LUNDS UNIVERSITET

INSTITUTIONEN FÖR PSYKOLOGI

Still making faces, now also a dynamic one: An updated study of parameterized three-dimensional models of emotional facial expressions

Patrik Bratkovic

Supervisor: Phd. Åse Innes-Ker

PSYM99 essay autumn term 2008

Abstract

With constant upgrades in 3d-modeling software, even previously functional and validated animated facial expressions can be updated. The present work seeks to provide such an additional upgraded set of tools for researchers to use interested in e.g. emotional facial expressions, and contains two experiments. One where static facial expressions are tested in three different ways and a second experiment which seeks to investigate whether there is a difference between 5 groups of smiles that develop differently. In the first experiment the creation of the stimuli is described; participants evaluated the created animated expressions and photographs of human actors on: multiple Likert-scales for intensity, a task where they freely were to indicate what they had seen, and a multidimensional scaling approach where all expressions were compared. Results indicate that the expressions were qualitatively similar. In the second experiment, participants rated the genuineness of 5 groups of differently developing smiles; either polynomial or linear. The results indicate that one group was rated significantly worse than the other. All results are discussed in the context, and future research propositions are given.

Keywords: animated, dynamic, emotion, faces, facial expressions, Poser, stimuli

Acknowledgments:

This paper would not have been made without the aid and much valuable feedback from my supervisor, Phd. Åse Innes-Ker. Special thanks to additional feedback concerning the facial expressions and action unit creation from Phd. Fredrik Björklund, and Sofia Söderin, a former international security agent (Long Island MacArthur Airport - ISP). Further, I would also like to extend special thanks to Leonard Loos, Jakob Persson, and Anders Nilsson whose support and conversations I could not have been without.

Table of Contents

Still making faces, now also a dynamic one: An updated study of parameterized three-dimensional models of emotional facial expression

When one tries to make oneself understood in a remote airport, the words might not suffice in conveying the message, but perhaps a facial expression might? Where a potential language barrier stops progress, a sad face might let the check-in attendant know that you are feeling low, perhaps because of a missed plane, a lost gateway, or if no tickets are at hand, that you are a forgetful person. A smile in that very situation might signal progress in understanding, or perhaps an attraction. The interpretation of its intent however, would be much more difficult to assess. One aspect could be whether it is considered as genuine, since there seems to be a limit to how quickly it can begin and end (Krumhuber and Kappas, 2005; Ekman and Friesen, 1982; Hess and Kleck, 1990). Another aspect is that the smile might give away that its elicitor is understanding of a situation that might have occurred (Spencer-Smith et al., 2001), or be interpreted as a social smile. And the duration of the smile, or how the smile develops, might also have an impact on how the smile is perceived; whether it is considered a genuine or not might depend on e.g. how short or long the smile is. On the other side of the emotion spectrum an angry expression and the possible aftermath of its display might lead to airport security being called to the site. Facial expressions thus allow humans a mean to communicate. Our species get a chance to test what these facial expressions mean, how they are perceived, whether we are being deceived and much more.

Most of the research on facial expressions has been made with help of human actors; the stimuli are most often still images of facial expressions. Their main strength is indeed their humanity, but even though human actors posing for such expressions are undoubtedly invaluable, they present some difficulties. Firstly, they are idiosyncratic; while the recognition of some facial expressions is argued to be universal, it could be interesting to modify an actor for testing purposes, which is not easily achieved with humans. Secondly, dynamic facial expressions could prove to be difficult to capture – and perhaps even worse to modify. Animated expressions thus lack in the harshest area, by not being human actors; what can be gained is however large enough to create new animated stimuli.

The present paper seeks to create and validate an additional set of tools for researchers who are interested in e.g. facial expressions and emotion – and all possible branches of sciences where they would be of interest. In a second experiment it intends to utilize the knowledge acquired from creating the animated facial expressions to add dynamic aspects to a smile, and test whether different trajectories are equally good. While the presently available tools developed in Poser 4 (Curious Labs, Santa Cruz, CA) are sufficient for experimental purposes, and have successfully been utilized (Spencer-Smith et al., 2001; Krumhuber and Kappas, 2005; Wieser, 2007), newer Poser editions are improving the possibilities of including e.g. more realistic detail into expressions, and Poser 7 (Smith Micro, Aliso Viejo, CA) will be used to create the stimuli. Morphology could be created or changed at demand, and when complete the researcher could alter the skin-colors as they please, or have realistic textures applied to simulate skin, scarring etc. It would also bridge difficulties with limited availability of the older software, used in creating the older set, as well as proper support for it. It also seeks to create a set of dynamic smiles in their onset-phases to better understand whether there is a difference between the ways smiles develop.

Making and recognizing faces are we?

The universality discussion. Facial expressions are, in most cases, welcome tools for humans to dispose of in social situations. Ranging from the possibility to greet someone with an expression of happiness, a smile, to angry expressions where disapproval is communicated and signals potential danger. Recognizing the potential expressions held for aiding in the survival of species, Darwin argued that expressions had an evolutionary basis; that facial expressions were biologically coded (Darwin, 1872, in Ekman 1973). This theory has been adopted and developed, most notably, by Ekman’s research on the theory of universality regarding facial expressions. Potentially lending support to this theory are results indicating the possibility of expression recognition occurs by infants, and that some expressions are preferred by them (Farroni, Menon, Rigato and Johnson, 2007). It should however be noted that they seemingly recognized, or rather found interest in, happiness in contrast to neutral and anger expressions; the recognition might have been influenced by the surroundings which were, presumably and hopefully, joyous. Yet, this would indicate that they are able to discriminate between expressions already at this early age. Speculatively, the universality would not have to include only humans, as the anger expression often comes with bared teeth in other species such as canines, felines, and monkeys as well, and can serve as a mean to warn the potential attacker, or as a way of intimidation. Further, being an expression that can signal a warning of potential danger and perhaps impending death, anger is quicker recognized and e.g. isolated in a crowd (Hansen and Hansen 1998; Öhman, Lundqvist, and Esteves, 2001; Juth, Lundqvist, Karlsson, Öhman, 2005), than a happy expression; most likely being a very welcome feature for airport security. There are however indications that we are drawn to threatening faces, which could account for the results; it remains a bit paradoxical due, and is still under investigation. On another part of the spectrum, a smile might signal interest, enjoyment, perhaps an opportunity to spread one's genes, in contrast to anger's threat based communication (Moskowitz, 2005; Fecica and Stolz, 2008). Or it could serve as a simple mean to make the day more pleasant at the mentioned airport.

Further, research shows strong indications that there are at least 6 'basic expressions', where happiness is included, that could be recognized universally (Ekman and Friesen, 1969; Ekman, 1972; Moskowitz, 2005; Ward, 2006). Recently, contempt has garnered interest as a potential universally recognized expression, and following the interest there are indications that contempt should be included (Ekman and Friesen, 1986; Ekman and Heider, 1988; Matsumoto, 1992; Matsumoto and Ekman, 2004). Disagreements on whether the recognition is universal do exist (see e.g. Russell, 1994), however in a sense it could be argued that they are not necessarily contending with the universality theory. This does not mean that theories on cultural, or rather social, construction should be disregarded. Instead, it could be argued that the theory of universality provides perspective to such theories and allows for plausible rationalizations to the interpretation of facial expressions as socially constructed. Display rules could serve as an example, which in short corresponds to the adjustment of expression exposure allowed according to a cultural norm (Ekman, 1972; Ekman and Friesen, 1975). Display rules could be used to mask an expression or evoke an expected expression which in turn could perhaps be interpreted as not corresponding to the theory of universality. Yet, even though display rules differ amongst cultures, the recognition in universality studies is continuingly above chance, i.e. that the display of an expression is willingly hampered in favor of a norm while the recognition remains; studies on athletes, from various countries and cultures, spontaneous facial expressions also indicate that the recognition could be considered as universal (Matsumoto and Willingham, 2006).

There are however other aspects that could perhaps be argued as less compatible with universality, such as the theory of an in-group advantage; ratings of facial expressions seem to be more accurate for members of the same cultural groups in tests made by several observers (Elfenbein, Mandal, Ambady, and Harizuka, 2002; Elfenbein and

Ambady, 2003). It should however be noted that they cannot always be contending, as an experiment using a set of facial expressions with actors similar to the group, or from it, would not exclude e.g. in-group theory simply because it would not test for differences between groups; it would not be applicable. Further, some methodological criticism has been raised against contributions to the in-group theory, more specifically against the hypothesis itself in the above mentioned study for disregarding other possibilities, and the results for the stimuli used, as well as a more recent study for only using left hemifacial composites (Matsumoto, in Hess and Phillipot, 2007). Results of the latter stem from using the left side of a face mirrored to the right side, creating the appearance of a symmetric expression. However, an experiment with such mirrored expressions would be overseeing the possibility of e.g. how, hemifacially due to asymmetry, facial expressions differ when made spontaneously or voluntarily (Ekman and Friesen, 1982; Hager and Ekman, 1985), which the critique brought forth (Matsumoto, in Hess and Philippot, 2007).

The discussion on universality has been great for methodological advances. Some of the research that would strengthen the theory of the universality of facial expressions has also been subject to critique; more accurately, the methodological approach has been the critiqued, and it was suggested that the levels of chance for e.g. free choice tasks would be adjusted for the probabilities of judgment depending on the valence of the expression (Russell, 1994). And that alterations should be made to forced choice formats much often used for the gathering of data, which commonly allowed for single responses only (Niedenthal, Krauth-Gruber and Ric, 2005). This format could indeed be problematic as it could arguably lock the participants in their responses to the list that is being used to gather the response, especially if the participants are only allowed to pick one expression. Revised experiments have however been made to adjust to Russell’s critique, and these have produced continuingly higher than chance results, by allowing the participants to rate e.g. the intensities of all expressions per expression. Critique concerning it still being a forced response method has been amended favorably as well, with e.g. allowing the participants a free choices response format which also tests for higher chance levels (Ekman and Rosenberg, 1995). Thereby, while the arguments in favor of and against the universality theory are immensely interesting, they have set up increased demands on the methodology used to test the recognition of facial expressions and various aspects of the facial expressions, which should provide beneficial resources for testing of emotional facial expressions that are not captured in photographs as well; this should allow the testing of e.g. malleable animated faces as well.

