10th Semester
Aalborg University Copenhagen
Medialogy

Cheap Emotion

Using a Webcam and 2D Motion Capture
to Portray Emotion

by
Ósk Hilmarsdóttir

Preface

This is a 10th semester Medialogy project at Aalborg University in Copenhagen. The profile is Technical Visualization and the project was produced in Reykjavík, Iceland during the fall semester of 2010.

I would like to thank the following contributors for their valuable input in this process:

Arnar Gunnarsson at CAOZ for all the meetings, counseling and inspiration throughout the semester.

Orkustofnun, for generously providing me with a calm and private surrounding to conduct my tests, and full access to the welcoming staff.

Thomas Heller for assisting me with the tracking and composition in Maya.

Albert Sölvi Óskarsson for his continuous support and for taking the role of the actor.

Hilmar Sigvaldason, my father, for allowing me to consult him with various issues throughout the entire process, for helping me design and build the headgear and finally for tirelessly pushing me towards my goals.

_____

Ósk Hilmarsdóttir

Introduction

Throughout my Master studies I have chosen to specialize in visualization. I did this because I am interested in movies and cartoons; effects and animation. I find human behavior fascinating, and peoples' ability to express their emotions via body language or the slightest changes in their facial expressions is very intriguing to me. Animators have made a profession out of exploring this silent language; exaggerating it, caricaturing and finally presenting it to viewers for their amusement.

In the fall of 2009, when it was time for me to choose a subject for my Master Thesis, the opportunities were countless due to my broad interest so I sought help in narrowing down my choices and contacted an animation company in Reykjavik called CAOZ[1]. CAOZ is an award winning 3D animation studio, that produces films, commercials and visual effects. They were about to start production on television series where the main character was very cartoony and extreme, and were eager to find a way to make the process faster and cheaper for them. We talked mainly about facial animation; its importance and application, and how beneficial it would be if there was an easy way for it to be produced faster than now. This could have been a great goal for my project, but when it came time for me to start development, the production of the series had been postponed until further notice.

My interest in facial animation had been raised at this point and I met with CAOZ again to talk about their suggestions to alternative options for my project. Again we talked about the importance of good facial animation in order to convey the desired feeling to the viewer, as well as animation solutions today. CAOZ mainly focuses on "cartoony" animation and they pointed out that there are no animation solutions today that automatically create cartoony animation. I was intrigued to know why this is the case and realized that in order to find out, I would have to explore what it really means for an animation to be considered cartoony.

While researching the term cartoony, I learned many things about animation, what makes it good, and how characters can be "brought to life" and emit their emotions to the viewer. I also gained knowledge about how state of the art animation is created, often involving highly advanced and expensive motion capture equipment. This made me curious to see whether good facial motion capture could be created with the use of much cheaper and more available equipment. And as there is no objective measurement as to what makes animation good or bad because the quality of such animation is very subjective, I set the minimum requirement that the character "came to life" by emitting emotion.

[1] www.caoz.is

I would first need to look into how the leading animation companies produce their animation, regardless of whether they use motion capture or not, in order to understand their techniques and learn from them. Also, I wanted to understand what it is that makes animation cartoony and why, even with all the technology today, this can not be automatically created. Therefore I formulated my *Initial Problem Statement* as follows:

> *What does it mean for animation to be cartoony and what are the main methods of professional facial animation today?*

Preliminary Analysis

All kinds of animation are used in media today, from simple stop motion, to technical and complex motion capture. Sometimes the animation is so realistic that the viewer is not even aware that the action on screen is animated or acted, whether it is a robot in *Transformers* or the face of Benjamin in *The Curious Case of Benjamin Button*, where his face was CGI based on Brad Pitt's facial movements.

Using animation allows the filmmaker to think outside of the box, to perform actions that might not even be possible in the real world. Chris Patmore, a graphic designer, describes this in his book *The Complete Animation Course: The Principles, Practice and Techniques of Successful Animation:*

> *Animation makes the impossible possible, not only with characters but also with camera angles. [...] We can put the characters and the camera anywhere we want and have them do whatever we want. We can change the light, and the weather, and make mountains out of molehills. We can achieve things that cannot be done just with the written word.* [2 p. 15]

In this chapter, I will look into what it means if an animation is said to be cartoony. As 2D Disney animation is a classic example of cartoony animation, I will go through the Disney's animation principles that were created in the 1930's and used as a guideline for all Disney's animation. I will review projects that were based on these principles and look into if they are still necessary in todays 3D computer animation. To conclude on this topic I will ask myself the question of whether it is possible to identify and isolate what it is that makes animation cartoony, in order to then apply it to any animation and making it cartoony.

Later in this chapter I will look into computer generated facial animation; the different ways of performing it and what is happening in this field in the media industry.

What is cartoony?

When making a cartoony drawing of a person, the artist identifies the persons' unique characteristics and then exaggerates them in his drawing. These characteristics are not necessarily only based on looks, they may also be based on personality, habits or mood. The drawing may not be anatomically correct, and some body parts might even be distorted in a way that they would not function in the real world. But it works. We suspend our disbelief and accept the character as a part of a story.

Figure 1: U.S. president Obama as a cartoon character. Pictures taken from [28] and [29].

The same can be done for animation. Animation is merely a sequence of still images shown in succession, creating the illusion of movement. Robert D. Field, the writer of the book *The Art of Walt Disney* stated that "*Only when a drawing surrenders it independence and becomes part of a moving sequence may it be said to justify itself. Only then may art be said to enter into the animated picture."* [1 p. 223]

Making cartoony animation is in general the same procedure as in making cartoony drawings; the characteristics are identified and then exaggerated. However in animation, not only the looks can be animated, but also the character's movements. This exaggeration needs to be carefully planned, as the excessive use of it may make the animation look chaotic and senseless. Robert D. Feild describes this in *The Art of Walt Disney*:

> *"It may depend, as was the case in the* Bambi *characters, on a happy compromise between the simplification of their formal structure and an exaggeration of their unique behavior, yet without loss of their actual shapes. Or the Studio may indulge in out-and-out caricature and exaggerate form to the limit of recognizability. In this case it will be a question not merely of overemphasizing some subtle quality of behavior but also of taking full advantage of a new opportunity offered by the moment - that of caricaturing the action as well as the form." [1 p. 230]*

The action the character is to form needs to be analyzed before it is performed, just as the character needs to be analyzed before being drawn, in order to eliminate all irrelevant details and take advantage of any characteristics that are unique to that situation. [1 p. 231] The goal is to stay true to nature and extract the essence from the situation, although exact realism is not necessary. [1 p. 242] Subtlety is important, as often the action alone is necessary to convey the main theme, and pantomime is powerful in animation, as it relies more than any other on visual appeal. [1 p. 246]

The Disney Principles

In the 1930's Walt Disney set up drawing classes at his studio, where animators studied the motion of human figures and animal, both in live action and on film, in order to analyze their actions. This analysis became important to the process of animation, as some of the animators applied the lessons of those classes to production animation, resulting in a more sophisticated look. The animators continually strove to refine procedures based on their learnings from these classes, resulting in isolated guidelines that were perfected and taught to new animators. These guidelines are today known as the fundamental principles of traditional animation and they are important to this project in order to understand what it is that constitutes a cartoony animation. [4]

Although some of these principles mean exactly the same thing in 3D animation as well as 2D, the application of others differs because of the difference in medium. [4] Following are these rules, followed by an assessment of whether they are directly relatable to making animation cartoony.

1. *Squash and stretch*

   This is the most important and the most widely known principle. An object's rigidness can be portrayed by its movement, making this rule very important for facial animation, "*not only for showing the flexibility of the flesh and muscle, but also for showing the relationship of between the parts of the face.*"

   This rule is very important to cartoony animation, and some characters may even look as if they are made out of rubber rather than skin and bones if it is used excessively. However, as most living things tend to squash and stretch when moving, this principle is not directly relatable to cartoony animation without being exaggerated. This is explained in the book *The Art of Rigging* by Kiaran Ritchie, Jake Callery and Karim Biri:

   "Early uses of squash and stretch resulted in cartoon characters who's forms appear to be composed of a gelatinous substance. When one describes an animation as being 'cartoony' they are usually referring to the liberal use of squash and stretch. Exaggerated deformations can be used to convey the feeling of weight and substance. Proper use of squash and stretch is fundamental to believable animation. [...] Squash and stretch can refer to not only the wobbly, squishy classical animations, but also proper muscle and skin deformations on modern digital characters. Ultra realistic creature rigs need to maintain volume throughout extreme deformations. This is the modern twist on squash and stretch." [24 p. 41]

2. *Timing*

   The timing of an action can express an object's weight and size and even portray emotional meaning. Heavy objects need a greater force to get them moving and therefore accelerate and decelerate slower than lighter objects. Likewise, the smaller an object is, the less inertia it has and therefore it can move faster. "*The way an object behaves on the screen, the effect of weight that it gives, depend entirely on the spacing of the poses and not on the poses themselves.*"

This rule is not directly relatable to making animation cartoony, but rather making it good and understandable for the audience.

3. *Anticipation*

   Anticipation directs the audiences' attention and prepares them for an action they might otherwise miss because of its speed or placement on the screen. It can also be an indicator of weight, when for example a character is picking up an object, the heavier the object, the more exaggerated anticipation.

   Anticipation is necessary both in live films and in animation, in order for the audience to understand the characters' actions. However, it by itself is not directly relatable to making animation cartoony, without it being exaggerated, making the exaggeration the more important factor.

4. *Staging*

   The purpose of proper staging is to unmistakably present an idea to the audience, guiding their attention to the right place at the right time. *"An action is staged so that it is understood; a personality is staged so that it is recognizable; an expression so that it can be seen; a mood so that it will affect the audience."* [p. 38]

   This rule can not make animation look cartoony, however it is very important in order to obtain good and understandable animation.

5. *Follow through and overlapping action*

   Follow through is the termination of an action, in the same way anticipation is the initiation of one. For example when a character throws a ball, his arm continues moving after having released it, following the action through. The speed of the follow through action depends on the size, weight and rigidness of the object. This means that the actions overlap, making the objects seem more natural and the action more interesting, as it *"maintains a continual flow and continuity between whole phrases of actions."* [p. 40]

   Even this rule needs exaggeration in order to be relatable to cartoony animation. By themselves, follow through and overlapping action contribute to making the animation look natural and realistic, but it is only when exaggerated that they contribute to making the animation look cartoony.

6. *Straight ahead action and pose-to-pose action*

   These are two different ways of animating. In straight ahead animation the animator does not plan the character poses ahead, instead he starts at the beginning and works his way straight ahead through each movement in a chronological order. This can produce fresh and lively action, which is rich in spontaneity. Pose-to-pose animation is planned ahead, with determined key poses and in-between poses that morph from one key position to the next.

This principle can not be related to cartoony animation, as both straight ahead action and pose-to-pose action can, and can not, result in a cartoony animation.

7. *Slow in and slow out*

   If an action is performed with the same constant speed from start to finish, it tends to look robotic. The action becomes more lifelike when it is slowed in and out, meaning in 2D animation that the in-betweens are not spaced equally, but grouped closer to each extreme.

   As with the follow through and overlapping action, this principle is indeed important for the animation to look good, but in itself is not directly relatable to making it cartoony.

8. *Arcs*

   Arcs describe the visual path of action from one extreme to another. They make the animation *"smoother and less stiff"* and they are rarely completely straight, except for perhaps a falling object, although even then the object might rotate.

   This principle is a good tool to make beautiful animation, but unless the arcs are exaggerated, it does not relate directly to making the animation look cartoony.

9. *Exaggeration*

   The point of exaggeration is not to distort objects and shapes so that they become unrealistic, but rather to *"go to the heart of anything or any idea and develop its essence, understanding the reason for it , so that the audience will also understand it. If a character is sad, make him sadder; if he is bright, make him shine; worried, make him fret; wild, make him frantic."* [p. 41] The exaggeration must be balanced, if only one object is exaggerated in a fairly lifelike scene, it will seem unrealistic and stick out.

   This is the key principle to making animation cartoony. By combining this principle with almost any of the others, cartoony animation can be created.

10. *Secondary action*

    Secondary action is an action that is directly drawn from a primary action and adds interest and realistic complexity to the scene. It should compliment the primary action and not dominate the scene. Facial expression can be a secondary action, and then it must be presented either before or after the primary action, or the body language, in order for it to be remarkable.

    As with *follow through and overlapping action*, this principle is good for making good and understandable animation, but can not be directly related to making the animation cartoony, without being exaggerated.

11. *Appeal*

Much like live actors have charisma, animated characters have appeal. Appeal is not the same as cuteness or cuddly kittens, but rather what people like to see. It can be the quality of a design, the charm or magnetism of the film or even its simplicity. In character animation, and especially facial animation, the movement of the right and left side of the body or face should never be completely mirrored, or the result becomes stiff and unappealing.

This principle can not be directly relatable to making animation cartoony, as any animation, regardless of how true to realism it is can sport appeal.

These principles are a guideline to how to make a good animation; a piece that is beautiful, understandable and easily relatable for the audience. However, only one of them can be directly related to making animation look cartoony; *exaggeration*. Even *squash and stretch* does not create the cartoony look unless it is exaggerated and as previously mentioned, the exaggeration needs to be based on the essence of the character or the moment in order to compliment the scene.

The Disney Principles in 3D animation

Four university students at the University of Wisconsin at Madison attempted to incorporate traditional animation techniques into procedural animation. They narrowed their project down to focusing only on the most commonly known and used principle, the *squash and stretch*, and created a simulation system of both kinematic and dynamic techniques, to geometrically squash and stretch colliding objects. They found that they could easily define the parameters of objects that make them appear light or heavy, rigid or soft, and concluded that this increased the expressive power of procedural simulations. Also, they reasoned that the most important aspect for future work would be to focus on stylized motion, rather than shape, as almost all researchers have focused on aspects of shape rather than global motion. [21 p. 6]

Figure 2: A simple game, similar to Breakout, shows the squash and stretch of the ball. Picture taken from [21 p. 6].

In another project from a collaboration between the University of Washington, Microsoft Research and University of California, Berkeley, a "Cartoon Animation Filter" was created. This filter takes arbitrary input motion signal and adds to it an inverted, smoothed and (sometimes) time shifted version of the acceleration of the signal. [22 p. 1] The input signal can be motions ranging from hand drawn trajectories, to motion captured curves, to simple animations within PowerPoint presentations. Their results, although they would not suffice for off-line animation systems, showed that the filter added simultaneously *exaggeration*, *anticipation*, *follow-through* and *squash and stretch* to a range of motions, and could be used in less professional settings, such as child focused animation tool. [22 p. 5]

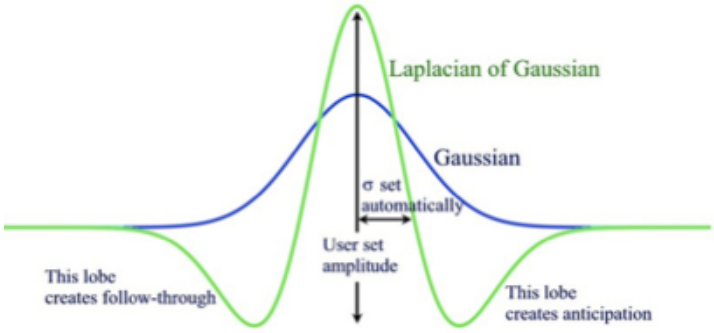


Figure 3: The Cartoon Animation Filter. Picture taken from [22 p. 1].

Frank Thomas, one of Disney's animators using the hand-drawn method, wrote an article in 1984 about whether classic animation could be duplicated to the computer and wondered if it even should. He felt that in computer animation, when applying the squash and stretch rule, the animators failed to realize the importance of the relationship between the various parts of the face and the body. He emphasized that the importance of this rule was in the related moves, how the shrug of the shoulders affects to the curve of the back or a lowered eyebrow pushes the eye to the cheek. He also outlined that some 2D characters, such as Donald Duck and Goofy, could have bones or not have them, depending on their actions. [23 p. 22] He stressed that in animation, the essence of the action must be clearly visible:

> *Animation cannot afford ambiguity. It has to speak quickly and directly to its audience. General kinds of movement are not enough to sustain a film of this type and the moves that are used must be completely convincing, natural and fluid.* [23 p. 24]

His opinion was that the computer should not aim to copy work that has been done by artists; "*rather the artist should search for artistic ways to use capabilities of the computer*". [23 p. 25] He concludes by predicting that an awkward duplication of Bambi or Snow White will not be a breakthrough in the use of computer generated endearing characters, but rather something new and vital, regardless of how different it is to classic animation.

Possible solution

Throughout this chapter I have concluded that exaggeration is the key factor to making animation cartoony. However, to stay true to the *staging* principle, not every single aspect in a scene should be exaggerated at once, but rather should it be used to underline important features, such as mood or characteristics. Therefore I ask myself, can all character animations be made cartoony, using the same method?

The exaggeration of the essence of a moment, or a feeling, can be done in many ways, and facial animation might not always be the best option to express this. Outer-body elements are often well suited for this, for example if the character is in love, hearts flow in the air around him, if he is furious with anger, a thundercloud appears above him and if he is very happy with himself, he glows. Colour changes can also emphasize feelings; a red face means that the character is very angry, while if he's green, he feels sick. All of these are great ways of exaggerating the essence of the moment, and making the character cartoony, but it is not directly to the character animation in itself.

Just as comics, animated films or shows are often described as cartoony, based on the way they look, rather than the animation of the characters. For example, in The Flintstones, Fred has got very exaggerated facial features, so it doesn't really matter how truly to human anatomy his face is animated, his oversized mouth makes all his facial movements seem cartoony.

Figure 4: Fred Flintstone has cartoony features and a big mouth. Picture taken from [30].

