusion problem for the markers. Furthermore it doesn't need as much post-processing as the optical system does. This is partly because the magnetic system uses fewer markers than optical system. But the configuration of the markers however is not as simple to change as in optical system (Menache, 2000).

The biggest problem is with the usage of magnetic field and electric current in a form of metal conductivity (Vlasic, 2007). If the system is using direct current (DC), tracking will be very sensitive for disturbances if there are ferrous metals, like iron or steal, nearby. On the other hand, alternating current (AC) tracker isn't so sensitive for ferrous metals but it is very sensitive to aluminum, copper and carbon steel. Some systems have algorithms to compensate these interferes but this problem limits the tracking a lot in different locations and stages (Menache, 2000). It also has a very high power consumption which decreases its portability (Vlasic, 2007).

Magnetic trackers are used in real-time purposes in the entertainment industry. It's been used for live television and performance. In these performances, the actor is constrained by cables and smaller capture area than in optical systems (Menache, 2000)

2.2.3 Electromechanical motion capture system

Electromechanical motion capture system is one of the earliest methods in capturing human motion straight from the performer's movement (Trager, 1999; Vlasic, 2007). This suit is a structure of potentiometers and goniometers which measure angular and rotational change from the major joint locations in human body (Menache, 2000; Vlasic, 2007). Potentiometers are used in electronics industry when variable voltage change is needed for example volume control in radios (Menache, 2000). This method can also be called as prosthetic motion capture or exo-skeleton motion capture because of the skeleton look-a-like suit attached to the performer.

Figure 5. Electromechanical motion capture system

The biggest drawback of this system is the assumption of human joint characteristics. These electromechanical systems assume that human bones are connected by simple hinges (Menache, 2000; Trager, 1999) and the system doesn't take account the rotational movement of joints. It also affects to the positions of measuring sensors. The configuration of the sensors cannot be changed very easily without a major modification to the suit (Menache, 2000).

This method is fairly cheap way to capture human motion. During the performance, the sensors are never occluded but the suit can be very uncomfortable due to the amount of hardware needed (Menache, 2000). Data coming out from sensors is simple so there is no need of extensive post processing. Suit is also very portable, so it can be moved from location to location.

2.2.4 Acoustic motion capture system

Acoustic motion capture system uses the time of flight of an audio signal to track the whereabouts of transmitters strapped to various parts of the performers body (Vlasic et al., 2007). The system uses a triad of audio receivers because the distance calculated from the performer by receivers is triangulated to provide a point to the 3D space (Trager, 1999).

Most of these systems use ultrasonic pulse emitters to trigger the measurement. By using ultrasonic sound for the measurement, the performer isn't disturbed by the sound (Vlasic et al. 2007). With a spread-spectrum ultrasonic sound the acoustic motion capture performs very well because of the lack of occlusion problem which is bigger problem with optical motion capture systems (chapter 2.2.1) (Vlasic, 2007; Trager, 1999). The use of sound can in some situations also cause different and unexpected sound reflections from walls and floor that affect the accuracy of the measurement (Trager, 1999).

The acoustic system is not without its own problems. System is equipped with cables that can cause problems to various types of performances. Most current acoustic systems are not portable and they can only handle a small number of markers attached to the performers body (Vlasic, 2007). This becomes a problem if the personality of the performance is important (Trager, 1999). Another problem is with the capture area size. The size is limited by the speed of sound in air and the number of transmitters (Trager, 1999).

2.3 Applications of motion capture

Motion capture equipment has been developed primarily for entertainment purposes. Other major markets are medical industry, sport and even law (Menache, 2000). The equipment can also be used for research and development, for example, ergonomic design or for safety tests to automobiles with crash test dummies (Menache, 2000).

The first application for motion capture was gait analysis as earlier stated. Etienne and Muybridge studied animal motion in series of photographs taken. This yielded after about a decade later to a medical science where motion capture data could be used for creating biomechanical data. Data could be used for analyzing gait and in orthopedic area for recognizing joint movement and designing prosthetic joint and limbs (Menache, 2000). With motion capture, we can separate different little mechanisms from our phases of walk cycle.

Motion capture differs from normal videotaping in one crucial way. Motion capture can represent the movement in a three-dimensional space. This is why in medical sciences motion capture is often called three-dimensional biological measurement (Menache, 2000). Afterwards a specialist can examine gait from numerous different angles. With motion capture it is easier to detect abnormalities and determine a treatment. Motion capture and gait analysis is also used for rehabilitation of patients with joint problems (Menache, 2000).

The entertainment business is one of the biggest user and certainly the fastest growing market segment for motion capture (Menache, 2000). But it is the video gaming industry that uses motion capture the most widespread. The motion capture is the most well accepted and understood in gaming business (Menache, 2000) where as television and movie industries are still quite infants with this technology. Gaming industry was the first segment of entertainment to really use motion capture for character motion animation.

In animation, motion capture is widely used to animate changes to motion trajectories that must appear as natural as possibly (Zordan et al. 2002). Some movements are almost impossible to create looking like natural without motion capture or the motion will be perceived as flawed (Zordan et al. 2002).

One of the major breakthroughs in motion picture industry was the movie called Avatar by James Cameron in 2009 (Twentieth Century-Fox Film Corporation, 2009). In this movie, the director used a method called performance capture in which the performers' movement and facial expression were transformed into CGI (Computer-generated imagery) avatar character in real time (Sagar, 2006). Visual appearance was totally computer-generated but the result looked photorealistic. Before all this, motion capture was used only for adding digital extras and digital crowds. It has been used also for doing digital stunt actions where the action itself is too dangerous or impossible to perform by live human being or the stunt performer need to look the same as the star of the movie (Menache, 2000).

In sports, motion capture is being used to improve the performance of different athletes. Especially in golf, there are lots of professional studios and analysts willing to help improving the swing (Menache, 2000). The improvement of motion capture equipment has led to undisturbed data collection in sport fields that were impossible earlier, for example, in swimming. Another important advantage of motion capture against videotaping can be noticed in fast-pace sporting. Normal videotaping can only record 30 frames per second, time in which all the motions may not be fully captured (Menache, 2000). With motion capture we can record much higher frequencies (OptiTrak, 100 frames per second (NaturalPoint, 2010)), so more detailed motion can be captured.

Motion capture can also be used for reconstructive videos of event for example in crime investigation. Videos can be used to aid the jury to understand the situation in the scene (Menache, 2000). It can also be used as a substitute for photographs which can be too horrific to show the jury (BBC, 2010). In one of the most widely known murder trial, the O.J. Simpson's trial, motion capture was used to reconstruct the events of the murder scene. The data was not actually used for evidence but it showed what can be done with the motion capture (Menache, 2000).

2.4 Facial recognition

Facial expressions are a very strong communication medium that is meant for social situations in everyday life (Atkinson et al., 2004). The reason why they are so strong is the fact that we all know the expressions very well. A person can recognize others expression very quickly, usually with only one glance. Fast recognition is necessary because, according to Izard (1991), some facial expressions only last ½-seconds. Usually the expressions last over 1/3-seconds and less than 10 seconds (Izard, 1991).

One very important tool for facial expression communication is the direction of gaze and the difference between direct and averted gaze. Gaze combined with facial expressions can have significant information value of either approach or avoidance (Adams and Kleck, 2003). They also have a significant role in facilitating social communication (Haxby, Hoffman and Gobbini, 2000 and 2002). Facial expressions, such as joy and anger (approach oriented emotions), can be processed more rapidly if there is direct gaze presented. Respectively, emotions are interpreted as avoidance oriented emotions if presented with averted gaze (Adams and Kleck, 2003).

The recognition of individual is based on the perception of different facial structure aspects. These aspects can be the changes in facial expressions and the movements of eyes and mouth (Haxby, Hoffman and Gobbini, 2000). Human visual system is divided into two pathways according to Bruce and Young (1998, as cited by Haxby, Hoffman and Gobbini, 2000), magnocellular and parvocellurar. Facial perception is usually done from the coarse information in the magnocellular pathway. For proper recognition, it has been noticed that lightness has much more importance to it than for example hue (Haxby, Hoffman and Gobbini, 2000).

Facial expressions and the recognition of those expressions have a huge role in the study of emotions in human-computer interaction (HCI) (Afzal et al., 2009). The facial information has a significant advantage over other physiological measurements, because it can be detected and analyzed unobtrusively in real time. It also can be made totally automatically in real-time and requires only a simple video camera (Afzal et al., 2009).

As in every motion capture and recognition technique, the facial recognition has also it flaws. We can very easily occlude face from cameras view area or the lighting conditions might not be suitable for proper recognition. One of the biggest problems with facial recognition is the mapping of the facial expressions to emotional states (Afzal et al., 2009). These mapping tools are much researched but they are not included to this thesis.

3 EMOTIONS

In this chapter we examine the question of how emotions and moods are created in the brain. This is one of the most central questions in affective neuroscience. The answer might not be so simple. There are a numerous different opinions and theories created in past decade. They should all be considered in order to understand the phenomenon completely.

We start by reviewing the very basis of the emotional studies from William James and Charles Darwin, the pioneers of emotional research. After this, we move on to more advanced emotional theories. We also examine briefly different brain regions which are involved in processing emotions.

These emotional theories might not be purely psychological but also physiological, which means that test subject experiences changes for example in heart rate, muscle tension or brain activity. That’s why we also need to have ways to measure the emotions in different ways.

Emotions are not the most logical phenomenon. They have taken a lot of time to be theorized. Still today, we don’t know exactly how they are created and interpreted. Even the Star Trek character Mr. Spock thinks that emotions are useless because they are not logical.

3.1 Early theories

One of the first theories of emotions where published in 1872 by Charles Darwin (Gross, 1996). It was the first real attempt to research emotions. In his study, Darwin argued that emotional expressions are not depended of race or culture of people but they are accompanied through borders of race and even species. Another notion in his work was that animal emotions are quite homogenous to human emotions (Dalgleish, 2004). Darwin approached the problem with emotions by comparing the similarities of behavior in other species (Gross, 1996).

Soon after Darwin’s theory, a well-known American psychologist and philosopher William James and Danish physician and psychologist Carl Lange had started to work their theories of emotions without knowing the others work in late 1890’s. Their theories acclaimed that our emotional reaction is the result of perceived bodily changes, not the cause (Gross, 1996; Dalgleish, 2004). In other words, when some situation causes a reaction in our body, we interpret this reaction as emotion. Widely used example is that if we meet a bear in woods, we are frightened and run. The James-Lange theory suggests that we are frightened because we run. Figure 6 presents the James-Lange theory thoroughly.

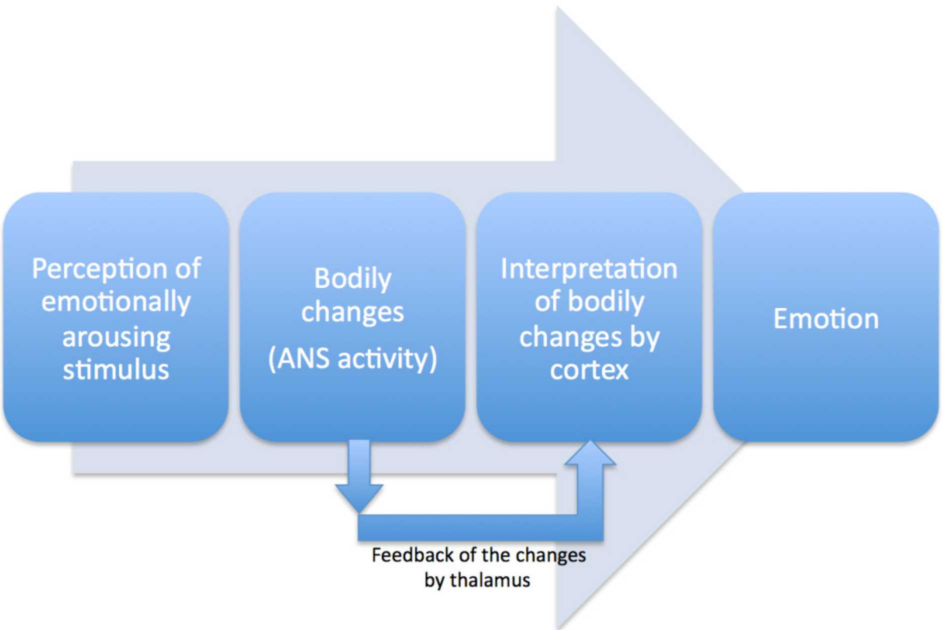


Figure 6. The James-Lange theory

Another interpretation of this theory is that if we deliberately alter our behavior we could control emotional experiences (Gross, 1996). Some studies that Gross (1996) presents, have found that stimulating some facial expressions actually alters the autonomic nervous systems (more about ANS in chapter 6.) activity. For example in Levenson, Ekman and Friesen (1990) study, stimulated anger and fear increased heart rate but with anger skin temperature increased whereas with fear it decreased. Happiness decreased heart rate but didn't affect to skin temperature.

Walter Cannon challenged the James-Lange theory several years later, acclaiming that the theory had four major faults (Cannon, 1929, as cited by Gross, 1996; Dalgleish, 2004). These faults were:

1. Assumption that every emotion has a unique set of physiological changes in our body and with these changes we could label each emotion.
2. Physiological arousal is insufficient to generate emotions.
3. Arousal may not be necessary.
4. Some bodily changes are too slow for some emotional experiences.

This criticism from Cannon led to his own investigations with Philip Bard. According the Cannon-Bard theory, the autonomic nervous system reacts to all emotional stimuli the same way (Gross, 1996). The subjective emotions are nearly independent of the physiological bodily changes. When emotion-producing stimulus is perceived, thalamus sends impulses to cortex and to hypothalamus. The theory suggests that humans feel the emotion first and only after that act upon it (Dalgleish, 2004). With the bear example, if we see a bear in woods, we first feel the emotion and only after that act accordingly. The Cannon-Bard theory is visualized in Figure 7.

A few decades later, the Cannon-Bard theory was put under review by Stanley Schachter in 1962. Schachter argued that Cannon was wrong in thinking that emotions and bodily changes are independent (Gross, 1996). He was convinced that the James-Lange theory was right claiming physiological changes as the cause of emotion. In other words, bodily changes are preceding the experience of emotion.

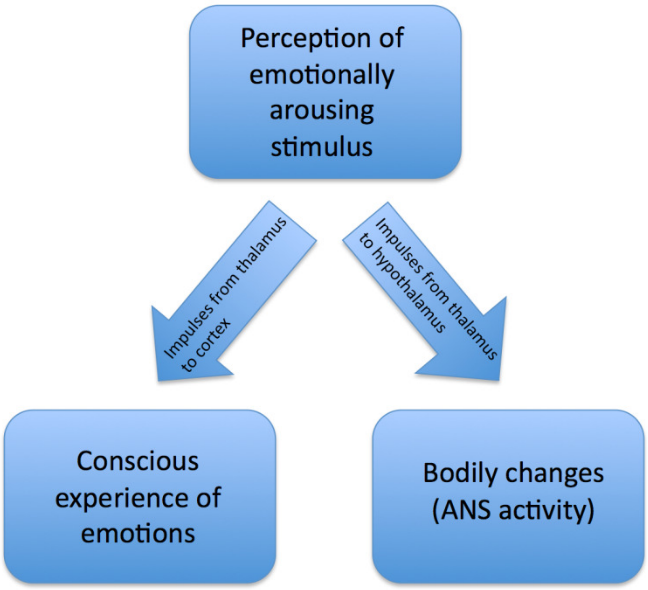


Figure 7. Cannon-Bard theory

Schachter's theory is often referred as the two factor theory of emotion. Schachter thinks that arousal (1st factor) is necessary for emotion experience (Gross, 1996). The difference between Schachter's and the James-Lange theory is the Schachter's believe that we have to decide which emotion we are feeling (Gross, 1996). This means that we have to cognitively interpret (2nd factor) or label the arousal. Schachter's theory is presented in Figure 8.

Additionally, a few theories that have been done in the 1980's are Lazarus's cognitive appraisal theory and Zajonc's affective primacy theory. Lazarus's theory argues that there is a minimal cognitive process preceding the emotional experience, though the process can be unconscious and automatic, whereas Zajonc's theory suggests, that emotional responses can happen without cognitive thinking (Gross, 1996).

Now a days, most of the theorists would agree that some sort of cognitive processing is a necessity for most of the emotions (Gross, 1996). The biggest problem lays with specifying more precisely these cognitive processes that are crucial to the emotion generation. After that we must identify the neural network for emotion and cognition, and the interaction between them (Gross, 1996).

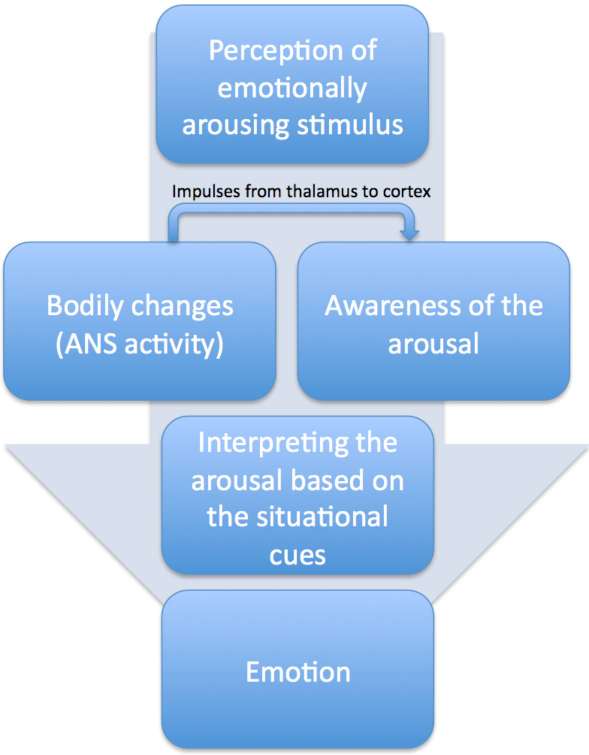


Figure 8. Schachter's theory

3.2 Psychology of emotions

Most theories of emotion acknowledge that emotion is not a simple phenomenon. They cannot be described completely by person telling about his/hers experience (Izard, 1977). Even sophisticated electrophysiological measurement tools cannot give definitive description of emotions. According to Izard (1977) a complete definition of emotions must take into account the experience of emotion, the processes that occur in our body and the patterns of emotions that can be observed from facial expressions.

Emotions are necessary in human daily life. They control nearly everything that we do. They are very important for our intelligence, for decision making in rational situations and in social situations and interactions (Picard, 1997). Also our memory, learning, and creativity are functioning based on what we feel. Picard (1997) argues in hers book *Affective Computing* that the key component for creating truly intelligent computers is the emotions. For long time scientists have concentrated to problem solving, learning and also to the *uncanny valley* in creating a believable human analog and intentionally ignoring the emotions (Picard, 1997).