Another applicable method, which has successfully been used in e.g. marketing but emotion research as well (see Green, Carmone and Smith, 1989; Halberstadt and Niedenthal, 1997; Spencer-Smith et. al., 2001), is the Multidimensional Scaling (MDS) approach. It is a set of statistical technique which measure similarities or dissimilarities in stimuli. Likert-scales can be used to assess a comparison between two stimuli, through simply asking the participant to rate the perceived similarity (or dissimilarity). The resulting data is then inserted into a matrix, which in turn is submitted to processing through a computer program. The processed data can be set up on a diagram with two-dimensional axes, and corresponds to the stimuli and the relationships between the stimuli in terms of distances; a closer distance (depending on the type of multidimensional scaling) will indicate similarity, and larger distances between stimuli will indicate dissimilarity. To further explain how the MDS functions one could imagine the distances between cities. The different distances between all assessed pairs are then compared to each other and should, if executed correctly, produce a map in which the position of the cities correspond to their positions on a 'real' map. It could thereby be argued that the multidimensional scaling produces cognitive maps for the perceived relationships between stimuli.

The dynamic aspects of facial expressions. Returning to the airport and the smile previously mentioned it would be quite safe to assume that the conveyed expression would be dynamic – change in intensity over time. It might however be difficult to gain a further understanding of such smiles since emotional facial expressions are most often tested as static images. Thereby, and in vast contrast to its static counterpart which a simple keyword search in a scientific database can reveal, there is comparatively little research on dynamic emotional facial expressions. Concerning that there are cases where static expressions are not recognized but dynamic are (Collignon et al., 2008), that such expressions enhance the effect of the emotional experience (Sato and Yoshikawa, 2007), or that they might lead to increased mimicry (Sato, Fujimura, and Suzuki, 2008), and indications of enhanced neural activity for brain-damaged patients even when taking the dynamicity itself into account (Sato, Kochiyama, Yoshikawa, Naito, and Matsumura, 2004), there has been some recent interest in these dynamic aspects. It should however still be noted that while the check-in attendant would likely recognize it as a smile, the research is still so scarce it would be more difficult to explain the characteristics of the smile due to its dynamic nature. The naturalness of the smile as dynamic has been tested (Sato and Yoshikawa, 2004), or rather how natural a smile is perceived based on onset, with onset being the development from a neutral intensity to an intended intensity of an expression. Depending on interpretation of natural however, with the baseline being something occurring in nature, the outcome of the experiment would be too general to interpret as just occurring in nature, or worse, as probable to occur in nature. The results could then exaggeratedly be attributed to the explosiveness in the jaw region for the model of the smile, or the dexterity of that very model – being able to voluntarily move muscles very slowly. Considering that natural naturally incorporates both fake and genuine smiles, it could indeed be problematic. Instead, Ekman and Friesen earlier proposed there should be a distinction between acted and spontaneously made smiles (1982). Backtracking to the previous paragraph, without such a distinction the overseeing part of the in-group theory experiments might have never been detected.

If glancing back at the naturalness of a dynamic smile, one could instead make use of the duration of the onset/s to investigate e.g. the interpreted genuineness of a smile. The smile seems to be constrained by time; smiles are, seemingly, sensitive to the duration of its development, and there are indications that smiles are judged as more genuine depending on how long they develop (Krumhuber and Kappas, 2005; Ekman and Friesen, 1982). The onset of the smiles in Krumhuber and Kappas experiment were tested with a linear onset, up to slightly above half a second, with the longer onsets of smiles seen as more genuine than the shorter. Another study explored the perception of dynamic expressions, including smiles, and found similar results to the aforementioned study regarding onset (mentioned as velocities), and the effect different onsets had on how realistic the participants found them (Hoffman, Traue, Bachmayr, and Kessler, 2006). The latter study found an optimal velocity of circa 830 ms for happiness, whereas the first study only went as far as circa 530, a velocity matched by another study as well (Schmidt and Cohn, 2001). The half-second velocity was also rated as significantly more genuine as an onset of circa 130 ms, and the study also found that the largest difference in genuineness was between circa 130 and 230 ms (Krumhuber and Kappas, 2005); longer onset durations generally seemed to be perceived as more genuine. There are also indications that expressions could have multiple onsets, and thereby apexes and offsets as well and thus develop in phases (Hess and Kleck, 1990). Such phases seem to rise in frequency in deliberate expressions and could be a source for the added time that comes with a non-spontaneous expression. An intriguing question that can be posed is whether still differently developing smiles could also be seen as genuine, as smiles could potentially develop in other ways, e.g. beginning slowly and then, in a quicker manner, reach their apex – highest point. This would translate to a twofold onset without actually pausing at the apex. Other smiles could potentially begin rather quickly, and then slowly reach the apex. Interestingly, especially in relation to the aforementioned study, there are some indications of lip corner movement that occurs in a polynomial manner (Schmidt and Cohn, 2001). Of course, polynomial does not necessarily mean anything else than a straight line which 'moves' either higher or lower. The timing should then not affect smiles in these ranges, and whether problems are to be found with such smiles, it would rather be due to problematic aspects of the smile itself.

Hypotheses

This paper primarily seeks to offer an updated version of the Poser faces first constructed in Spencer-Smith et als., *Making faces: Creating three-dimensional parameterized models of facial expression* (2001). The advantage would be to have controllable up to date faces, with improved textures and possibilities to manipulate creasing skin in further detail than before. The newly created expressions would also be easily compatible with recent editions of Poser, ranging from Poser 6, Poser 7 and Poser Pro .

Thereby, in its first experiment the paper seeks to validate the new animated expressions through different methods in comparison to photographs of human actors making the same expressions; through forced choice, free choice, and multidimensional scaling the paper intends to investigate whether the newly created animated facial expressions could suffice against non-animated images. For the first stage of validation, intensities, we posit that the forced-choice multiple Likert-scales will yield an equal or better result for animated images than for photography's. For the second stage of validation, free-choice, we posited that the animated expressions would be rated as better than chance, except for contempt which would correspond to earlier research (Ekman and Rosenberg, 1995). For the third and last stage of validation, we posited that the multidimensional scaling approach will yield distances relatively close to each other. Thus, all in all we posit that the animated poser images will suffice or generally be better when compared against its photography counterparts. In addition, and using the same methods bar the free-choice, this iteration of the poser faces further seeks to add the facial expression for 'contempt' to the previously sevenfold family of animated facial expressions.

In the second experiment, involving dynamic smiles created with aid of the happy model in the first experiment, we posited that there would be no significant difference in ratings between the durations of dynamic smiles in the 'onset' phase due to their range which included ~260 ms to ~560 ms – in accordance to Krumbacher and Kappas (2006) results, and extending that range to circa 860 ms; results found by Hoffman et al. (2006) to indicate optimal velocity for happiness. It also seeks to investigate whether dynamic smiles with a very low amount of phases – i.e. that don't develop linearly - are equally good as linear smiles if they occur within the aforementioned durations, potentially allowing future investigations of such smiles. We thus posited that the different trajectories of smiles beginning slowly or quickly are equally good, and seek to disprove it by allowing participants to rate all available onsets in wholly randomized presentations.

Experiment 1: making and validating animated facial expressions

Introduction to the experiment

The creation of the animated facial expressions in Poser will be treated in the beginning of this section. It then continues to the validation of these facial expressions. The first and second part of the validation were conducted simultaneously and will thereby be reported together. The first part of the validation investigates whether there would be a difference between animated facial expressions and photographs of containing the same expression. Eight female and 8 male expressions created in Poser 7 were tested with 8 female and 8 male photographs of the same expressions. These were rated using multiple Likert-scales where participants had the option to rate the intensity of perceived emotion/s in the shown images; a forced choice experiment with the option to choose more than one expression. This method has been extensively used, and even though it locks the participant to the responses, they are at liberty to choose how many, or how few, expressions they want.

The second part of the validation used the same stimuli as the first, but investigated whether there would be a difference in judgment when participants were allowed to freely note what emotion they perceived the person expressed/felt in the images; a free choice experiment. To avoid experimenter bias, a smaller task where judges categorized synonyms acquired in the free choice task was also carried out. The judges' categorizations were then used as containers to assess the frequency of the participants' responses. This method is an evolution of the forced choice experiments described above, but requires extra steps in getting to know what the participants see in the images.

This third and final part of the validation examines the assessed similarities between the animated expressions and the photographs using a multidimensional scaling approach. Participants were given the task to make pair-wise comparisons of all possible pairs of expressions, separated by gender, to be able to explore similarities between the sets of stimuli. This part of the validation uses a multiple comparison method, and seeks to illuminate the relationships between photographs and animated images; it serves as a mean to understand whether the stimuli are rated as similar using data from all available stimuli.

Making the faces

Smith Micro's Poser 7 was used to create the animated facial expressions; reason being that there had been previous success in using the program for creating Action Units and from them facial expressions. Poser is software that allows creation or manipulation of 3D models and the creation of both stimuli such as still images and their dynamic counterparts - movies. The face area in the software enables precise alterations, and numerous possibilities to model the 3D model in a desired manner, either by pre-programmed parameter dials, or by freehand alterations.