Given that cartoony animation is linked with the exaggeration of unique characteristics, it becomes clear that what is cartoony for one character might not have the same effect with another if they do not share the same physical features. [34] As an example, a character with a very small mouth would not come across as very cartoony if he were to have the same cartoony animation as one with a very large mouth. The "cartoony-ness" of the animation of the large mouthed character would disappear in the transition, suggesting

that there is no single definable way, that works for all characters, of making animation cartoony, as it depends so greatly on the look of the characters.

Facial Animation

There are two main ways of creating computer generated facial animation. First the animator can manually animate each keyframe, a very time consuming task that provides the animator with perfect control over the whole process. Secondly, there is motion capture, where the animation is completely automated although the results can be tweaked afterwards, and it is also possible to have the animation morph between pre-determined facial positions. The morphing can be based on other factors, and not only on motion capture, for example computer vision or even speech and text analysis.

Following is a review of what big media producing companies are using today.

Keyframe Animation - Pixar

Pixar Animation Studios is an Academy Award winning computer animation studio, renowned for its ability to tell stories by using only body language. Pixar has used keyframe animation since its incorporation in 1986, and still today, even with the latest techniques in motion capture available, keyframe animation remains their way of animating. Using their own software, the animators choreograph key poses for the characters and Pixar's proprietary computer software creates the in-between frames. For Pixar, making a full feature film is a very time consuming project, where the phrase *"You can't rush art."* applies perfectly. [25]

Figure 5: Sulley and Mike, two of many memorable characters from Pixar's creation. Picture taken from [31].

Motion Capture - Image Metrics

Image Metrics is an animation company founded in The United Kingdom in 2000 by Ph.D's in computer vision. It is a leading provider of performance-driven facial animation, that relies only on the video footage of an actor's performance. It needs no markers, makeup or special equipment, other than a video camera, meaning that the performance can be captured in a sound studio at the same time the voice-over is being recorded. [20]

Figure 6: "Emily", Image Metric's photo-real demo, created in partnership with the University of California's ICT Graphics Lab, was seen as a new achievement in believable computer graphics at SIGGRAPH 2008. Picture taken from [20].

Their trademarked program studies the footage, analyzing the change in colour for each pixel over time, creating a set of algorithms which can then be translated into motion. The animators find the key expressions, such as smiles and frowns, and teach the software how to recognize them. These key poses are then combined with the image analysis data, resulting in animation curves that can be tweaked with any standard 3D application, such as Maya or 3ds max. [20]

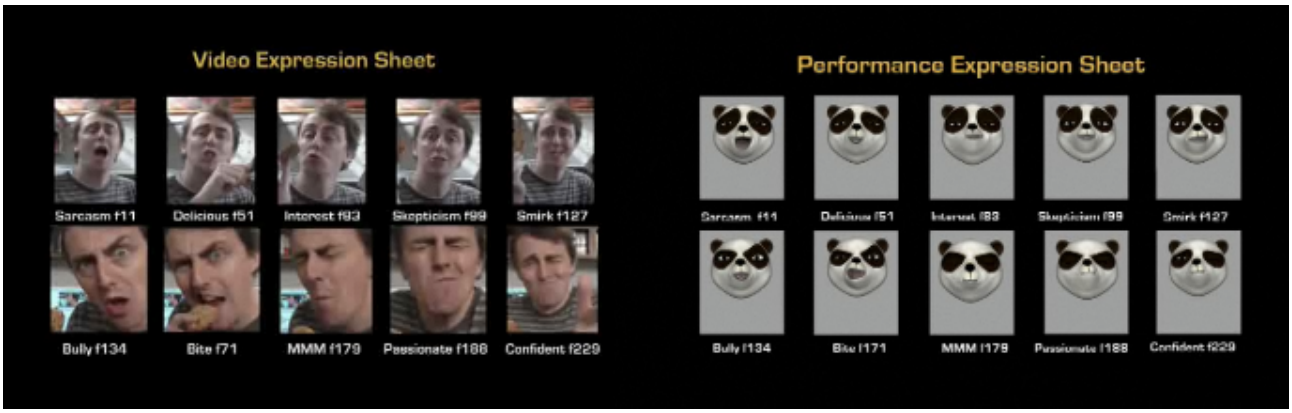


Figure 7:Video and Expression Sheets are created and then combined with image analysis data, creating realistic animation. Picture taken from [5b].

Image Metrics offers solutions both for the production of films and computer games, and has made facial animation for numerous high end creations, such as The Curious Case of Benjamin Button, Grand Theft Auto IV and even a music video for Black Eyed Peas. [20]

Morphing - Softimage and Voice-o-matic

Autodesk Softimage is a high performance visual effects and 3D character animation application with a built in advanced animation toolset called Face Robot. Face Robot offers automatic lip syncing based on an audio file, which is paired with a text file. The audio is analyzed, spotting the phonemes and visemes[2], and then in accordance with the text file it creates animation for the character's lips, jaw and tongue. [18]

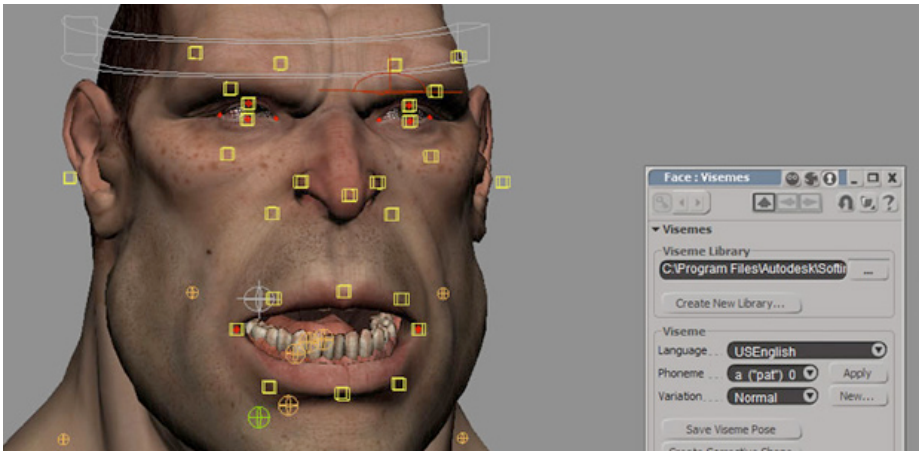


Figure 8: Lip syncing in Face Robot. Picture taken from [18].

The results can be tweaked and manipulated by the animator, adding additional animation to give the character emotion and personality.

Voice-o-matic is a lip sync animation plugin available for Autodesk 3ds max, Autodesk Maya and Autodesk Softimage. It analyzes speech from an audio file and based on the timing of the phonemes, it generates lip positions that can easily be tweaked by the animator. It is applicable for both blend shapes and bone rigs and can be used for all languages. [16]

Figure 9: Since 2008, The Garfield Show has been animated with the use of Voice-o-matic in order to trim production time and allow the animators to focus on character emotions and performance. Picture taken from [17].

[2] Not all phonemes are visually distinct, and the ones sharing the same visual attributes are grouped into visemes.

This sort of software can save animators valuable time as it makes it possible for them to change the characters' "acting", without interrupting with the lip sync. The lip sync can be made quickly with this program and then the animation expressing feelings and emotions can be added on top of that. This layered animation makes it easy to change the characters' acting if the director wishes so, because the emotional animation is in a separate layer. [19]

Motion Capture - Weta Digital

Weta Digital is a New Zealand based visual effects company, which recently won an Academy Award for Best Visual Effects for James Cameron's Avatar. In Peter Jackson's movie, King Kong, they broke the geometric connection between the actor and King Kong's CG face by fitting motion capture movement to an expression space, which was then mapped to another expression space, defined by gorilla movements. [27]

Figure 10: King Kong. Picture taken from [32].

As previously mentioned, this morphing kind of animating can be based on various inputs. In one project a group used 3D scans of a head, combined with optical motion capture, in such a way that the timing of the motion capture controlled the morphing between the scanned poses. [26]

Every method for computer facial animation has its benefits and disadvantages. Keyframe animation is accessible for any person that has a computer and a 3D application such as Maya. Motion capture however is more complex and usually consists of multiple high speed cameras and a quantity of markers to be tracked from multiple angles.

Problem Statement

Previously in this chapter I concluded that exaggeration that is complementary to the mood in a scene or a character, is the foundation for calling films cartoony. This is however a rough estimation and I am aware that this is a subjective matter, what one sees as cartoony, another might not. The division between realistic and cartoony features is not

always clear, and does not only apply to drawn characters, regardless of whether they are made in 2D or 3D; think for example of Jim Carrey in The Mask. As his character puts on his mask, he behaves almost like a Looney Tunes character and his behavior can best be described as cartoony, with extreme exaggerations underlining his mood or situations.

While researching this term, cartoony, I have learned about animation, rules how to make it good and understandable for the viewer, as well as how it is performed in the industry today. Keyframe animation has remained relatively similar over the last years, while motion capture techniques and equipment evolve quickly. Most professional motion capture productions rely on very expensive equipment of the highest technology to track an actor's face from all angles. This creates very realistic and accurate animation, which the viewer easily empathies with, as the motion is based on human facial expressions.

I want to know if the same can be done using much cheaper and easily accessible equipment. Instead of using numerous high definition cameras, I want to see if I can create motion capture with cheap equipment that is available for anyone. And I want the motion capture to be good enough so that the viewer can understand an emotion that the CGI character emits through his facial expressions. Hence, my problem statement is:

> *How can facial motion capture be created, using cheap equipment to capture the motion, in such a way that emotion can be read from the facial expressions?*

Analysis

The outcome of this project will be an animated clip, where I have used motion capture on a character's face, based on footage that was recorded with cheap and available equipment. I will then put the animation to the test and see whether the actors emotion comes through in the animation.

In this following chapter I have identified the most important aspects that I need to consider before moving on to the design phase. First I will look into a few aspects of facial motion capture that are necessary in order to create good animation. Second, I will discuss emotion and empathy and how emotional contagion can be measured. Next I will look into 2D tracking and its comparison with 3D tracking in order to understand the advantages and disadvantages of each. As I will not be buying an expensive motion capture system with a built in tracking software, I will address my tracking options as well as software options, to put the scene together and tweak the results. In regards to marker placement, I will research how others have done it and make note of important aspects in this regard.

My findings are based on published articles, books, official information from companies in the industry and expert interviews with Arnar Gunnarsson, the head of CG at CAOZ[3].

Facial Motion Capture

Before conducting a facial motion capture, there are a few things that need to be considered, which can improve the quality of the animation. According to Arnar Gunnarsson, CG supervisor at CAOZ, there is no specific standard in professional animation, that is, there is no certain checklist or scale that can be applied to the animation to assess whether it is good enough. The quality depends on the future plans for the project, whether it is meant for big budget film or a small television production, and its measure is very subjective. [15] The most important areas in facial animation are the eyes and the mouth, because the viewer points his or her attention primarily to these areas, similarly to when interacting with another person. [15]

When creating an impression of a human face, it is important not to fall into the *uncanny valley*. The uncanny valley is a concept coined by Masahiro Mori in 1970 and it details that as a robots appear more humanlike, our sense of their familiarity increases, until it drops down in a valley. [14] This is not conclusively related to robots, but also applies in drawings, photos, puppets, 3D characters and other creations that resemble humans. The uncanny valley is plotted in a graph in figure 11, as familiarity versus appearance.

[3] The interviews can be seen in the Appendix.

When we look at a machine, or an industrial robot, it looks fine to us, it looks familiar and the same goes with stuffed animal, it looks familiar. But when we see a corpse, we feel a sense of unease. It looks so much like a human but there is something missing and it looks uncanny. The valley and peaks of this graph are exaggerated when motion is added, as motion is usually an indicator of life. An example of this exaggeration is a prosthetic hand. It looks very similar to a normal hand, and yet it looks a bit off and therefore we find it uncanny. If the fingers start to move, our sense of unease is increased. [14]

Mori advises designers to try and avoid falling into this valley by aiming their design towards the first peak rather than the second. Although the second is higher, the risk of falling into the uncanny valley is much higher.

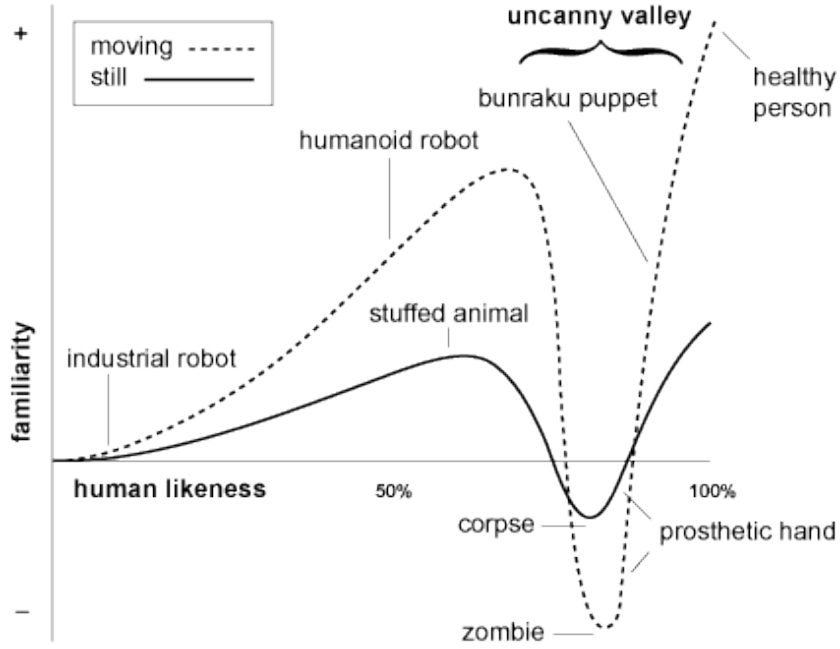


Figure 11: The uncanny valley. Picture taken from [14].

To avoid a vacant and dead stare for the character, it is important to include saccades in addition to coordinated directional movement of the eyes in a gaze, according to the authors of *Anatomical Considerations in Facial Motion Capture* [3]. Saccades are *"depiction of the brief flitting simultaneous movements of both eyes in the same direction"* and are vital to give the character the "*spark of life*". [3] In people, these movements improve focus and object resolution and scan relevant features in the surrounding area and are perceptibly related to the distance of an object from the viewer. The further away the character is staring, the less frequent, slower and lower amplitude saccades are and vice versa. Although these movements are hard to track without the use of contact lenses or optical systems to track eye movement, they are vital to communicate not only emotion, but also to implicate distance between characters. [3]

One important thing to consider when choosing an actor for a motion captured face is that proportional differences between genders and age types may influence the outcome. For example, women blink more often than men and have different skull proportions. For this

reason, proportional differences between the actor and the CG model can cause problems when the difference is substantial, however it might be highly satisfactory if the CG character is meant to be grotesque and/or non-human, such as Gollum in Lord Of The Rings. [3] Also, symmetry should not be the goal for any animation:

> *"Asymmetry of human faces and facial expression is the rule, not the exception, and therefore asymmetrical models and movements are crucial in creating believable and sympathetic human characters."* [3]

Emotion and Empathy

The problem statement elicits that the viewer should be able to see emotion in the CGI character's facial expressions. Emotion differs from memory and attention in two important ways. First, emotional stimuli are categorized as either being negative or positive, and second, emotional stimuli "*predispose people to bivalent behavior (e.g., approach or withdrawal) toward the stimuli.*" [13 p. 3]

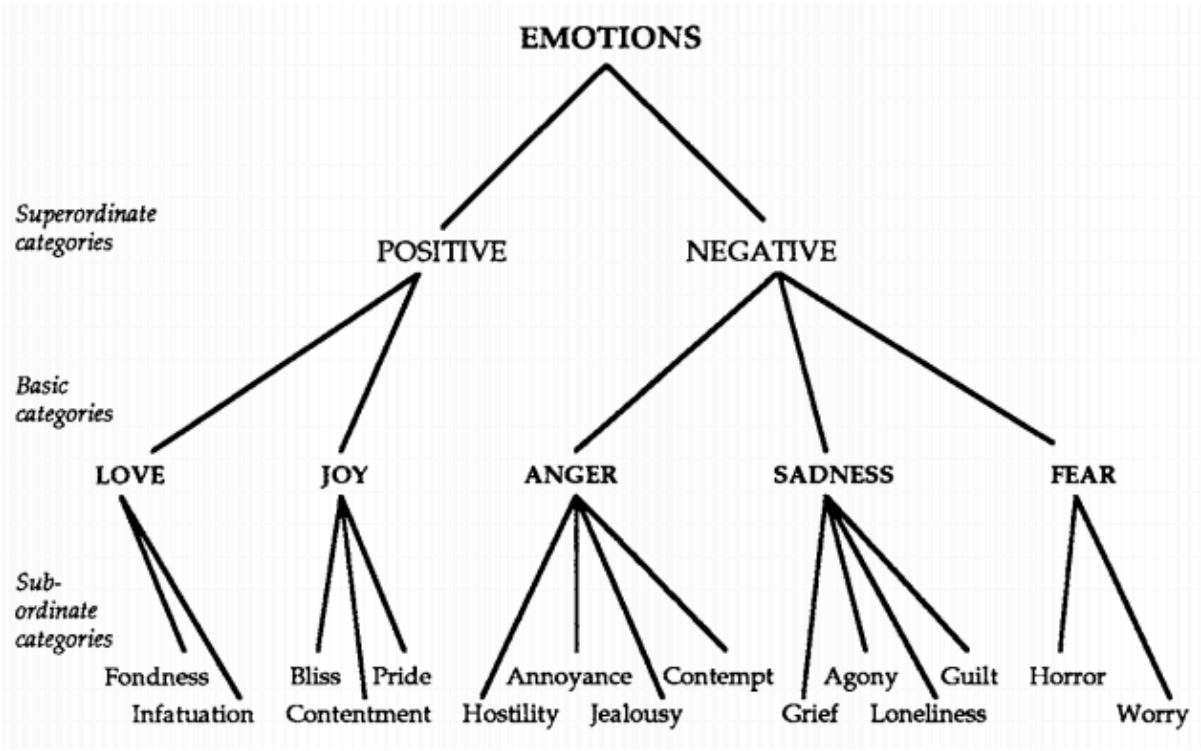


Figure 12: An emotion hierarchy. Picture taken from [13 p. 3].