Emotions have been ignored because they are being considered inherently non-scientific (Picard, 1997). They are also considered to have negative influence to research's logical thinking. This is why emotions have mainly a sideline role in scientific research (Picard, 1997). Emotions are studied by carrying out practical user experiments and at the same time measuring test subject physiological or psychological indicators. Emotions are considered to be as products of evolution (Partala, 2005).

The word *emotion* has many different interpretations and many synonyms describing the same effect. They are often used interchangeably and without any clear definitions. The most common terms used in scientific research are affect, emotion and mood (Zimmermann et al. 2003). Zimmermann et al. (2003) defines that the term affect can be used for generalizing both emotion and mood. Emotion has a specific cause and effect. It is triggered by stimulus or preceding thought followed by short duration of intense experience which the person in well aware of it (Zimmermann et al. 2003). A mood is much less intensive than emotion. It also tends to be more in the background and last longer.

There are many different theories defining human feelings and how emotions are created. Wilson's (2000) theory is dividing the emotion into three layers of human behavior. The top layer consists of reactions and momentary emotion that are small bursts of feelings. The next level down is the mood. Wilson (2000) defines mood as "prolonged emotional states caused by the cumulative effect of emotions that is signals of punishment and reward". Below both of these layers is human personality. The personality is always present in every situation unless some momentary emotion or mood overrides it (Wilson, 2000).

Figure 9 represent Wilson's (2000) theory. Different layers of behavior have different levels of rise and fall. The highest in the layers has also the highest priority. If there is no reactions or momentary emotions present, our behavior is based on our current mood. And if our mood is below a specific threshold, we base our behavior on our underlying personality (Wilson, 2000).

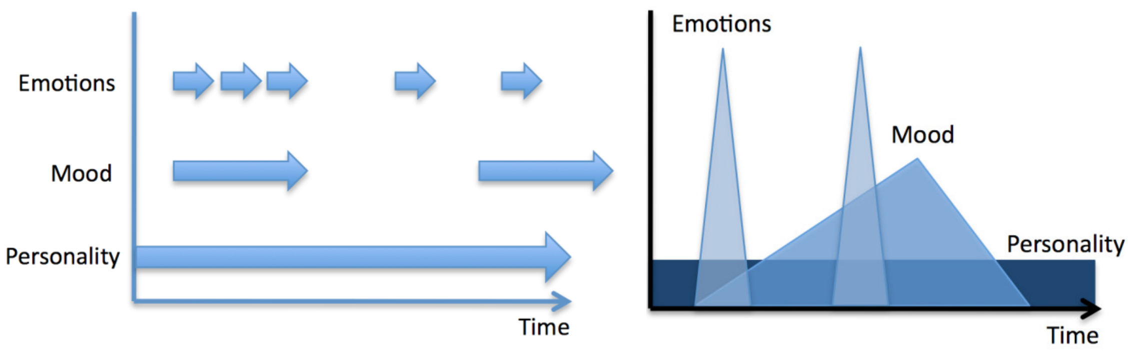


Figure 9. Wilson's theory

Another well known theory of emotion presented in the scientific literature is by Arne Öhman from 1987. He constructs his theory to the assumption that emotions are multifaceted structures (Öhman, 1987). His theory is based on to the assumption that all emotional phenomena's are caused by emotionally meaningful situation. After the stimulus emotions can show themselves in three different ways. Emotion can occur in some verbal delivery or it can be felt in some physiological response or it can be seen in changes of behavior (Öhman, 1987).

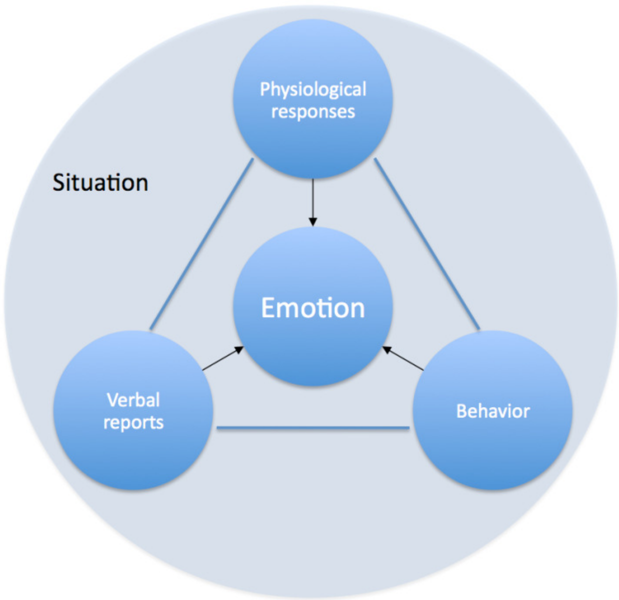


Figure 10. Öhman's theory

According to Gross (1996), there are three components that each distinct emotion has. From these three categories, the second and the third can also be categorized together as bodily reactions. Gross's theory (1996) is significantly based on the Öhman's theory.

1. The subjective experience
2. Physiological changes
3. Associated behavior

Emotion theories can be divided into two categories based on their level of abstraction (Partala, 2005). Discrete emotions can be seen as a sort of packages which have specific and very noticeable bodily and facial reactions (Tomkins, 1967). Discrete emotions can be detected already from very early age. Dimensional emotion theory assumes that emotions can be defined in continuous dimensions. These two theories are considered as compatible and complementary to each other (Partala, 2005). This means that discrete emotions can be presented as subspaces of dimensional emotions.

Next chapters presents a little more about both of these theories and a little about how to measure emotions in different physiological ways.

3.3 Discrete emotion framework

Discrete emotions are assumed to have unique experiential states that are someway noticeable and measurable (Barrett, 1998). In other words, the emotional transition for example from anger to happiness is not seamless. The emotional space between different emotions is categorized into discrete areas. According to some theories, for an emotion to be a discrete emotion, the emotion should be experienced separately from one another for some proportion of the time (Barrett, 1998).

Discrete emotion theory is sometimes also referred as differential emotion theory or DET (Izard, 1977). The theory is based on five key assumptions by Izard (1977):

1. Ten fundamental emotions that are the principal motivational system for human.
2. Each fundamental emotion has specific unique motivational and phenomenological properties.
3. Some fundamental emotions lead to more inner experiences and different behavioral consequences.
4. Emotions interact with each other by amplifying, activating and attenuating another emotion.
5. Emotion processes interact with other processes in human body.

Another well known theory of discrete emotions is the theory of six basic emotions by Paul Ekman (1992). In this theory, Ekman suggested there are six universally recognizable facial expressions that can be associated for a specific emotion. These six emotions are anger, disgust, fear, joy, sadness and surprise. The facial expressions of these emotions can be seen in Figure 11 below.

Figure 11. Ekman's six basic emotions: Happiness, surprise, fear, sadness, disgust and anger (Ekman, 1992)

Izard (1971 and 1978, as cited in Matsumoto, 2006) claimed that six basic emotions should also include interest-excitement and shame-humiliation expressions. Other emotions or expressions suggested to be basic include contempt, embarrassment and pride (Matsumoto, 2006). Ekman revised his own theory later on and suggested that there should be 15 basic emotions instead of the six (Ekman, 1999). These 15 emotions were amusement, anger, contempt, contentment, disgust, embarrassment, excitement, fear, quilt, pride in achievement, relief, sadness/distress, satisfaction, sensory pleasure and shame (Ekman, 1999).

The recognition of these 15 emotions was supported by the development of a theory of characteristics. Ekman (1999) distinguished eleven characteristics that could help in defining basic emotions. The first three were used in distinguishing one basic emotion from another and the rest of eleven characteristics for distinguishing emotions from moods, emotional traits and emotional attitudes (Partala, 2005).

Another theory closely related to Ekman's basic six -theory, is the theory of primary and secondary emotions. Plutchik presented this theory in 1980 (Gross, 1996). As the primary emotions Plutchik used emotions that corresponded Ekman's, plus acceptance and expectancy. The idea of the theory was that every single emotion could be derived from these eight emotions. Figure 12 describes the emotion wheel that presents the theory visually (Gross, 1996). Plutchik believed these primary emotions are biologically and subjectively distinct in every human (Gross, 1996).

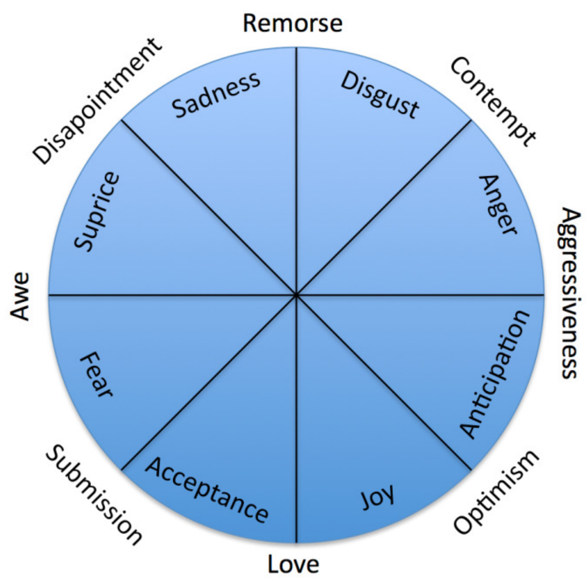


Figure 12. Emotion wheel

3.4 Dimensional emotions framework

Most researchers agree that emotions have at least two qualities (Barrett, 1998). These qualities are valence, or pleasantness, and arousal. With these two qualities, emotions can be presented in a two-dimensional affective space. The valence dimension presents how pleasant some experience is ranging from negative to positive leaving the neutral zone at the center of the dimension (Partala, 2005). The arousal dimension uses scale from calm to excited or highly aroused for representing experienced arousal. This dimension also has neutral zone at the center (Partala, 2005). Bradley and Lang first introduced this dimension scale in 1994.

These two dimensions are often completed with a third dimension. The dominance dimension presents the degree in which a test subject feels that she/he dominates the situation (Partala, 2005). It ranges from "in control of the situation" to "controlled by the situation" from end to end, leaving the neutral zone at the center (Partala, 2005).

From these three dimensions, valence and arousal are the most frequently used in emotional research (Partala, 2005). Valence can be seen as reflector of presence of an appetitive of aversive motive systems in a way of behavioral approach and withdrawal whereas arousal dimension presents the intensity of the emotional stimulus (Lang et al., 1995, as cited by Partala, 2005).

The next figure (Figure 13) represents some discrete emotion categories positioned into dimensional emotion space. The faint scale in the middle marks the neutral zone for each dimension.

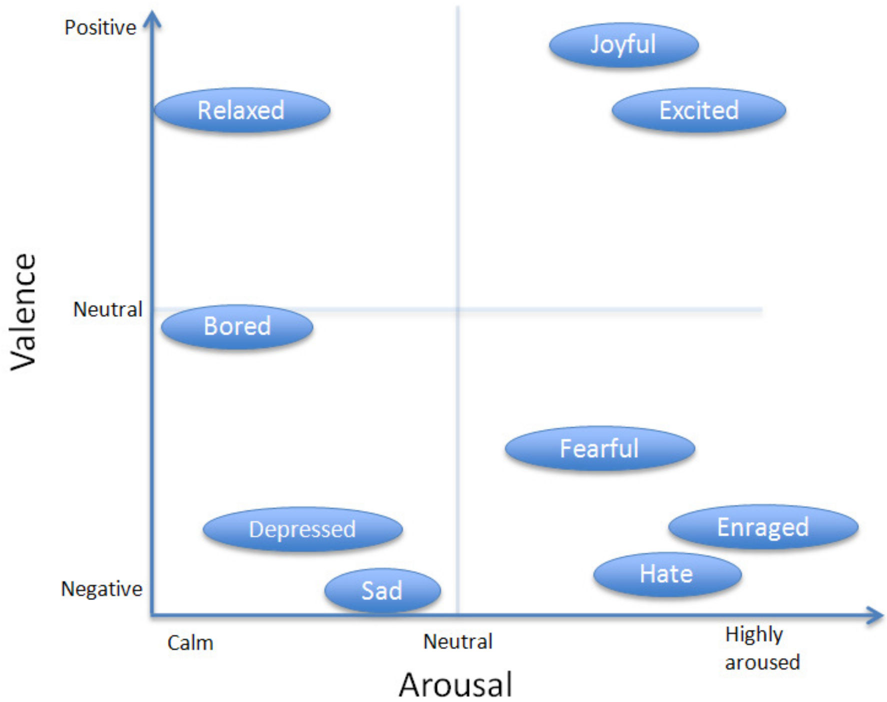


Figure 13. Discrete emotion in dimensional space (Partala, 2005)

3.5 Controversy of emotional theories

There are few advantages with using dimensional emotion theory rather than discrete emotion frameworks in experimental research. First problem is when test subjects are asked to report their emotional states using the discrete emotions approach. This method forces test subject into a forced-choice response which often oversimplifies the felt experience (Partala, 2005).

Picard (1997) points out that if test subjects are allowed to report their emotions and feeling totally freely, they usually report more thoroughly the experiences and with many different emotions. This brings another problem with interpretation. Each test subject can report a specific emotion in different ways and researcher must know how each of them thinks. For example, merry, jolly, or glad can all mean the same emotion which is basically happy. Researcher must do a lot of work for cross-referring these reported emotions. There are many researches already done with these cross-reference models but they are not discussed here.

4 EMOTIONAL CIRCUITS AND HOW TO MEASURE THEM

There are many psychological theories of emotions and how they are created. But psychological research is not the only way to get information about emotional responses. We can measure many different physiological indicators, which are present when emotions accrue, from our body. So what really happens in our body when sudden emotional rush emerges? How does our body react for example to fear or happiness? What parts of our body react to it and what can we measure?

The brain has obviously a big role in creating and interpreting emotions. The more we collect information about the representation of emotions, the more we can eliminate the nature of emotional processes (LeDoux, 1995). With this knowledge, we can choose correctly between different hypotheses that are explaining emotional processes. Furthermore, we can complete psychological theories with the findings about the neural basis of emotion. This might bring new insights to the functional organization of emotion (LeDoux, 1995).

This chapter presents the emotional circuits in our body. In other words, how emotion are created. The chapter may not be the core of this thesis but it is still very important to understand how emotions are created in our body, how they affect our body and how they might affect our way of moving and gesturing.

4.1 Emotional circuits in our body

4.1.1 Nervous systems

The brain isn't the only part of human body that is involved in creating emotional information. The involvement includes the entire human nervous system (Partala, 2005), which is a part of human psychophysiological system. The nervous system is a network of neurons that transmit electrical signals all around our body. Our nervous system consists of two parts: the central nervous system (CNS) and the peripheral nervous system (PNS) (Partala, 2005). Brain and the spinal cord are the two parts of CNS. The peripheral nervous system consists of the sensory-somatic nervous system and the autonomic nervous system (ANS).

The main function of CNS is to receive sensory input from the environment. From CNS the information is passed on to the limbic system and the cortex. Partala (2005) notes that the limbic system is often described as the seat of emotion. The limbic system contains the hypothalamus, the hippocampus and the amygdala (LeDoux, 2000). Some studies have shown evidence that the limbic region, especially amygdala has an important role in the processing of emotional information (LeDoux, 2000), specially the processing of the emotional fear. The limbic system is presented in next chapter figure 15.

Autonomic nervous system is mainly functioning below level of consciousness, meaning that it is mostly involuntary. The ANS transmits the electrical impulses from CNS to internal organs (LeDoux, 2000; Partala, 2005) whereas sensory-somatic nervous system is responsible for moving limbs and receives information from our senses. ANS has the vital role in regulating emotion creation in our body.

ANS is further divided to the sympathetic nervous system and the parasympathetic nervous system. These systems work as a counter balance for each other. When sympathetic nervous system is activated our heart rate, blood pressure and pupil size increases. It prepares our body for "emergency status". Parasympathetic nervous systems function is to lower our heart rate etc. and bring our body back to the normal status (Partala, 2005).

4.1.2 Brain areas involved in processing of emotions

Amygdala seems to be one of the most important brain regions for emotion and for processing social signals of emotion (Dalgleish, 2004; Cardinal et al., 2002). Amygdala is very important also with recognition of faces. Some studies have shown that if amygdala is damaged, it can lead to impairments in the processing of faces and other social signals (Jacobson, 1986 as cited by Dalgleish, 2004). Damage can also lead to state where patient cannot learn to predict aversive events. This symptom is also known as fear conditioning (LeDoux, 1995 and 1996). The fear conditioning is shown to act without subject consciousness.

Another important region of brain involved in processing emotions is the prefrontal cortex. The prefrontal cortex is the very front of the brain, just beneath the forehead. It has been discovered that prefrontal cortex meditate our conflicting thoughts, predicts the future events, makes choices between right and wrong and even suppresses our emotional and sexual urges (Dalgleish, 2004).

The prefrontal cortex is also been called in some cases the reward center of the brain because it expects a reward when heading to some future goal. In some studies, the activation of the prefrontal cortex with control subjects is shown to elevate sweating (skin conductance) in situations with possible reward. In the same situation, the patients with prefrontal cortex damage the skin conductance showed no elevation (Bechara, Damasio et al. 1994 as cited by Dalgleish, 2004). These different parts of brain involved in emotion creation and recognition are presented in figures 14 and 15.

Hypothalamus, like the prefrontal cortex, is also involved in the processing of rewarding stimuli. The hypothalamus is responsible for different activities of the ANS. It has been shown that the hypothalamus is a part of an extensive rewarding system which also involves amygdala and ventral striatum (Dalgleish, 2004).