In creating the anger, disgust, happiness, sadness, and surprise stimuli excerpts from Ekman and Friesen's Facial Action Coding System (FACS) was used as a reference (1976), as well as previously created animated faces which also had used FACS as reference (Spencer et al., 2001). This allowed the stimuli to be created from the muscle-movements, or simply 'action units' as Ekman and Friesen call them, which, when activated independently, form the skin in different ways (ibid). These units might be independent, but they can be seen as parametric modules, especially when it involves the making of facial expressions: some will fit with others without a hitch, some need to be adjusted before fitting, and some could be argued to be incompatible. The FACS can be used as a key to identify, or decode, expressions into these independent AU's, or vice versa code AU's to form facial expressions. Contempt was modeled after Matsumoto's example of contempt (2005), and for the neutral expression the default Poser expression was used.

A total of 16 AU's were used in the paper, the same amount as the study by Spencer et al. (2000), however a total of 25 AUs per gender were created. Contempt was created partially using the AU's, but mostly through freehand modeling. The AUs, and from them the expressions, were created in Poser 7. Poser is, as previously mentioned, a 3D software centered around the creation/modification of realistic figures. The figures consist of a large amount of polygons, with a texture laid upon it. To ease the understanding, imagine that you have the measles, but that every dot on your body is a vertex, which can be moved around. These movable vertices are interconnected, see Appendix A for an example (p. 40), and have a layer of 'skin' stretched across them and it has been fastened in each and every one of the vertices. The malleability that comes with such a model, as well as the possibility to replace the skin – textures – can be very helpful in the creation of stimuli of very different kinds. A morphological difference which could require, if photographs are used, a search for a fitting and willing stimuli has the possibility to instead be created by the researchers. Further, camera-angles can be positioned arbitrarily, the focus of the camera can be altered, and the possibilities to add/remove lightning sources and alter their direction are numerous. Returning to the vertices, the head on the Poser figures contains a very large number of them; the large number of vertices in the head area allows detailed alterations in e.g. the face. Alterations can be made by freehand, in other words moving the measles around, or by the use of pre-selected parametric groups. Such groups can include e.g. blinking and opening the mouth. These groups can move large amounts of vertices at once, like the open mouth, or small amounts like the blink of an eye, and can be seen as a basis for creating expressions for normal users. Just like the measles, they do not have to be uniform in magnitude. A 1.0 open mouth will be more open than the moderately open 0.5, or less open than the skin-tearing 2.0 open mouth. Such groups, or rather morph targets, can also be user created and gain scalability since they too become parametric - like the pre-selected groups. Hence, an AU can be created by the tools in poser, and from the result a morph target can stem. Further morph targets can then be created to resemble AU's. They all gain the ease of application and malleability through being parametric. After the necessary AU's have been created, one could begin to 'build' facial expressions from these building blocks.

To further ease the understanding how the AU's are created in the 3d software, an example of how the complex AU 6 + 13(a) looks like, and a bit on how it was created. Beginning with AU 6, this action unit e.g. pushes skin upward and slightly inward from the temple and cheeks, can cause crow's feet around the eyes, and may lower parts of the eyebrows to a small extent – a kind of squinting look. To illustrate this description, Sydney™

from Poser can assist by a short set of simplified steps.

Figure 1. Six steps toward Action Unit 6

On the leftmost image, a tendency of crow's feet has been already been modeled by moving the vertices on her face. In the second step, the model has received a slight lowering of the outer brow, while in the third step the entire brow has been slightly lowered, and in the fourth the vertices above the center of the inner brows were pulled up. The fifth step involves a slight pull up and tightening of the lower eyelids, which is accentuated in the sixth step where a weak squint was added.

While the previous AU is situated in the top half of the face, AU 13 involves another set of muscles – those surrounding the mouth – even though AU 6 can be involved if AU 13 is very strong. Action unit 13 involves the pulling of the lip corners while the rest, the red part of the lips, stays put. The nasolabial furrow deepens as a result of this pulling and also results in slightly puffy cheeks, due to the angle at which the corners are pulled. The upper lip seems to become flattened. The leftmost image has already received some work to the lips and the dimples around the lip corner, as well as a slight truncation of the red part of the lips around the corners. The second step involves a slight pull outwards on the right lip corner, whereas in the third step this has been mirrored to include the left side.

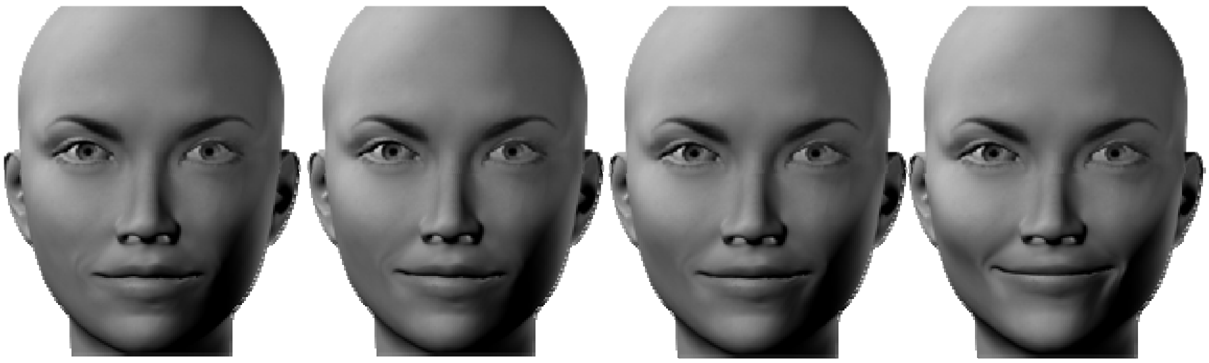


Figure 2. Four steps toward Action Unit 13

The third step also illustrates a problem with the mirroring as the artifacts, jagged edges, on the left side of the lips are still present – artifacts not present on the right side. These have been taken care of in the fourth step by using Poser's built in tools that can cover areas of vertices to make them smoother. The fourth step also involved further pulling the lip corners outwards, and at an angle, which accentuated the cheeks which had been puffed in step 1. When the separate AU's were finished, they were made into morph targets. Since the morph targets are the sum of the changes made in each of the AU's, the morph targets can then be set independently with ease and to whichever value the user chooses – do note that by increasing a morph target one increases or decreases all the small changes contained by it proportionally. The below images illustrate two versions of AU 6 + 13, with the leftmost having both AU 6 and AU 13 set to 1 – the original setting for the AUs. When these AUs are both set to this default, the cheek that has been lifted receives changes additionally from both AUs, and as a result is quite strongly lifted.

Figure 3. Touch up on AU 6 + 13

Thus, these AUs could not both be set at 1, preferably none of them set to 1, and needed slight tweaking to more properly fit to the FACS version of AU 6 + 13. The second image is the result of the tweaking. This tweaking is possible due to the aforementioned parametrical nature that the morph targets assume, and can be very precisely altered. Thereby, it would be possible to further exaggerate the AUs, even into a caricature of the combination, or lower their intensity to illustrate a weak, beginning, or ending combination; there are several possibilities that arise from the ability to precisely tune the intensity of the action units with smooth transitions in animations being one.

Validation: Part 1 and 2 – Multiple Likert-Scales and Free choices

Method

Participants

A total of 35 participants (20 females and 15 males) contributed to the completion of the first step in validating the facial expressions. The majority of the participants were students of *Lunds Universitet* (introductory course to psychology and marketing psychology), and a small minority were from an IT consulting company. No one was reimbursed for participating.

Thirty-one participants completed the first part of part 2 (19 females and 12 males) by freely choosing to supply information about what they perceived from the slides. Due to the nature of this experiment, partially measuring related words and synonyms to emotions, it also had its own task to sort and categorize the answers from the free choice part into emotion categories. A total of 3 (2 females and 1 male) participants completed this synonym-assignment, and thus served as judges of this sorting task.

Material

The stimuli consisted of 32 grayscale images. 14 images were photographs from the Karolinska Directed Emotional Faces (KDEF) (Lundqvist, Flykt and Öhman, 1998). We chose this set due to that Pictures of facial affect (PoFA) is quite well known, and considering that the vast majority of participants would be students of psychology, we tried to avoid the set partially presented in the book that everyone in the introductory course has to study. Unfortunately, neither KDEF nor PoFA include the contempt expression, and the female photograph for contempt was extracted from “Scalar ratings of contempt expressions” (Matsumoto, 2005), whereas a male counterpart for that image was trawled through the internet (picsearch.com). Albeit the male image has a slight downwards tilting chin, it is the closest to Matsumoto’s example without being e.g. a celebrity, or another person the participants might have been subjected to previously. There was 1 individual - 1 of each gender per expression - per image in the photography case; reason being that the KDEF did not include any contemptuous expressions, leading us to diversify the amount of faces instead of breaking them off with two different persons for the missing expression. For validation purposes, this was not incorporated in the Poser animated images - these were of two individuals. These individuals were created and rendered with 8 expressions – anger, contempt, disgust, fear, happiness, sadness, surprise and neutral – aided by a component table for action units (Parker and Waters, in Spencer-Smith et. al., 2001), using Poser's freehand sculpting tools for recreating contempt, and the neutral expression was the default Poser face with removed crow feet that are set remarkably high as a standard. Finally, both animated images and photographs were desaturated (removing the color, but not turning it to true grayscale), cropped to match each other, and resized to 504x504 pixels. A LGE T1 laptop was used for the experiment and PowerPoint 2007 (Microsoft, Redmonds, WA) was utilized to display the stimuli.

The material in the first part of the free choices-assignment was the same as in Experiment 1. The second part consisted of 184 words and small phrases extracted from the first part of Experiment 2. DMDX, a millisecond-accurate presentation program (Forster and Forster, 2003), was used, and a script written for the program to display the words and phrases was created.