People's ability to notice and understand other's emotions is called empathy. Empathy comprises both cognitive and more primitive emotional elements, and can be described as congruent reactions of one person to the observed emotional experiences of another. [11 p. 132] Cognitive models suggest that when people listen to other's description of an

emotional experience, they remember similar experiences and feelings, which produce a similar emotional response. The more primitive emotional elements regard the basic process of emotional contagion. [11 p. 132]

Since Darwin first stated that emotion appear to be contagious in 1972, this has been a field of study for many researchers. The term "emotional contagion" was coined, and is defined as humans' tendencies to automatically synchronize and mimic expressions, movements, vocalizations and postures of the person they are interacting with, and consequently converging emotionally. [11 p. 132] This tendency to mimic other people's expressions does not seem to be learned, as it is apparent in neonates and usually happens without conscious processing. It contributes to understanding and experiencing the other person's affect. [11 p.133]

Different patterns of facial muscle activity are associated with different emotions and, although emotions are not solely generated by facial, vocal or postural feedback, the face plays an important role in the experience of emotions. Emotions are influenced by these sources of feedback to a certain degree, and therefore "*spontaneous mimicry should contribute to emotional contagion*". [11 p. 133] For example if a person receives a smile, and without thinking reacts with a smile (spontaneous mimicry), that person might feel a positive emotion (emotional contagion).

People are not all equally susceptible to emotional contagion, and the reasons lie in their background and personality characteristics. [11 p. 134] A scale, called *The Emotional Contagion (EC) Scale*, can estimate people's susceptibility to emotional contagion by their measured frequency with which "*emotional stimuli elicit an emotional expression characteristic of the eliciting emotion*". [11 p. 134] The higher the score on this scale, the more susceptible to emotional contagion that person is said to be. This scale can be seen in figure 13.

The Emotional Contagion (EC) Scale

1 If someone I'm talking with begins to cry, I get teary-eyed.
2 Being with a happy person picks me up when I'm feeling down.
3 When someone smiles warmly at me, I smile back and feel warm inside.
4 I get filled with sorrow when people talk about the death of their loved ones.
5 I clench my jaws and my shoulders get tight when I see the angry faces on the news.
6 When I look into the eyes of the one I love, my mind is filled with thoughts of romance.
7 It irritates me to be around angry people.
8 Watching the fearful faces of victims on the news makes me try to imagine how they might be feeling.
9 I melt when the one I love holds me close.
10 I tense when overhearing an angry quarrel.
11 Being around happy people fills my mind with happy thoughts.
12 I sense my body responding when the one I love touches me.
13 I notice myself getting tense when I'm around people who are stressed out.
14 I cry at sad movies.
15 Listening to the shrill screams of a terrified child in a dentist's waiting room makes me feel nervous.

Note. Happiness items = 2, 3, & 11; Love items = 6, 9, & 12; Fear items = 8, 13, & 15; Anger items = 5, 7, & 10; Sadness items = 1, 4, & 14.

Figure 13: The Emotional Contagion (EC) Scale. Picture taken from [11].

According to this scale, those who are more responsive to self produced cues, and therefore more susceptible to emotional contagion, will show less stability when giving evaluations in situations where their emotional expressions are affected by mimicry. [Emotional Contagion and Social Judgmen p. 189]

2D Tracking Compared to 3D Tracking

Most facial motion capture systems are based on multi camera setups, where all angles of the face are tracked and calculated into 3D posture and motion. This setup is ideal for detailed captures, but is more complicated than 2D tracking. The recordings of the cameras need to be synchronized and the lighting adjusted so that each camera can see all markers facing towards it at all times. Although 2D tracking can suffice to capture fairly accurate facial mimic, the 3D tracking is more accurate, as it does not leave out depth information of the facial expressions.

Figure 14: Multi camera optical motion capture setup. Picture taken from [33].

In 2005, two researchers from Taiwan presented a solution to track facial movements in 3D by using one camera. To do this, they used two mirrors, located on each side of the actor, projecting the profile images into the camera, which was located in front of the actor. [12] The angles of the mirrors needed to be set before the shoot, and then the result could be applied to a CGI character's face. Although the actor's head either needed to be stabilized before the shoot, or complicated algorithms performed to eliminate head movement, the result was successful, resulting in a detailed 3D tracking in near real time. [12] Figure 15 shows the setup in this project.

Figure 15: One camera catches three perspectives of the actor's face via the use of mirrors. Picture taken from [12].

As 2D tracking requires much less of expensive equipment or algorithms to solve tracking points, it is a feasible option for smaller productions with a limited budget. The fact that it can not track the depth of the character's facial expressions does not mean that the outcome of the animation needs to be without it. First, as used by Image Metrics, expression sheets can be created based on the actor's poses, that can later be combined

with the motion capture data. Second, the z-axis movement can be animated afterwards by hand, using keyframe animation. Although keyframe animation is more time consuming than motion capture, this solution merely requires a small amount of z-axis tweaking, while still entailing the accurate timing of the motion capture, as well as the majority of movements needed for lip sync.

Tracking options

In my past, when I have had to track 2D motion, for example to stabilize a video, I have used *Adobe After Effects*. For each tracked point in After Effects, three variables are defined; a *search region*, a *feature region* and an *attach point*. The feature region surrounds a distinct visual element, which is preferably an object in the real world. This region defines the element in the layer to be tracked and After Effects must be able to clearly identify this feature throughout the track, despite changes in light, angle or background. The search region defines the area After Effects will search in order to locate the tracked feature, which needs to be distinct within the search region only, not the entire frame. The attach point, or the target, determines the control point to synchronize with the moving feature. [8]

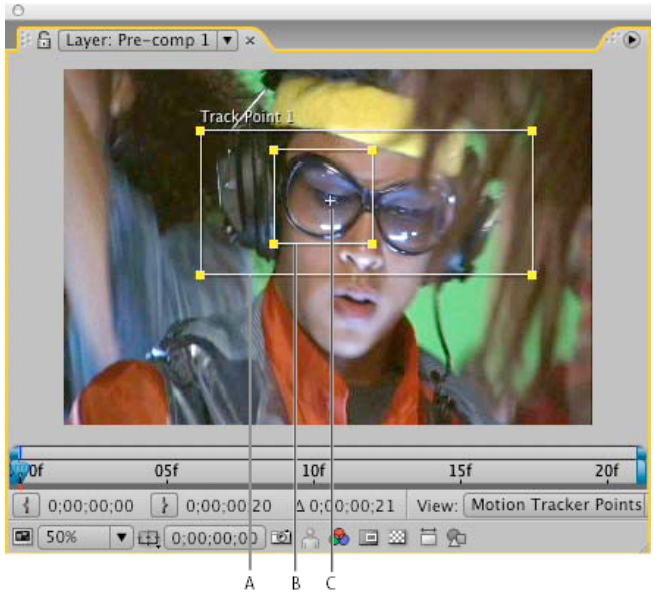


Figure 16: Tracking in After Effects. A is the Search Region, B the Feature Region and C the Attach Point. Picture taken from [8].

3D programs can however also be used for 2D tracking. *Autodesk Maya Live* was designed for matchmoving in Maya, so that a CGI object could be added realistically in a 2D real footage scene, even with moving camera. Objects are tracked in the live action scene and then a solver calculates the camera position for a given image. [9 p. 215] The tracking method is very similar to the one in After Effects. An outer box determines the range to search for the target pattern, inner box defines the pixel pattern to match, and a cross-hair defines the target point. [9 p. 224] The procedure when matchmoving using Maya Live is to first track the 2D points, then to solve and compute animated camera, but it is also

possible to use the tracking information without solving for camera movements. This results in points that move on a 2D plane, that can then be used in whichever way needed, for example to control the x and y movements of joints on a CGI character's face.

Autodesk Matchmover is another camera tracking application that allows the user to export camera data, based on 2D tracking of video or film, to any animation or compositing program, making it easy to add a CGI object into a real footage scene. [10] Matchmover is built into Maya 2011, and Maya Live is built into Maya 2009, but they both work in similar fashion and can both, as well as other 3D matchmoving softwares, produce 2D tracking that can be used as controls for a CGI character's face.

When the tracking information is in place, they can be added to the character in any 3D application that offers character animation, such as Maya, 3ds Max or Softimage.

Camera

Professional optical motion capture cameras are very precise, fast and costly. *Vicon*[4] offers cameras that can shoot up to 2,000 frames per second and capture up to 16 megapixels with the frame rate of 120 frames per second. This speed and accuracy makes it possible to track every single movement of a marker, almost regardless of how little it may be. The marker count can also be very high without negatively affecting the tracking results, *PhaseSpace*'s[5] motion capture system can track up to 256 unique active markers at the same time.

Is all this speed and accuracy necessary? What if the goal is not to produce an anatomically correct replica of a human face and expressions but perhaps a cartoony character that needs to emit emotion? The resolution and frame rate for webcams are drastically lower than in professional motion capture cameras, and yet they can capture lip sync and emotions. A cheap webcam can track up to 30 frames per second and has resolution on the range from 0,3 up to 3 megapixels based on the supply from Logitech[6], Trust[7] and Creative[8].

The frame rate of a webcam video is however based on the computer it runs on. If the computer has a weak processor, is low on disk space or has many programs running in the background, this will lower the frame rate.

---

[4] http://vicon.com/

[5] http://www.phasespace.com

[6] http://www.logitech.com

[7] http://www.trust.com

[8] http://www.creative.com

Mapping The Face

In most parts of the body, skeletal elements create movement around joints. This is not the case for the facial movements, as they are the result of *"highly variable thin sheets of muscle which attach -- not to bone -- but from skin to skin."* [3] Our skin covers these small muscles and deforms as they are activated and deactivated. Mapping these movements with visual markers is not easy because the skin slides on top of the bones. This means that the placement of the markers needs to be carefully thought out in regards to which spots have the most influence on the surrounding areas, or as it says on Face Robot's help website: *"The human face has a finite number of useful landmarks to capture. The areas between are only reacting to the movement of these points, albeit in a very complex manner."* [5]

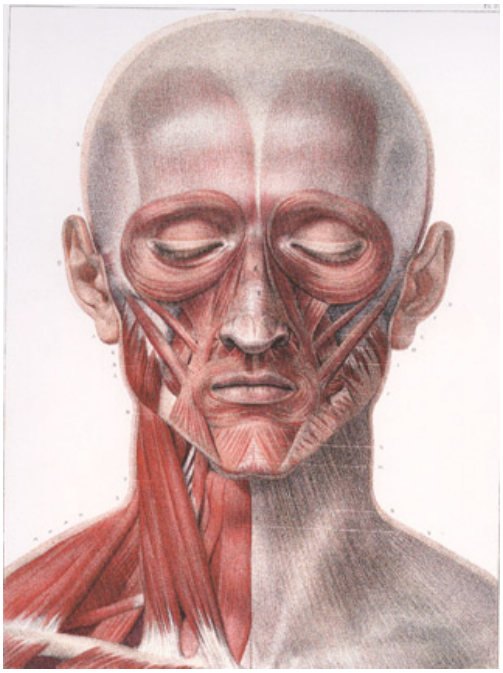


Figure 17: Illustration of the facial musculature. Picture taken from [3].

Face Robot supports motion capture, and provides a guide as to where the visual markers should be placed on the face in order to capture all movements. For Face Robot, 30 to 35 markers need to be placed on the critical areas of the face, and within the program there is a "Soft Tissue Tuning" function that accurately moves the face in response. In Face Robot's marker placement guide, three markers, one placed on the bridge of the nose and two on each side on the top of the forehead, are used as stabilization markers, to define the global movement of the head. [6] The placement of the markers is based on 3D tracking and can, as well as the suggested marker naming convention, be seen in figure 18.

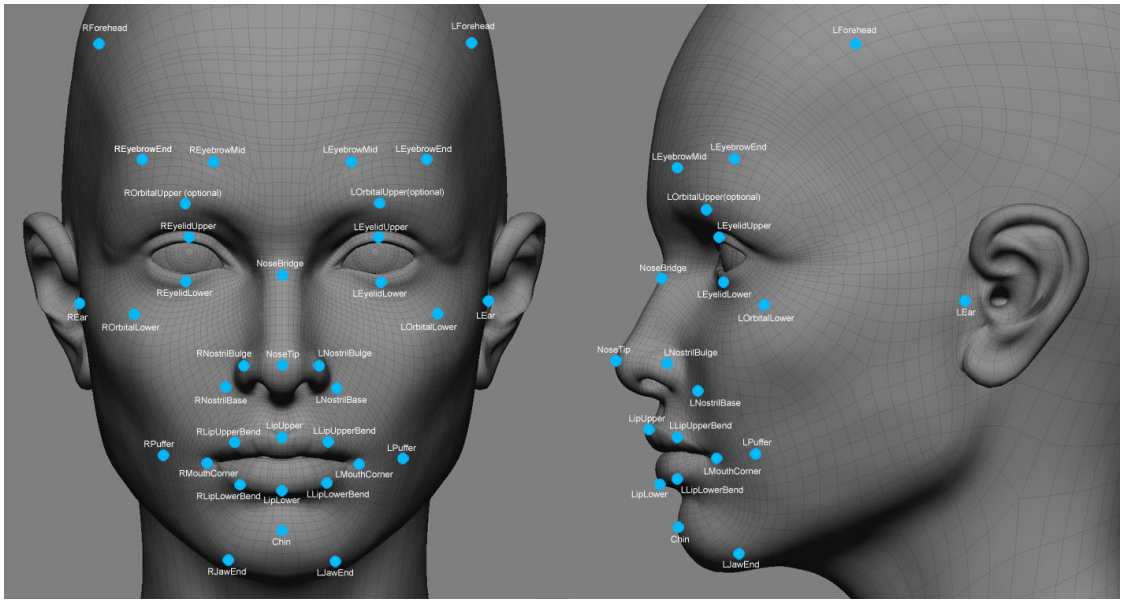


Figure 18: Face Robot's marker placement and naming convention. Picture taken from [36].

PhaseSpace is a company founded in 1994 that specializes in developing products, both hardware and software, for optical motion capture. They use IMPULSE cameras and LED markers so that the risk of swapping between markers during tracking, as well as marker occlusion, is not a threat. [7] Below is a picture of the marker setup for PhaseSpace that is, as well as Face Robot, based on 3D tracking.

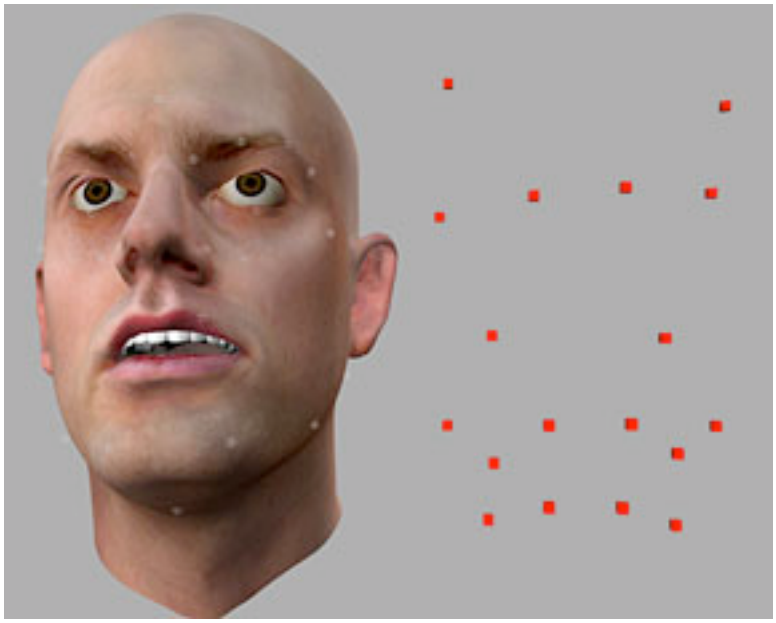


Figure 19: Marker placement in PhaseSpace. Picture taken from [35].

Some parts of the face are more important to track, in order to obtain a "lifelike" character, than others. For example, the movement of the jaw is not nearly as important as the motion of the lips in order to produce believable speech. Minor changes in the configuration of the lips, as well as interactions of the tongue and lips with the teeth and oral cavity, decide if readable movement of speech can be achieved. [3]

Requirement Specification

The following is a list of basic requirements, based on the conducted analysis, my project needs to fulfill in order to be able to answer the problem statement.

Narrative
- This project needs to result in a video where a CGI character's face and expressions are clearly visible to the viewer.
- The character must emit emotion through facial expressions, in a manner that it can be seen and understood by the viewer.

Character
- The CGI character needs to have facial features that are able to emit emotions.
- The actor from whom the motion is captured, needs to be the same gender as the CGI character.
- The actor needs to be capable of portraying emotion via facial expressions.

Recording
- The recording equipment needs to be cheap and easily accessible.
- Markers, that are clearly distinguishable from the actor's skin, need to be placed on the face of the actor, according to previously defined marker position.
- The lighting must make it possible for all markers to be visible to the camera at all times, in all poses the actor makes.
- The recording must be stable, or done in a way that makes stabilization afterwards possible.

Post-Production
- The markers need to be tracked in a tracking program.
- The tracking information must be combined with the character in a 3D application.
- The facial mesh of the CGI character must be controlled by the tracking data.