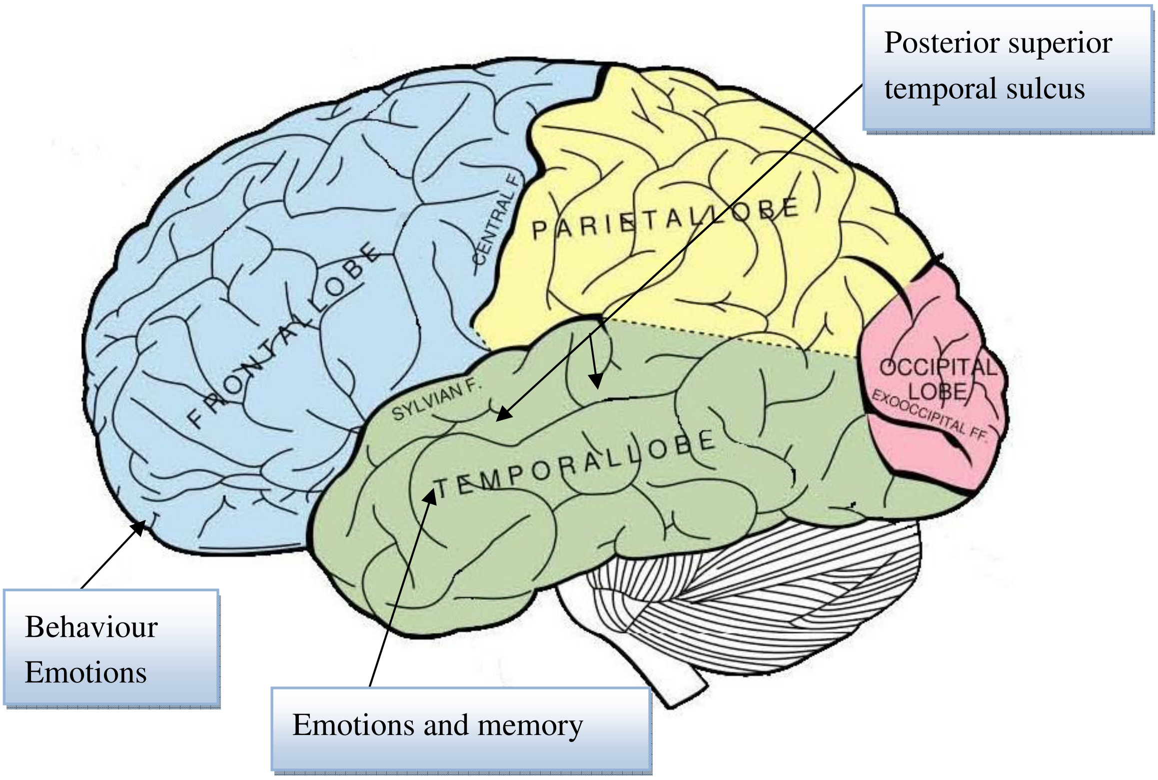


Figure 14. Different brain regions affecting to emotions processing

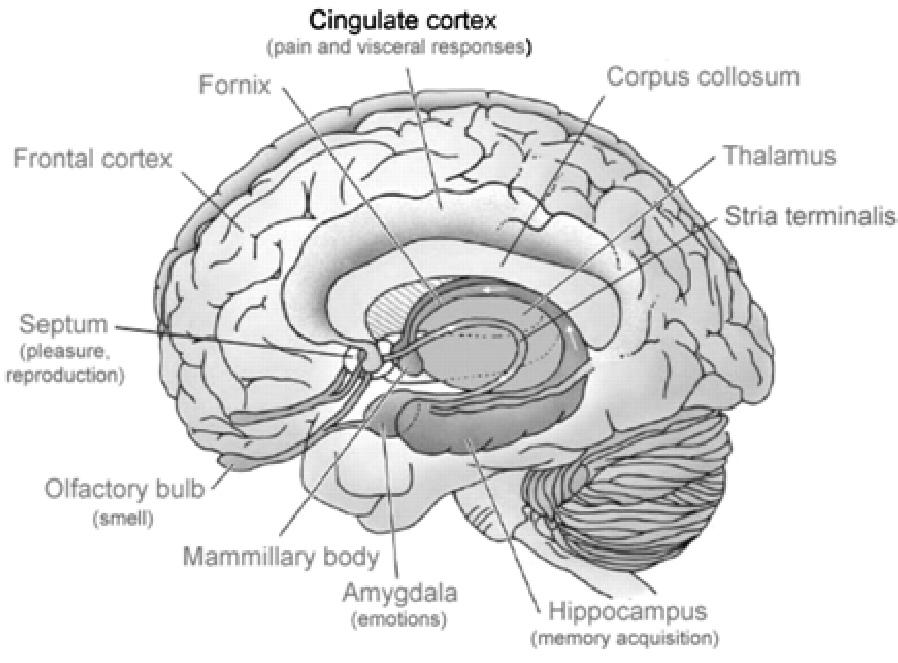


Figure 15. The limbic system

4.1.3 Brain areas involved in perception of biological motion

Perception of biological motion influences in indentifying, interpreting and predicting the movements and actions of others (Grèzes et al., 2001). Especially the facial movements create strong neural responses that are all different from each other (Puce and Perrett, 2003). These movements can be such as deviations in eye gaze are very important for determining individual's social attention whereas mouth movement represents different types of prosody (Puce and Perrett, 2003).

Grères et al. (2001) suggest that perception of biological motions is served by a specific neural network. These neural networks can be studied with electroencephalography (EEG) or functional magnetic resonance imaging (fMRI). Research of biological motion includes a use of point-light displays as a stimulus. Test subjects can recognize male from female with no more than 12 point-lights when portraying a biological motion (Grossman et al., 2000). We can even recognize friends from strangers by just looking at their point-light movement (Cutting and Kozlowski, 1977). These types of researches are discussed in chapter 5.

Is has been shown that perceiving biological movement is highly dependent of the orientation. If we for example invert the motion or turn it upside down, the recognition of emotions in the point-light displays becomes more challenging (Grossman and Blake, 2000).

Biological motion is a complex and highly evolved motion pattern. It is also a good example of the sophistication of pattern analysis in the brain (Bientema and Lappe, 2002). Brain links seamlessly the perception of motion with the perception of form together in point-light displays. These qualities involve different part of cortical processing streams (Bientema and Lappe, 2002). Motion perception is analyzed in the areas of the dorsal stream whereas form perception analysis is located in the ventral stream (Bientema and Lappe, 2002; Vaina et al., 2001). Dorsal stream is also called the where- or how-pathway and ventral stream the what-pathway.

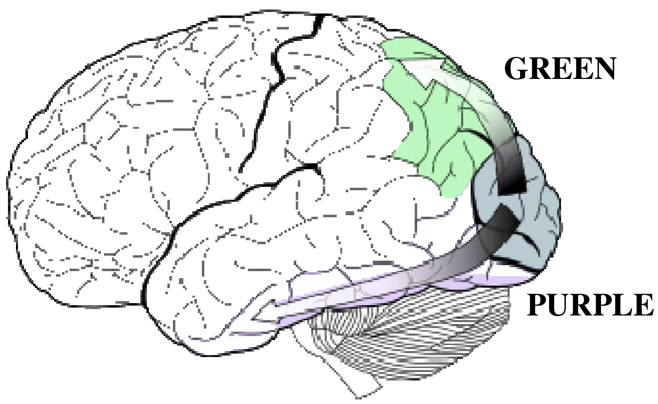


Figure 16. Ventral (colors, purple) and dorsal (visual, green) streams

Many studies indicate that viewing biological motion activates specialized visual mechanism in the posterior superior temporal sulcus (STSp) (Grossman and Blake, 2001 and 2002; Grossman et al., 2001; Puce and Perrett, 2003; Grèzes et al., 2001; Vaina et al., 2001). The STSp is located in temporal lobe of the brain which is located in both sides of the brain.

The posterior superior temporal sulcus activates during viewing biological motion. If the motion vectors are scrambled in space, the activation can be bigger than in normal situation (Grossman and Blake, 2001). The STS can also activate if test subject only thinks of the motion that draws a mental image of the motion. In this case, the level of activation is weaker (Grossman and Blake, 2001). Grossman et al. (2001) discovered that in the right hemisphere the activation was more frequent than in the left hemisphere. The activation of right hemisphere in 8-month-old infants suggest that the neural substrates for recognizing biological movement begins to develop at around 8 months of age (Hirai, Hiraki, 2004).

Beauchamp et al. (2003) researched how different brain regions react to human motion and to man made manipulable objects such as tools, i.e. saws or hammers etc, moving with their characteristic natural motion. They showed these actions to test group with full-light motion (normal video) and with point-light motion. They found that the superior temporal sulcus (STS) reacted strongly to all human motion regardless of if the motion were shown in full-light or point-light display. This reinforces the meaning of the STS in recognizing the biological movement. They also found another area called middle temporal gyrus (MTG) and inferior temporal sulcus, which reacted strongly to tool videos regardless of the display (Beauchamp et al., 2003).

4.1.4 Theory of structure and image based object recognition

As we concluded, our brain has multiple ways to analyze visual images. Theories of structure and image based object recognition are two radically different theories that have been proposed to explain how human can recognize different objects. These both theories might be entirely plausible, it just shows that our brain has multiple ways of analyzing visual output (Ware, 2000).

Image based object recognition proposes that we can see different objects by matching the visual image of an object with a rough snapshot-like image in our memory, whereas structure based object recognition analyzes an image in terms of primitive 3D form. For example, we know that a chair has three to four legs, seat and back rest. If we would see this kind of form, we would immediately recognize that form as a chair. In image based recognition, we recognize the chair from its outlines (Ware, 2000).

4.2 Emotion measurement

There are many ways we can measure emotions from a human. These methods do not actually measure the emotion but fluctuations in human physiological outputs. In other words, the methods measure emotional responses from our body actions. A nervous person might sweat a lot. A scared person has faster heart rate. A happy person tends his/hers facial muscles more and so on.

Emotions affect to our body through our nervous systems. The system works sometimes voluntary and sometimes involuntary. We cannot control all of our emotions. Figure 17 presents how different emotional outputs affect and how voluntary they are.

These psychophysiological measurements are not a part of my thesis but they are very important for whole emotions research. This is why they are presented only briefly.

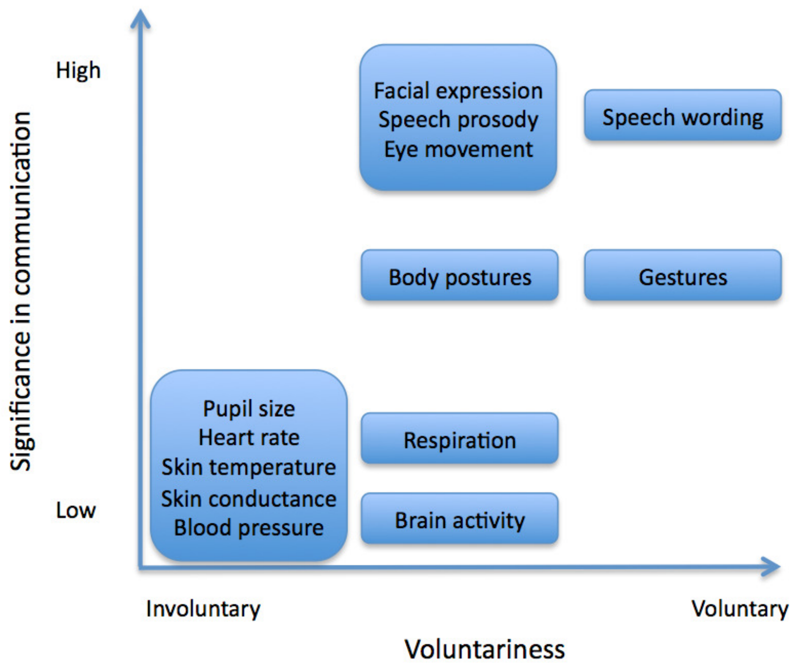


Figure 17. Voluntariness of different communication channels

4.2.1 Electrocardiography (ECG)

One of the most commonly used method for measuring heart rate is electrocardiogram (ECG or EKG). As mentioned before, heart rate is related to ANS activity and studies have shown that heart rate can change between positive stimulus and negative stimulus (Anttonen and Surakka, 2005). ECG measures heart's electrical activity during the pumping action with three electrodes attached on body. From ECG we can measure heart rate and inter-beat intervals where we can calculate the heart rate variability (Haag et al., 2005). If the heart rate variability is low, that usually indicates a state of relaxation. On the other hand, increased heart rate variability indicates mental stress or frustration (Haag et al., 2005). Variability and heart rate represents closely dimensional emotion theory's arousal axis.

4.2.2 Galvanic skin response (GSR)

Galvanic skin response - also referred to as electrodermal activity (EDA) or skin conductance response (SCR) or psychogalvanic reflex (EDR) – measures changes in skin resistance. Resistance fluctuates when glands on the skin produces sweat (Haag et al., 2005). Galvanic skin response needs a careful reference measurements and calibration before the actual tests, because temperature changes can affect with the results. GSR has also been shown to have correlation with ANS and dimensional arousal (Haag et al., 2005, Partala, 2005). Therefore it can be a good indicator of stress or differentiate between anger and fear.

4.2.3 Electroencephalography (EEG)

Brain activity can be measured with electroencephalography or EEG. It measures both frequency and amplitude of electrical activity generated in the brain using electrodes attached to the scalp (Partala, 2005). The activity is caused by neurons firing prior to some task or emotion. There are studies that have shown that EEG asymmetries over the frontal cortex are relatively grater in the left side of prefrontal cortex than the right side. This asymmetry happens with emotion related to joy, interest and anger which have a behavioral tendency of approach (Coan et al., 2001, as cited by Partala 2005). As a conclusion, the asymmetry is inversed with behavioral tendency of withdrawal like sadness and fear (Coan et al. 2001, as cited by Partala, 2005).

4.2.4 Electromyography (EMG)

Facial expressions have a significant role in human to human communications. EMG measures facial muscle activity or frequency of muscle tension (Haag et al., 2005). High muscle tension in specific muscles tells us of high stress level (Haag et al., 2005). Facial recognition is often used in entertainment business with optical motion capture systems. It can also be used for learning purposes and ubiquitous computing (Partala, 2005).

The facial electromyography is used for measuring the dimensional valence axis. Studies have shown that increased activity at the zygomaticus major and corrugator supercilii muscles associates with positive and negative emotion (Partala, 2005). Zygomaticus major is responsible for the muscle that produces smile by drawing the lip corners up and corrugator supercilii knits and lowers the brows for frowning. Orbicularis oculi activity is noted to be particularly high with positive valence and high arousal emotion (Ravaja, 2004). Orbicularis oculi is totally voluntary and responsible of closing our eyelids, but there is also involuntary muscle tissue which is a part of sympathetic ANS. Activation of this muscle tissue is thought to be involved with expressions of enjoyment smile and genuine pleasure (Ekman, Davidson and Friesen, 1990).

4.2.5 Psychological interview

One of the most commonly used methods of measuring the emotional response to stimulus is to simply ask what the test subject feels. Test subject can be whether asked to express feelings, moods and emotions totally freely or he/she can be forced to make a choice between different predetermined alternatives. Both methods have their pros and cons that were presented earlier.

The free choice method can also be carried out different ways. One way is to allow the test subject to express him-/herself in writing or orally. Another way is to use for example the dimensional emotion model and ask the test subject to estimate his/her feelings and emotions with valence, arousal and dominance scales. Scales can be divided for example into nine steps.

5 RECOGNITION FROM MOTION

Biological motion information tells a lot about the actor actions, intentions and emotions (Troje, 2002). Pollick (2004) presents two complementary approaches to study the human understanding of the movements of others. The first approach uses visual motion perception as a starting point and examines the human movement with traditional psychophysical methods. The other approach specifies different brain regions in processing of human motion and from that extends to organizing these regions (Pollick, 2004).

Pollick (2004) continues with listing three common explanations offered for the human capability to recognize point-light displays. The first two of those dealt with the structure of the human body and the third with the kinematics of the human movement.

Many researchers have concluded that the recognition of the biological motion relies on recruiting higher level of cognitive processes to the interpretation of the motion signals (Pollick et al., 2002; Shiffrar and Freyd, 1993; Thornton, Pinto and Shiffrar, 1998).

5.1 Biological motion in point-light displays

The first attempts for isolating pure biological motion were made by Gunnar Johansson in 1973 (as cited by Atkinson et al., 2004; Hill, Jinno and Johnston, 2003). He devised a technique called point-light displays (sometimes referred also as patch-light display) to study the perception of biological movement. With this technique he could minimize or totally eliminate all static form information from stimuli but still retain the motional information (Atkinson et al., 2004). As an opposite to point-light displays, a full-light display presents the whole body image and motion and corresponds to a normal video recording.

Point-light display technique differs from full-light displays in a way. In point-light display, the moving character is presented by a small number of illuminated dots or patches (Atkinson et al., 2004). These stimuli are produced by filming reflecting dots attached to actor joints with multiple cameras (Hill, Jinno and Johnston, 2003). This technique is widely based on the optical motion capture technique. The dots can be placed to joints or to actor's facial area if facial expressions are needed to capture.

When point-light display is static, the dots appear as a random series of dots. But if the dots are moving, we can immediately recognize the form as human motion (Atkinson, 2003; Heberlin et al., 2004). Point-light displays are very useful in numerous research areas but they have a few disadvantages associated to facial expression recognition. Hill, Jinno and Johnston (2003) claims that with natural facial movements, point-light displays only provide a limited sampling and representation of the motion information available. They also argue that if the point-light display is not explicitly normalized, the display will contain residual cues to spatial configuration along side with motion information.

Different brain regions have been noticed to react differently if test subjects are shown point-light displays or full-light displays (Beauchamp et al., 2003). Even infants can recognize the human form and decode the biological motion from point-light displays (Beauchamp et al., 2003; Hirai and Hiraki, 2004).

It has been shown that perception of a structure in a point-light display doesn't require prior knowledge of individual features or their local relations (Bertenthal and Pinto, 1994). This means that perception of form, when specified by biological motion, foregoes the perception of different elements and local relation like joint angles (Bertenthal and Pinto, 1994).

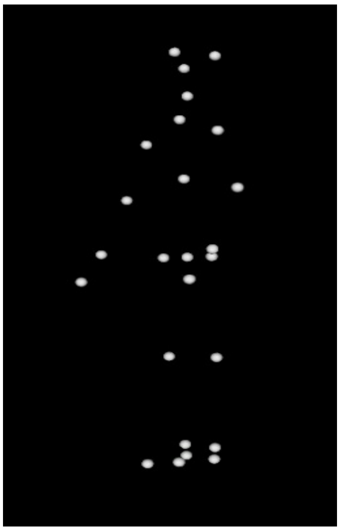


Figure 18. PL-display

5.2 Person recognition

The most researched and often cited area of recognition is the recognition of gender from gait or motion (Pollick, 2004; Pollick et. al, 2002). Pollick's (2004) review had found that gender recognition from the side view and with point-light walkers yielded a range of 46% to 72% correct answers with median of 65% correct.

Head and facial movement provides lots of valuable information about the actor's identity. Hill and Johnston (2001) discovered that test subjects could in fact indentify impressions between individuals and between males and females. Their findings suggest that from rigid movements of head test subjects can identify individuals and from non-rigid motions is used for recognition of sex. They also showed the stimuli backwards and upside down which lead to reduction of the identification performance.

There are many researches considering the gait analysis and how humans can recognize a friend from a stranger or male from female by just looking at the biological motion (Kozlowski and Cutting, 1977; Little and Boyd 1998). But some studies argue that the experiments held in laboratory conditions differ substantially from the realistic conditions (Jacobs and Shiffrar, 2005).

Humans have been found to predict the effects of actions more reliably if the actor is the test subject oneself (Knoblich and Flach, 2001; Loula et al., 2005). This result collides with theories claiming that perception and action systems are totally separated from the cognitive system. The test subject knows one's action because when recording the motion, the same system plans the action and afterwards predicts the action's effects (Knoblich and Flach, 2001). Some studies suggest that representation of biological movement may be based upon the movement limitations of the human body (Kourzi and Shiffrar, 1999). Troje (2002) also presents that the dynamic part of the motions contains more information about the gender or sex of the actor that structural cues that are motion-mediated.