Procedure

The participants were either in classrooms in Lunds Universitet, or in the conference room of the IT consulting company. They were read a short set of general instructions, received specific information on their answering booklets and then a final general set of instructions on a projector. They were told that some of the faces would be animated, i.e. had been created, and some were photographs of faces. Further, they were told that their viewing time was limited, and that every image would be on screen for 10 seconds followed by a 1 second blank – completely white image – before the next image would be loaded. During this time they were to answer how intense they perceived the expression, or expressions if several, shown. The participants rated each image for the perceived accuracy of each of the eight expressions: anger, contempt, disgust, happiness, sadness, surprise and neutral. This judgment of the participants perceived accuracy was made on multiple (8 per slide) 7 point Likert-scales ranging from 1-7 where 1 was, in relation to an expression, *not at all* (Swedish: inte alls), and 7 indicated that the participant perceived the intensity as *very strong* (Swedish: väldigt stark). The question order of the Likert-scales per slide, i.e. the order of perceived expressions per face, was randomized across participants.

The procedure for the free choices part was in general similar to the one in Experiment 1, with some important differences. The participants got different instructions and a different booklet. This time, the participants were instructed to note what they thought the person in the image felt. They were told to use one word, several words, or a phrase per emotion; if they saw more than one emotion in the expression, they were instructed to include them all. They were at liberty to indicate whether they had seen something else than an emotional expression, or whether they had seen no expression at all. Finally, they were instructed to optionally indicate the perceived intensity of what they saw, with 1 being no intensity ranging to 7 which indicated a very strong intensity. The participants were thanked for their cooperation, debriefed, and then excused.

The second part, the judges' synonym-assignment, in the free choices part consisted of participants either doing the test in my home, or on their own computers with support over the internet, or telephone, as to completely set up DMDX and to execute the script that contained the test. The first screen in DMDX after the execution of the script greeted the participants with an instruction of how to commence with the test. The participants were told that 184 words and small phrases were included, and that these were divided into 4 blocks, i.e. 46 units per block – which were randomized within the blocks by the program. They were shown a response-sheet with numbers connected to 8 emotions on the right hand on the screen, and instructed that they should respond according to these numbers, or answer 0 in case they perceived that the presented unit did not fit any of the emotions. They were further instructed to look for the best match, and respond as quickly as possible – even though the test was not time constrained. The instructions were made clear to assure them that we were interested in what they saw, and that there is no correct response but theirs. They consented to the experiment by pressing spacebar, which also started the test. In the case where they performed the test in my home, I left the room if they had no further questions.

Results and discussion – Multiple Likert-scales

Female faces. The responses in regard to the perceived expression of female faces are presented in Table 1. The table contains both Poser animations (top half) and photographs (lower half), and the responses have been averaged across the 35 participants. Both settings share the characteristic that there, for some expressions, are indications of

TABLE 1
Mean Intensity Ratings for the Female Poser Character and Photographs (n = 35)

	Rating							
Expression	Happy	Angry	Disgust	Surprise	Contempt	Sadness	Fear	Neutral
				Poser images				
Happy	**4.86**	1.06	1.09	1.17	1.46	1.00	1.00	1.17
Angry	1.00	**5.09**	*1.29*	1.06	1.43	1.46	1.00	1.09
Disgust	1.03	*4.26*	**4.69**	1.03	2.14	1.00	1.17	1.06
Surprise	1.17	1.00	1.00	**4.06**	1.03	1.43	*1.2*	2.31
Contempt	1.37	1.46	1.00	1.29	**4.06**	1.17	1.09	1.94
Sadness	1.03	1.11	1.00	1.06	1.03	**4.60**	1.71	1.60
Fear	1.03	1.06	1.06	*2.11*	1.09	1.63	**4.83**	1.00
Neutral	1.06	1.17	1.00	1.20	1.23	1.00	1.00	**5.31**
				Photographs				
Happy	**4.83**	1.00	1.00	1.37	1.11	1.03	1.00	1.40
Angry	1.00	**3.06**	*1.74*	1.40	1.68	1.57	1.34	1.03
Disgust	1.09	*2.17*	**4.43**	1.14	1.06	1.31	1.17	1.03
Surprise	1.94	1.06	1.03	**4.46**	1.57	1.17	*1.07*	1.14
Contempt	1.46	1.09	1.09	1.00	**3.71**	1.14	1.06	2.40
Sadness	1.00	1.40	1.31	1.20	1.11	**4.86**	1.46	1.23
Fear	1.03	1.00	1.20	*2.66*	1.43	1.97	**3.17**	1.00
Neutral	1.09	1.11	1.00	1.11	1.40	1.29	1.00	**4.83**

Intended expressions are set to bold
Most common confusions are set to italics

additional perceived expressions. The intended surprise expressions share fear as their second strongest emotion, which could be understood from the perspective that surprise per se is a neutral expression, but shares action units with fear. Additional mixes of perceived expressions include the commonly found relatively high levels of anger intensity for disgust. Interestingly, the intensity of disgust for the anger expression is perceived as remarkably weaker than anger for the disgust expression – for both photographs and poser images. In general, the similarities of the intended expressions indicate that the animated facial expressions created here should more than suffice for our purpose. Finally, there is a noticeable difference between the anger expressions; the anger expression for the photographs is noticeably lower than for the poser images, commanding an almost 2 point difference in intensity.

A significant difference as per the average judged intensity of the facial expressions was found; a repeated measures ANOVA computed on the participants' intensity judgments revealed that Poser images were rated as significantly better than photographs [$M = 4.693$, $SE = 0{,}086$; $F\ (1, 33) = 16.749$ $p < .001$ $\eta^2 = .337$]. While this is not a part of the test procedure per se, it should be noted that much of the difference could be attributed to the very noticeable difference between the anger expressions. And eliminating anger for both photographs and animated images yielded a non-significant result if the same p-value was used [F (1, 33) = 6.671 p = .014[1] η^2 = .168]. Participants' gender did not yield a significant main effect.

Male faces. Again, one can see indications of multiple perceived expressions in e.g. disgust which overlaps to anger. And yet again, what seems to be a possible perceptive overlap is not perceived for the anger expression. The most noticeable difference between the Poser images and the photographs occurs in the sadness images, where the photograph was rated more than 2 points lower than the Poser image. The difference could perhaps be argued to stem from a perceived ambivalent expression by the model – e.g. that the model either has a special way to express sadness or was unsure on how to portray sadness.

TABLE 2
Mean Intensity Ratings for Male Poser Expressions and Photographs (n = 35)

	Rating							
Expression	Happy	Angry	Disgust	Surprise	Contempt	Sadness	Fear	Neutral
				Poser images				
Happy	**5.46**	1.11	1.00	1.09	1.09	1.37	1.06	1.03
Angry	1.00	**4.82**	*1.06*	1.00	1.97	1.14	1.06	1.63
Disgust	1.00	*2.51*	**4.71**	1.23	2.09	1.09	1.14	1.14
Surprise	1.03	1.00	1.00	**4.09**	1.09	1.57	*1.71*	1.60
Contempt	1.63	1.00	1.00	1.54	**3.94**	1.06	1.00	2.65
Sadness	1.03	1.69	1.00	1.09	1.14	**4.54**	1.2	1.34
Fear	1.00	1.03	1.00	*1.03*	1.09	3.03	**4.66**	2.09
Neutral	1.20	1.00	1.00	1.14	1.37	1.43	1.03	**5.46**
				Photographs				
Happy	**5.6**	1.29	1.00	1.17	1.11	1.00	1.00	1.23
Angry	1.3	**3.54**	*1.17*	1.29	2.09	1.11	1.00	1.74
Disgust	1.03	*1.91*	**5.54**	1.37	1.51	1.2	1.09	1.03
Surprise	1.03	1.03	1.00	**5.11**	1.03	1.09	*2.09*	1.34
Contempt	2.34	1.14	1.09	1.06	**3.89**	1.00	1.00	1.49
Sadness	1.51	1.06	1.14	1.46	1.03	**2.49**	1.83	1.58
Fear	1.03	1.14	1.17	*2.17*	1.17	1.37	**3.46**	1.00
Neutral	1.00	1.17	1.06	1.00	1.23	1.54	1.23	**3.86**

Intended expressions are set to bold
Most common confusions are set to italics

There is a significant difference between the male Poser images and photographs; the male Poser images are on average judged as significantly more intense than the male photographs [M = 4.711, SE = .0834, F (1, 33) = 24.042, p < 0.001 η^2 = 0.421]. Most of the difference can be attributed to the large difference in the sadness expressions. However, it is also in part due to the fear expression where the Poser image is 1.2 points stronger, but this low is partially canceled out by a lower score on the Poser images, circa 1

[1] p-value set to 0.01

point, for the surprise expression. Using the same idea as for the female faces, eliminating the weakest expression, we instead get a non-significant result for the same p-value [F (1, 33) = 5.805, p = 0.022 η^2 = 0.150]. Finally, participants' gender does not yield a significant main effect.