Design

In the following chapter I will discuss the design considerations that are necessary to be able to answer the problem statement. First I will decide on the requirements for *The Character* and actor, followed by reasons for the *Narrative* design. Then I will move on to more technical aspects when choosing the design for the *Hardware*, *Software* and *Marker Setup*.

The Character

In order to save time and keep focus on the animation, I decided not to model a character from scratch myself, but rather download a free and available model from the internet. My requirements for the character were that it would have the same basic facial features as a human, in order for the motion capture to move the face mesh appropriately to my actor's movements. However I did not want the character to look too realistic, in order to stay clear from falling into the uncanny valley.

On a website called *Rigging 101*[9] I found a few models that were free to download for non-commercial use. Three characters matched my search criteria:

Figure 20: The three characters that I could have used as my motion capture model.

Although I preferred the look of the character wearing glasses, I decided not to use him as his eyes are very small, which might decrease his ability to use them to express a feeling to the viewer. The simplicity of the blue character appealed to me, as I felt it might focus the viewers attention to the facial expressions rather than his facial features. However I decided to go with the grey character, not only because of his exaggerated "eyebrows", which I felt could underline facial expressions, but also because his looks appealed to me, especially his detailed teeth and eyes.

[9] http://student.vfs.com/~m07goosh/rigging101/index.htm

Because I chose a male-looking character, as argued in the Analysis, the actor needed to be male as well, also in order to match the voice with the visuals. I needed someone who had experience in acting, to ensure that he could portray an emotion in a clear manner, and preferably he would have facial features that made it easy for me to place markers on tricky areas such as the eyelids.

Albert Sölvi Óskarsson is not a professional actor, but he has performed in several theatrical productions and received positive reviews for his acting. He is also used to performing, being a saxophone player, and therefore I decided to cast him as my actor.

Figure 21: Albert, my actor.

Narrative

I decided that the video clip should be very simple, in order to maintain the viewer's full focus on the character's face and expressions. Therefore I decided that the clip should be of the character speaking directly to the camera, so that only the character's face was visible on screen, with neutral background.

When choosing an emotion for the actor to portray, I decided to choose a positive emotion rather than negative, for the simple reason that I wanted my working environment throughout the semester rather to be positive than negative and thereby keep my spirit up. Therefore, selecting from the basic categories of emotions, as seen in figure 12, I decided to choose *Love*.

In order to make it easy for Albert to display love, I decided to have him talk about something that he loved, while being filmed, and to direct the reply to the camera. Albert has a cat that he loves dearly and therefore I decided that the narrative of the video would be him describing his cat to the camera.

Because the emotion was to be visible through the facial expression and not heard by speech, I decided to ask him to keep the answer as objective as possible, and leave out subjective adjectives. Basically, I wanted the text to be as neutral as possible so that he would rely solely on facial expressions to emit the emotion of love.

Hardware

The main goal of this project is to see whether facial motion capture can be created with the use of cheap and accessible equipment. Therefore I decided that the recording equipment would be a single camera, and that the tracking would therefore be solely in 2D.

Camera

I really wanted to focus the project on keeping it simple and cheap and therefore I decided to borrow a cheap three year old webcam that my father owns and uses for online calls. It is a Creative WebCam and costs about 100 DKK, although probably less.

Headgear

In order to capture reliable tracking information, either the camera needed to follow the actor's face, or the actor would have to keep his face still and aimed towards a static camera. As I found it very limiting for the actor and his expressions to keep his head still while performing, I decided that the camera should be mounted to the head with headgear.

The purpose of the headgear was to stabilize the camera so that it constantly aimed to the face, allowing the actor to move his head around freely during his performance. It needed to be tightly fitted to the head, so I chose to use the insides of a construction helmet that fitted neatly around the head as a base. On this, on each side of the head, I planned to attach two wooden bars that on one side had the camera attached and on the other a small weight, in order for the gear not to tip forward.

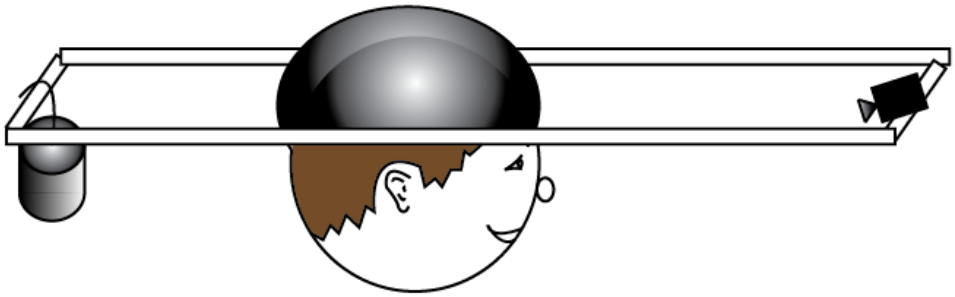


Figure 22: The headgear design.

Software

I needed software for a few different tasks. First I needed a capture software to get the information from the camera to the computer, secondly I would need a tracking program to track the physical markers and lastly I needed software to apply the motion capture data to the character.

Capture Software

The webcam came with a supporting software, called *Creative WebCam Center*, but this software only allows the recording of visuals, and no audio. In Windows, there is a built in capture software called *Windows Movie Maker* that can be used to record audio and video at the same time, although the image quality is not quite as good as when using the Creative Center. However I chose to use Windows Movie Maker in order to capture the video and audio at the same time, with the same software.

Tracking Program

The tracking options are listed in the Analysis and out of these I decided to use Maya Live because I am familiar with working with Maya, and also because I would like to get familiar with the workspace in Maya Live for future projects.

Cleanup and afterwork software

I decided to use Maya to apply the 2D tracking information to the character's face because, as previously stated, I am familiar with working in Maya and I seek every opportunity to learn more and become more skilled in the program. I decided to create as many joints as there were physical markers, and attach them to the face of the character, at places that corresponded with the physical marker setup. I would then need to paint the weight of each joint to the facial mesh in such a way that they influenced similar areas as the physical markers on the actor's head.

The eyes and jaw would need to be keyframe animated after the motion capture had been applied to the character, and as stated in the Analysis, it was important to add saccades movements to the eyes to make them more realistic. There is no good way of tracking the eyes and jaw while recording with the equipment I planned on using. The same goes for all motion in the z-axis direction. However I did not want too much animation in the z-axis because I wanted to rely on the 2D tracking as much as I could, and therefore only add motion in the z-axis where I deemed necessary.

I decided to render the animation as an image sequence that is later put together using Quick Time Player or a similar application, because this method has provided good results in the past. I would then add the sound to the visuals in Final Cut Pro, as well as introductory text, because Final Cut is renowned for producing better render results with audio than Maya.

Marker Setup

To find the optimal placement for the markers, I had the actor perform very extreme facial poses, in order to spot which areas of the face were the most influential.

Figure 23: By asking Albert to perform extreme facial poses, it was easy for me to see which areas of the face were more influential than others.

As the two examples of marker placement guide reviewed in the Analysis were both based on 3D tracking, they could only serve as a reference and not a final template.

Figure 24 shows the chosen marker placements and it is noteworthy that during the facial expression test, it became clear the ears would need a marker each, as the actor often used them in his expressions.

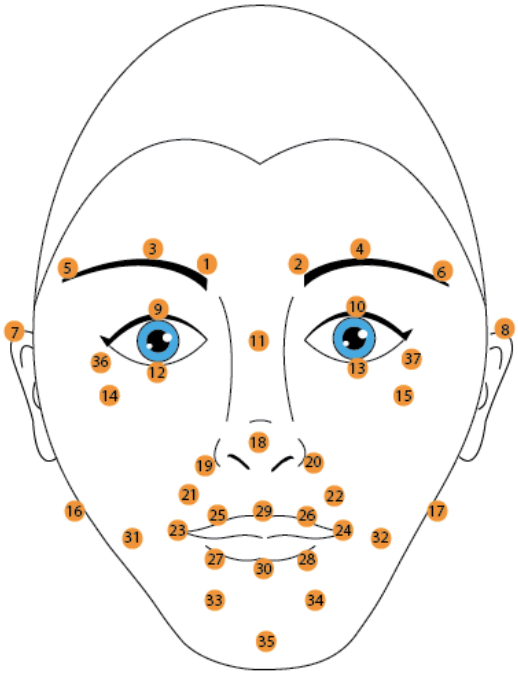


Figure 24: The designed marker placement.

I decided that I would not need any stabilization markers placed on still areas of the face, such as the top of the forehead and on the tragus on the ear, because the camera was to be mounted to the head tightly.

To design the markers, I performed minor tracking tests with different markers. A briefing of these tests can be seen in the Appendix. The design that gave the best result was gluing black paper circles on white plastic squares, as can be seen in figure 25.

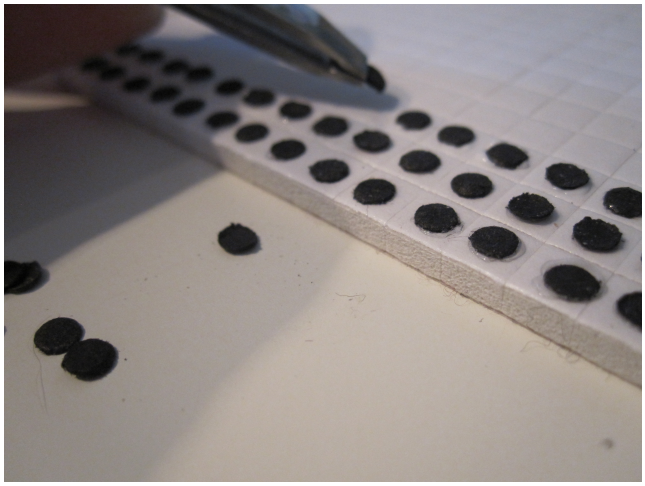


Figure 25: The design of the markers.

Implementation

After having conducted the shooting, major problems occurred that forced me to do another shooting. Following is a description of what went wrong in the first *Shooting* and how I fixed it in the second, as well as some interesting or problematic areas that arose during the production in *Maya*.

Shooting

In the first shooting I used Windows Movie Maker to capture the data recorded by the webcam. However, the sound was not synced perfectly with the visuals, which in itself would not be a big problem because this can be adjusted in Maya. The bigger problem was the motion blur and noise of the footage, which made it very hard for Maya Live to track the markers.

After having manually tweaked the tracking information so that it was roughly accurate, I used the tracked points to control the x and y movement of locators via expressions, which the facial joints were then parented to, applying the motion capture to the character's face. With the tracked points controlling the character's face, it became clear how unstable and rough the motion capture was, due to the shaking of the headgear mounted on Albert's face. I tried to stabilize the video in After Effects, but prior to the shooting I had not placed any stabilization markers on Albert, making it hard to find two spots on his face that were completely still. The data was not useable because of how unstable it was, and therefore I decided to conduct another shooting.

Figure 26: The first shoot.

As can be seen in figure 26, the headgear was made of wood, which was not very solid, causing it to bend and jiggle every time Albert moved his head during speech in the first shoot. For the second shoot, I stabilized the headgear by attaching more wood to strengthen the construction. As the headgear was very large, if Albert was not careful, it bumped against the wall behind him when he moved. However this could easily be avoided so I decided not to change the overall design of it, only strengthen it. The improved headgear can be seen in figure 27.

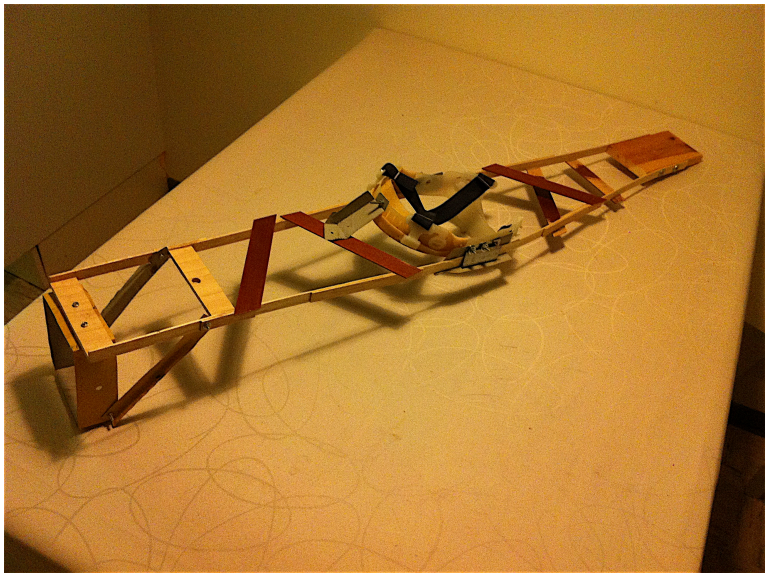


Figure 27: The renovated headgear.

However, this was not enough to get a stable outcome. The headgear was so tightly fitted to Albert's head that its weight pushed his forehead up- and downwards when he spoke. The movement was now much less than in the previous version of the headgear, and only vertical. As I could not come up with a solution to go around this, I decided to place four stabilization markers, two on the top of the forehead and two on the tragus on the ears. These markers are number 36 to 39 in figure 28, which shows the final marker placement.

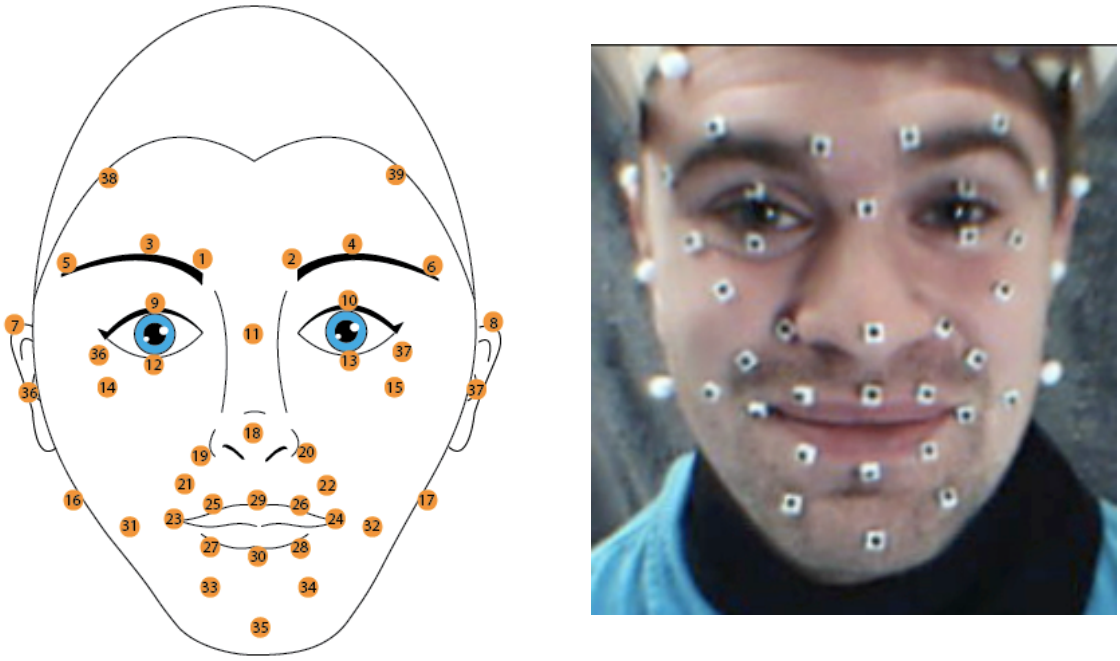


Figure 28: The final marker placement.

Stabilizing the video this way reduced the motion blur from the previous shot remarkably but I also decided to change from using Windows Movie Maker to Creative WebCam Central. I had chosen Windows Movie Maker for the capture because of the option of being able to track audio and visual at the same time, and in the first take it turned out to be out of sync, so I might just as well use another software for capturing the audio. I decided to use Quick Time Player and the microphone on my computer to record the sound and add it to the visuals in Maya.

Because of the stabilization markers, stabilizing the video in After Effects was more successful than after the first shooting. I used the markers on the tragus of the ears to stabilize, as the ones on the forehead often moved because of the headgear.

Maya

The tracking in Maya Live was much more stable than after the first shoot and there were no problems regarding this stage, except for a bug in the software that inhibits a tracking scene to be reopened in track mode once it has been closed. However the tracking information is not lost, so I made sure to finish the tracking before closing the program and that way I would not loose any information.

I created one extra joint and located it at the top of the character's head. I used this joint to stabilize the parts of the body that were not meant to be influenced by the motion capture. This can be seen in figure 29.

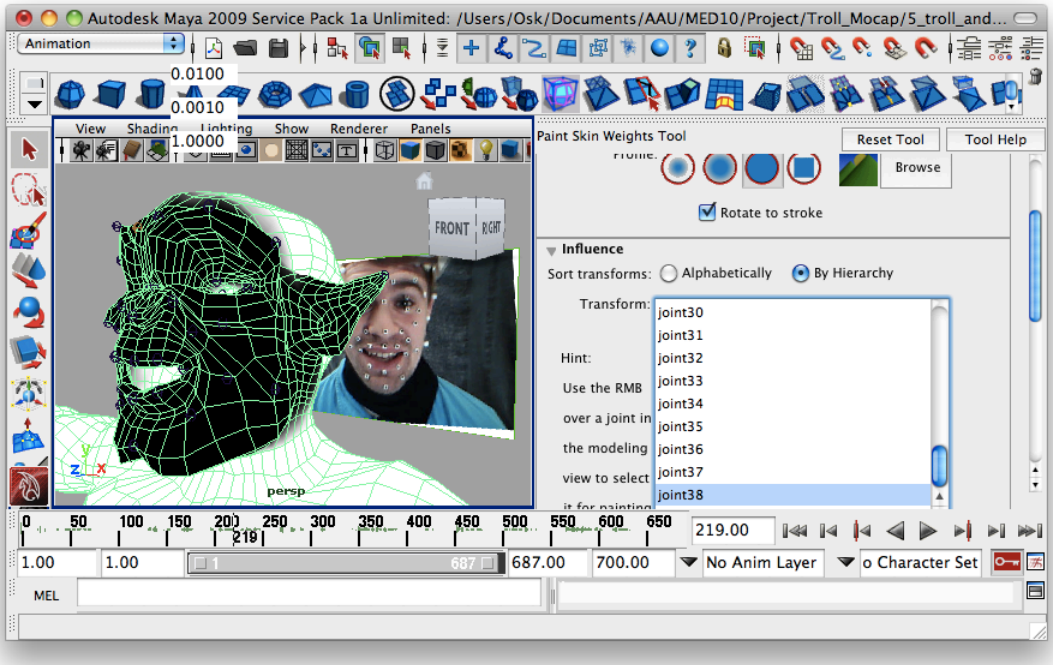


Figure 29: Painting the weights in Maya. Joint 38 is the stabilization joint for all areas that were not meant to be influenced by the motion capture.