The role of posture as an emotional communications medium has been poorly studied if related to research of facial and vocal expressions (Coulson, 2004). Some evidences have already been found that test subjects can read emotions from static body postures with good level of accuracy (Coulson, 2004). With point-light motion data, recognition rates were only 30% for one emotion, which were still above chance (Walters and Walk, 1988, as cited by Coulson, 2004). With full-light display, the rates were above 70%.

5.2 Emotion recognition

The role of emotions affecting to human movements has been research for long. Some studies have even tried to interpret emotions from stylized dance movements with conclusions that of the six basic emotions that were captured, the emotional anger was the most reliably indentified (Pollick et al. 2001).

Test subjects can recognize with high probability emotional content of dance movements which are varied in four emotional states (Camurri, Lagerlöf and Volpe, 2003). The most common mix-up was between anger and joy.

Humans can recognize from point-light display whether the movement style specifies to vulnerability in case of a physical attack (Gunns, Johnston and Hudson, 2002). Same experiment also confirmed that different gendered point-light display walkers can be recognized to male and female. The hard to attack walker was recognized from a longer stride length, swinging foot movement, a larger range of arm swing, higher energy, lower constraint, a faster walk, and a relatively heavier body weight than easy to attack walkers whereas easy walkers moved more gesturally (Gunns, Johnston and Hudson, 2002).

Pollick et al. (2001) examined how the visual perception of emotions can be seen from simple arm movements in point-light displays. They used two actors to perform and capture drinking and knocking movements with ten different emotional states. Motion capture was done with optical motion capture system which recorded the three-dimensional positions of actor's arm.

After the capture, Pollick et al. (2001) showed these motions to test subjects. The point-light stimuli were shown in normal way and in other experiment the displays were altered so that the stimuli were shown upside down and randomly phase shifting the points. In the latter test, the distortion should decrease the performance of the recognition.

The results showed that the test subjects can answer correctly 30% of the time with a range of 15% with strong-emotion and 50% with afraid-emotion (Pollick et al., 2001). It is well above the rate of chance. The other experiment resulted a correct answer rate of 14%. The range was from 4% with afraid-emotion and 34% with excited-emotion.

However, some studies have shown totally different result. For example Pollick et al. (2002) research showed a failure to recognize gender from biological motion. This was determined because of inability to use the available information. This concludes that human performance in this kind of studies can vary dramatically (Pollick et al., 2002). This lead to a suggestion that the gender information is carried in the form of the movement and it varies across actions and limb segments.

6 RESEARCH QUESTIONS / HYPOTHESIS

The main research question that we explore in this thesis is whether emotions can be perceived in point-light displays with a single actor and without social interaction. Which emotional states are recognized more easily and which are less accurately.

The second part of this research is to find out how much different attributes affect on the recognition rate. We concentrated on the orientation of point-light displays, the gender of both the actor and the test subject. We also examine the possible learning process that might occur during the test.

Some of the emotions might be easier to recognize than others, for example angry might be easier than weak. We hypothesize that weak, afraid and sad will be mixed up between each other, and also strong and angry will be easily confused together.

We hypothesize that there might be some learning process during this experiment in a way that our test subjects learn to recognize emotions more accurately toward the end. In other words, the recognition rate of the first 40 stimuli will be lower than the latter 40 stimuli. We also hypothesize that the distribution of the wrong answers will decrease with the latter 40 stimuli.

The question if the gender affects somehow the recognition rate is a difficult one. We would like to think that female participants would concentrate more on these kinds of tests and get better results than male participants.

We also hypothesize that our test subjects can recognize emotion from point-light displays better that just guessing the right answer. There are many emotion cues in the motion so the recognition is assumed. The extent of the recognition isn't so easy to estimate. The past researches (Pollick, 2004) suggest a recognition rate of 15-50 percent of all emotions and stimuli with normal oriented point-light displays. We would be pleased if all emotions could be recognized with rate over the level of change (10 percent, 10 different possibilities of answering). From previous studies and references we can already tell that emotions can be recognized from point-light displays. Our goal is to verify these results and find out the scale of recognition rate. In other words, how accurately and in what conditions emotions are recognized from the point-light displays. These results will give a good ground for further studies which are presented in chapter 10.

7 RESEARCH METHODS

In this chapter we give a detailed picture of the research done for this thesis. We'll try to be as thorough as possible so that these tests can be reproduced afterwards again.

Our experimental research consists of two different parts. The first part is the capture of motion from two actors with different emotions. The second part deals with the experiment with test subjects and emotional point-light displays.

Different experiments were performed simultaneously to investigate how emotions are perceived from point-light displays of human motion. In the first part of the experiment we presented the actions with ten different emotional states and measured how accurately the participants can categorize the emotion from motion. In the second part of the experiment we displayed the movement in upside down as Pollick et al. (2001) and Clarke et al. (2005) research. We examined the effect of gender and learning in recognition from PL-displays.

7.1 Movement collection

For motion capture, we used OptiTrack's optical motion capture system that contains 12 infrared cameras. We obtained the movement and position data from whole body. That means that the actors had 31 reflective markers on their body tracking their motion. Figure 19 presents the placement of these markers.

Figure 19. Actor wearing the motion capture suit and markers

We used two volunteer actors (one male, one female) from Aalto University student union's acting group called Teekkarispeksi. They both acted two different acts with 10 different emotional states. These acts were basic walking action and knocking at door. Due to our small capture area, the walking action was limited to a couple of steps.

Emotional states that the actors played were the same as in Pollick et al. (2001) research. These emotions were afraid, angry, excited, happy, neutral, relaxed, sad, strong, tired and weak. These emotional states are placed quite nicely in every side and corner of the dimensional scale (Figure 20) so wide range of emotion was covered. The placement of strong and weak emotion in valence-arousal scale is quite controversy. Their proper placement should be in the dominance scale with is not presented in Figure 20.

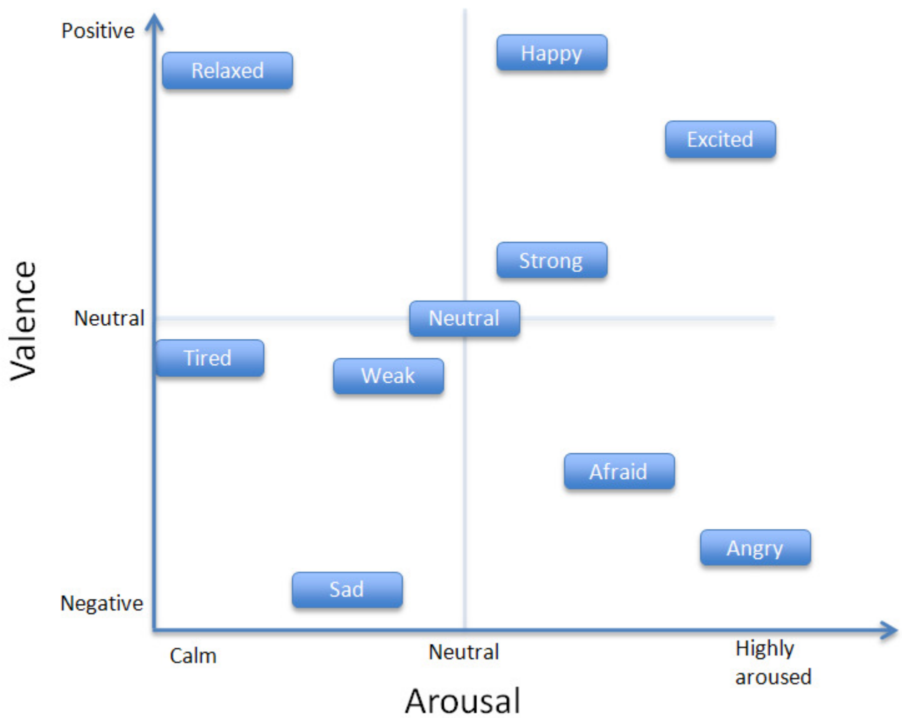


Figure 20. Used emotions mapped into dimensional space

Actors were asked to perform a small pre-act before the act that we wanted for the experiment. This way we got our actors better in the mood. The whole act consists of the pre-act, walking and finally the knocking. Before the knocking the actor approached the "door", stayed still for a second and then started the knocking action. This way we could better edit the scene afterwards. Walking might have not been ended before the knocking, if it was done with seamless motion.

7.2 Psychological tests

The psychological tests were held in at the beginning of April 2010. We had 35 test subjects, 16 males and 19 females (mean age 24.9 years) and they all volunteered to serve as subjects. All the participants were naive to the purpose of this experiment and had never previously encountered this kind of experimental stimuli. They were not paid for participate in this experiment.

A PHP-questionnaire form was made for running this test online. Test subjects were collected via contacts in different social media. More detailed instructions were sent via email. This method might have brought more uncertainty to the results but this way we could acquire more test subjects and more data which might decreased the uncertainty of statistical analysis.

The test subjects were informed that some of the PL-displays would be distorted. They were told what they were looking at. One test sample of PL-displays was shown prior to the actual test. This was done to get the test subjects familiar with the PL-displays. They were also informed explicitly that there was no right or wrong answers in this research. This way we could decrease the feeling of competition.

One of the biggest challenges was to form the research question for test participants so that everyone could know what we wanted them to answer. The test participants were finally told to describe the emotion of the character played in display with one of the given adjectives. These adjectives were the same as in Pollick et al. (2001) research and the same we used in capturing the motion with different emotional states (afraid, anger, excited, neutral, happy, relaxed, sad, strong, tired and weak).

Gathering data with interview based methods is time consuming. You get more control for the situations but it takes more time. The order of the stimuli must be randomized preferably to every test subject. Randomizing erases the possibility for different stimuli to affecting each other. This way we get more accurate data from our test subjects.

The order of the PL-displays was randomized for each test subject by the PHP-software. The test subjects didn't know if the next display would be right way up or upside down. Each display was shown to the test subject two times, so each test contained 80 PL-display stimuli (2 actors x 2 orientations x 10 emotions x 2 repetitions).

In the PHP-questionnaire form, the test subjects were first asked to fill out his/hers age and gender. They could start the display whenever they wanted. Below the display were 10 adjectives to choose from. After all stimuli, the test stopped automatically. The questionnaire form is presented in appendix A.

7.3 Equipment

7.3.1 Motion capture system

Our OptiTrack optical motion capture system is manufactured by NaturalPoint, Inc. The company is specialized to high quality image tracking technology and different kinds of unique computer control devices (NaturalPoint, 2010). Our system uses 12 FLEX:V100R2 infrared cameras (Figure 21) that produces up to 100 FPS pre-processed video and full grayscale video.

Figure 21. OptiTrack optical infrared camera

Cameras have an onboard MJPEG image compression unit so that the movement in front of the cameras can be viewed in real time and with multiple cameras. Cameras have a resolution of 640 times 480 pixels. A standard 16 millimeter (5/8 inch) marker can be tracked up to 7 meters with precision of 0.1 millimeter (NaturalPoint, 2010).

There are a few possibilities how these cameras can be placed to the capture area. Six cameras is the minimum amount of cameras for full body motion capture (NaturalPoint, 2010). It is recommended that eight or more cameras to be used for improved tracking and much larger capture area. Our cameras were setup according to 12-camera truss setup which is presented in Figure 22.

System is controlled with ARENA™ motion capture software which can provide natural looking motions, animated characters and even some special effects (NaturalPoint, 2010). It automatically detects the actors' markers assignment and can place a skeletal model on top of those markers.

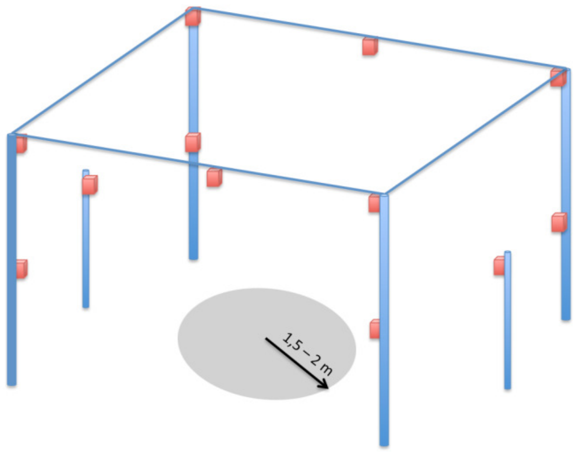


Figure 22. 12-camera truss setup

7.3.2 Displaying the point-light displays

Actors' movements were converted to point-light motion with right way up and then turning each display upside down. Each movement was about 3-4 seconds long. PL-displays were shown as white dots moving against black background and 25 points on actor's body were shown. With normal orientation the angle of view was from up-left position from which we could get as informative view as possible (Figure 18). The upside down displays were made by just turning around the normal displays. Both actors were right-handed so the knocking action was visually more informative from this angle. Displays were converted into DVD-format with resolution of 720 x 576 and 25 frames per second.

7.4 Statistical analysis methods

For statistical analysis of the data, we used the Student's t-test and the Wilcoxon signed-rank test. The t-test calculates the means of to data groups and assesses whether these groups are statistically different from each other.

The formula of the t-test can be seen as a ratio. It is also a metaphor of signal-to-noise ratio. The difference of two group means is divided by the variability of the groups.

$$t = \frac{Difference\ between\ means}{Variability\ \ of\ groups} = \frac{\bar{X}_T - \bar{X}_C}{SE(\bar{X}_T - \bar{X}_C)} \quad (1)$$

Usually the top part of the equation is easy to calculate. X_T and X_C are the mean values of two different data groups which we want to examine. These two groups are normal distributed. The bottom part is called the standard error of the difference (SE) and it is calculated by taking the variance of each group and dividing it with the number of people in the group. Variance can be calculated by simply taking the square of the standard deviation. The equation is shown below.

$$SE(\bar{X}_T - \bar{X}_C) = \sqrt{\frac{var_T}{n_T} + \frac{var_C}{n_C}} \tag{2}$$

$$t = \frac{\bar{X}_T - \bar{X}_C}{\sqrt{\frac{var_T}{n_T} + \frac{var_C}{n_C}}} = \frac{\bar{X}_T - \bar{X}_C}{\sqrt{\frac{STDEV_T^2}{n_T} + \frac{STDEV_C^2}{n_C}}} \tag{3}$$

The t-test result is used to look up in a table of significance to test whether the ratio is large enough so that the groups are really different. If the P-value (from table of significance) is smaller than 0,05, it means that the difference has "statistical significance". The t-test was used for examine the difference between orientations and comparing recognition rates against the level of chance.

The Wilcoxon signed-rank test is an alternative to the paired Student's t-test. It also compares the differences between group measurements, but this method can be used when the data cannot be assumed to be normally distributed. The Wilcoxon signed-rank test can be used for binomial data, which was the case also in our experiment. The test can be performed as following five steps:

1. Sort the data.
2. Discard the data points that are exactly equal to specified center (C) (usually the mean value).
3. Rank the remaining data points (N) by the absolute values of their deviation from C. The smallest deviation ranked 1. Assign mean ranks to tied points.
4. Calculate the sum of the ranks (S) for all the values that are above C and for the values below C. Define D as the smaller of the two rank sums.
5. Compare S to a table to define the significance level p.

The Wilcoxon signed-rank test suits better for data that has small differences. We used it to recognize the significance of differences between different emotions, orientation and the gender of the actor.

8 RESULT AND ANALYSIS

The data were obtained from 35 test subjects that fully completed the test. One set of data had to be excluded from the results because of misinterpretation of the test. We had overall 70 data samples for each video which meant that each emotion had 140 data samples (Male + female actors) per orientation (2 actors x 35 test subjects x 2 repeats = 140). The order of the point-light display stimuli was totally randomized by the PHP-software so every test subject got unique sequence of stimuli.

Test subjects used on average 28 minutes for doing this test set which included 80 videos.

8.1 Results

8.1.1 The overall recognition rate

First we present the overall emotion recognition rate with normal and upside down orientation. Figure 24 shows the percentage of correctly recognized emotion from 140 answers that each stimulus had. The figure is presenting the recognition rate of the normal orientation display with descending order. The upside down orientation's rate is marked on the right side of normal oriented results. The results are also presented in Figure 23.

By purely guessing the answers, the results should yield an emotion recognition rate of 10 percent because of the 10 emotion that the test subject could choose from. The level is marked into the Figure 24 with black line.

Figure 23. Recognition rates and percentages for different orientations

	Normal orientation		**Upside-down orientation**	
	Correct answers	%	Correct answers	%
Tired	62	44 %	54	39 %
Neutral	40	29 %	52	37 %
Angry	32	23 %	30	21 %
Excited	26	19 %	40	29 %
Happy	22	16 %	18	13 %
Sad	22	16 %	11	8 %
Relaxed	21	15 %	11	8 %
Weak	19	14 %	19	14 %
Strong	17	12 %	21	15 %
Afraid	16	11 %	11	8 %

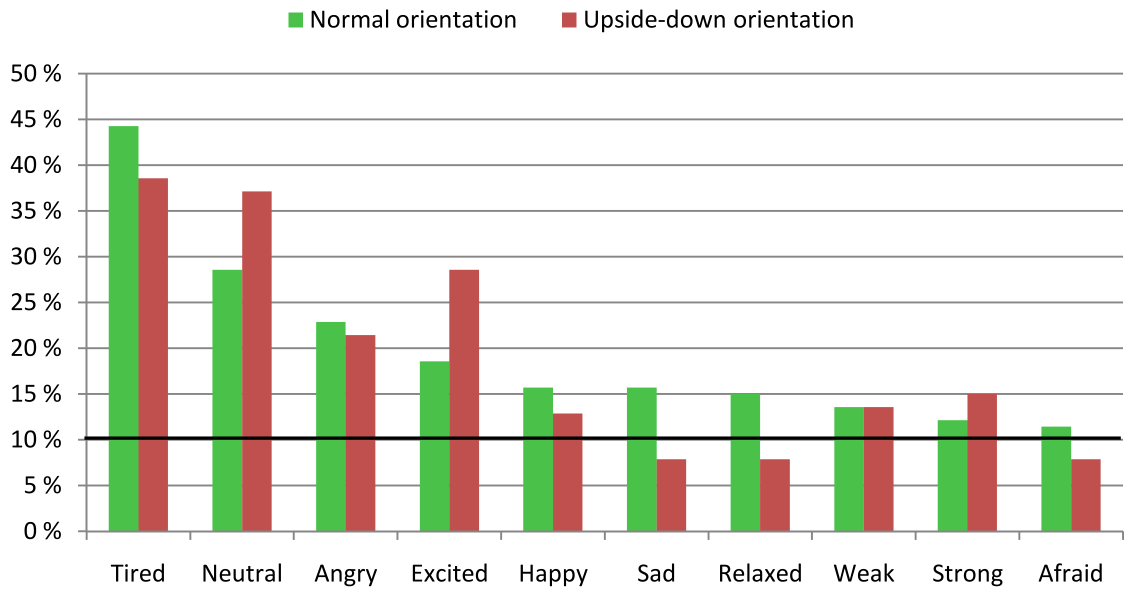


Figure 24. Recognition percentages arranged in descending order by normal orientation

8.1.2 Effect of gender in motion capture

We must also assess question of are there differences between male and female actor's acts? Can either one of these actor's stimuli be recognized more accurately that the other?