The results indicate that the intensities for both female and male Poser images are significantly larger than their photographic counterparts. In both cases however, a notable part of the difference in average judged means could be explained by deviations in single expressions. Interestingly, and in both female and male cases, it was, overall, the poser images that were judged as significantly more intense in their intended expressions. It indicates that action units hold the potential to successfully be manipulated, and composed, through Poser into complete facial expressions. Some expressions are however seemingly not as good as one think they should be. Male sadness and the angry female photography are such examples. Data in table 2 can suggest that there were some confusion data between the intended expression of sadness, photography version, and the participants' responses to it on e.g. fear and neutral which could be problematic due to that sadness is normally not confused with either (Schlosberg, 1952; Rosenberg and Ekman, 1995). Also, a quick investigation of the respondents' frequency of rating intensity at 4 or above reveals a very low result at that

Table 3
Frequencies of Respondents Indicating an Intensity of 4 or Higher

§Results in %	Expression							
cutoff at 4	Anger	Contempt	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Hum. female	40	57,2	71,4	48,6	71,5	77,2	82,7	82,8
Anim. female	85,7	80,1	82,9	88,6	91,5	88,6	80	74,4
Hum. male	57,1	62,8	91,4	45,6	91,4	62,8	22,9	88,5
Anim. male	82,9	80,1	88,6	85,7	88,6	94,3	88,6	65,7

cutoff; only 22.9% of the participants rated sadness at 4 or above. Possible reasons for the photography version of sadness' deviation will be further illuminated in the remaining two validation parts. The table also serves as a mean to explain the relatively low intensity the female photography angry expression received; the participants did generally not rate it at 4 or above intensity.

One should, however, perhaps not rule out order effects for intensity, e.g. adding a propensity towards neutrality instead of sadness, as they could have negatively influenced e.g. the sadness expression in the photographs' case; following the quite intense angry Poser expression, the participants might have been affected by an already ambiguous expression – as could be evident by the distribution of the judgments since sadness is normally not confused with the expressions included in the text. For the female photography low-scorer, it would be difficult to attribute the 2 point difference in judging the female angry expressions to order effects in the same manner as per the male sadness; the low scoring female photograph is the beginning of the entire test. However, it could of course be argued that the participants could assume a careful approach as it is in the beginning, and the order would thus matter. Regardless, the results are still indicative of that the intended expressions have been judged at least equally well, if removing the weak photography's and their counterparts, or better in favor of the poser expressions.

Results and discussion – free choices

Whereas frequencies were a quick way of assessing how many rated e.g. an intensity of 4 or above in the first experiment, in this second experiment they were the point of interest. Since participants were to indicate with words what they perceived, and only asked to optionally indicate the intensity, the discrepancy between frequency of ratings and actual attachment of intensity to ratings lead to much missing data. The frequencies however, which were our main aim to investigate in this part of the validation, were relatively sound with a few noteworthy dips.

Table 4
Particpants' Free Choice Responses According to Categories Derived From the Judged Sorting Task (n = 31)

Ratings in percentages	human female	animated female	human male	animated male	Chance Criterion 1	Chance Criterion 2
Anger	41.9ab	93.5abc	38.7a	71abc	.2	.33
Contempt	16.1	12.9	32.3	35.5a	.2	.33
Disgust	45.2ab	45.2ab	87.1abc	38.7a	.2	.33
Fear	51.6ab	71.0ab	67.7ab	32.3	.2	.5
Happiness	93.5ab	90.3ab	96.8ab	80.6ab	.5	N/A
Sadness	77.4ab	77.4ab	19.4	61.3ab	.2	N/A
Surprise	61.3ab	83.9abc	74.2abc	74.2abc	.33	.5
Neutral	35.5	45.2	45.2	58.1	N/A	N/A

a Different from Chance Criterion 1, $p < .05$.
b Different from Chance Criterion 1, $p < .01$.
c Different from Chance Criterion 2, $p < .05$.

When testing whether the indicated perceived emotions differed from chance, we compared them to two sets of chance criterions through binomial tests – for results see table 4. The first criterion is concerned with the possibility of choosing any similar emotion (Rosenberg and Ekman, 1995); e.g. the chance of choosing a negative emotion, of which there are five, leaves each one with a 20% chance of being chosen. Happiness stands alone, and gets a stringent chance of 50% due to the chance of choosing a positive emotion or negative emotion is considered as 1 in 2. Surprise is sometimes augmented with a positive or negative emotion by the participants, and includes these additional probabilities (ibid); surprise thus receives a chance of 33%. Due to the randomly chosen photographs, the intermixing of animated images and photographs, as well as the responses being judged instead of computed through a lexicon, the test employs a more lenient approach to chance levels than the experiment which served as inspiration – see Rosenberg and Ekman (1995). The second criterion depends on the probability to pick an expression which is commonly confused with other expressions (ibid). The neutral expression was omitted from these calculations.

Glancing at the ratings on the previous page, the angry female photography and the male sadness photography received much less correct indications than their animated counterparts. Only about 7% of the participants did not indicate that the female angry animated image was in fact angry, whereas this number was circa 58% for the photography; and similarly 25 did not indicate sadness for the male photography of sadness whereas 12 did not indicate sadness for the animated version of the expression. Other noteworthy low scores are the male animated contempt, disgust and fear; these range between circa 32 and 39 % according to the judges' responses. However, contempt doesn't rate higher for the photography version, as expected, but disgust and fear do. Instead, the male photography versions have their own problems with sadness being the most noteworthy, receiving only circa 19% correct responses. Female contempt ratings were equally problematic at sub 20% for both conditions. When applying the criterions, only the animated male contempt was above chance levels. At the same time, that very animated male was the only one that failed all criterions for the fear expression. It is quite difficult to assess the low recognition of the fear expression, especially with the intensity score in experiment one in mind, which was higher than the photography-version. Particularly when fear should not need a label that reminds the participants. It could perhaps be an order effect, considering that it succeeds a human fear expression, but it should be minor in relation to the interest of this recognition experiment (Rosenberg and Ekman, 1995). It could then, of course, also mean that the male animated fear needs further tweaking. Other notable results where animated images yielded higher results were the angry expression and the neutral expression.

In general, the results are similar to those in part one, with a few exceptions – even though they are of course not directly comparable. Contempt receives a battering, and oddly enough receives raw, i.e. pre-judged, ratings such as cocky and horny – a treat for Freudians to analyze. The results for contempt are however in line with previous results of contempt on non-forced choice tasks (ibid). Other responses suggested that some participants were not entirely serious about their task as e.g. cross-eyed which, whatever criterion one uses, one would be hard pressed to confirm as an actual expression. Such results could be ascribed to a suboptimal response sheet, and its instructions. But such critique would dismiss the actual responses, and while making it more informative might seem as a good idea, too much control would hamper the intention of it being a free choice task. It could be argued that the judges were too stringent; in fact, the judges dismissed a vast amount of responses as non-expressions; being callous against the judgments is however hardly helpful, even though some synonyms, according to a dictionary, were not judged as such. Free-choice tasks as such receive weaker scores than their forced response counterparts, and even though some expressions received quite impressive scores, this could very well serve as an explanation.

Validation: Part 3 – Multidimensional Scaling

Method

Participants

10 participants contributed to the third experiment; 4 were female and 6 male. 4 completed the experiment in my home, 2 on their own computer, and 4 in a lab on *Lunds Universitet.* None were paid for their participation.

Material

The content of the images shown in this part of the paper are the same as in Experiment 1 and 2. There are however important differences. As the Multidimensional Scaling depends on comparisons, the participants would see all possible orders in order to pick how similar the expressions are, including a carbon copy of the same image, which was most easily achieved by extensive copy and paste to get the images side by side. The images were to be displayed in DMDX, primarily on computers with a 1280x1024 resolution – remote participants were instructed on how to reset the monitor, or excused if their monitor did not fit our demands. The images were then resized to remove the possibility that the faces would accidentally be cropped, and to avoid that the faces would block the scale used as a reference – which could happen for wide-screen monitors had they not been resized. The composite image was resized to 709x357 pixels. A total set of 256 images (16x16 for all possible comparisons) was created for the female, and 256 for the male images, giving a total of 512. Further, for the images to display in DMDX, scripts were made to show the female comparisons, male comparisons, and two different training scripts. The training scripts included an instruction and selected images from the experiment.

Procedure

If the participant completed the experiment on their own computer, they were first instructed on how to set up DMDX and TimeDX (a part of the DMDX package), and to start the program and a version of the experiment, through a script. The order of the experiment, whether the participants began with male or female faces, was arranged by notifying the participant on which script to run. The participants received their instructions on the first screen that came up; participants received information that they were going to see a large number of faces that would come up side by side, and that they were to judge how similar these faces were on a scale from 1 (highly dissimilar) to 9 (highly similar). They were further instructed that the images would amount to 256 images of females and 256 images of males and that these would not mix – i.e. either female or male faces per part of 256 images. These faces would come in blocks of 16 images. Finally, they were instructed that we were interested in their opinion – i.e., no correct answer exists – and that they both begin the experiment and consent to it by pressing the spacebar.

Results and discussion

The acquired ratings for the animated images and the photographs were submitted to a scaling analysis (PROXSCAL – PROXimity SCALing) in the SPSS software. Given that the participants were rating all possible pairs, the data turns out as asymmetric; mirrored pairs do not necessarily yield the same similarity score, and thus the matrix was not symmetrical. The data was then symmetrised within the software, and PROXSCAL also allows for similarity input leading to that input of the ratings was untouched prior to its entry into the software. Since data was aggregated, the responsible algorithm for analysis was simple Euclidean. The proximity of the data, how different variables or stimuli are related and rated in relation to each other, is then calculated. This way, the similarities of stimuli are measured and the participants comparisons of all available pairs, is analyzed to determine how they relate and rate these images to each other.

The output is presented as distances between the animated images and photographs can in our case be viewed as a map where the participants' answers have decided where the landing strips and their control towers should be if continuing on the airport example. A closer proximity between stimuli, e.g. happy photograph and happy animated face, indicates that they, in this case, are more similar to each other than they are to other stimuli. Their directions are also indicated on the plot and indicate which other stimuli they are related to.