As previously mentioned, the joints were parented to locators that had the x and y movement of the tracked points. To do this, in the Expression Editor I created an expression for each locator that stated:

locator_n.translateY = trackedPointShape_n.locationY;
locator_n.translateX = trackedPointShape_n.locationX;

In this example, n denotes the number of locator, which was the same as the number of joint, as well as the tracked point.

As I wanted as little keyframe animation as possible to make the most out of the motion capture, I only tweaked the z-axis movement of joints near the mouth and on the top eyelids. Before tweaking the eyelids, every time the character blinked, the eyelid intersected the eye. For the mouth, I pulled joints 21 to 24 a bit backwards when the character smiled, because it looked more natural that way. Also, joints 33 and 34 needed z-axis animation, because sometimes the jaw would intersect with the face mesh and require to be pulled forward.

Some of the tracked data needed to be tweaked as well, especially the points controlling the lips. When Albert smiled, his lips got thinner, which reduced the distance of markers on the top and bottom lip. This made the lips of the character intersect, and therefore I needed to tweak the y-location of these points whenever his mouth was closed.

Testing

In the following chapter I will explain how I decided to test my project, why I chose to do it this way and review the results. I conducted *Expert* testing, and to *Answer the Problem Statement,* I performed direct observation test in a controlled environment, as well as obtaining quantitative data through questionnaires. Each of these method answered questions regarding different aspects of the project.

Asking Experts

Arnar Gunnarsson, CG supervisor at CAOZ, has been very helpful to me throughout the semester, providing me with inspirational suggestions and interviews. He can be considered as an expert in his field, and when the project was ready, I decided to ask him to give me his professional opinion.

Objectives

The first objective for asking an expert was to get a professional judgement of my animation. As previously mentioned, there is no check list or a scale that an animation can be measured up against in order to judge whether it is good enough. Therefore I wanted to get a professional review of my animation and its creation, and also to put it in perspective with the industry today. My goal was to receive constructive critique that would both benefit this project as well as my future work.

As I mentioned in the introduction of this report, CAOZ was planning the production of television series, where they would need to produce a lot of material in a short period of time, and were looking for solutions in order to speed up the process. Although the production of the series was canceled, my second objective for seeking their advice was to find out whether, if they would indeed produce the television series, my project could be seen as a solution if it were further developed.

Methods and Procedure

In order for the expert to fully grasp the purpose of my project and how it was all set up, I decided that a demonstration and a walkthrough of the process would be fit to begin with. For the demonstration I first wanted to explain to the expert all the equipment I had used for the creation, and then show him the rendered video as often as he wanted. This would then be followed by a walkthrough of the setup within Maya, in such a way that he could play around with it and become familiar with it.

This demonstration and walkthrough was then to be followed by a semi-structured interview, which was focused around his assessment of the project. This included asking

him for ideas how to improve it, as well as predictions about whether a further developed prototype could be beneficial to the company.

Results

Arnar payed full attention during the demonstration, and the walkthrough proved that he understood the setup.

He said that the outcome looked better than he had expected when he heard what equipment I had used. He felt that he could see that the character was thinking as he spoke and that he looked fleshy, which he described as natural movement of the skin. He felt that the face was very wobbly, which might either be because of the webcam used for tracking or because the tracking equipment was not stable enough. According to him, this wobble of the skin was more obvious because the eyes were very still in the sockets. He suggested that I would add more eye movement, drift or saccades, in order to make the eyes come more alive. Also, he pointed out that the audio and visuals were not completely in sync, and that I would have to push the audio five to ten frames backwards for it to match.

He deduced that the video clip as it was could not be used directly in a movie production, and that it would need some kind of stabilization first; perhaps by using blendshapes. He added that this method could go a long way in the production process and might be good as a guide for additional animation. *"Paint by numbers"* he called it, and said that the benefit of using it that way was that the animators would not need to make any guesses about timing or poses, the information was already there and only needed tweaking.

In regards to the production of the motion capture, Arnar said that there were many ways of conducting it, and the way I had done it was reasonable and functional.

He concluded with the words "you can get away with this, it works" and added that this sort of production was especially good when mass quantity of material needed to be produced in a small period of time. The tongue could also be automatically animated with a sound file, a lip sync animation method I described in the Analysis. When I asked him if a solution such as this project would be an option for CAOZ if they would resume production of their previously planned television series, he replied that it would definitely be worth taking into consideration.

Briefs of all interviews with Arnar can be seen the Appendix.

Answering The Problem Statement

After the interview with Arnar, I tweaked the result a bit according to his comments. I added more saccades to the eyes, both to make the character "more alive" and to decrease how wobbly the mesh was compared to the still eyes. Also, I added more subdivisions to,

or smoothed, the character, which called for minor extra tweaking of the z-location of the joints, as the gums then sometimes intersected the chin.

With the expert testing out of the way, the project was ready for general audience and a verification or refute of whether my project would be the answer to the problem statement. To do so, I performed two rounds of tests. The first round consisted of two qualitative tests that were initially supposed to serve as a valid answer to the problem statement. However, I was not satisfied with the clarity of the results so I decided to obtain qualitative data, hence I added a small quantitative test. I will include and expand on all three tests however, as each of them provided me with valuable insights to my project.

Qualitative Tests

As the title of this section implies, the objective for these tests was to see whether my project could rightfully give an answer to the problem statement, which was:

> *How can facial motion capture be created, using cheap equipment to capture the motion, in such a way that emotion can be read from the facial expressions?*

The requisite for this to be true, was that I could show that the viewer felt that emotion was emitted from the character's facial expressions. I planned two tests to provide qualitative information regarding this.

First Qualitative Test - Methods

First of all, I wanted to ask on a questionnaire what emotion, if any, the character emitted. In order to keep the question and answering options clear, I decided to form the answer possibilities visually rather than verbally. That way they would be asked to draw a circle, on the image below, around what they felt the character experienced during the video.

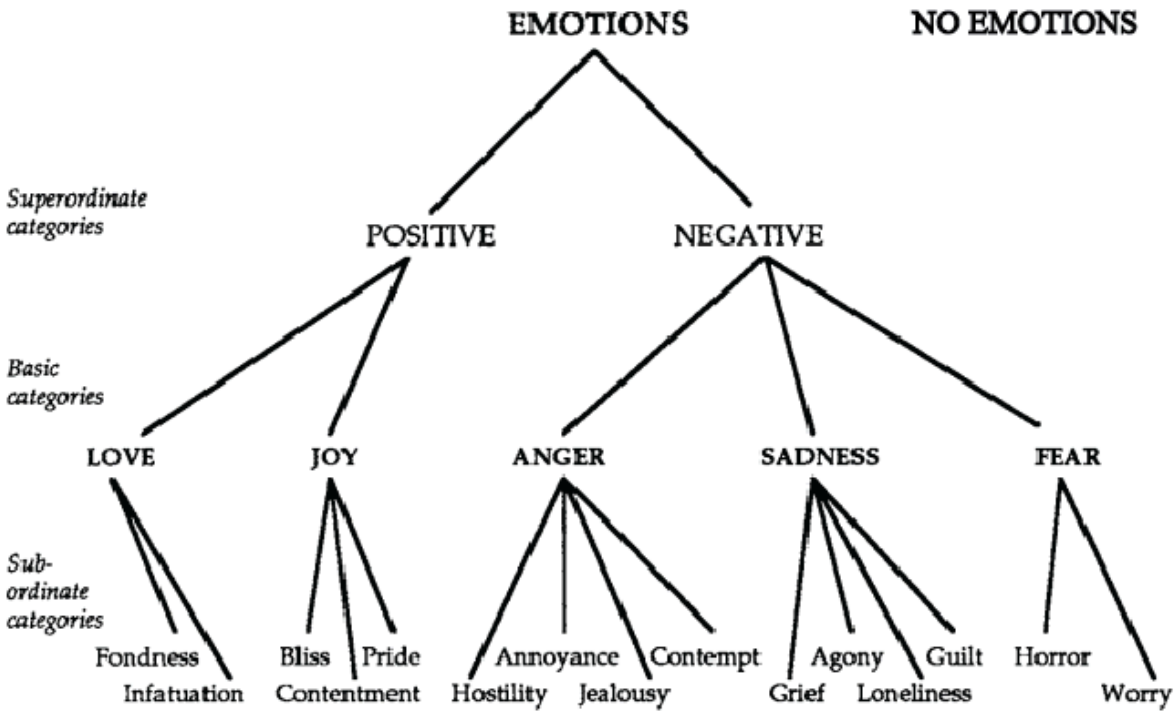


Figure 30: This image is the same one as in the Analysis, but with an added "NO EMOTIONS" option.

I decided to allow the test participants to draw as many circles as they saw fit, and in any level of category, to increase their expressive freedom. I also wanted to know what the test participants based their assumptions on, in order to see how big of a roll the facial expressions played when emitting the emotion. Lastly, I decided to ask if they had any written comments, because it is my experience that this freedom of expression for the test participants often reveals both interesting aspects of the project and testing scenario, and can also bring to light any flaws or elements that I failed to consider in my production.

Second Qualitative Test - Methods

As explained in the Analysis, according to Hatfield, emotional contagion is visible from the facial expressions of people who score highly on the Emotional Contagion Scale (ECS). This means that if a person, who is susceptible to emotional contagion, witnesses another person emit emotion, the emotion is copied onto that person and can even be seen in his or her facial expressions.

Based on this theory, I reasoned that if my character does emit emotion, and is watched by a person who is highly susceptible to emotional contagion, this emotion should be visibly mirrored in the facial expressions of that person. And oppositely, if the person is susceptible to emotional contagion and does *not* visibly react by displaying similar emotion as the character was to emit, this means that the character did not successfully emit emotion. However if the person is not susceptible to emotional contagion, regardless of this gives me no information about whether the character emitted emotion or not.

A way for me to test this was to measure test participants' level of emotional contagion and compare the results with their facial expressions while watching the video. I decided to conduct a direct observation in a controlled environment, as well as capturing the test participants' facial expressions on video tape. That way I could compare the results after having calculated the ECS score and eliminated those who scored very low, as those instances were not useful to me.

Practical Issues For The Qualitative Tests

The prerequisites for the test participants were that they needed to be able to see and interact with people, as well as understanding English. The language used in the video is rather simple and people in Iceland are used to hearing English on the television, and therefore I did not regard this as a concern. I did not want to test on children because their social skills have not been fully developed yet.

The ECS test is quite personal and I was concerned that the test participants might be uncomfortable revealing such personal information to a stranger. Therefore I chose to stress the fact that their answers would be kept strictly confidential and give them personal space when filling out the questionnaire.

I decided to translate the questions to Icelandic so people would have an easier time filling out the form. Also because some of the questions were rather personal and I did not want to risk them skipping a question because they did not understand it.

Procedure Of The Qualitative Tests

I asked the Director General at Orkustofnun (National Energy Authority), where I was once employed, if I could perform the test there, as I was familiar with the surroundings there and knew that the staff was friendly. The Director General welcomed me and offered me a conference room for my setup, as well as sending all staff members a notice of my purpose there and encouraging them to assist me.

I situated the equipment so that a camera was aimed at the test participants' face, who sat directly in front of the computer and there was a screen located behind the test participant's chair, which reflected the video to the camera. This was convenient for me as it meant that I did not need to specifically signal to the camera that the video had started, which would otherwise have been necessary for the later inspection of the test participants' facial expressions.

Figure 31: The setup for the first two tests.

As a test participant walked in the room, I asked him or her to sit down and explained that I would ask them to watch a video while being video taped. I would then ask them to fill out a questionnaire, but first I needed to ask for their signature for a consent form that grants me permission to use the information acquired during the testing session for my project.[10] I played the video and as it is rather short, I asked the participants if they wanted to watch it again or move on to the questions. I then handed them the questionnaire, with questions for the first test on one side and the second on the other, and offered them candy for their effort. When they were done filling out the form, I thanked them for the help and put the questionnaire in a secure place.

At the end of the day I went through all the videos captured of the test participants' faces and noted whether they had changed their facial expression during the course of the video.[11]

Results For The First Qualitative Test

26 people people participated in the test, 13 females and 13 males of all ages, at the average of 49,6 years. I had told them that they could draw a circle around any of the emotions, or absence thereof, they felt fit, and some drew circles around only the basic category, some drew circles in all categories and others only in the sub-ordinate categories. Two persons answered that they felt the character had not experienced any feeling during the video. One of those two approached me later during the day and confessed to having misunderstood the questions and directions and to probably not having payed any

[10] The consent form can be seen in the Appendix.

[11] Pictures of peoples expressions during the video display and my notes regarding them can be seen in the Appendix.

attention to the video. That same person thought, and shared this opinion with one other test participant, that the character himself was a cat, talking about himself.

Many of the test participants became uneasy when I told them that I would videotape them while they watched the video and suspected that I would play a prank on them. Some of the female participants also became very aware of their looks, adjusted their hair and asked if this (the camera recording) was really necessary.

No participant marked *Negative* in the superordinate category, however one marked *Loneliness* in a subordinate category under the basic category of *Sadness*. Nineteen marked that they felt the character had experienced *Positive* emotion, and in the basic categories twelve circled *Love* and fourteen *Joy*. In the subordinate category under *Joy*, fourteen circled *Contentment* and twelve *Pride*, and under *Joy*, fourteen circled *Infatuation*. This distribution can be seen in figure 32.

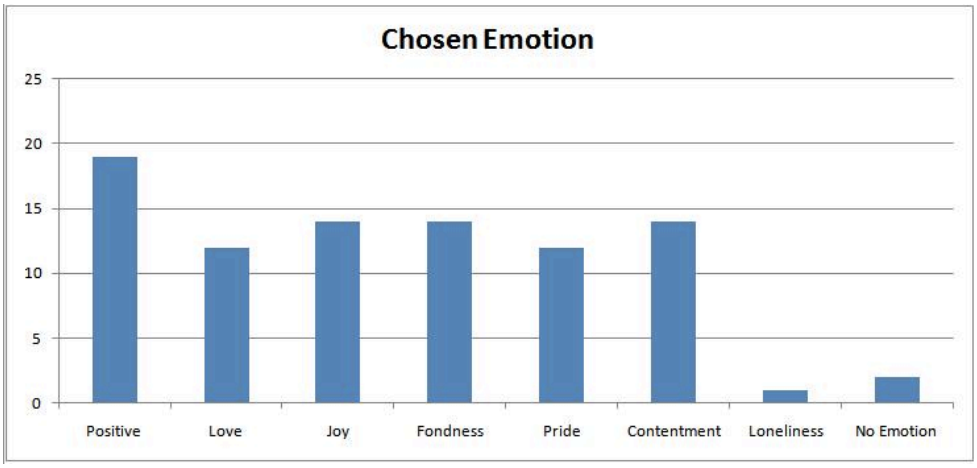


Figure 32: This graph shows the number of people who circled these emotions. Those emotions who did not get circled are not included in the graph for the sake of clarity.

When answering what they based their assumption on, eighteen marked that they based it on the character's words, eighteen based it on his facial expressions, fifteen on his tone of voice and one based it on something else. This is demonstrated in figure 33.

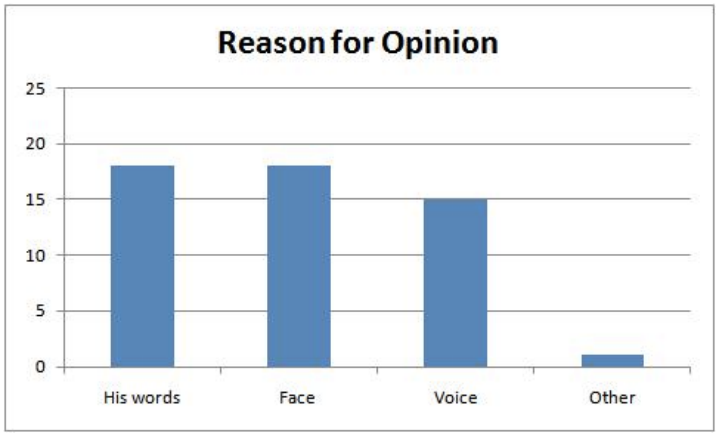


Figure 33: This graph shows the number of people who chose these factors as basis for their opinion, in regards to the character's emotion.

Five people wrote written comments. One stated that the character was nice, despite not being very pretty while another felt that the character was too made up for his opinion. One said that the character both seemed happy and joyful and also rather proud. Pity, sympathy and care were the feelings one participant felt from the character and one described him as a teacher, educating about his cat, that is, its behavior and strange colours.

Results For The Second Qualitative Test

A few of the participants admitted after the test was done that when they started watching, they were not sure of what they should be paying attention to, so they tried to carefully memorize what the character was saying.