These results might indicate that some emotions are easier to recognize from male actor than female actress. There might be emotions that are especially recognizable from the movement of specific gender. Anger might be more manly emotions whereas excited could be related more often to emotions with females. The recognition rate should be evenly distributed between male and female actor, if the stimuli are not statistically different. The results are presented in Figures 25 and 26. Both figures are arranged in descending order with male actor's recognition rate.

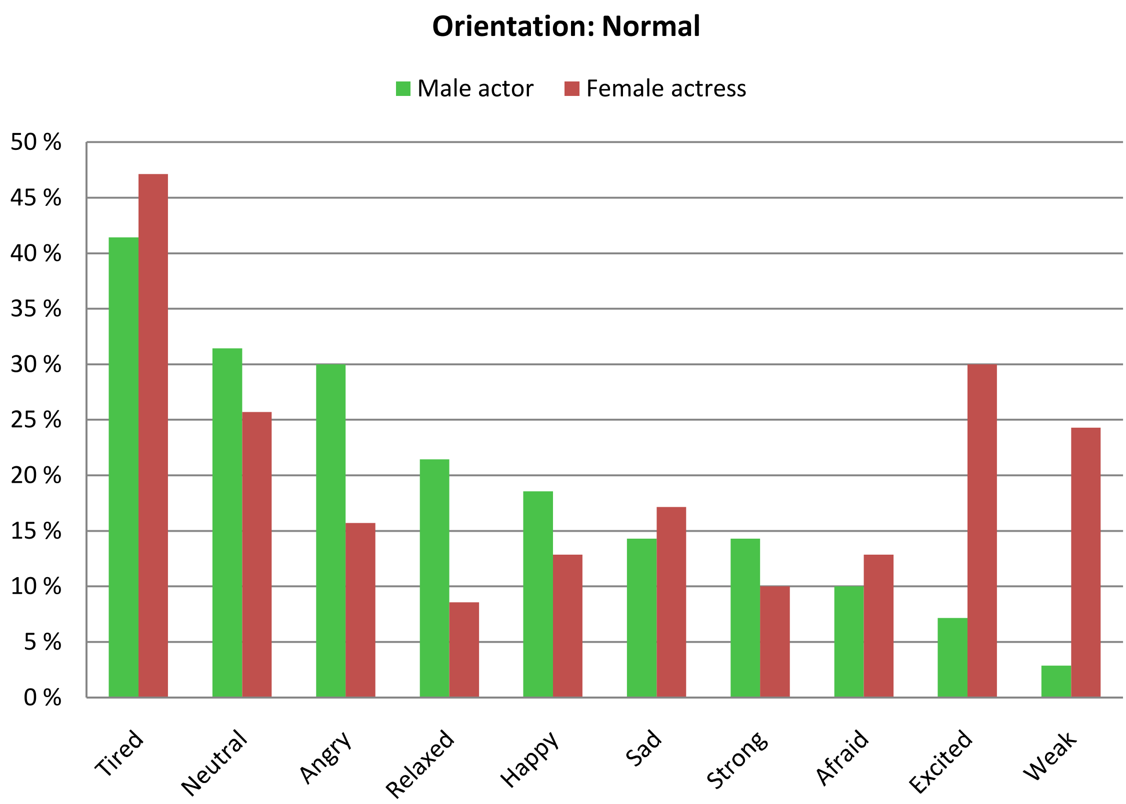


Figure 25. Recognition rates of male and female actors with normal orientation

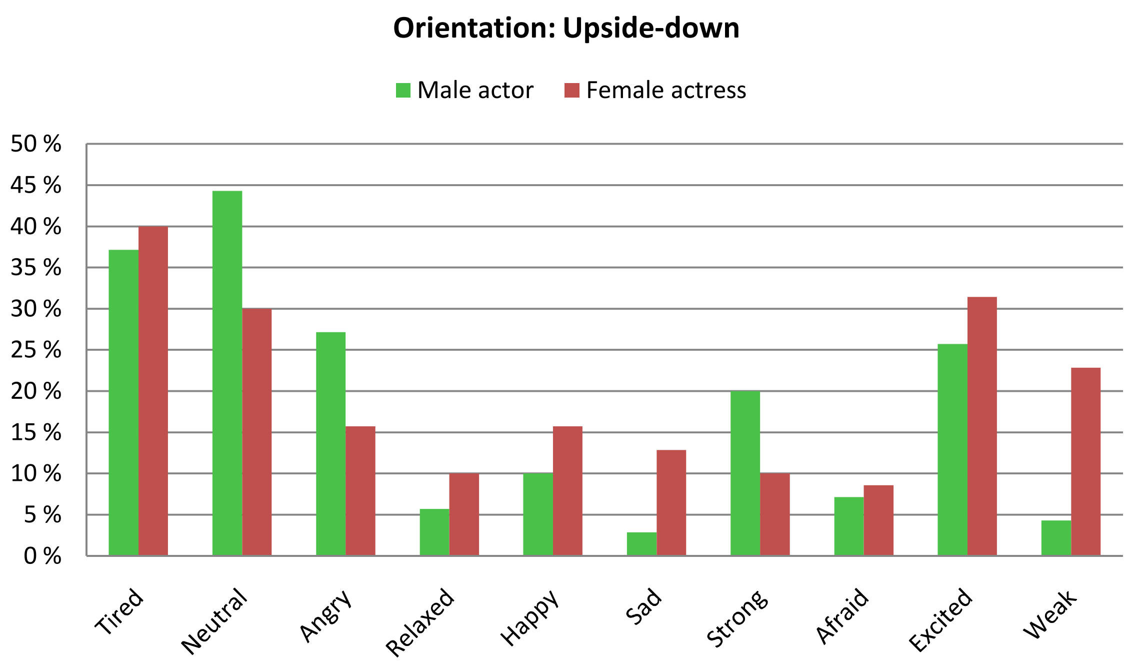


Figure 26. Recognition rates of male and female actors with upside down orientation

8.1.3 Test subject's gender dependent recognition

Next, we assess the test subject's gender dependent recognition rate. There might be difference of recognition when female test subjects are watching female actress's moving and vice versa. On the other hand, theory suggests, that there might not be any major difference between male and female test subjects (Pollick et al., 2002). These results are presented in Figures 27-30. The bar graphs are arranged according to descending order of female actress's normal oriented displays, male subject's recognition rates. Figure 31 below the figures 27-30 presents the average recognition rates for both gender test subjects watching male and female actor's acts.

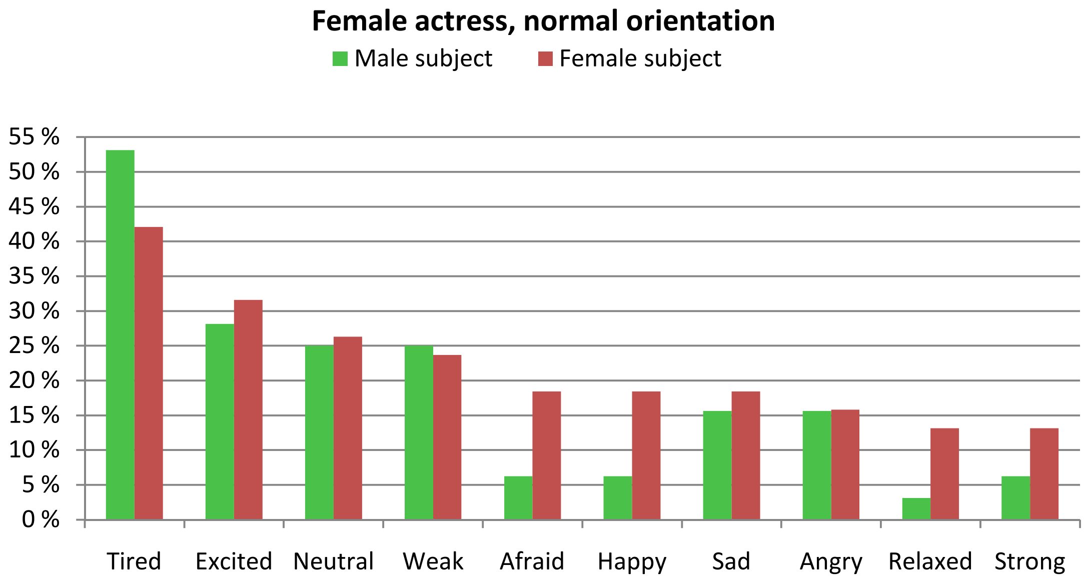


Figure 27. Female actress's recognition rates with normal orientation

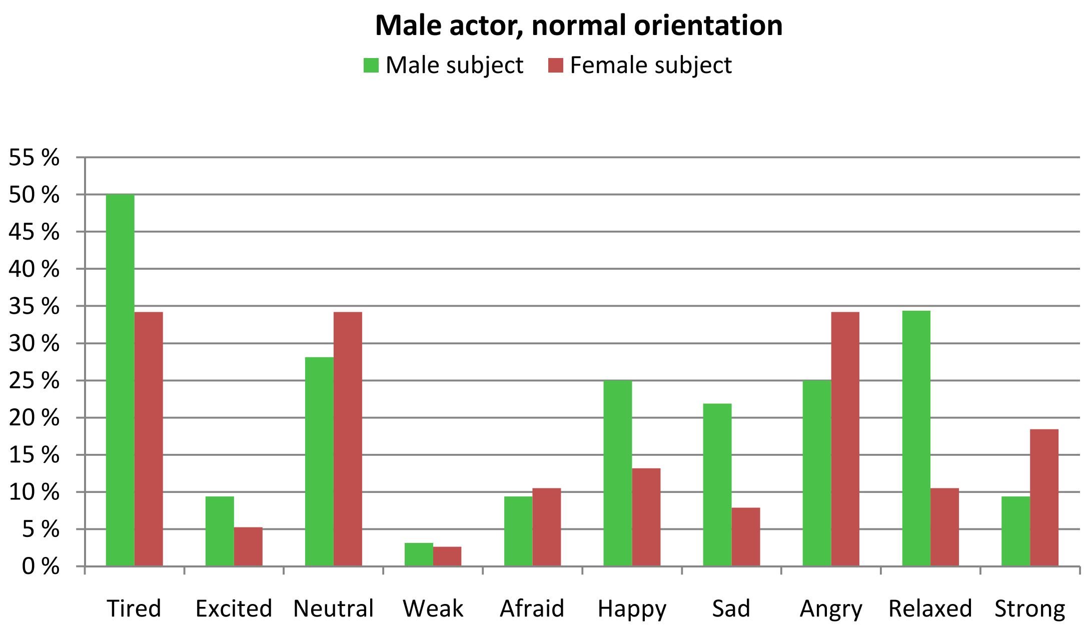


Figure 28. Male actor's recognition rates with normal orientation

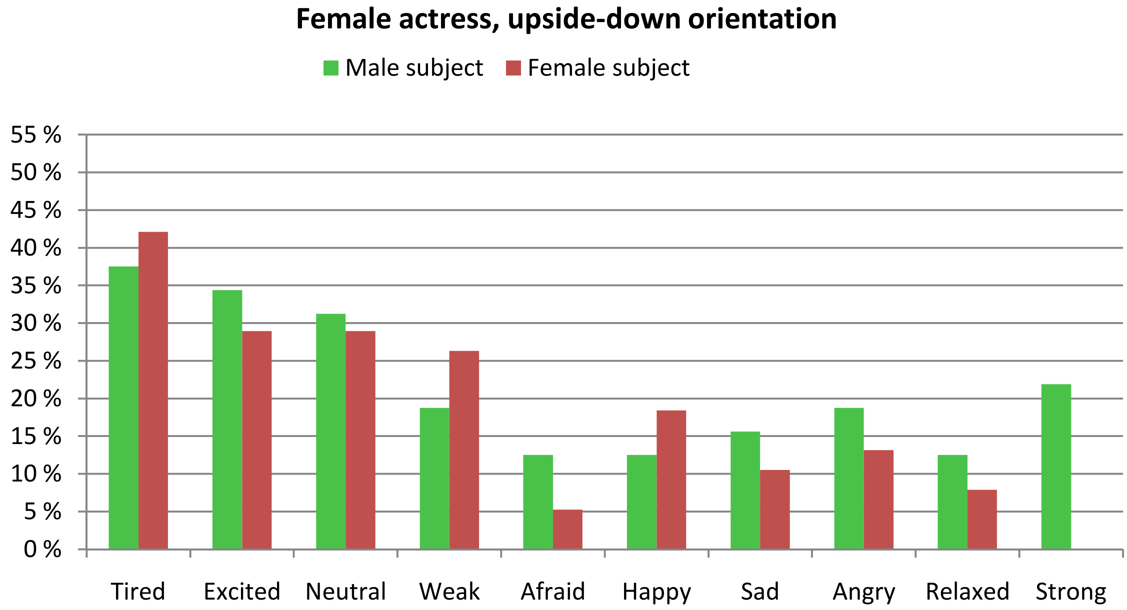


Figure 29. Female actress's recognition rates with upside down orientation

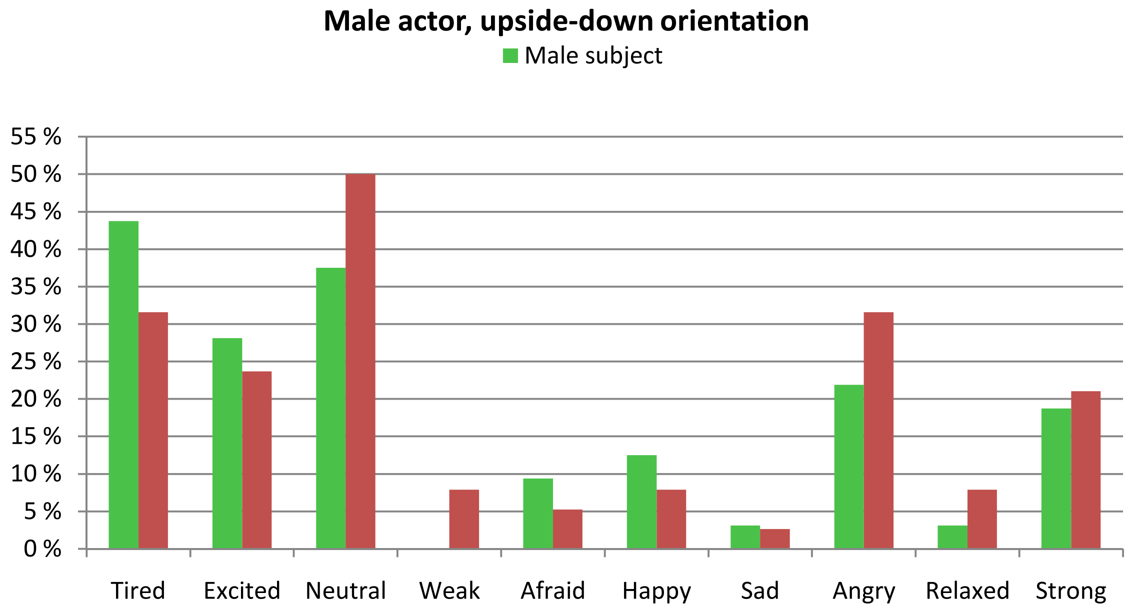


Figure 30. Male actor's recognition rates with upside down orientation

Figure 31. Average recognition rates for different orientation, gender of the actor and the test subject

	Normal orientation		Upside-down orientation	
Test subject \ Actor	**Female**	**Male**	**Female**	**Male**
Female	22 %	17 %	18 %	19 %
Male	18 %	22 %	22 %	18 %

8.1.4 Learning during the test

Finally, we assessed the possibility of learning process during the test. Do the participants learn to recognize the emotions from point-light displays more efficiently? Are the first 40 displays more difficult to recognize than latter 40 displays?

Since the order of displays was randomized over the whole test of 80 display stimuli, we couldn't get evenly distributed number of different emotion displays to first and latter 40 set of displays. This is why the statistical analysis of the data is not possible. The results are presented in figures 32 and 33.

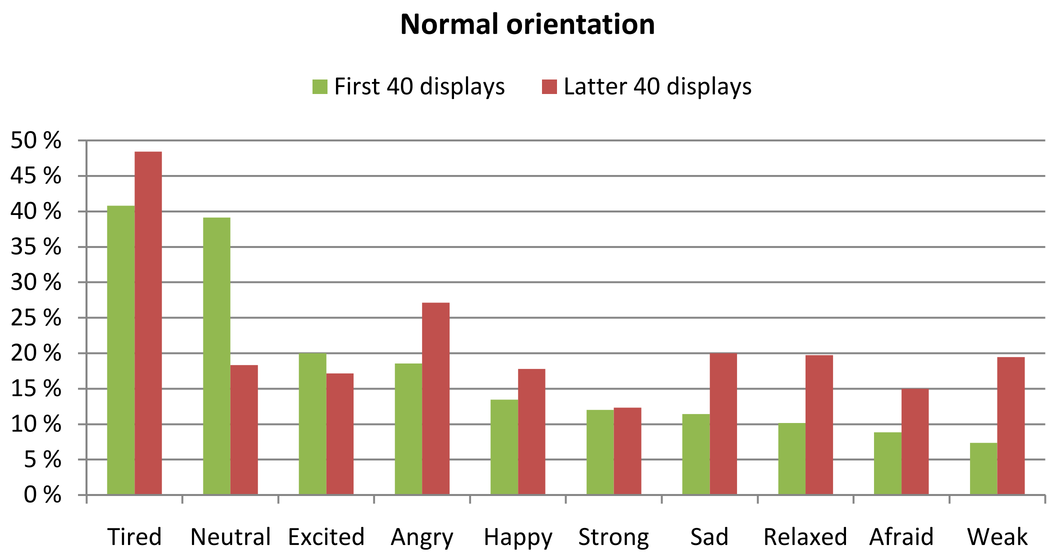


Figure 32. Learning process during normal oriented displays

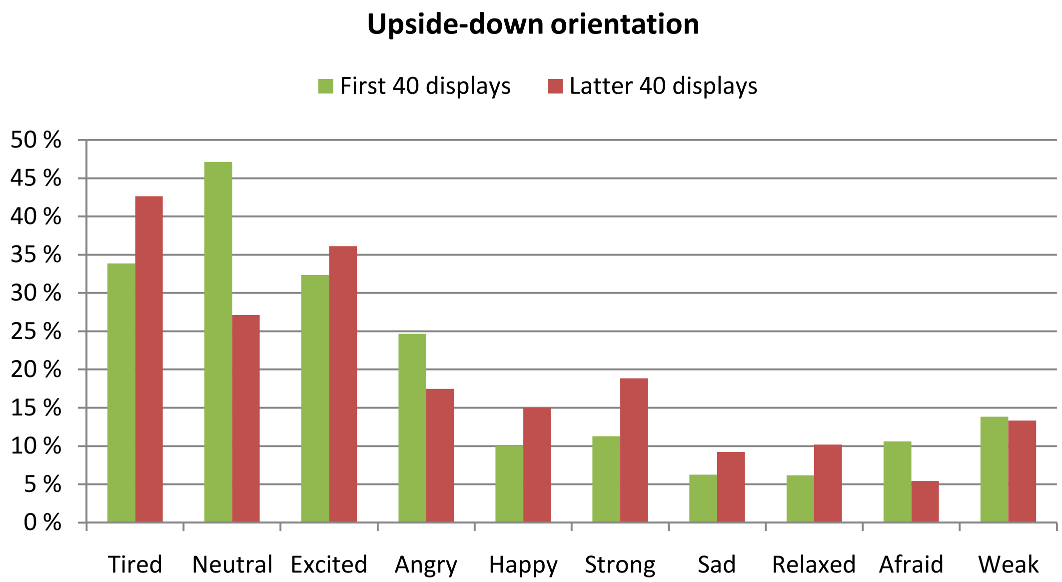


Figure 33. Learning process during upside down oriented displays

8.3 Analysis

8.3.1 The overall recognition

First we will start the analysis by looking the overall results of our experiment.