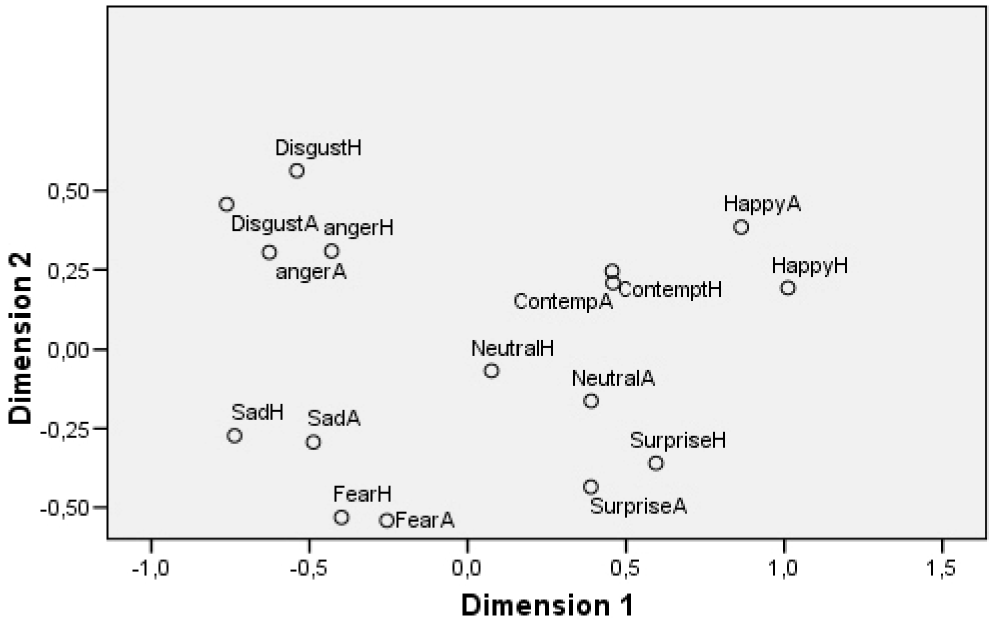


Figure 4. Derived multidimensional scaling configuration for female Poser images and photographs. The A-suffix corresponds to an animated expression, and H for human.

A separate MDS analysis of the male and female expressions was computed, and the resulting plot for the female configuration is shown in figure 4. The Kruskal stress value for the solution is .11. The configuration of the male faces is shown in figure 5, on the following page, and its Kruskal stress value is .14. These configurations can serve as a tool to analyze the relationships between the facial expressions. The male 'map' indicates that the photography of sadness is not only perceived as sad, but as a concoction of sadness, neutrality, a touch of fear and oddly enough happiness and contempt. It seems odd since the last two proximities are not illuminated by experiment 1 where they both received null ratings on intensity.

But it does indeed indicate a weakness in the expression which seemingly stems from ambiguity, as the relationship to the other expressions not only tells of a potential indecisiveness on the participants part but more so that they were simply not sure what they saw. Or in more critical terms, what they figured that they were expected to see. The actor portraying the male sadness does indeed look ambivalent, but was picked at random from a set of previously validated photographs. Obviously, this anomaly could serve as a strength to the Multidimensional scaling exploratory approach to validating, e.g. expressions, since it provides valuable information about the relationships between the expressions through how the participants have perceived them – i.e. rated them across all pairs.

Further, it serves as a fine way to understand the relationship between anger and disgust both in the male and female configurations, as they have effectively separated themselves from the rest of the stimuli. And apart from the male sadness, the points of the expressions in the plot are relatively close together, or in the same domain.

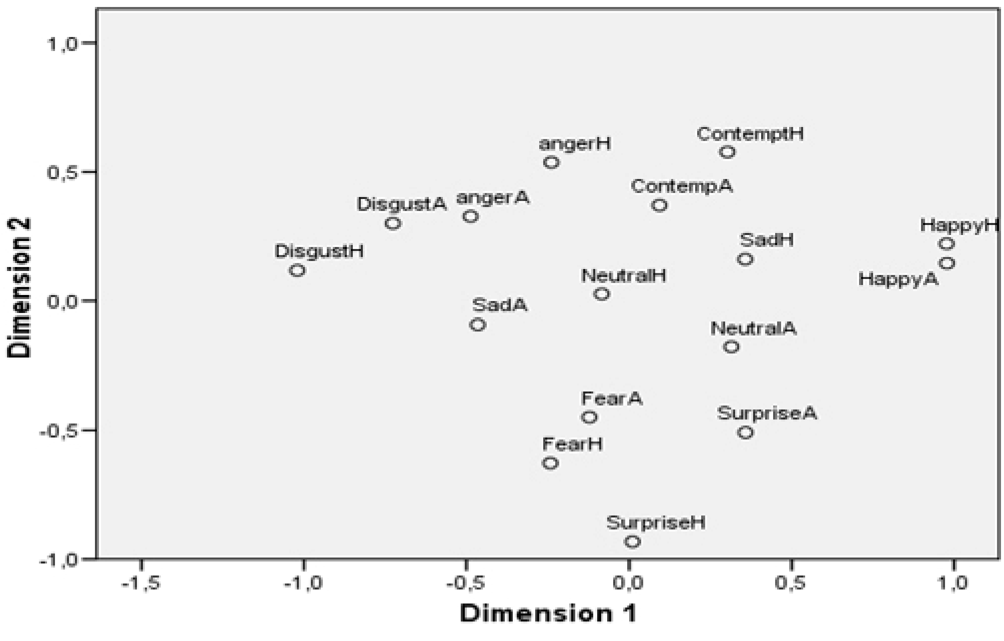


Figure 5. Derived multidimensional scaling configuration for male Poser images and photographs. The A-suffix corresponds to an animated expression, and H for human.

Experiment 2: Start that smile

Introduction to the experiment

In this experiment we make use of the techniques used in the validation to create a dynamic emotional facial expression – a smile – which reaches its maximum intensity in 5 different ways. We examined primarily whether the development of linear and polynomial smiles of different orders and configurations would be deemed as equally genuine, and secondarily, whether the different durations would not impact the ratings. The morph target for 'smile', the synthesis of action units 6 + 12 was used on a male character in Poser. It does not differ from the male character used in the validation in any way and the morph target is the same as used for happiness in the beginning of this paper.

Participants

The participants, who amounted to 60, were collected from the grounds of *Lunds Universitet*, and all were students. Thirty-five were women, and 25 men. No one was paid for their contribution, although they could help themselves to gingerbread cookies.

Material

A set of 70 dynamic smiles were created at 99 frames per second; a total of 99 images were shown per second in the renderings. These were divided between full intensity and 2/3rds of the full intensity, and shall here forth be mentioned as a total of 35 smiles unless clarification of potential differences is necessary. The smiles were generated so they would range from an intensity of 0 (neutral) to 1 or 0.667 (intended intensity) in 26 to 86 frames in increments of 10 frames; yielding a range of approximately 260ms to 860ms. The mentioned duration between a neutral expression and the intended expression can be referred to as the onset of a smile (Krumhuber and Kappas, 2005). 7 of the 35 smiles developed linearly. Additionally, 14 smiles resemble polynomials of the second degree, where 7 begin slowly and erupt to their maximum intensity quickly, and 7 begin quickly only to have a slower path to their maximum intensity; they basically consist of two onsets differing in speed. The remaining 14 smiles are similar to polynomials of the third degree, and are essentially an extension of the second degree polynomials mentioned in the last sentence. These new smiles however contain a bridged pauselike section which allowed us to either hold the beginning of the smile for a longer time, or make it begin quicker and pan out more slowly towards its maximum intensity. Examples of the trajectories can be found in Appendix

B (p. 41).The onset was preceded by 1 second with an intensity of 0 – i.e., a neutral expression – and succeeded at the peak of the expression, intensity of 1 or 0.667, with so many frames that the total clip amounted to 2.5 second. All work on the animation was done in Poser.

To finalize the stimuli, an image with the same background color as the expressions was rendered in Poser. The rendered smiles and the mid-section were then moved to Windows Movie Maker (Microsoft, Redmond, WA), in which some extra work would be done to finalize the stimuli. The mid-section was inserted between two copies of the same animated dynamic smile, and made into a clip. Both the mid-section and the stimuli lasted for 2500ms, yielding a total of 7500ms per clip – i.e. stimuli, mid-section, and a repetition of the stimuli.

Procedure

Participants were led to a lab and received instructions that they would see 40 smiles on a computer; they did either the lower intensity test or the full-intensity version. 5 smiles were in a test-block, enabling the participants to get familiar with the procedure; one smile from each group of smiles was in the test-block. The participants were told that they would assess the genuineness of the smiles by pressing a number from 1 to 7 with 1 being very false. They were further told that there would be two dark grey columns near the top of the black screen, and the smiles would be shown between these. The smiles would be shown two times, with a 2.5 second break between them, and after the second time the smile was shown, a scale would pop up, allowing them to rate the genuineness of the smile. After they had made their choice, a new presentation would commence. There was no time limit to the test, although they were encouraged to answer quickly. After the final smile had been shown, they were asked to complete a voluntary post-experiment questionnaire. Completing the questionnaire or declining it lead to a debriefing of our intentions with the experiment, and the participants were excused.

Results and discussion

For the ratings on the 35 smiles a Repeated Measures Analysis of Variance (ANOVA) was computed, with the four factors groups of smiles (5), durations (7), as well as smile intensity (1 and 0.667) and on the participants' gender (female and male). There was a significant main effect for type of smile: F (4, 224) = 4.150, p = 0.003 η^2 = 0.069. Further, there was a quadratic trend: F (1, 56) = 10.688, p = 0.002 η^2 = 0.160, indicating a difference between the types of smiles – a trend which culminated in one set and then declined, as evident from table 4. There was no main effect for intensity, or for participants' gender. The smile group found to be less genuine was of the phase variety which began slowly, arrived to an apex and basically paused there, to finally burst into full intensity; it contained two onsets

TABLE 4
Means and Standard Errors (n = 60) for Genuineness as a Function of Onset Development (Group)

Smile group	*M*	*SE*
Beginning slowly, with pause	3,651	0,141
Beginning slowly, without pause	3,935	0,139
Beginning quickly, with pause	4,077	0,13
Beginning quickly, without pause	3,966	0,13
Linear	3,88	0,125

and an apex during the master onset phase. While Hess and Kleck (1990) mention that genuinity falls with additional onsets, a sign of a deliberate smile, the top scoring smiles contain the same amount of onsets and the extra apex. This lastly mentioned group however bursts in the beginning and then slowly reaches full intensity. Durations did not affect the ratings of the participants in a significant manner as predicted.