The second test was a comparison of the result from the ECS and changes in participants' facial expressions while watching the video. As previously mentioned, the higher the score on the ECS, the more susceptible to emotional contagion a person would be said to be. The participants rated sixteen statements on the scale from one to four, where four meant that this statement was always true for them, and one meant that it was never true for them. The scale can be seen in the Analysis. All participants scored at least 2,2 on the scale, denoting all of them to be susceptible to emotional contagion. This is displayed in figure 34.

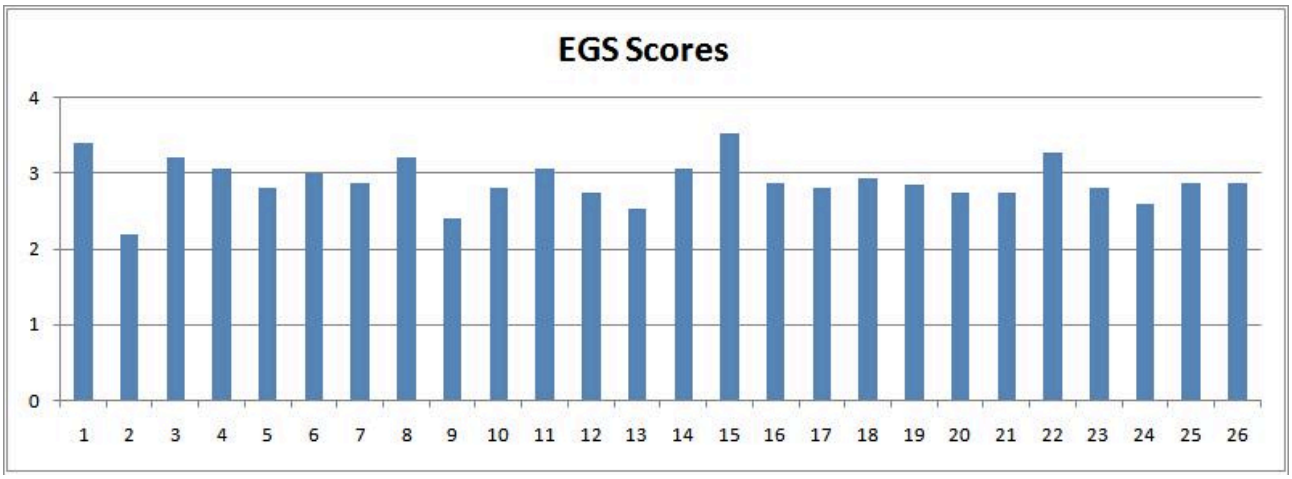


Figure 34: This graph shows the EGS Score for all test participants.

The scale can be divided into five categories; happiness items, love items, fear items, anger items and sadness items. Figure 35 displays the results for these categories.

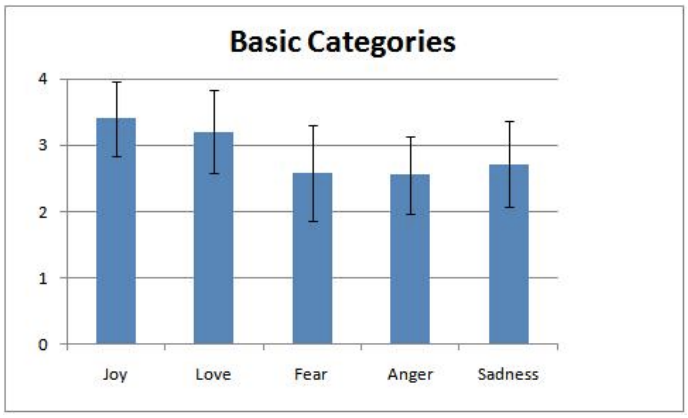


Figure 35: This graph shows the average score for each basic emotion.

Many of the participants were eating candy while watching the video, so it was sometimes hard to recognize changes in their facial expressions. Twelve of the participants showed little to no change in their facial expressions during the video, seven people smiled vaguely and seven smiled in an obvious manner. Ten participants showed signs that they were focusing, by either leaning forward, tilting their head downwards or peering their eyes. A display of the changes in the participants' facial expressions can be seen in the Appendix.

Quantitative Test

In this test I wanted to obtain quantitative data because, although the first two tests provided qualitative information regarding the project, I felt they were too open to give a clear answer to the problem statement. I wanted clear results of whether the facial expression alone was enough for the user to see emotion in the character.

Quantitative Test - Methods

The tone of the actor's voice and the words he used had the possibility of influencing the results in the two qualitative tests, despite the neutral design of the narrative. In order to isolate the facial expressions, I decided to make another test where the sound was turned off. This time, instead of basing the result entirely on what emotion (if any) the viewer felt that the character experienced, I decided to make a comparison between the original video and the motion capture.

The chosen emotion for this project was Love, however I had yet not tested to see whether this emotion was emitted from the original video. I decided to perform two simple identical tests, to see what emotion (if any) was emitted from each video, and to compare the results. I argued that if the results were homogenous, I could conclude that the problem statement had been fulfilled.

In order for the results to be clear, and to minimize confusion for the participants, I wanted the questions to be precise, and only include the basic category of emotions to select from.

I decided to upload the videos without sound to the internet and create an online survey with focused questions. I could then promote the survey, in order to get test participants to take it, on a popular Icelandic forum called ER[12], where a big variety of people meet to chat about current events. This way I would get participants of various ages from all classes of the society.

Quantitative Test - Results

Out of the 124 test participants, 12 (9.7%) felt that the character had not experienced any emotion, and equal amount of participants felt the same with Albert. 51 (41%) felt that the character experienced Love while the character spoke and 57 (46%) felt he experienced Joy. For Albert, 44 (35%) felt that he experienced Love and 47 (38%) Joy. 4 people felt that the character experienced sadness, but none of them felt he experienced Anger or Fear. 18 (15%) participants felt that Albert had experienced Sadness while he spoke, 2 that he had felt Anger and 1 Fear. These results are shown in Figure 36.

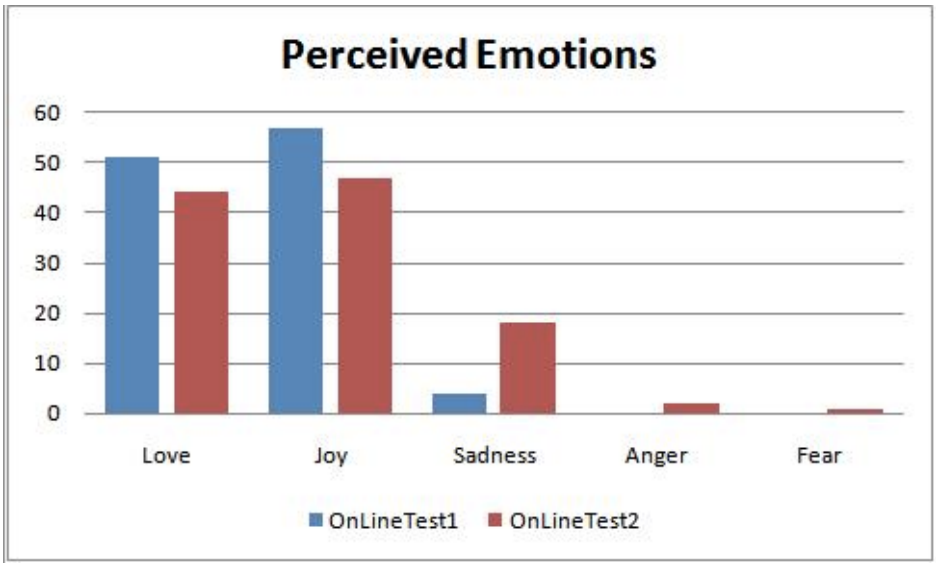


Figure 36: This graph shows which emotion the test participants thought the character was experiencing during the video. OnLineTest1 is the motion capture character and OnLineTest2 is Albert.

[12] www.er.is

Discussion

In this chapter I will review interesting aspects of the project, as well as my decisions throughout the process. Also, I will reflect on the procedure of tests and the results I obtained. First I will address the *Test Results*, including all three testing scenarios, followed by a quick summary of my findings regarding the term *Cartoony*. Lastly I will discuss the *Technical Aspects* of the project, both in regards to hardware and software.

Test Results

The concept of this project is rather simple and I wanted to obtain both valuable qualitative data as well as clear quantitative data.

Expert Test

As previously mentioned, when I first started thinking of a project idea for my Master Thesis, the people at CAOZ suggested that I would focus on finding a way to make the process of animation faster for them. At this time they were about to start production of television series, which would require a lot of material in a short amount of time, that later got canceled. However, even if I knew my project would not be used professionally, I still wanted a CAOZ spokesperson to judge it as a prototype that could later be developed if they gather that it had potential.

The main flaw Arnar saw in the project was how wobbly the mesh was. This could however be improved greatly in further development, by using a better camera and constructing a more solid headgear that would not influence the movement of the forehead. Arnar said that a product like my this could save animators a lot of time, because it could produce a base, if not more, for them, on top of which they could then build their animation. The main question I wanted to have answered, was whether my project was something that they would consider using in the future, if they for example restarted the development of the previously mentioned television series. To this, Arnar replied that if it would be further developed and made more stable, than it would definitely be taken into consideration at CAOZ.

Qualitative Tests

The qualitative tests were initially meant to answer the problem statement. However, while processing the information they provided, I realized that they were too open to serve as an answer to the problem statement, however they did shed light on interesting aspects of the project.

First Qualitative Test

As can be seen in Figure 36, where viewers were asked what emotion Albert experienced, it is obvious that his interpretation of Love was easily confused with Joy. Although Love was chosen as the emotion to be tested for in this project, it could not be expected that the character only emitted Love as emotion, but rather that it was a combination of Love and Joy, as Albert did in the original video. This changed the test parameters, as now the criteria was not to emit Love, but a similar mix of Love and Joy as Albert had done in the original video.

Although this test indicated that the character had emitted emotion, mostly positive emotion under the categories of Love and Joy, the results did not rule out the possibility that this opinion was also based on the tone of the character's voice or the words he used.

I can think of no good reason as to why some of the participants misunderstood the test procedure or the narrative of the video. As previously mentioned, the camera made some of the participants uneasy, and perhaps they did not pay as much attention to the video as they would have if they had not known that they were being taped. However this assumption is merely a guess and can not be validated.

Second Qualitative Test

All the participants scored above the middle value of the ECS, which denotes all of them susceptible to emotional contagion. Therefore I deduced that all of them should have responded to the video with positive body language. 14 of the 26 participants showed positive change in their facial expressions and twelve showed little to no change. Therefore this test does not prove, nor disprove, that the character emitted emotion. However, this test did not produce very reliable results. There were a couple of factors that I should have taken into consideration before starting the test procedure.

First, it might have been better if I had explained to the test participants in beforehand what I was about to test, rather than afterwards. As previously stated, some of the participants admitted that because they were not sure what the purpose of the test was, they wanted to be sure to remember everything the character said, in case I would quiz them afterwards. This might have made the participants more focused on remembering the speech rather than seeing the character as a person who emitted emotion. This assumption is supported by the fact that ten participants showed clear signs of focusing, by either peering their eyes, leaning forward or tilting their head down.

Secondly, the candy that was intended for the test participants as a gratitude token for helping me, and meant to be eaten while filling out the questionnaire, was available to them throughout the entire testing session. This meant that they were eating candy while watching the video, which sometimes made it hard to distinguish if their facial expression was changing and therefore severely effected the results from this test. Also, some

participants were nervous that they were being filmed, which might have been distracting for them.

Quantitative Test

As can be seen in figure 36, Albert and the character were seen as experiencing similar emotions during their speech. The character scored slightly higher for both Love and Joy, while Albert was seen as sad by some. These results show that the emotion in the original video is preserved throughout the process and is clearly visible to the audience via his facial expression.

Cartoony Animation

I first became interested in the concept of cartoony animation when CAOZ informed me that there was no automatic way of creating cartoony animation available today. Through my research I found that some have tried to invent applications to make this automation possible, but all to a limited degree. I concluded that cartoony is a word that can describe not only animation but also looks of a character and even the acting of a real person. In my reasoned opinion, cartoony is a word that describes it when features of a scene or looks of a character have been exaggerated in a way that compliments the narrative of the story. This exaggeration needs to be carefully thought out in order for it to be complimentary and not confusing, and therefore I predict that this can not be automated in the near future. A computer does not understand a narrative, nor is it capable of thinking up a solution as to how key features of a scene can be exaggerated in a flattering way.

Technical Aspects

In general, the technical execution of the project was surprisingly free of major hindrances, and I believe this is due to the thorough research I conducted before I started designing.

Hardware

Although the headgear fulfilled its basic requirement, of constantly making sure that the camera was aimed at the actor's face, in retrospect I can see possibilities for improvement. The wood I used was not very strong, causing it to wobble a lot the first shoot before I strengthened it. Also, it was rather big, forcing the actor to be careful in his movements, which reduced the freedom in his expression. This could perhaps have been prevented if the headgear had not extended as far backwards as it did, and that instead a weight would be situated close to the back of the head.

The camera I chose was the cheapest I could find and there is no doubt that a better camera would have increased the odds of higher quality tracking. However, I do not regret my choice because, although it was rather extreme using the absolute cheapest I could find, I am inspired by the outcome. Motion capture is by many regarded as technique that

requires very expensive and technical equipment that only productions with big budgets can afford. If facial motion capture can be created with the use of a cheap webcam, such as in this project, it becomes more available for smaller productions that can perhaps develop standard methods in the future.

Software

In this project, I wanted the motion capture to be the main focus and therefore did not want much keyframe animation, except for bare necessities. However, this result could also be combined with the use of blendshapes in order to achieve a more polished animation. For example, adding blendshapes that adjust the lip poses might increase the quality of the lip sync.

Conclusion

The goal of this project was to find out whether facial motion capture could be created with the use of cheap equipment, in such a way that emotion could be read from the character's facial expressions. I created a wooden headgear that mounted a cheap webcam on an actor's head so that it pointed towards his face at all times. By using Maya Live to track the physical markers, and Maya to use the tracked points to control the character's face mesh.

I performed three types of tests on this project; an expert test, two qualitative tests and one quantitative test.

The result from the expert test was that this project was a good idea as a tool to quickly create a foundation for the animation, that can later be tweaked and perfected. This sort of project is therefore good for animation companies that need to produce a lot of material in a short amount of time.

The first qualitative test showed that people saw various feelings, mostly positive, in the character's expressions, and based this opinion mainly on the facial expressions and the words used. However this test did not view the facial expressions separately, so although emotion was emitted from the character, this test did not prove that this was due to the character's facial expressions.

The execution of the second qualitative test was flawed in many ways and this affected the test results in a manner that rendered them unreliable.

The quantitative test showed that the emotions perceived by viewers in the original video, which the motion capture was based on, are still perceived in the final outcome. This means that not only is emotion perceived in the animation, but the emotion from the original video is preserved throughout the process and still visible in the final result.

Future Perspectives

When thinking of further development of this project, one important improvement would be to change the design of the headgear. As mentioned in the Discussion, the actor has limited mobility for his expressions, which could be fixed by decreasing the size of the headgear, so that it only extends forward from the head, with a small weight at the back of the head.

Because the results were very positive in this project, regardless of the low quality equipment, it would be interesting to see how much the quality of the end result would increase by using a slightly more expensive equipment. Perhaps the result would be so good that it could be used in small productions without much tweaking.

Bibliography

[1] D. Feild, Robert. *The Art of Walt Disney.* The Macmillan Company. New York, 1942.

[2] Patmore, Chris. *The Complete Animation Course: The Principles, Practice and Techniques of Successful Animation.* Barrons Educational Series Inc. New York, 2003. ISBN: 0-7641-2399-8.

[3] Rega, Elizabeth A. and Sumida, Stuart S. *Anatomical Considerations in Facial Motion Capture*. In *Computer Graphics: Building Bridges - Science, the Arts & Technology.* Volume 43, Number 2. ACM SIGGRAPH. 2009.

[4] Lasseter, John. *Principles of Traditional Animation Applied to 3D Computer Animation*. In *Proceedings of the 14th annual conference on Computer graphics and interactive techniques* (SIGGRAPH '87), Maureen C. Stone (Ed.). ACM, New York, NY, USA, pp. 35-44.

[5] Autodesk Inc. *Motion Capture - FaceRobot - Wiki*. Website: www.softimage.com. Visited: November 2nd, 2010. Location: http://facerobot.wiki.softimage.com/index.php/Motion_Capture.

[6] Autodesk Inc. *Motion Capture in FaceRobot - FaceRobot - Wiki*. Website: www.softimage.com. Visited: November 2nd, 2010. Location: http://facerobot.wiki.softimage.com/index.php/Motion_Capture_in_Face_Robot.

[7] PhaseSpace. *PhaseSpace Motion Capture | Products*. Website: www.phasespace.com. Visited: November 11th, 2010. Location: http://www.phasespace.com/productsMain.html.

[8] Adobe Systems Incorporated. *Adobe After Effects CS4 * Motion tracking overview and resources.* Website: www.adobe.com. Visited: November 2nd, 2010. Location: http://help.adobe.com/en_US/AfterEffects/9.0/WS3878526689cb91655866c1103906c6dea-7c5da.html.

[9] Autesk. *Getting Started With Maya Unlimited*. Website: www.autodesk.com. Visited: November 9th, 2010. Location: http://images.autodesk.com/adsk/files/gettingstartedmayaunlimited2009.pdf.

[10] Autodesk. *Autodesk MatchMover Online Help: About Autodesk MatchMover*. Website: autodesk.com. Visited: November 9th, 2010. Location: http://download.autodesk.com/us/maya/MayaMatchMoverOnlineHelp/index.html.

[11] Doherty, R. William. *THE EMOTIONAL CONTAGION SCALE: A MEASURE OF INDIVIDUAL DIFFERENCES*. In *Journal of Nonverbal Behavior* 21(2), Human Sciences Press, Inc, New York, NY, 1997.