Figure 24 shows the recognition rates of different emotions for normal oriented displays and upside down oriented displays. The stimuli are arranged to descending order with normal oriented displays. Tired-stimulus has been the most recognized emotion with recognition rate of 44 % (Figure 23). The upside down oriented tired-stimulus was recognized with rate of 39 %, being the highest rate with upside down orientation. The lowest percentages were with normal oriented displays 11 % (afraid) and with upside down oriented displays 8 % (sad, relaxed and afraid).

All normal oriented display's recognition rate stayed above 10 %, which was the level of chance (average rate of recognition by purely guessing). With upside down orientation sad, relaxed and afraid stimuli fall under the 10 % level. The average recognition rate for normal oriented displays was 19,79 % and for upside down oriented displays 19,07 %. The t-test shows that both of these results are significantly better than the level of chance (normal oriented: $t(34)=9{,}0071$, $P < 0{,}0001$, two tailed; upside down: $t(34)=8{,}8918$, $P < 0{,}0001$, two tailed). From these results we can conclude that overall recognition rate for both orientations were above the level of chance.

A few anomalies can be detected from the Figure 24. The recognition rate of neutral, excited and strong stimuli have higher percentages in upside down orientation than in normal orientation. With excited stimulus, the percentage difference is 10 % and with neutral stimulus almost 9 %. The t-test shows that the differences' P-value with excited stimulus is $t(34)=1.9834$, $P = 0.0554$ (two tailed) and with neutral stimulus $t(34)=1{,}4148$, $P=0{,}1662$ (two tailed). Excited stimulus' difference is not quite statistically significant whereas neutral stimulus' difference doesn't have any significance.

Another interesting result can be seen if we review all the answers that our test subjects have given to each stimulus category. We can also detect anomalies in answering patterns, e.g. the test subjects have answered more often "neutral" than "sad". To assess more thoroughly the recognition rate, we must consider mix-ups with the stimulus displays. In other words, some stimulus might have been mixed-up with other stimulus. A confusion matrix (Figure 34) shows in percentages all the responses that our test subjects gave to each emotion stimulus. The correct answer rate is shown in the diagonal of the matrix.

		Stimulus									
		Afraid	Angry	Excited	Happy	Neutral	Relaxed	Sad	Strong	Tired	Weak
Response	Afraid	9,64 %	2,50 %	3,21 %	3,93 %	13,57 %	2,50 %	10,00 %	2,50 %	3,57 %	29,29 %
	Angry	2,14 %	22,14 %	21,43 %	12,86 %	3,21 %	10,00 %	2,86 %	18,57 %	6,79 %	1,07 %
	Excited	10,71 %	12,86 %	23,57 %	27,50 %	3,21 %	11,07 %	2,86 %	9,64 %	3,21 %	5,36 %
	Happy	10,71 %	10,00 %	13,57 %	14,29 %	4,64 %	8,93 %	3,57 %	8,21 %	1,43 %	3,21 %
	Neutral	44,29 %	20,71 %	18,93 %	20,36 %	32,86 %	25,00 %	19,29 %	27,86 %	6,43 %	12,86 %
	Relaxed	8,21 %	6,07 %	4,29 %	7,14 %	7,86 %	11,43 %	6,79 %	7,50 %	7,50 %	2,86 %
	Sad	1,07 %	1,79 %	0,71 %	0,36 %	6,43 %	4,29 %	11,79 %	2,86 %	13,93 %	21,07 %
	Strong	4,64 %	14,64 %	11,43 %	7,50 %	4,64 %	7,50 %	4,29 %	13,57 %	2,86 %	1,07 %
	Tired	4,29 %	8,21 %	2,14 %	3,57 %	12,86 %	14,64 %	24,29 %	7,50 %	41,43 %	9,64 %
	Weak	4,29 %	1,07 %	0,71 %	2,50 %	10,71 %	4,64 %	14,29 %	1,79 %	12,86 %	13,57 %

Figure 34. Confusion matrix to all responses (regardless orientation). If stimulus has more responses than the correct one, the cell is in red. If stimulus has responses more than 10 % (the level of chance), the cell is in yellow.

From the confusion matrix we can observe how different stimuli have been mixed-up with each other. As early estimated in theory Pollick et al. (2001), sad and weak stimuli have been mixed-up with afraid and tired stimuli. For instance the weak-stimulus has been confused with sad and afraid. There are more responses to angry and neutral stimulus than the correct answers in strong stimulus. Angry, excited, neutral and tired stimuli have been recognized most conclusively since there are no responses that overcome their rate. The worst mix-up is with afraid and sad stimulus. Three responses overtake those categories' correct answer.

One discovery from the confusion matrix is the amount of neutral responses. Almost all stimuli have been mixed-up with the neutral stimulus for some extent. This might implicate that the test subjects have been confused during the experiment and they have used this response as a "don't know"-answer. Whenever a test subject has not recognized enough from the stimulus, they have answered "neutral" to it. The increase of neutral recognition rate in upside down oriented displays (Figure 24) also suggests that the neutral answer might have been used more often than the other answer possibilities.

We also wanted to know if there are differences in how test subjects answer in normal oriented and upside down oriented displays. For this we need two confusion matrices more, shown in Figures 35 and 36 below. If we compare these to matrices we can notice the increased amount overtaking confusions in upside down oriented displays (Figure 36). The change of orientation might have some effect to the amount of confusion. Relaxed, afraid, sad, happy and weak stimuli have much worse answering distribution in upside down orientation than in normal orientation. In other words, some other stimulus' response rate overtakes the correct answer's percentage. In upside down oriented displays (Figure 36), tired stimulus is more often confused to weak stimulus than sad stimulus which is opposite to normal oriented displays (Figure 35).

		Stimulus									
		Afraid	Angry	Excited	Happy	Neutral	Relaxed	Sad	Strong	Tired	Weak
Response	Afraid	11,43 %	0,71 %	3,57 %	6,43 %	12,86 %	4,29 %	7,86 %	3,57 %	2,86 %	34,29 %
	Angry	2,86 %	22,86 %	19,29 %	12,14 %	2,14 %	7,14 %	2,14 %	17,86 %	8,57 %	1,43 %
	Excited	11,43 %	11,43 %	18,57 %	22,86 %	1,43 %	12,86 %	0,00 %	5,71 %	0,71 %	6,43 %
	Happy	10,71 %	9,29 %	15,71 %	15,71 %	4,29 %	11,43 %	2,14 %	6,43 %	0,71 %	2,14 %
	Neutral	42,14 %	22,14 %	15,71 %	17,14 %	28,57 %	19,29 %	17,14 %	34,29 %	4,29 %	7,86 %
	Relaxed	8,57 %	5,71 %	7,14 %	9,29 %	9,29 %	15,00 %	7,86 %	9,29 %	9,29 %	2,14 %
	Sad	1,43 %	2,14 %	0,71 %	0,71 %	7,86 %	2,14 %	15,71 %	2,14 %	16,43 %	25,00 %
	Strong	4,29 %	15,71 %	17,14 %	9,29 %	5,00 %	5,71 %	4,29 %	12,14 %	2,14 %	0,71 %
	Tired	4,29 %	8,57 %	1,43 %	3,57 %	13,57 %	18,57 %	27,14 %	7,14 %	44,29 %	6,43 %
	Weak	2,86 %	1,43 %	0,71 %	2,86 %	15,00 %	3,57 %	15,71 %	1,43 %	10,71 %	13,57 %

Figure 35. Confusion matrix for normal oriented display stimuli

		Stimulus									
		Afraid	Angry	Excited	Happy	Neutral	Relaxed	Sad	Strong	Tired	Weak
Response	Afraid	7,86 %	4,29 %	2,86 %	1,43 %	14,29 %	0,71 %	12,14 %	1,43 %	4,29 %	24,29 %
	Angry	1,43 %	21,43 %	23,57 %	13,57 %	4,29 %	12,86 %	3,57 %	19,29 %	5,00 %	0,71 %
	Excited	10,00 %	14,29 %	28,57 %	32,14 %	5,00 %	9,29 %	5,71 %	13,57 %	5,71 %	4,29 %
	Happy	10,71 %	10,71 %	11,43 %	12,86 %	5,00 %	6,43 %	5,00 %	10,00 %	2,14 %	4,29 %
	Neutral	46,43 %	19,29 %	22,14 %	23,57 %	37,14 %	30,71 %	21,43 %	21,43 %	8,57 %	17,86 %
	Relaxed	7,86 %	6,43 %	1,43 %	5,00 %	6,43 %	7,86 %	5,71 %	5,71 %	5,71 %	3,57 %
	Sad	0,71 %	1,43 %	0,71 %	0,00 %	5,00 %	6,43 %	7,86 %	3,57 %	11,43 %	17,14 %
	Strong	5,00 %	13,57 %	5,71 %	5,71 %	4,29 %	9,29 %	4,29 %	15,00 %	3,57 %	1,43 %
	Tired	4,29 %	7,86 %	2,86 %	3,57 %	12,14 %	10,71 %	21,43 %	7,86 %	38,57 %	12,86 %
	Weak	5,71 %	0,71 %	0,71 %	2,14 %	6,43 %	5,71 %	12,86 %	2,14 %	15,00 %	13,57 %

Figure 36. Confusion matrix for upside down oriented display stimuli

In next Figure (37), we have calculated each test subject's recognition rate for each stimulus. From these recognition rates we could calculate the standard deviation. We also used these individual recognition rates to compare them to the level of chance (in Figure 38).

The Figure 37 below shows the average recognition rate of each stimulus (regardless of the orientation and gender of the actor) and range of one standard deviation to plus and minus sides. Because percentages are not normally distributed, a few stimuli's standard deviation falls under the zero.

From Figure 37 we can conclude which stimuli have been recognized and which might have been somewhat guessing. The average recognition rate of tired and neutral stimuli is well above the level of chance as earlier established. Also angry and excited stimuli's deviations are mostly above the level of chance, so we can conclude that these stimuli the test subjects have recognized. For all other stimuli, we cannot say definitely that the test subjects would have recognized them, although the average recognition rate is above 10 % (except afraid being 9,64%).

Figure 37. Average recognition rates and their standard deviations

Figure 38. Average recognition rate, standard deviation and P-value from t-test

	Afraid	Angry	Excited	Happy	Neutral	Relaxed	Sad	Strong	Tired	Weak
Average	9,64 %	22,14 %	23,57 %	14,29 %	32,86 %	11,43 %	11,79 %	13,57 %	41,43 %	13,57 %
STDEV	10,54 %	14,57 %	16,26 %	17,45 %	15,48 %	10,66 %	13,54 %	16,43 %	19,12 %	11,09 %
P (2-tail)	0,842306	0,000021	0,000021	0,155474	0,000000	0,433521	0,440608	0,207146	0,000000	0,065145

From Figure 38 we can find the t-test results for each stimulus against the level of chance. These results support our previous conclusions. Angry, excited, neutral and tired are considered to be significantly different than the level of chance whereas all others are not considered as significantly different.

8.3.2 The effect of gender

Next we want to know, if there are differences between male and female actor playing in the display stimulus. We also wanted to know how the orientation affects to the recognition rate with male and female actors.

First we examine the Figure 25 with normal orientation. We can notice male actor has been recognized significantly (in percentages) better than female actress with angry and relaxed stimuli. Also happy, neutral and strong stimuli have higher recognition rate with male actor. Excited and weak stimuli have better rate with female actress. The difference between those categories is over 20 percent. Female actress has also been recognized better with tired, sad and afraid stimuli. One noticeable result is the amount of male actor's weak stimulus' recognition. Only 3 % of our test subjects have recognized this stimulus. More detailed numbers of both genders and orientations are presented in an appendix B.

Figure 26 shows the results with upside down orientation. We can still notice the difference in recognition rate between male and female actor with weak emotion. Other notable differences can be noticed with neutral, angry and strong stimuli to male actor's advance and the rest of the stimuli to female actress' advance. If we want to know which of these differences are significant, we must use the Wilcoxon signed-rank test to test them statistically. Results are shown in Figure 39 below.

Figure 39. P-values from Wilcoxon signed-rank test, male against female

Normal oriented, male against female										
	Tired	Neutral	Angry	Relaxed	Happy	Sad	Strong	Afraid	Excited	Weak
P(2-tail)	0,5892	0,5419	0,0293	0,1052	0,4413	0,6965	0,3733	0,7414	0,0044	0,002
Upside-down oriented, male against female										
	Tired	Neutral	Angry	Relaxed	Happy	Sad	Strong	Afraid	Excited	Weak
P(2-tail)	0,7263	0,234	0,1381	0,4354	0,3371	0,0784	0,1285	0,7441	0,5755	0,0074

The signed-rank test gives a two tailed probability as an answer. If the probability is under 0,05, the difference is considered to be statistically significant as in the t-test. From Figure 39, we can conclude which stimulus' difference is significant. With normal oriented displays, the difference between male and female actor is considered to be significant with angry (P=0,0293), excited (P=0,0044) and weak (P=0,002) stimuli. Relaxed emotion is close to being significant but not quite. With upside down orientation only weak stimulus' difference (P=0,0074) between male and female actor is significant. Sad's difference isn't quite enough to be significant. Angry and strong decreases even more.

8.3.3 The effect of orientation and gender

Next we must also assess the differences between the two orientations. When comparing the Figures 25 and 26, there has been a decrease of five percent in recognition rate with tired, angry and afraid stimuli in upside down orientation (Figure 26). With relaxed, sad and happy stimuli, the male actor's recognition rate has also deceased. With excited stimulus', the recognition rate has increased almost 20 % in upside down orientation. Biggest differences are with male actor's emotions recognitions. Finally, to support our earlier theory of neutral answers, there has been a significant (in percentages) increase in recognition rate with upside down orientation. This suggest that the neutral has been used as a "don't know" –answer. To support our visual findings, we calculated the P-values with the Wilcoxon's signed-rank test for the two orientations. The results are shown in Figure 40 below.

Figure 40. P-values from Wilcoxon signed-rank test, normal against upside down orientation

Male actor, normal against upside-down orientation										
	Tired	Neutral	Angry	Relaxed	Happy	Sad	Strong	Afraid	Excited	Weak
P(2-tail)	0,5961	0,1835	0,6965	0,0135	0,1098	0,0186	0,4009	0,5712	0,0105	0,5712
Female actress, normal against upside-down orientation										
	Tired	Neutral	Angry	Relaxed	Happy	Sad	Strong	Afraid	Excited	Weak
P(2-tail)	0,3953	0,6818	1	0,7441	0,6241	0,4715	1	0,4902	0,8572	0,8493

The biggest differences are in the male actor's stimuli with different orientations. This means that the change of orientation did not affect so much to the female actress's recognition rate. Male actor has a significant difference between orientations with excited (P=0,0105), relaxed (P=0,0135) and sad (P=0,0186) stimuli. Happy has also a notable difference but a P-value of 0,1098 means that it is not significant.

We can also go further with this analysis and see how our test subjects have answered to different stimuli. In other words we can make confusion matrices describing the distribution of answers. We made four of these matrices with two genders and two orientations. These matrices are presented in Figures 41-44 below.