Table 5
Means and Standard Errors (n = 60) for Genuineness as a Function of Onset Durations

Duration	*M*	*SE*
~260	3,755	0,152
~360	3,825	0,137
~460	3,908	0,133
~560	4,062	0,128
~660	3,868	0,134
~760	3,952	0,148
~860	3,943	0,155

The posited proposition that the tested smiles would not differ from each other was disproved. One group of smiles did in fact differ from the remaining four groups; a group which could perhaps best be described as hesitant smiles. The results however still indicate that the group of smiles performs worse than other. Even though one should perhaps assume that a more hesitant smile would yield lower scores due to its extra phases, the duration of the smiles is constant throughout the groups and the extra phases were assumed to add time to the actual duration of the expression. This could actually prove to be problematic, as the time constraint allows only for a minor burst in the last part of the smile. One would feel inclined to note that the onset which bursts into a smile in the begging should also suffer from a shortened onset, i.e. seen as more deliberate, but it is unscathed, and comes across as the top scoring group of smiles – albeit not significantly different against the remaining 3 groups. The main difference between these is that the smile in the latter group is quickly started and then ‘pushed’ to full intensity, providing the participant with more exposure. Then again, the increased exposure might not have anything at all to do with the higher genuineness ratings; apart from the differing group, the top scoring group was not significantly different from any other group. Whether the smile was seen as hesitant or not is a question which deserves to be posed, and the twofold presentation might have served as a mean of confusion, or a way to impress the participant with that opinion. Of course, all smiles were shown twice in a row, but this could easily be altered in two ways if required: either having only one presentation per smile, having two presentations but not in a row or both one presentation per smile and two presentations per smile as another condition.

General discussion

The intention of this paper was to provide the older Poser-faces with an upgrade, and thus invite researchers to use either more detailed animated static images of facial expressions, newer malleable models for research purposes or both as e.g. a complement to the already used Poser-faces. It further intended to investigate whether no- or low-phase smiles would be rated as equally genuine, and to investigate whether they would remain equally genuine across already accepted durations. In effect, the static part was intended as a validation-study of animated expressions created in Poser against photographs of human actors, and the dynamic part as an exploratory approach to Poser-made dynamic smiles.

The static animated images were crafted as building blocks from the very source of facial expression building blocks, Action Units (Ekman and Friesen, 1978). While the results are of course two-dimensional, these action units are malleable in three dimensions, and allow for ‘snap-shots’ from every angle imaginable with a multitude of possible lightning settings and e.g. quickly changeable complexions. When complete, the greatest asset of these animated AUs is the parametric nature they assume. Everything that has been done to one AU is easily altered by a dial, parametrically, after it has been transformed to a morph target. From these AUs, complex expressions can be made, and it is from the use of such AUs that the expressions used in this paper were created. The parametric nature of the AUs translates to the full expressions and they as well can be made into morph targets, allowing for very detailed alterations of e.g. intensity. The process of crafting animated AUs and from them expressions has previously been conducted by Spencer-Smith et. al., (2001); yielding results subsequently used in research on emotional facial expressions (see e.g. Krumhuber and Kappas, 2005; Wieser, 2007). The addition of contempt went well for both the forced response alternative and in the Multidimensional Scaling approach. The second part however, involving emotional labels where the participants were free to indicate whatever they saw fitting, yielded quite low results for both female expressions of contempt – photograph and Poser image. Actually, the male animated contempt was the only contemptuous expression to pass even the most lax chance criterion. Translated to the airport-setting, one might consider these results as lists of words or picture comparisons that the security can consult. Consulting these lists, they might be able to pick up the negative expression of contempt very easily. However, when they do not have access to such lists, confusion might arise and the contemptuous expressions might pass as something else, or unknown. These results were actually in line with prior free choice results for contempt, even though they follow Matsumoto's contempt (2005) instead of the contempt used by Rosenberg and Ekman (1995). The animated male fear interestingly did not manage to pass the chance criterions in the free choice task, but was rated as more intense than the photography fear in the multiple Likert-scale part. The multidimensional scaling approach also did not indicate any problems with the animated version. Yet, the seeming success in those approaches does not cancel out the low score it received in the free choice task, and it should at least be flagged as slightly problematic. Another animated expression which fared very well on both the multiple Likert-scales and the Multidimensional scaling, but should potentially be flagged is the male disgust. In addition, both the female photography and the female animated expression received quite low recognition frequencies, however not as low as the male animated disgust. The grand confusion source for disgust is usually the anger expression (Rosenberg and Ekman, 1995), and perhaps a more pronounced disgust could be considered, especially the wrinkles around the nose. There might have been a technical difficulty as well; while all expressions share camera and lightning settings, the wrinkles around the nose seem to have lost strength for the male disgust, which could be an artifact of the lightning, or even the choice to desaturate the images could have impeded the detail.

The female photograph and animated image also received relatively low results on disgust, but despite this slightly lower recognition rate the free choice experiment revealed an interesting aspect of the animated poser images; the animated poser images seem to receive slightly better ratings if female than their male counterparts when free choice was applied. When the forced choices method was applied, the scores were relatively equal. Logically then, fewer male animated images would be considered recognized as intended, and thus participants would have rated male animated images as far more intense than their female counterparts. However, including the frequency tables for part one regarding the cutoff at 4 reveals that the response frequencies are very similar across gender of the animated image. This leads to an assumption that when participants did not have the support of forced choices, they recognized the female animated expressions better than male. It should be stressed that this is merely an assumption, and that an extended study with free choices would, for purpose of extending this vague case, be quite interesting. Since the actor's gender seemingly does not matter (Ekman, 1989), the differing ratings could e.g. be due to a perceived difference between the models; a further study might thereby include additional models.

The male happiness expression was used to create a set of dynamic smiles. These smiles all shared a neutral starting point, and every other aspect of the morphology itself during the transition from neutral to fullblown expression. The only difference was which trajectory the smiles in each of the 5 groups assumed on its way to that peak. They could assume a linear path, begin slowly and end quickly, or begin quickly and end slowly. Both of the latter groups were subdivided into one that has an apex, a pause, in the middle of the smile, and then a burst in the slowly-beginning version or a slow completion in the quickly-beginning one. One of the groups of smiles did in fact differ from the other groups; it was a version with two onsets and one apex between them, beginning slowly and finally bursting out to a full-blown expression after a short pauselike 'break'. The duration of the onset, ranging from circa 260 to 860 ms, did however not have a significant effect on how these smiles were rated, confirming what had previously been found concerning duration. This brings about some interesting implications for the airport setting. Are the check-in attendants smiles genuine when they assist you, or are they considered as genuine due to a timing which generally works well. The large numbers of people that receive their services each day might train them in involuntary deceit, using a smile without a hesitation which potentially makes the customer feel better. From that perspective it is interesting that the group which was found to differ from the remaining four could at best be described as a hesitant smile, beginning slowly, reaching a short apex, and bursting out quickly.

There are some indications that additional phases add to the time it takes for a smile to develop (Hess and Kleck, 1990). Fitting such phases to the tested durations did seemingly not affect the way the participants rated the smiles, except in this one case. It could potentially mean that the slight hesitancy in the beginning of the smile could have caused the participants to respond in the manner they did. This would be supported by the ratings for the group of smiles with which it shares the amount of phases, albeit beginning quickly and reaches full intensity in a slow manner, since it was interestingly enough the smile that was rated as most genuine. Another aspect that should be taken into consideration is that the smile that 'failed' should only be considered as 'worse' than the smiles that matched it. Because it should of course be noted that there is nothing that indicates that the dynamic smiles used here would fare well if matched against real smiles – there is however of course nothing that indicates that they would fare ill either. This should however not necessarily be considered as a downside; cramming extra phases into the same time-duration might be difficult for an actor at will, but a computer program will surely accept such a task. Returning to the smiles in the experiment, the optional post-experiment questionnaires did not reveal anything noteworthy that could fit the specific lower-rating category, but could be considered as a treasure grove for further ideas and inspiration. Despite that the trajectories are the only aspect of the smiles within the groups that differ from each other, some respondents perceived more genuine smiles if: 'they smiled in a wide manner', 'they could see it around the eyes' or 'the eyes seemed happier'.

Of course, there are always further questions to be asked in this quite unexplored field of emotional facial expressions. And even though such smiles would perchance pass as genuine smiles, it could be interesting to investigate whether they are interpreted as something more than genuine, and obviously, if they are, to investigate what they are interpreted as; an experiment containing a free choice task where the participants would have to describe the genuineness of the smile, as well as what they see, and why they respond the way they do, could be quite interesting as a future attempt. An interesting attempt would be to try a validation from a phenomenological perspective, which could at a glance resemble a free choice task. It could however instead involve the participants in writing a story about the persons in the image, and then allowing the experimenter to see no more than the story in an attempt to see whether the perception of the facial expression, and what it signifies is shared. Of course, this idea would have to be developed further, and should at this point be considered as merely an idea.