[12] Lin, I-Chen and Ouhyoung, Ming. *Mirror MoCap: Automatic and efficient capture of dense 3D facial motion parameters from video.* In *The Visual Computer*. Kunii, Tosiyasu L. Springer, Berlin, 2005. pp. 355-372.

[13] Hatfield, Elaine; Cacioppo, John T.; Rapson, Richard L. *Emotional Contagion.* Cambridge University Press, Editions de la Maison des Sciences de l'Homme, Paris. 1994.

[14] Mori, Masahiro. *The Uncanny Valley.* In *Energy*, 7(4), pp. 33-35. Translated by Karl F. MacDorman and Takashi Minato. 1970.

[15] Gunnarsson, Arnar. Personal Interview performed by author. December 13th, 2010. Notes from the interview can be found in the Appendix.

[16] Di-O-Matic, Inc. *Automated lip sync animation in Autodesk Maya: any rig, any language! Voice-O-Matic (Maya Edition)*. Website: www.di-o-matic.com. Visited: September 1st, 2010. Location: http://www.di-o-matic.com/products/plugins/maya/VoiceOMatic/#page=overview.

[17] Di-O-Matic, Inc. *Animating The Garfield Show with Softimage and automatic lip synching.* Website: www.di-o-matic.com. Visited: September 1st, 2010. Location: http://www.di-o-matic.com/press/success/garfield.html

[18] Autodesk, Inc. *Autodesk - Autodesk Softimage*. Website: http://usa.autodesk.com. Visited: September 1st, 2010. Location: http://usa.autodesk.com/adsk/servlet/pc/index?siteID=123112&id=13571168.

[19] Gunnarsson, Arnar. Personal Interview performed by author. November 11th, 2010. Notes from the interview can be found in the Appendix.

[20] Image Metrics, Ltd. // *About Image Metrics | Image Metrics*. Website: www.image-metrics.com. Visited: October 20th, 2010. Location: http://www.image-metrics.com/company.

[21] Chenney, Stephen; Pingel, Mark; Iverson, Rob; Szymanski, Marcin. 2002. In *Proceedings of the 2nd international symposium on Non-photorealistic animation and rendering* (NPAR '02). ACM, New York, NY, USA, 2002. pp. 133-138. DOI=10.1145/508530.508553.

[22] Wang, Jue; Drucker, Steven M.; Agrawala, Maneesh; Cohen, Michael F. *The cartoon animation filter*. In *ACM SIGGRAPH 2006 Papers* (SIGGRAPH '06). ACM, New York, NY, USA, 2006. pp. 1169-1173. DOI=10.1145/1179352.1142010.

[23] Thomas, Frank. *Can Classic Disney Animation be Duplicated on the Computer?* In *Computer Pictures*. Back Stage Publications (1984) pp. 20-25.

[24] Ritchie, Kiaran; Callery, Jake; Biri, Karim. *The Art of Rigging - Vomume I: A definitive guide to character technical direction with Alias Maya.* CGTOOLKIT. 2007

[25] Disney/Pixar. *Pixar How We Do It*. Website: www.pixar.com. Visited: November 9th, 2010. Location: http://www.pixar.com/howwedoit/index.html.

[26] Breidt, M.; Wallraven, C.; Cunningham, D.W.; Bülthoff, H.H.; Campbell, N. *Facial Animation Based on 3D Scans and Motion Capture.* In *ACM SIGGRAPH Sketches Program*. 2003.

[27] Sagar, Mark. 2006. *Facial performance capture and expressive translation for King Kong.* In ACM SIGGRAPH 2006 Sketches (SIGGRAPH '06). ACM, New York, NY, USA, Article 26. DOI=10.1145/1179849.1179882.

[28] Mother Jones. *Obama: #1 Most Liberal Senator? | Mother Jones*. Website: motherjones.com. Visited: December 6th, 2010. Location: motherjones.com/mojo/2008/01/obama-1-most-liberal-senator

[29] Matzoball Online Entertainment LLC. Website: http://estergoldberg.typepad.com/. Visited: December 6th, 2010. Location: http://estergoldberg.typepad.com/.a/6a0105349ca980970c010535e1845f970c-popup.

[30] Mr. Blog'sTepid Ride. *President Obama Nominates New Supreme Court Judge.* Website: http://bmj2k.wordpress.com. Visited: December 6th, 2010. Location: http://bmj2k.wordpress.com/2010/05/11/president-obama-nominates-new-supreme-court-judge/.

[31] ProProfs. *Monsters, Inc*. Website: www.profprops.com. Visited: December 6th, 2010. Location: http://www.proprofs.com/quiz-school/story.php?title=monsters-inc_5.

[32] Telegraph Media Group Limited. *QI: Quite interesting facts about Kings - Telegraph.* Website: www.telegraph.co.uk. Visited: December 6th, 2010. Location: http://www.telegraph.co.uk/culture/qi/4176490/QI-Quite-interesting-facts-about-Kings.html.

[33] Autodesk Inc. Motion Capture Camera Placement - FaceRobot-Wiki. Website: http://facerobot.wiki.softimage.com. Visited: October 20th, 2010. Location: http://facerobot.wiki.softimage.com/index.php/Motion_Capture_Camera_Placement.

[34] Gunnarsson, Arnar. Personal Interview performed by author. September 1st, 2010. Notes from the interview can be found in the Appendix.

[35] MetaMotion. PhaseSpace Facial Motion Capture. Website: www.metamotion.com Visited: October 20th, 2010. Location: http://www.metamotion.com/PhaseSpace/facial-motion-capture.html.

[36] Autodesk Inc. Facial Mocap Marker Placement. Website: http://facerobot.wiki.softimage.com/ Visited: October 20th, 2010. Location: http://facerobot.wiki.softimage.com/index.php/Facial_Mocap_Marker_Placement.

Appendix A - Interviews

September 1st. 2010

Meeting with Arnar Gunnarsson

What is the status of the television series project?

Would I get any access to characters?

Can they perhaps meet me at least twice more before Christmas and once after Christmas to give me feedback and perhaps test my project?

The project is in a standstill, and its future has not been planned.

FaceRobot is a cool software that is great for doing motion capture. However this can not be done for cartoony looking characters or animals.

In the animation industry today, we need a solution like FaceRobot offers but also for cartoony characters.

Using EMG for motion capture can produce a realistic outcome, however the muscles in the face are very interrelated and therefore it might be hard to get a good result.

In cartoony animation there are no bones or muscles, think of Mickey Mouse.

I could get characters and real production material from them.

Voice-o-matic is a plugin in Softimage that creates lip syncing blendshapes based on a sound file. The lip sync in Garfield is made this way.

In animation, it is optimal to use as few controllers as possible, that are as flexible as possible. Basically what you want is few controllers that are capable of much control.

It is often a problem that there is a visual difference between scenes because many animators are working on the same movie. This could be improved by using fewer and more powerful controllers.

In Thor, the movie they are producing, there are over 100 morph targets.

Overlapping blendshapes is a problem in Maya but not in FaceRobot.

CAOZ are slowly switching from using Maya to using Softimage.

November 11th. 2010

Meeting with Arnar Gunnarsson at CAOZ

Explain to him my conclusion regarding the term cartoony.

Voice-o-matic

Is lip sync to ambitious?

Using facial motion capture with 1 or more cameras + voice-o-matic?

They will be working on Egils Saga in 2011, for a Hungarian company. The movie was shot using motion capture. The movie will be very arty and brutal. CAOZ will then animate the face and hands, by kind of rotoscoping.

They are thinking of doing the animation in layers. First create a lip sync layer (perhaps with a program such as voice-o-matic) and then create a separate layer with emotions. That way it is easy for the director to ask for different emotion because the animator doesn't need to do as many alterations.

He has not tried working with voice-o-matic. Using a software such as voice-o-matic on top of motion capture would be rather pointless because you get the lip sync with the motion capture. Then you would just create a rig in beforehand and then apply the tracked data to it.

In Digital Domain - the company who did the facial animation for Benjamin Buttons, they used some substance that covered the skin and captured Brad Pitt's movements.

Meta Motion and Phase Space are leading companies in facial animation.

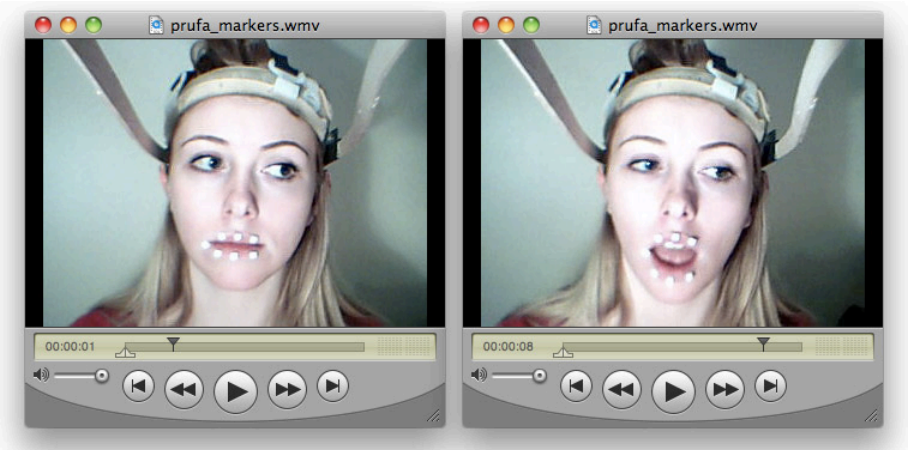
Lip sync is the easy part, the hard part is emitting emotion.

With facial motion capture you could use the same capture for different faces.

December 13th. 2010

Expert Testing - Arnar Gunnarsson

Lip sync: The sound is 6-10 frames ahead of the animation.



The eye movements: You can sometimes see that the eyes were animated afterwards. Would be better if I added some drift to the motion and saccades, otherwise it is good.

“You can see that the character is thinking and searching for words.”

The reason for the eye movements is that “skinkluster” needs to go above “blendshape” in the input list for trollBody.

“Looks remarkably good, however it is very wobbly.”

In further development, blendshapes could be created that stabilized the middle of the face, therefore making it easier on the eyes that the character is so wobbly.

The character is fleshy - the face mesh behaves like human skin, this is because of the motion capture, much harder to create this feeling using blendshapes.

“You can go a long way with this.”

This would be worth considering if it were further developed (to reduce the wobble) to use in television series.

Whether it is “good enough” depends on the demands set for it. This could obviously not be used in a high budged movie, but if more stable, maybe some smaller production.

This could well serve as a guide for animators, for them to animated on top on. Then they would only need to “paint by numbers”, that is, fill in the gaps with their animation. There would be no need for them to make guesses, because the timing and major poses would already be in place.

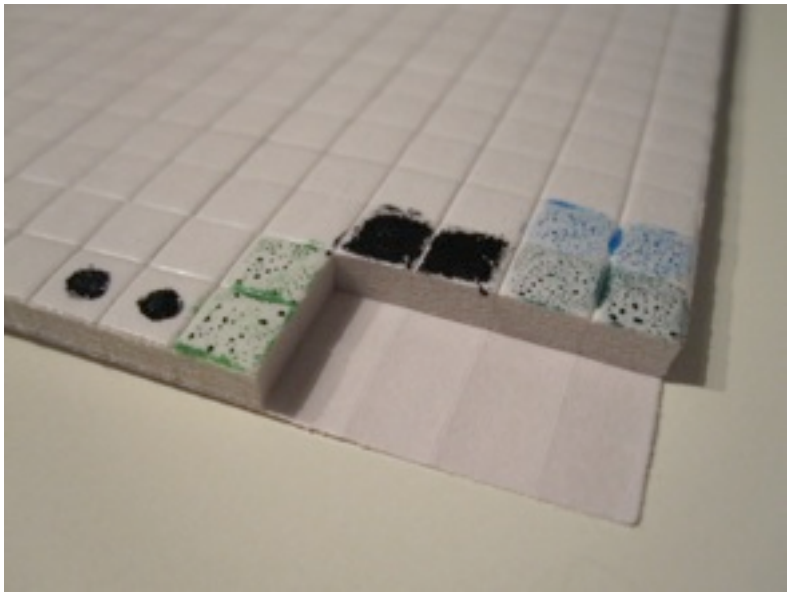
The most important parts of the animation is the mouth and the eyes because the viewer darts his eyes between those two places.

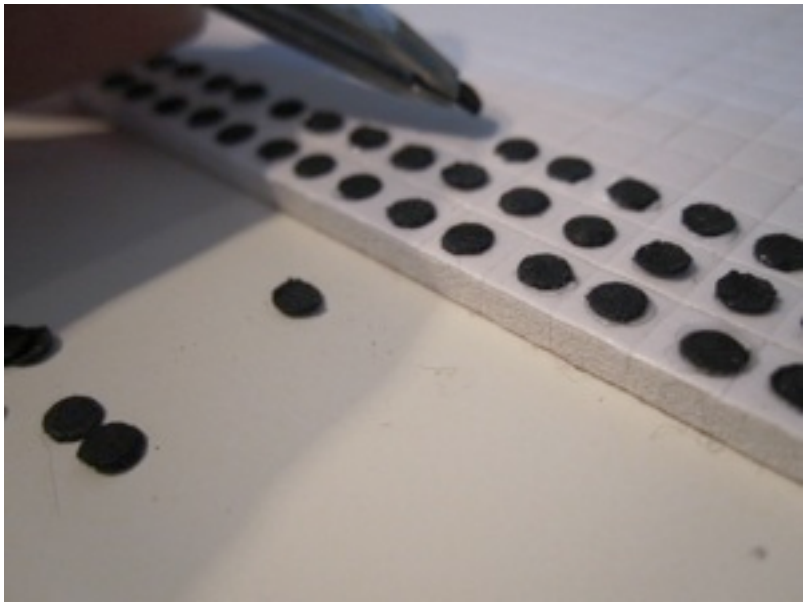
“You can get away with this.” (the animation)

This solution is particularly beneficial when a lot of material needs to be produced in small amount of time.

The higher the quality of the outcome (how stable it is and well captured) the more useable it becomes, as it will need less tweaking or “paint by numbers” for the animators.

The tongue could be triggered with a sound file, for example with voice-o-matic.



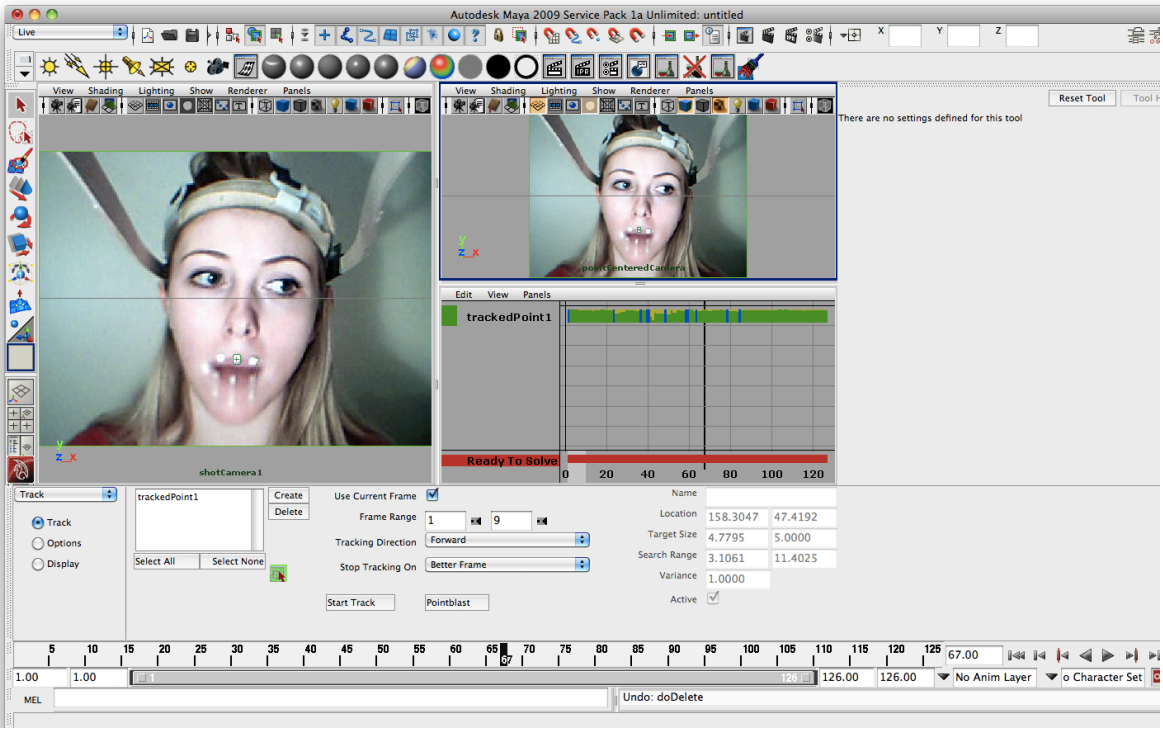


Appendix B - Tracking Tests

First tracking test

Goals
- To test the camera and see if the output is clear
- Test the headgear, see if it is functional and stable
- Test the markers, see if they are appropriately sized
- See if Maya Live can track the markers

Notes
- I am too white! The markers blend with my skin colour in extreme movements.
- Lighting is important. When the light shines directly on my face, it turns white.
- The size of the markers seems fine and Maya Live can track them, however the tracking needs adjustments both in fast movements and when my skin is so white that it blends with them.
- Headgear worked fine, didn't block my forehead and wasn't even heavy.
- Next time I should be more tanned, one way or another.
- Try green markers next time, to see if it balances out my "whiteness".
- One of the markers fell off and I didn't notice. Next time, be sure to attach them so they stick and pay attention if they fall off!