		Stimulus									
		Afraid	**Angry**	**Excited**	**Happy**	**Neutral**	**Relaxed**	**Sad**	**Strong**	**Tired**	**Weak**
Response	**Afraid**	**10,00 %**	0,00 %	0,00 %	0,00 %	2,86 %	1,43 %	1,43 %	2,86 %	0,00 %	34,29 %
	Angry	2,86 %	**30,00 %**	27,14 %	22,86 %	4,29 %	10,00 %	4,29 %	31,43 %	14,29 %	0,00 %
	Excited	11,43 %	17,14 %	**7,14 %**	12,86 %	2,86 %	20,00 %	0,00 %	10,00 %	1,43 %	8,57 %
	Happy	8,57 %	11,43 %	12,86 %	**18,57 %**	5,71 %	22,86 %	1,43 %	10,00 %	0,00 %	1,43 %
	Neutral	42,86 %	20,00 %	15,71 %	12,86 %	**31,43 %**	12,86 %	32,86 %	21,43 %	5,71 %	5,71 %
	Relaxed	11,43 %	7,14 %	8,57 %	15,71 %	17,14 %	**21,43 %**	14,29 %	8,57 %	15,71 %	1,43 %
	Sad	1,43 %	1,43 %	0,00 %	1,43 %	4,29 %	1,43 %	**14,29 %**	1,43 %	18,57 %	42,86 %
	Strong	7,14 %	10,00 %	25,71 %	14,29 %	10,00 %	8,57 %	8,57 %	**14,29 %**	2,86 %	1,43 %
	Tired	1,43 %	2,86 %	2,86 %	0,00 %	17,14 %	1,43 %	18,57 %	0,00 %	**41,43 %**	1,43 %
	Weak	2,86 %	0,00 %	0,00 %	1,43 %	4,29 %	0,00 %	4,29 %	0,00 %	0,00 %	**2,86 %**

Figure 41. Confusion matrix of male actor with normal orientation

		Stimulus									
		Afraid	**Angry**	**Excited**	**Happy**	**Neutral**	**Relaxed**	**Sad**	**Strong**	**Tired**	**Weak**
Response	**Afraid**	**12,86 %**	1,43 %	7,14 %	12,86 %	22,86 %	7,14 %	14,29 %	4,29 %	5,71 %	34,29 %
	Angry	2,86 %	**15,71 %**	11,43 %	1,43 %	0,00 %	4,29 %	0,00 %	4,29 %	2,86 %	2,86 %
	Excited	11,43 %	5,71 %	**30,00 %**	32,86 %	0,00 %	5,71 %	0,00 %	1,43 %	0,00 %	4,29 %
	Happy	12,86 %	7,14 %	18,57 %	**12,86 %**	2,86 %	0,00 %	2,86 %	2,86 %	1,43 %	2,86 %
	Neutral	41,43 %	24,29 %	15,71 %	21,43 %	**25,71 %**	25,71 %	1,43 %	47,14 %	2,86 %	10,00 %
	Relaxed	5,71 %	4,29 %	5,71 %	2,86 %	1,43 %	**8,57 %**	1,43 %	10,00 %	2,86 %	2,86 %
	Sad	1,43 %	2,86 %	1,43 %	0,00 %	11,43 %	2,86 %	**17,14 %**	2,86 %	14,29 %	7,14 %
	Strong	1,43 %	21,43 %	8,57 %	4,29 %	0,00 %	2,86 %	0,00 %	**10,00 %**	1,43 %	0,00 %
	Tired	7,14 %	14,29 %	0,00 %	7,14 %	10,00 %	35,71 %	35,71 %	14,29 %	**47,14 %**	11,43 %
	Weak	2,86 %	2,86 %	1,43 %	4,29 %	25,71 %	7,14 %	27,14 %	2,86 %	21,43 %	**24,29 %**

Figure 42. Confusion matrix of female actress with normal orientation

		Stimulus									
		Afraid	Angry	Excited	Happy	Neutral	Relaxed	Sad	Strong	Tired	Weak
Response	Afraid	**7,14 %**	4,29 %	2,86 %	0,00 %	1,43 %	0,00 %	2,86 %	0,00 %	2,86 %	27,14 %
	Angry	2,86 %	**27,14 %**	38,57 %	22,86 %	8,57 %	20,00 %	5,71 %	35,71 %	8,57 %	1,43 %
	Excited	8,57 %	22,86 %	**25,71 %**	28,57 %	8,57 %	18,57 %	8,57 %	21,43 %	11,43 %	7,14 %
	Happy	10,00 %	7,14 %	5,71 %	**10,00 %**	4,29 %	8,57 %	7,14 %	14,29 %	1,43 %	5,71 %
	Neutral	51,43 %	24,29 %	15,71 %	21,43 %	**44,29 %**	22,86 %	32,86 %	2,86 %	15,71 %	18,57 %
	Relaxed	5,71 %	7,14 %	1,43 %	5,71 %	10,00 %	**5,71 %**	7,14 %	2,86 %	5,71 %	2,86 %
	Sad	1,43 %	0,00 %	1,43 %	0,00 %	4,29 %	4,29 %	**2,86 %**	1,43 %	8,57 %	31,43 %
	Strong	4,29 %	2,86 %	7,14 %	7,14 %	7,14 %	11,43 %	7,14 %	**20,00 %**	4,29 %	0,00 %
	Tired	4,29 %	2,86 %	1,43 %	4,29 %	11,43 %	5,71 %	20,00 %	1,43 %	**37,14 %**	1,43 %
	Weak	4,29 %	1,43 %	0,00 %	0,00 %	0,00 %	2,86 %	5,71 %	0,00 %	4,29 %	**4,29 %**

Figure 43. Confusion matrix of male actor with upside down orientation

		Stimulus									
		Afraid	Angry	Excited	Happy	Neutral	Relaxed	Sad	Strong	Tired	Weak
Response	Afraid	**8,57 %**	4,29 %	2,86 %	2,86 %	27,14 %	1,43 %	21,43 %	2,86 %	5,71 %	21,43 %
	Angry	0,00 %	**15,71 %**	8,57 %	4,29 %	0,00 %	5,71 %	1,43 %	2,86 %	1,43 %	0,00 %
	Excited	11,43 %	5,71 %	**31,43 %**	35,71 %	1,43 %	0,00 %	2,86 %	5,71 %	0,00 %	1,43 %
	Happy	11,43 %	14,29 %	17,14 %	**15,71 %**	5,71 %	4,29 %	2,86 %	5,71 %	2,86 %	2,86 %
	Neutral	41,43 %	14,29 %	28,57 %	25,71 %	**30,00 %**	38,57 %	10,00 %	40,00 %	1,43 %	17,14 %
	Relaxed	10,00 %	5,71 %	1,43 %	4,29 %	2,86 %	**10,00 %**	4,29 %	8,57 %	5,71 %	4,29 %
	Sad	0,00 %	2,86 %	0,00 %	0,00 %	5,71 %	8,57 %	**12,86 %**	5,71 %	14,29 %	2,86 %
	Strong	5,71 %	24,29 %	4,29 %	4,29 %	1,43 %	7,14 %	1,43 %	**10,00 %**	2,86 %	2,86 %
	Tired	4,29 %	12,86 %	4,29 %	2,86 %	12,86 %	15,71 %	22,86 %	14,29 %	**40,00 %**	24,29 %
	Weak	7,14 %	0,00 %	1,43 %	4,29 %	12,86 %	8,57 %	20,00 %	4,29 %	25,71 %	**22,86 %**

Figure 44. Confusion matrix of female actress with upside down orientation

These matrices are constructed same way as before. If we first take a glance of normal oriented matrices and compare the differences between male and female actor (Figures 41 and 42). Male actor's emotions have been recognized with varying rates of recognition (Figure 41). Angry, neutral and tired emotions have been recognized without bigger confusions. Most of the confusion with male actor comes in afraid, excited, sad and weak stimuli. The female actress has been recognized accurately with excited and tired stimuli (Figure 42). Overall, the female actress has been recognized better than the male with normal orientation. With upside down stimulus, male actor has same stimuli at the top (Figure 43). The biggest difference is the increase of confusion in sad, relaxed and weak emotions while female actress's recognition confusion (Figure 44) amount has been nearly unchanged.

There are a few observations that are interesting. The amount of neutral answers with afraid-stimulus is significant (in percentages). In all figures (41-44) the amount of answers exceeds 40 % and goes even over 50 % with male actor upside down orientation. Another observation is the significant (in percentages) decrease of neutral answers with strong stimulus in male's upside down oriented matrix (Figure 43). Happy stimulus has been quite often been confused with excited stimulus.

We also want to examine which stimuli in which orientation and gender of actor has a statistically significant difference. If we combine the orientations, we can examine the effect that gender had and vice versa. We used the t-test and got the P-values for all the stimuli. The results are shown in Figure 45 below.

Figure 45. T-test results for different combinations

	Male and female combined, normal against upside-down orientation									
	Afraid	**Angry**	**Excited**	**Happy**	**Neutral**	**Relaxed**	**Sad**	**Strong**	**Tired**	**Weak**
P (2-tail)	0,324375	0,76785	0,055446	0,422144	0,166223	0,048265	0,009478	0,512886	0,324375	1
	Orientations combined, male against female actor									
	Afraid	**Angry**	**Excited**	**Happy**	**Neutral**	**Relaxed**	**Sad**	**Strong**	**Tired**	**Weak**
P (2-tail)	0,608686	0,013866	0,032907	1	0,123675	0,422144	0,130027	0,057554	0,556159	0,00001384

We first examine the upper part of Figure 45 in which we have compared orientations regardless of the gender. When comparing the two orientations, sad and relaxed stimuli have enough small P-value (Sad: P=0,00948; Relaxed: P=0,04826) for the difference to have significance. Excited stimulus has P-value (P=0,05545) just on the threshold so we can say that there is little significance in its difference between the orientations. The lower part of Figure 45, we have a comparison of gender of the actor regardless the orientation. Under the threshold fall angry (P=0,01387), excited (P=0,03291) and weak (P=0,000014) stimuli and strong stimulus (P=0,05755) is just on the threshold. Angry, excited, weak and strong stimuli have statistically significant difference between genders of the actor.

Finally we calculate if there are differences between orientations or the gender of the actors regardless of the emotion/stimulus. We combine test subjects' responses into four categories based on two orientations and two genders of the actors. After that we can calculate the t-test for each possible combination. The difference between male and female actor is not statistically significant with normal oriented displays (t(34)=0,6357 P=0,5295) nor is it with upside down oriented displays (t(34)=0,5766 P=0,5680). We also couldn't find difference between orientations with male actor (t(34)=0,4047 P=0,6882) or with female actress (t(34)=0,3448 P=0,7324). If we further on combine the results we can calculate overall P-values for difference between orientations and between different genders. The t-test results show that there are no difference in orientation recognition (t(34)=0,5606 P=0,5788) or in gender of the actor (t(34)=0,8212 P=0,4173).

Next we analyze the data for the effect of participant's gender. For this data, the statistical analysis was impossible because of the experiment setup, different number of male and female test subjects and total randomizing of the display stimuli. So we settle for only visual analysis of the data and try to find some trends based on these bar graphs. We can also use past results to estimate the significances of the differences.

In Figures 27-30 we have bar graphs presenting data of how male and female test subjects have answered to differently oriented and gendered stimulus. First examine the Figure 27 with female actress and normal orientation. We see quite evenly distributed recognition rate between male and female test subjects. Exceptions can be noticed with afraid, happy and relaxed stimuli. These stimuli have notable difference. We can compare these differences to already examined differences and estimate that all of them might have statistically significant difference. The male actor's recognition rates (Figure 28) have been arranged into same order as female actors. We notice the differences in excited, weak, happy and relaxed stimuli between genders. Excited and weak stimuli's recognition rates have decreased over 20 % whereas male subjects' rate for happy and relaxed is increased the same amount.

Next we are comparing the results of upside down oriented recognition (Figures 29 and 30). The recognition rates between male and female subjects are quite evenly distributed with only one exception. Female test subjects haven't recognized the upside down presented female actresses performed strong stimulus (Figure 29). They've had this display shown 38 times and none of those times the answer was correct. With male actor the same situation is with weak stimulus (Figure 30). None of male subjects have recognized it correctly. Other notable difference is the increase of neutral and angry recognition rate and a decrease of female subject's weak recognition rate when comparing different actors in upside down orientation (Figures 29 and 30). But considering the past results, we estimate that there are no statistically significance differences between these stimuli.

We can also compare the results from each gendered actor with different orientations. The biggest differences with female actress are in afraid and strong stimuli (Figures 27 and 29). There is somewhat significant increase (in percentages) with afraid –stimulus' female subject's recognition rate and significant difference with both gendered subjects in strong stimuli when comparing these with different orientations. Overall the recognition rates haven't changed significantly between orientations based on past results. But when we are comparing the results of male actor's recognition rates we see a lot more variation (Figures 28 and 30). Excited, happy, sad and relaxed stimuli rate has decreased notably. Neutral and strong stimuli increase is about 10 %. We estimate that the decrease of excited, male subject's sad and relaxed stimuli are statistically significant, based on the past results. None of the increases were so big that they could be significant.

8.3.4 Analysis of learning

The best way to get some confirmation to the suggestion of neutral response being the confusion answer is to study how the responses have changed from the first period of the test to the latter period.

Figures 32 and 33 present bar graphs of possible learning process during our experiment. In other words, we want to know how the recognition rates changes from first part of 40 displays to latter part of 40 displays. From Figure 32, we notice how almost all “latter” bars are above the “first” bars. This indicates that there has been learning in test subjects during the experiment. The decease of neutral rate also suggests that the test subjects haven’t just answered “I don’t know” but they’ve seen something on those displays. Also the upside down orientation (Figure 33) has same kind of trend with neutral answers decreasing. These finding just indicates a possibility for learning but there is no statistical evidence of the difference between the parts.

9 CONCLUSIONS

In this master thesis, we used optical motion capture system to collect “a knocking at the door” motion with 10 different emotions. Captures were converted into point-light displays which were shown to our test subjects as a stimulus. The test subjects were asked to describe the actor’s emotion or overall feeling with one pre-given adjective. A PHP-questionnaire form was used to collect the data.

Overall we got good results out of our experiment. We obtained an overall recognition rate of 19,79 % for normal oriented displays and 19,07 % for upside down oriented displays. Both of these rates of recognition were significantly better than the level of chance. This means that our test subjects have recognized emotions from point-light displays. The difference between orientations was surprisingly small. The difference was determined not being statistically significant.

The best recognized stimulus was tired in both orientations. This might be caused by a little bit longer stimulus display. Tired, as an emotion, is longer lasting so it might be easier to recognize from longer stimulus. The lowest recognition rate was with afraid – stimulus with normal orientation. Sad, relaxed and afraid stimuli had the lowest recognition rate with upside down orientation.

The second best recognized stimulus was neutral. Neutral, as an emotion, is very simple and it doesn’t hold any specific motional cues. Whenever our test subject was confused with the display they might have answered “neutral” as “I don’t know” -answer. Confusion matrices show the amount of neutral answers with every stimulus. This suggests that our theory is right. In future studies, the neutral emotion should be considered to be left out.

The confusion matrices gave us an insight of how our test subjects answered to different stimuli. There we could conclude that for example weak was confused often to afraid and sad was confused to tired, weak and neutral. We also analyzed the effect of male and female actor’s performance. Male actor’s angry and relaxed stimuli were recognized better than female actress’. Female actress’s stimuli were recognized better with excited and weak stimuli. There was not a big difference between orientations. Male actor was already strong body built which might have affected to his performance in weak emotion. In future studies, we should also consider having many different sizes of actor whom which we capture the motion.

Based on the analysis, we claim that only four emotions were recognized. These four stimuli were angry, excited, neutral and tired. With all other emotions, the deviations or confusion matrices showed too widespread responses that we would think they were not recognized properly. The test situation was not controlled in any way so it will bring the most of the uncertainty to the results.

We can also conclude that there were some significant differences between male and female actors, but since we had only one of both genders, we cannot say definitely if the difference was caused by the gender. To get more reliable result, we have to capture motions from more than five male and five female actors. We also do not have any statistical proof for the effect with participant’s gender. There were visually significant differences between male and female participants with some stimulus but the lack of actors forces us not to draw any definite conclusions.

To the final research question of learning, we cannot conclude anything definitely. We can just visually inspect these learning charts and note that latter 40 display’s recognition rate has a trend of being little bit better than the first 40 display’s. This would suggest that learning has happened during this experiment.

10 POSSIBLE FUTURE STUDY

We believe that if we'd had more controlled experiment environment, the recognition rates would have risen even higher. Now our test subjects could do the experiment as they liked and as fast as they could. These tests might be useful to run again with few modifications. First of all, more controlled environment. Test subjects would have to come to the test "site" themselves. Secondly, we'd need to capture the performance of more actors with less emotion categories. We could capture same emotions but we should use only a few of those. We would leave neutral, strong and weak emotions out of the experiment with test subjects. Finally we'd had to show these stimuli to great number of subjects, to find out what emotions they really see. Only after this we could start the larger scale experiments.

This thesis will be a good starting point for further studies in Aalto University's aivoAalto project. These results will give a baseline measurement for more advanced research like fMRI brain scanning research or more thorough psychological interviews.

Further studies of this research area might include capturing motion from persons who has noticeable abnormalities in their gait or postures. These persons usually have a long-term sickness like Parkinson's disease or might have an artificial join. The captured motion could then be shown to medical professionals in order to develop their skills to recognize sicknesses more accurately.

Another possibility for further study might also be to use physiological measurement tools like EEG, GSR, EMG or ECG to measure what the test subject feels during certain point-light display. This would need a radical change to the test configuration. Each test subject would have to be measured individually and the length of the stimulus and interstimulus would have to be longer. With this setup, there is a possibility that the test subject would not feel anything (or anything measurable) during the display.

We could also use traditional video to capture the motion and then show it to the test subjects. In this case, we should blur the facial expression out of the picture if we want to see how posture, gait and body movement is intrepid to different emotions. We also need to somehow improve the change of interpreting the emotions right. This might be achieved by adding the sound of knocking to the PL-display. Or we might play only the sound to the test subjects and they would have to recognize the emotion from the sound.

One very interesting sideline to this research could be the research of uncanny valley and purely animated PL-displays. As parallel to my thesis, another thesis in my research group is trying to achieve different emotional states by filtering PL-data and trying to get it to resemble the ones with actor playing the scene. We could also alter the PL-data away from human look-a-like by distorting the appearance somehow.

One last potential further study might be to research the amount of data needed to recognize the emotion. This basically means that we'd reduce the amount of light points and see how the rate of recognition decreases. We hypothesize that we can recognize emotions, from data such as mine, by using only two light points (one in the elbow and another in wrist).

This research area has potential for many different studies. If I would pursue to PhD-studies, I probably would like to study how colors affect to emotion in different cultures and different situations. Another possibility would be neuroleadership combining the brain research and social behavior leadership studies.

REFERENCES

Adams, R. B., & Kleck, R. E. (2003). Perceived gaze direction and the processing of facial displays of emotion. *Psychological Science,* 14(6), 644 647.

Afzal S., Sezgin T. M., Gao Y., Robinson P., (2009). Perception of Emotional Expressions in Different Representations Using Facial Feature Points.

Amaya, K., Bruderlin, A. and Calvert, T. (1996). Emotion from motion. In W. A. Davis& R. Bartels (Eds.), *Graphics Interface* '96 (pp. 228–229).

Anttonen, J., Surakka, V. (2005). Emotions and heart rate while sitting on a chair. *In Proceedings of Conference on Human Factors in Computing Systems (Portland, OR, 2005.).* ACM Press, New York, 491–499.

Atkinson, A.P., Dittrich, W.H., Gemmell, A.J., Young, A.W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception* 33:717–46.

Barrett, L. F. (1998). Discrete emotions or dimensions? The role of valence focus and arousal focus. Cognition and Emotion, 12, 579–599.

Beauchamp, M.S., Lee, K.E., Haxby, J.V., Martin, A. (2003). fMRI responses to video and point-light displays of moving humans and manipulable objects. *J. Cogn. Neurosci.* 15:991–1001

Beintema, J.A., Lappe, M. (2002). Perception of biological motion without local image motion. *Proc. Natl. Acad. Sci. USA* 99:5661–63

Bertenthal, B.I., Pinto, J. (1994). Global processing of biological motions. *Psychol. Sci.* 5: 221–25

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.

British Broadcasting Corporation, BBC, (2010). Scottish crimes to be reconstructed in 3D. Available on-line [http://news.bbc.co.uk/2/hi/uk_news/scotland/8457624.stm]. Cited in 31st January 2010.

Camurri, A., Lagerlöf, I., and Volpe, G. (2003). Recognizing Emotion from Dance Movement: Comparison of Spectator Recognition and Automated Techniques. *International Journal of Human - Computer Studies*, 59(1-2), 213-225.

Cardinal, R.N., Parkinson, J.A., Hall, J., Everitt, B.J., (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav*. Rev. 26, 321–352.

Clarke, T.J., Bradshaw, M.F., Field, D.T., Hampson, S.E., Rose, D. (2005). The perception of emotion from body movement in point-light displays of interpersonal dialogue. *Perception* 34:1171–80

Coulson, M. (2004). Attributing emotion to static body postures: Recognition accuracy, confusions, & viewpoint dependence. *J. Nonverbal Behavior*, Vol. 28, No. 2, 117-139.