Finally, it would be very simple to argue that animated images – and movies – could not be used as a template for investigating e.g. genuine smiles. The controllability however could make them infinitely more valuable in assessing aspects of genuineness than e.g. actors whose hallmark much often is the asymmetry produced by deliberate muscle movement which does not exist in genuine smiles to such degree – neural damage disregarded. And the idiosyncrasy also shows with the male sadness and female anger expression in the intensity part, as well as in the free choices task. Instead, the animated models can be turned into everything you might want them to be, except of course, real life actors. If the male fear was problematic, it is not necessarily so after a few tweaks; a quite unethical procedure to do on human actors – the ones in research. Yet, it is indeed a very problematic aspect that they are not real life actors. One could perhaps object that there is no way to know whether they smiles are deliberate or spontaneous. Actually, from that perspective, there is not much to it; they are of course deliberate, as they convey the experimenter's intentions and will. It is however the impression that the participant gets of such expressions which is interesting; the perception of what they experience in seeing the stimuli. In regard to that, it is also much easier to discuss problematic aspects of the stimuli. The participants might react strangely to them, or not take their assignment seriously. Even though that could potentially always be problematic, and the most affected expression in the emotional labels experiment was the photography of male contempt. Obviously, there is always more to be done, and a future study might include a lower intensity, or two lower intensities and an attempt to add additional stimuli to the collection of animated facial expressions. And the options seem endless. Malleable 3d-faces could in the future also be modified due to options which include being skinned with different textures, endless malleability thus gaining the ability to disregard or create morphological differences, asymmetry, or symmetry which would add possibilities of heavily controlled experiments. This in turn could, with little doubt, ease the probably needed future research.

REFERENCES

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008) Audio-visual integration of emotion expression. *Brain Research, 1242,* 126-135.

Ekman, P. (1972). *Universal and cultural differences in facial expression of emotion*. In J. R. Cole (Ed.), Nebraska symposium on motivation, 1971 (pp. 207-283). Lincoln, NE: Nebraska University Press.

Ekman, P. (1973). *Darwin and facial expression; a century of research in review*. New York: Academic Press.

Ekman, P., & Friesen, W. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica, 1,* 49-98.

Ekman P., & Friesen W.V. (1975) *Unmasking the face*. Prentice Hall, Englewood Cliffs, N.J.

Ekman, P., & Friesen, W. V. (1982). Felt, false and miserable smiles. *Journal of Nonverbal Behavior, 6,* 238–252.

Elfenbein, H. A., & Ambady, N. (2003). When familiarity breeds accuracy: Cultural exposure sand facial emotion recognition. *Journal of Personality and Social Psychology, 85,* 276-290.

Elfenbein, H. A., Mandal, M.K., Ambady, N., & Harizuka, S. (2002). Cross-cultural patterns in emotion recognition: Highlighting design and analytic techniques. *Emotion, 2,* 75-84.

Farroni, T., Menon, E., Rigato, S., & Johnson, M. H. (2007). The perception of facial expressions in newborns. *European Journal of Developmental Psychology*, *4,* 2-13

Fecica, A. M., & Stolz J. A. (2008). Facial affect and temporal order judgments. *Experimental psychology, 55,* 3-8.

Forster, K.I., & Forster, J.C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers.*

Green, P. E., Carmone, F. J. Jr., & Smith, S. M. (1989). *Multidimensional Scaling, Concepts and Applications,* Boston: Allyn and Bacon.

Hager, J.C., & Ekman, P. (1985). The asymmetry of facial actions is inconsistent with models of hemispheric specialization. *Psychophysiology, 23,* 307-318.

Halberstadt, B. J., & Niedenthal, P. M. (1997). Emotional state and the use of stimulus dimensions in judgment. *Journal of Personality and Social Psychology, 72,* 1017-1033.

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology, 54,* 917-924.

Hess, U., & Kleck, R. E. (1990). Differentiating emotion elicited and deliberate emotional facial expressions. *European Journal of Social Psychology, 20*, 369–385.

Hess, U., & Philippot, P., (Eds.), (2007). Group Dynamics and Emotional Expression. New York, Cambridge University Press.

Hoffman, H., Traue, H. C., Bachmayr, F., & Kessler, K. (2006). Perception of Dynamic Facial Expressions of Emotion. Lecture Notes in Computer Science. Heidelberg, Springer Berlin.

Juth, P., Lundqvist, D., Karlsson, A., & Öhman, A. (2005). Looking for foes and friends: Perceptual and emotional factors when finding a face in the crowd. *Emotion, 5,* 379–395.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska Directed Emotional Faces – KDEF.

Matsumoto, D. (1992). More evidence for the universality of a contempt expression. *Motivation & Emotion, 16*(4), 363-368.

Matsumoto, D., & Ekman, P. (2004). The relationship between expressions, labels, and descriptions of contempt. *Journal of Personality and Social Psychology, 87*(4), 529-540.

Matsumoto, D., & Willingham, B. (2006). The thrill of victory and the agony of defeat: Spontaneous expressions of medal winners of the 2004 Athens Olympic Games. *Journal of Personality and Social Psychology, 91,* 568–581.

Moskowitz, G.B. (2005). *Social Cognition: Understanding self and others*. New York: The Guilford Press.

Niedenthal, P.M., Krauth-Gruber,S. & Ric, F. (2006). Psychology of Emotion: Interpersonal, Experimental, and Cognitive Approaches. New York: Psychology Press.

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology*, *80,* 381–396.

Rosenberg, E. L., & Ekman, P. (1995). Conceptual and methodological issues in the judgment of facial expressions of emotion. *Motivation & Emotion, 19*, 111-138.

Russell, J. A. (1994) Is there universal recognition of emotion from facial expression? A review of cross-cultural studies. *Psychological Bulletin, 115,* 102-141.

Sato, W., & Yoshikawa, S. (2004). The dynamic aspects of emotional facial expressions. *Cognition and Emotion, 18,* 701–710.

Sato,W., Kochiyama, T., Yoshikawa, S., Naito, E., & Matsumura, M. (2004). Enhanced neural activity in response to dynamic facial expressions of emotion: an fMRI study. *Cognitive Brain Research, 20,* 81–91.

Sato, W., Fujimura, T., & Suzuki, N. (2008). Enhanced facial EMG activity in response to dynamic facial expressions. *International Journal of Psychopysiology, 70,* 70-74.

Schlosberg, H. (1952). The description of facial expressions in terms of two dimensions. *Journal of Experimental Psychology,* 44.

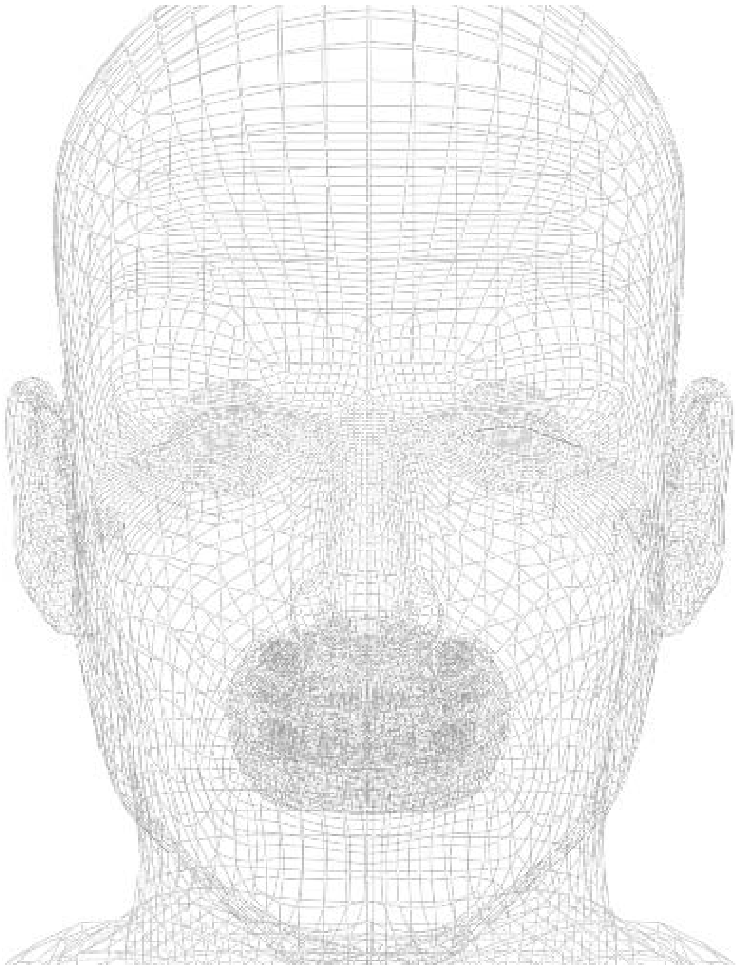
Schmidt, K. L., & Cohn, J. F. (2001). Dynamics of facial expression: Normative Characteristics and individual differences. In (IEEE, Ed.), *IEEE Proceedings of International Conference on Multimedia and Expo, August 2001, Tokyo, Japan* (pp. 728–731). Waseda University: Tokyo.

Spencer-Smith J., Wild, H., Innes-Ker , A. H., Townsend, J., Duffy, C., Edwards, C., Ervin, K., Merritt, N., & Paik, J. W. (2001). Making faces: Creating three dimensional Parameterized models of facial expression. *Behavior Research Methods, Instruments, & Computers, 33,* 115–123.

Wieser, J. M. (2007). Der Einfluss von emotionalem Gesischtsausdruck und Blickkontakt auf visuelle Aufmerksamkeitsprozesse bei sozialer Ängstlichkeit – Empirische Studien zu Augenbewegungen und Peripherpsysiologie. Retrieved December 15, 2008, from OPUS database.

Appendix A

Example of vertices forming a massive grid across an animated face and head in Poser (Smith Micro, Aliso Viejo, CA).



Appendix B

Example of trajectories of the onset/s of smiles in experiment 2.