Second Tracking Test

Goals
- To test different colours of markers
- Last time I just used Windows Movie Maker, now I want to try Creative WebCam Center

Notes

- There is a bug in Maya Live! If I close a scene I will not be able to open it again in tracking mode. Must be careful to finish all tracking in one session.
- The different colours were not good, the best one was white background with a black dot. (Because Maya crashed I don't have any screenshots)
- Creative WebCam Center has a bit better visual quality but it can not record audio.

Appendix C - Testing Material

The Emotional Contagion Scale

This is a scale that measures a variety of feelings and behaviors in various situations. There are no right or wrong answers, so try very hard to be completely honest in your answers. Results are completely confidential. Read each question and indicate the answer which best applies to you. Please answer each question very carefully. Thank you.

Use the following key:
4. Always= Always true for me.
3. Often = Often true for me.
2. Rarely = Rarely true for me.
1. Never = Never true for me.

1. If someone I'm talking with begins to cry, I get teary-eyed.
2. Being with a happy person picks me up when I'm feeling down.
3. When someone smiles warmly at me, I smile back and feel warm inside.
4. I get filled with sorrow when people talk about the death of their loved ones.
5. I clench my jaws and my shoulders get tight when I see the angry faces on the news.
6. When I look into the eyes of the one I love, my mind is filled with thoughts of romance.
7. It irritates me to be around angry people.
8. Watching the fearful faces of victims on the news makes me try to imagine how they might be feeling.
9. I melt when the one I love holds me close.
10. I tense when overhearing an angry quarrel.
11. Being around happy people fills my mind with happy thoughts.
12. I sense my body responding when the one I love touches me.
13. I notice myself getting tense when I'm around people who are stressed out.
14. I cry at sad movies.
15. Listening to the shrill screams of a terrified child in a dentist's waiting room makes me feel nervous.

Note: The higher the score, the more susceptible to emotional contagion a person would be said to be. Happiness items = 2, 3, & 11. Love items = 6, 9, & 12. Fear items = 8, 13, & 15. Anger items = 5, 7, & 10. Sadness items = 1, 4, & 14. Total score = all items.

Source: Doherty, R. W. (1997). The Emotional contagion scale: A measure of individual differences. Journal of Nonverbal Behavior, 21, pp. 131-154.

Concent Form

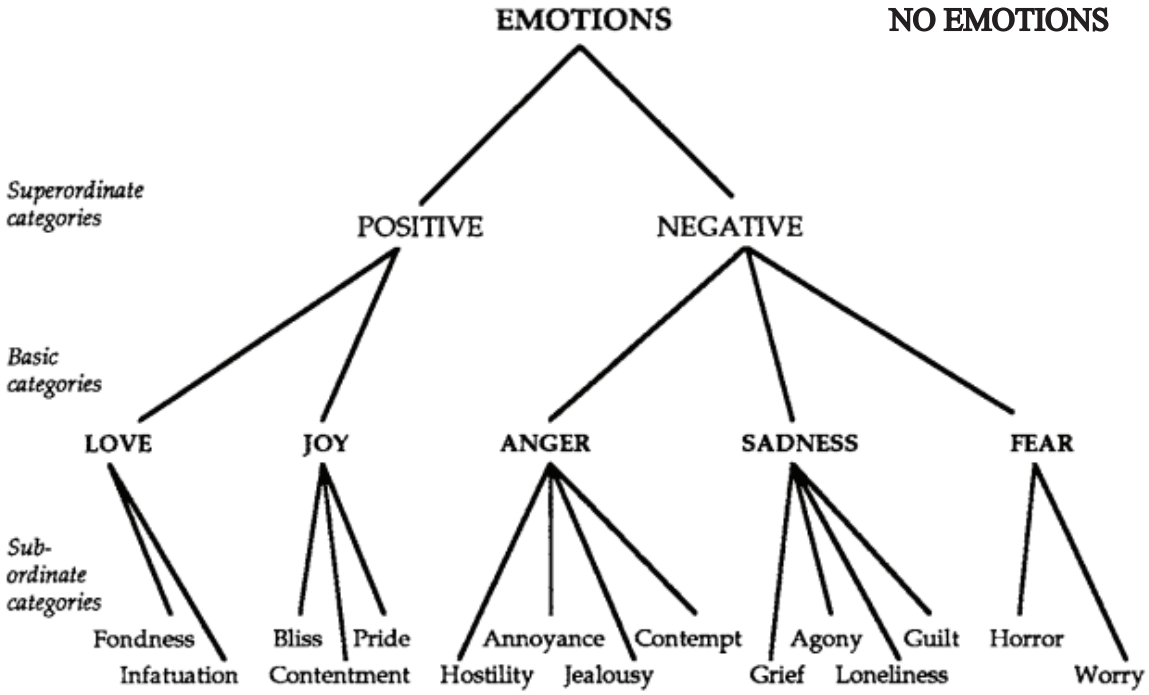
I hereby give the permission to use information acquired today with my help, whether it is written or captured on video, in the production of this project. I am aware that this use of information may entail public distribution of a report and/or a production video.

Please note that the results are strictly confidential.

Name:	Signature:
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____

Age: _____ Gender: ☐ M ☐ F

Please draw a circle around the emotion or emotions you think the **character** in the video experienced.



Please write a number next to the circle(s), on the scale from 1-5,
where 5 means that the character had an intense experience and 1 means weak.

What do you base your opinion on?

- ☐ His words
- ☐ His facial expressions
- ☐ His tone of voice
- ☐ Other, what? _____

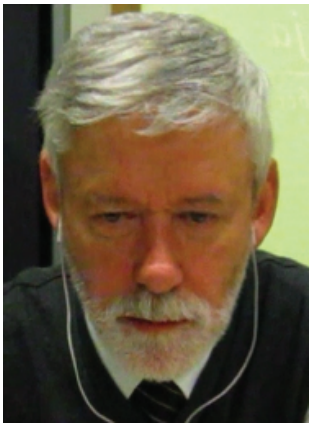
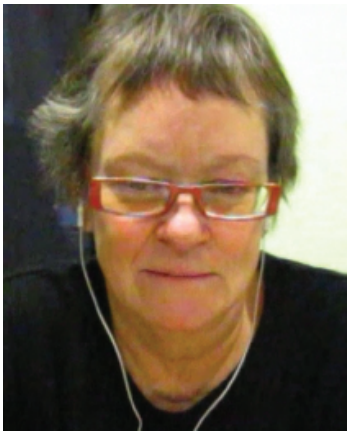
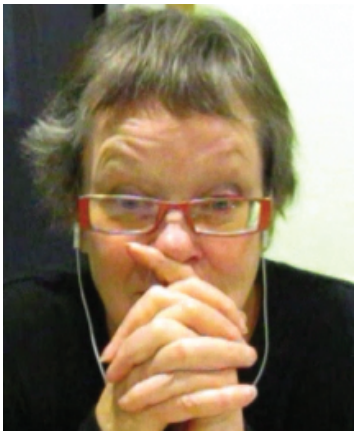
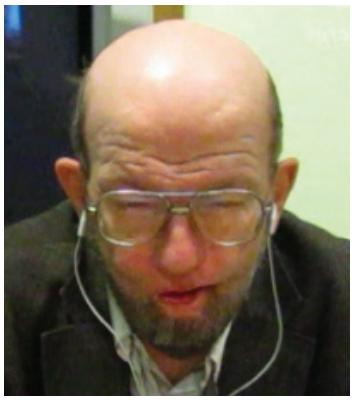
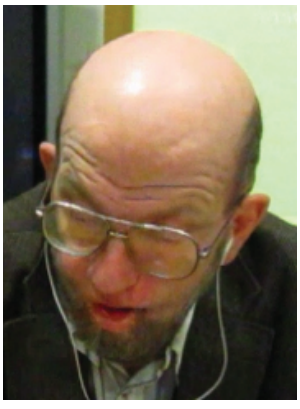
Do you have any comments?

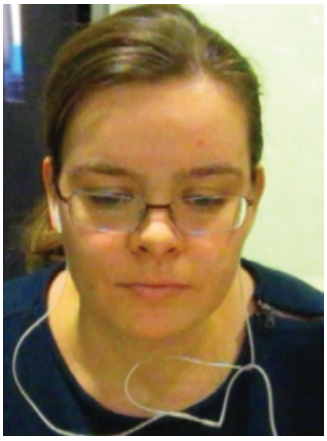
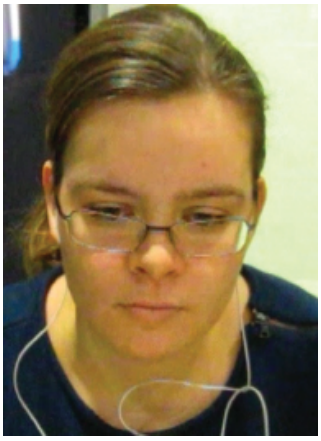
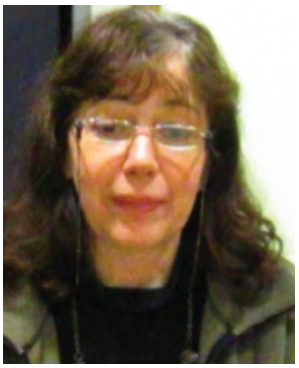
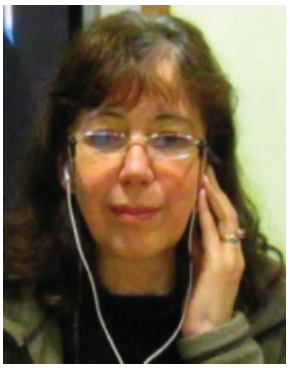
_____

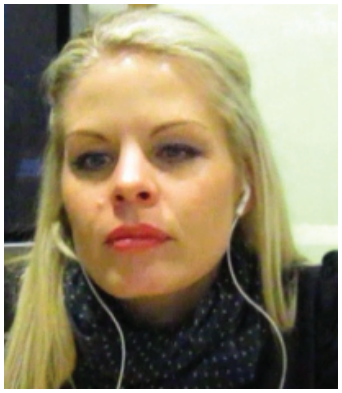
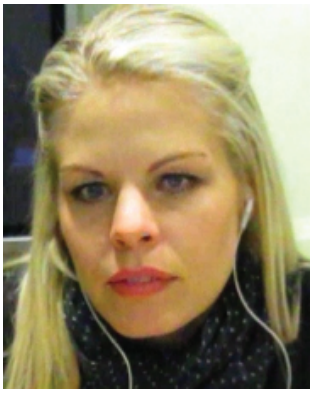
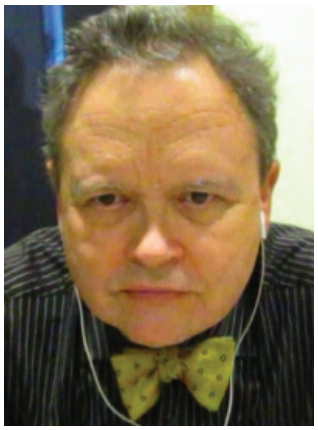
_____

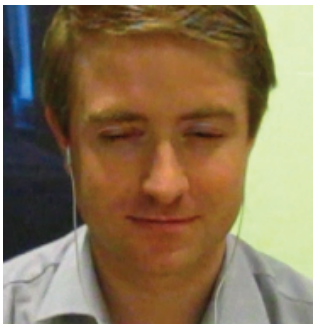
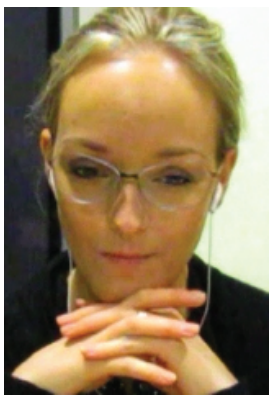
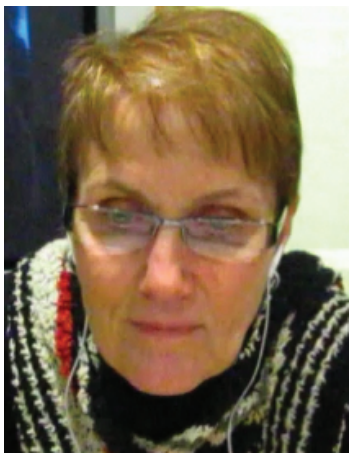
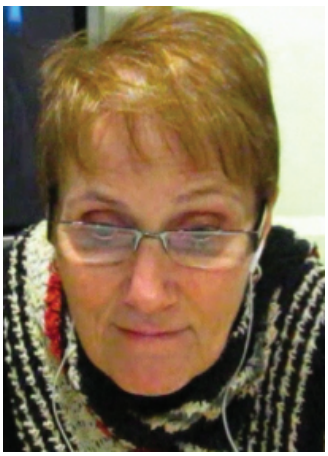
_____

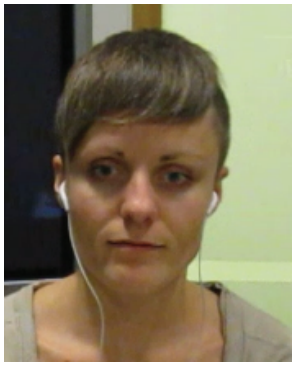
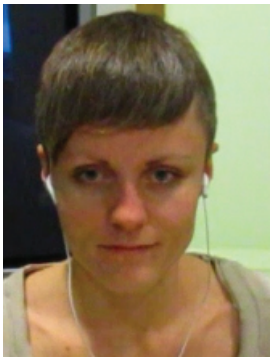
_____

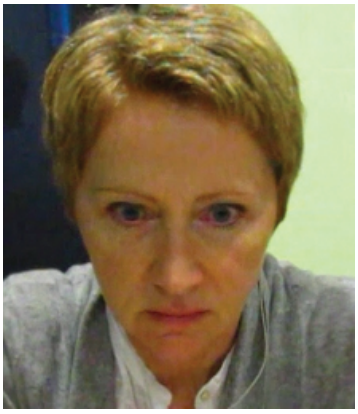
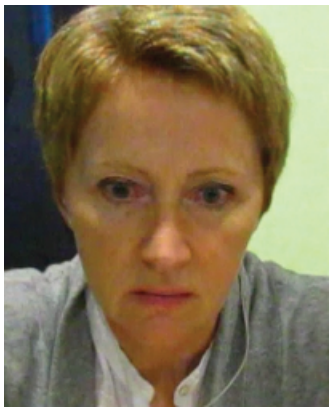
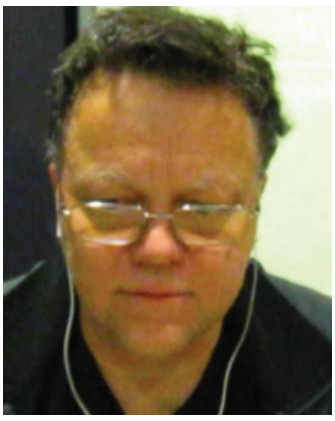
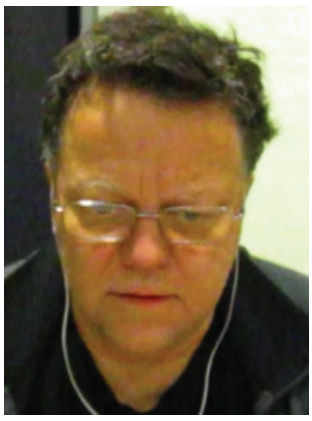
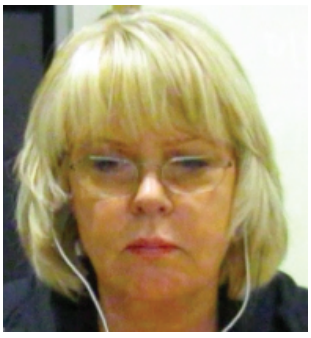
Watching the text at the start.	Watching the character.	Comments on expressions:
		Showed little to no changes in his expressions. ECS score: 3,4
		Smiled very vaguely towards the end of the video. ECS score: 2,2
		Raised her eyebrows and smiled. ECS score: 3,2
		Showed no changes. Did not watch the video, only listen. ECS score: 3,1

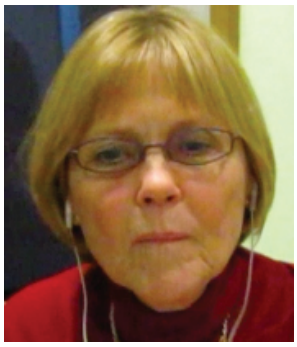
Watching the text at the start.	Watching the character.	Comments on expressions:
		Smiled constantly and raised eyebrows a couple of times. ECS score: 2,8
		Showed no changes in facial espression. ECS score: 3,0
		Vague smile throughout the video. ECS score: 2,9
		Giggled and smiled. ECS score: 3,2

Watching the text at the start.	Watching the character.	Comments on expressions:
		Smiled and tilted his head. ECS score: 2,4
		Smiled very vaguely. ECS score: 2,8
		Showed no changes other than looking a bit more focused. ECS score: 3,1
		Leaned forward and looked focused. ECS score: 2,7

Watching the text at the start.	Watching the character.	Comments on expressions:
		Leaned forward andpeered his eyes. ECS score: 2,5
		Smiled. ECS score: 3,1
		Smiled. ECS score: 3,5
		Raised her eyebrows and smiled vaguely. ECS score: 2,9

Watching the text at the start.	Watching the character.	Comments on expressions:
		Leaned forward and smiled vaguely. ECS score: 2,8
		Leaned forward and raised his eyebrows. ECS score: 2,9
		Tilted face downwards. ECS score: 2,8
		Smiled vaguely. ECS score: 2,7

Watching the text at the start.	Watching the character.	Comments on expressions:
		Lowered eyebrows, looked focused. ECS score: 2,7
		Tilted head back, otherwise no changes. ECS score: 3,3
		Opened his mouth slighly, looked focused. ECS score: 2,8
		Peered her eyes, looked focused. ECS score: 2,6

Watching the text at the start.	Watching the character.	Comments on expressions:
		Leaned forward and smiled vaguely. ECS score: 2,9
		Smiled. ECS score: 2,9