Cutting, J.E., Kozlowski, L.T., (1977). Recognition of friends by their walk: gait perception without familiarity cues. *Bull. Psychon. Soc.* 9:353–56

Dalgleish, T., (2004). The emotional brain, *Nat. Rev. Neurosci* 5, pp. 583–589.

Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne smile: Emotional expression and brain physiology II. *Journal of Personality and Social Psychology*, 58, 342–353.

Ekman, P. (1992). An argument for basic emotions. Cognition and Emotion, 6(3/4), 169-200.

Ekman, P. (1999). Basic Emotions. *Chap. 3, pages 45-60 of: Dalgleish, T., & Power, M. (eds), Handbook of Cognition and Emotion*. Sussex, U.K.: John Wiley & Sons, Ltd.

Grezes, J., Fonlupt, P., Bertenthal, B., Delon-Martin, C., Segebarth, C. & Decety, J. (2001) Does perception of biological motion rely on specific brain regions? *NeuroImage* 13, 775–785.

Gross, R., (1996). *Psychology: The Science of Mind and Behaviour*, 3rd Edition. Hodder and Stoughton, London.

Grossman, E.D., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., Blake, R. (2000). Brain areas involved in perception of biological motion. *J. Cogn. Neurosci.* 12:711–20

Grossman, E.D., Blake, R., (2001) Brain activity evoked by inverted and imagined biological motion. *Vis. Res.* 41, 1475–1482

Grossman, E.D., Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron* 35:1167–1175.

Gunns, R.E., Johnston, L., Hudson, S. (2002). Victim selection and kinematics: a point-light investigation of vulnerability to attack. *J. Nonverbal Behav.* 26:129–58

Haag, A., Goronzy, S., Schaich, P., Williams, J. (2004). Emotion Recognition Using Bio-Sensors: First Step Towards an Automatic System. *Affective Dialogue Systems, Tutorial and ResearchWorkshop*, Kloster Irsee, Germany, June 14-16.

Haxby, J.V., Hoffman, E. A., Gobbini M. I. (2000). The distributed human neural system for face perception. *Trends Cogn. Sci*. 4, 223–233.

Haxby, J.V., Hoffman, E. A., Gobbini M. I. (2002). Human Neural Systems for Face recognition and Social Communication. *Biol. Phychiatry.* 51:59-67.

Heberlein, A.S., Adolphs, R., Tranel, D., Damasio, H. (2004). Cortical regions for judgments of emotions and personality traits from point-light walkers. *J. Cogn. Neurosci.* 16:1143–58

Hill, H., Johnston, A. (2001). Categorizing sex and identity from biological motion of faces. *Curr.Biol.* 11:880–85

Hill, H., Jinno, Y., Johnston, A. (2003). Comparing solid-body with point-light animations. *Perception* 32:561–66

Hirai, M., Hiraki, K. (2005). An event-related potentials study of biological motion perception in human infants. *Cogn. Brain Res.* 22:301–4

Izard, C. E. (1977). *Human emotions*. New York, NY, USA: Plenum Press.

Izard, C. E. (1991). *The Psychology of Emotions*. New York, NY, USA: Plenum Press.

Jacobs, A., Shiffrar, M. (2005). Walking perception by walking observers. *J. Exp. Psychol.: Hum. Percept. Perform.* 31:157–69

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Percept. Psychophys*. 14:195–204

Knoblich, G., Flach, R. (2001). Predicting the effects of actions: interactions of perception and action. *Psychol. Sci.* 2:467–72

Kourtzi, Z., Shiffrar, M. (1999). Dynamic representations of human body movement. *Perception* 28:49–62

LeDoux, J. E. (1995). Emotion: Clues from the Brain. Annual Review of Psychology, 1995(46), 209-235.

LeDoux J. E., (1996). *The Emotional Brain*. New York: Simon & Schuster.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, 23, 155–184.

Leslie, M. (May/June 2001). The Man Who Stopped Time. *Stanford Magazine*. Available on-line [http://www.stanfordalumni.org/news/magazine/2001/mayjun/features/muybridge.html]. Cited December 8th 2009.

Levenson, R. W., Ekman, P., Friesen, W.V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. Psychophysiology, 27, 363-384.

Little, J.J, Boyd, J.E. (1998) . Recognizing people by their gait: the shape of motion. *J. Comput. Vision Res.* 1 2.

Loula, F., Prasad, S., Harber, K., Shiffrar, M. (2005). Recognizing people from their movement. *J. Exp. Psychol.: Hum. Percept. Perform.* 31:210–20

Matsumoto, D. R. (2006). Culture and Nonverbal Behavior. *Chap. 12 of: Manusov, Valerie, & Patterson, Miles (eds), The SAGE handbook of nonverbal communication.* Thousand Oaks, CA, USA: Sage Publications. http://www.davidmatsumoto.info/Books/Chapter1.pdf Cited February 5th 2010.

Menache, A. (2000). *Understanding Motion Capture for Computer Animation and Video Games*. Morgan Kaufmann.

NaturalPoint, Inc. (2010). Webpage online [http://www.naturalpoint.com/optitrack/] Cited February 23rd 2010.

Partala, T. and Surakka, V. (2003). Pupil Size Variation as an Indication of Affective Processing. *International Journal of Human Computer Studies*, 59(1-2), 185- 198.

Partala, T. (2005) Affective Information in Human-Computer Interaction. Academic Dissertation. Tampere University of Technology, Department of Computer Sciences. Tampere. 94s.

Picard, R.W. (1997). *Affective Computing*. Cambridge, MA: M.I.T. Press.

Pollick, F.E., Paterson, H.M., Bruderlin, A., Sanford, A.J. (2001). Perceiving affect from arm movement.*Cognition* 82:B51–61

Pollick, F. E., Lestou, V., Ryu, J., & Sung-Bae, C. (2002). Estimating the efficiency of recognizing gender and affect from biological motion. *Vision Research*, **42**, 2345-2355.

Pollick, F.E. (2004). 'The features people use to recognize human movement style', in *Gesture-Based Communication in Human-Computer Interaction*, eds., Antonio Camurri and Gualtiero Volpe, number 2915 in LNAI, 10–19, Springer.

Puce, A., Perrett, D. (2003) Electrophysiology and brain imaging of biological motion. *Phil. Trans. R. Soc. Lond.* B 358, 435–445.

Ravaja, N. (2004). Contributions of psychophysiology to media research: Review and recommendations. *Media Psychology*, 6, 193–235.

Sagar, M. (2006). Facial performance capture and expressive translation for king kong. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Sketches*, page 26, New York, NY, USA, ACM.

Shiffrar, M., Freyd, J. J. (1993). Timing and apparent motion path choice with human body photographs. Psychological Science, 4,379–384.

Thornton, I.M., Pinto, J., Shiffrar, M. (1998). The visual perception of human locomotion. *Cogn. Neuropsychol.* 15:535–52

Tomkins, Silvan S. (1964) *Affect, Imagery, Consciousness, Volume I.* London: Tavistock.

Trager, W. (1999). A Practical Approach to Motion Capture: Acclaim's optical motion capture system, *Character Motion Systems*, SIGGRAPH 94: Course 9. Available on-line [http://old.siggraph.org/education/materials/HyperGraph/animation/character_animation/motion_capture/motion_optical.htm] Referred in 11.01.2010.

Troje, N.F. (2002). Decomposing biological motion: a framework for analysis and synthesis of human gait patterns. *J. Vis*. 2:371–87

Twentieth Century-Fox Film Corporation (2009). Avatar movie page [http://www.avatarmovie.com/index.html]

Vaina, L.M., Solomon, J., Chowdhury, S., Sinha, P., Belliveau, J.W. (2001). Functional neuroanatomy of biological motion perception in humans. *Proc. Natl. Acad. Sci. USA* 98:11656–61

Vlasic, D., Adelsberg, R., Vannucci, G., Barnwell, J., Gross, M., Matusik W., Popovic, J. (2007). Practical motion capture in everyday surroundings. *ACM Transactions on Graphics 26*, 3.

Ware, C. (2000). *Information Visualization: Perception for Design*: Morgan Kaufmann.

Wilson I., (2000). The Artificial Emotion Engine, Driving Emotional Behaviour, In *Working Notes of the AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment*. Menlo Park, CA.

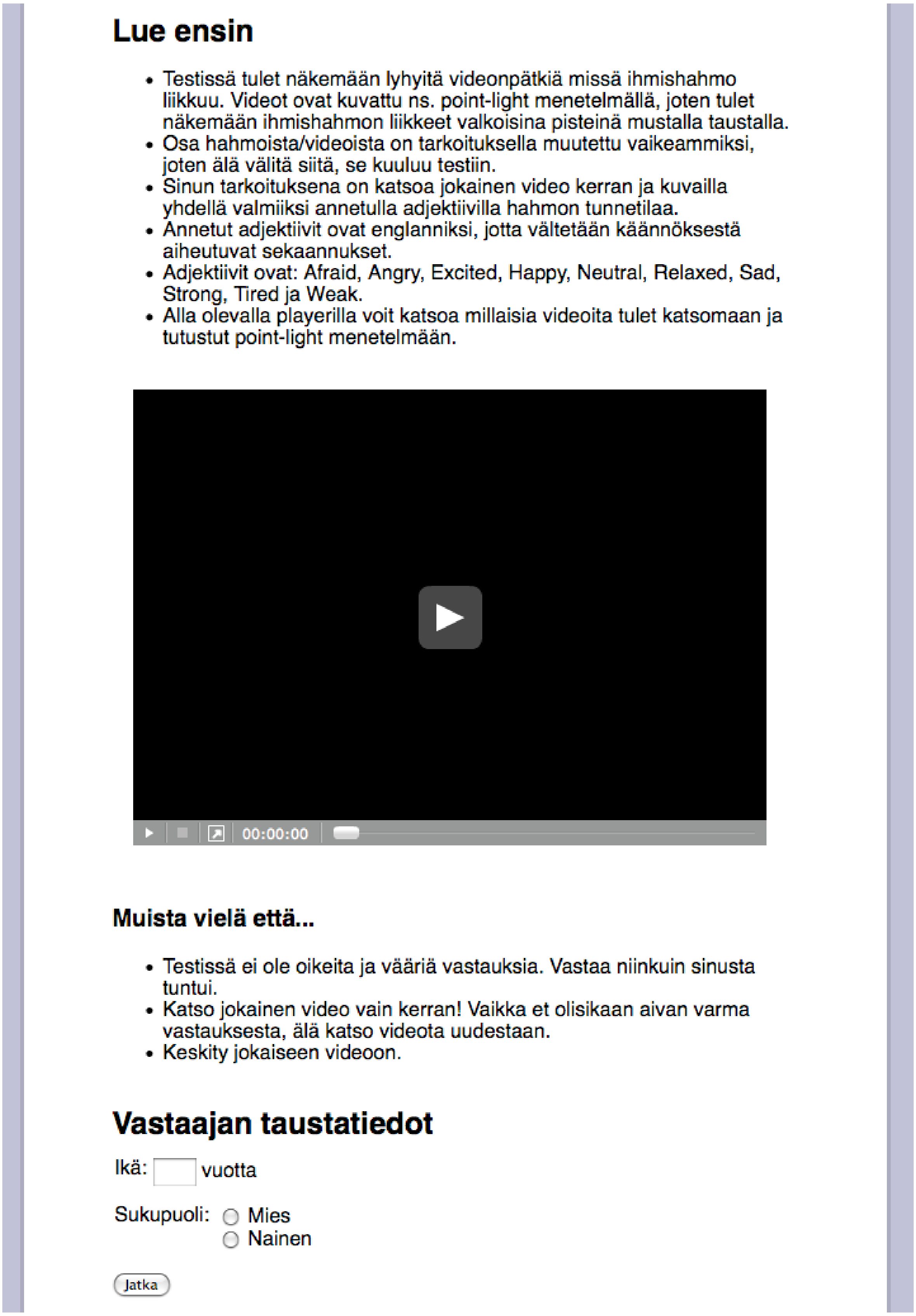
Zimmermann, P., Guttormsen, S., Danuser, B., and Gomez, P. (2003) Affective Computing - A Rationale for Measuring Mood with Mouse and Keyboard. *International Journal of Occupational Safety and Ergonomics.*

Zordan, V., Majkowska, A., Chiu, B., Fast, M. (2005). Dynamic response for motion capture animation. *Proc. ACM SIGGRAPH 2005 24*, 3, 697–701.

Öhman, A. (1987). The psychophysiology of emotion: An evolutionary-cognitive perspective: In P.K. Ackles, J.R. Jennings and M. G. H. Coles (Eds.) *Asvancces in Psychophysiology. Vol. 2*. Greenwich, CT: JAI Press. 79-127.

APPENDIX

Appendix A: The questionnaire forms



Lue ensin

- Testissä tulet näkemään lyhyitä videonpätkiä missä ihmishahmo liikkuu. Videot ovat kuvattu ns. point-light menetelmällä, joten tulet näkemään ihmishahmon liikkeet valkoisina pisteinä mustalla taustalla.
- Osa hahmoista/videoista on tarkoituksella muutettu vaikeammiksi, joten älä välitä siitä, se kuuluu testiin.
- Sinun tarkoituksena on katsoa jokainen video kerran ja kuvailla yhdellä valmiiksi annetulla adjektiivilla hahmon tunnetilaa.
- Annetut adjektiivit ovat englanniksi, jotta vältetään käännöksestä aiheutuvat sekaannukset.
- Adjektiivit ovat: Afraid, Angry, Excited, Happy, Neutral, Relaxed, Sad, Strong, Tired ja Weak.
- Alla olevalla playerilla voit katsoa millaisia videoita tulet katsomaan ja tutustut point-light menetelmään.

00:00:00

Muista vielä että...

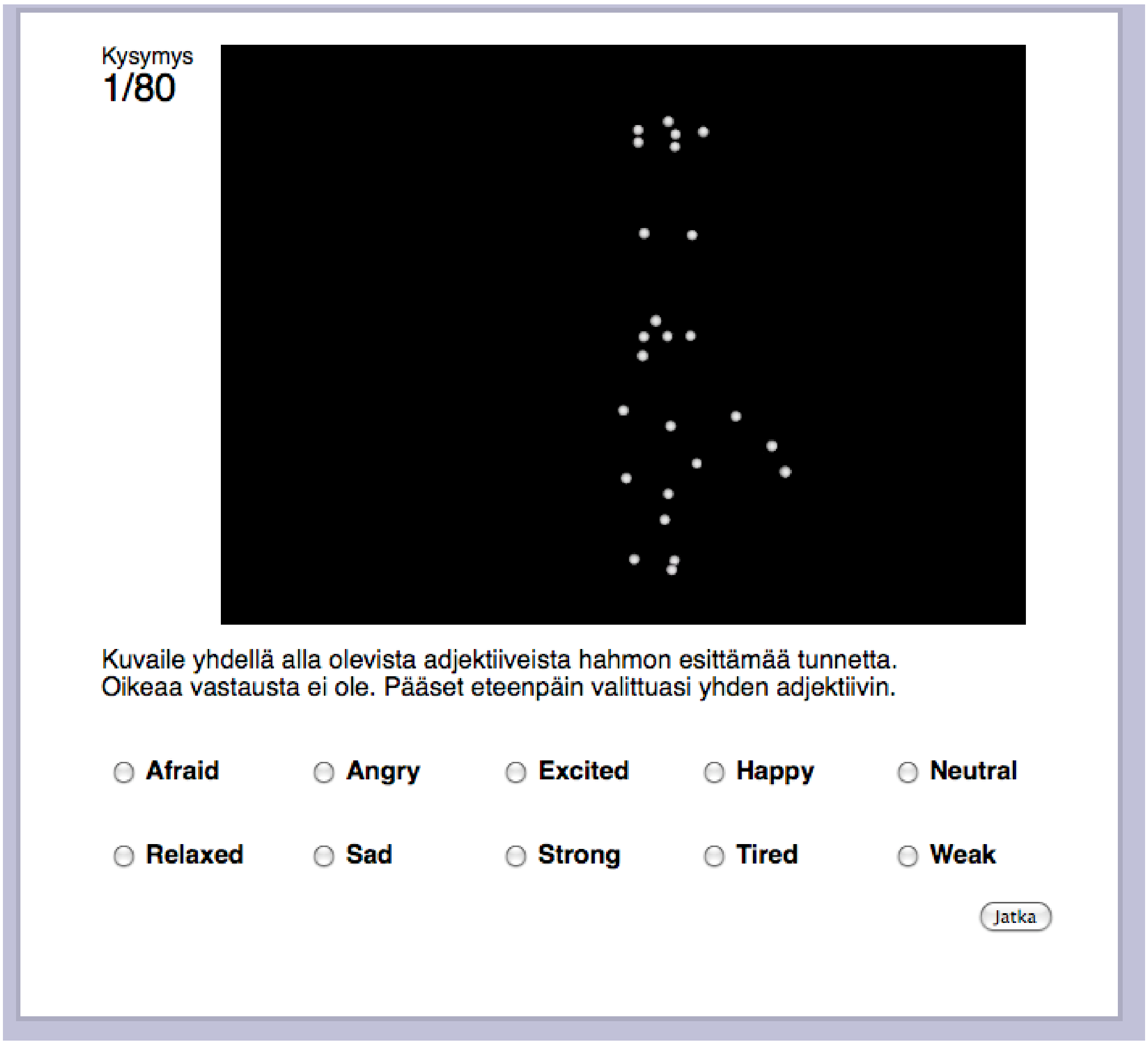
- Testissä ei ole oikeita ja vääriä vastauksia. Vastaa niinkuin sinusta tuntui.
- Katso jokainen video vain kerran! Vaikka et olisikaan aivan varma vastauksesta, älä katso videota uudestaan.
- Keskity jokaiseen videoon.

Vastaajan taustatiedot

Ikä: [ ] vuotta

Sukupuoli: ○ Mies ○ Nainen

Jatka



Appendix B: Recognition percentages of both genders in both orientations

	Normal orientation		Upside-down orientation	
	Male actor	Female actress	Male actor	Female actress
Tired	41,4 %	47,1 %	37,1 %	40,0 %
Neutral	31,4 %	25,7 %	44,3 %	30,0 %
Angry	30,0 %	15,7 %	27,1 %	15,7 %
Relaxed	21,4 %	8,6 %	5,7 %	10,0 %
Happy	18,6 %	12,9 %	10,0 %	15,7 %
Sad	14,3 %	17,1 %	2,9 %	12,9 %
Strong	14,3 %	10,0 %	20,0 %	10,0 %
Afraid	10,0 %	12,9 %	7,1 %	8,6 %
Excited	7,1 %	30,0 %	25,7 %	31,4 %
Weak	2,9 %	24,3 %	4,3 %	22,9 %