Real-Time User Assistance With Parameterized Avatar Face Generation

Chris Maltzan

Union College

807 Union Street

Schenectady, NY 12308

maltzanc@garnet.union.edu

4 December 2012

1 Introduction

Imagine that you have just witnessed a crime. You saw a man with a firearm run into a bank and a you saw him a minute later running out with a bag of cash. The police are probably on their way and the first thing they will ask you is, "Who robbed the bank?" It's a simple question, but not an easy one to answer. You don't know their names, so the best verbal description you might give could be, "Six feet tall, Caucasian, muscular, brown hair, dark jeans and a black jacket." That description might be helpful with weeding out suspects who did not commit the crime, but it probably won't help the police pick the suspect out of an entire city, much less the surrounding neighborhood.

Law enforcement officers often rely upon untrained witnesses to provide them with leads in investigations. Right or wrong, these frequently ambiguous leads can uncover circumstantial evidence that may lead to arrests of the guilty and innocent alike. Time spent following the wrong lead not only gives the offender time to flee the scene or destroy evidence, but may also lead to a false conviction[7]. Police sketch artists are tasked with assisting witnesses of a crime in creating composites of suspects. A sketch artist does not have the luxury of seeing what a witness saw and can only create a composite, or sketch, based on a witness's description. The witness attempting to express a suspect's physical appearance to others is the only person to have first hand knowledge of the desired final product. However, the witness's role in creating the composite is secondary to the professional sketch artist, who is working with second-hand knowledge of the suspect's physical appearance through the witness.

If there was a police sketch artist available, he or she might be able to assist you as you describe the physical features of the suspect, but this has its draw backs. First of all, you have to actually have a sketch artist. Secondly, there is a degree of error which could be considered unnecessary when you use a sketch artist. To understand this, we must first analyze the process of taking the idea of a physical description and sharing it with others. There are multiple steps in this process which can produce error, see Figure 1.

When the witness first sees the target, he or she must first commit the image to memory in order for it to later be recalled. If the witness interpreted the original incident incorrectly or was missing key details, this would be the first source of error. After the the police arrive and find a sketch artist, the witness would then be asked to describe the physical features of the suspect to the sketch artist. This event will

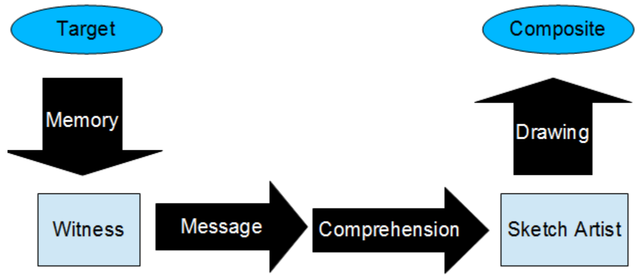


Figure 1: Sources of Error in Sketch by Sketch Artist

comprise two different sources of error. First, the witness may introduce error by giving a poor description of the suspect, even though they correctly remember his or her appearance. Secondly, the sketch artist may not completely understand the details described by the witness. The fourth source of error will come from the actual drawing the sketch artist produces. The sketch artist may simply make mistakes in his or her composite.

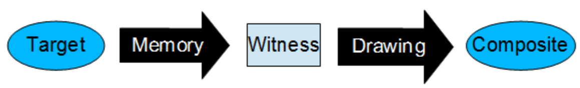


Figure 2: Sources of Error in Sketch by Witness

There is a lot of room for error in using a sketch artist. It would be of great benefit to law enforcement if it were possible to reduce the possible error in a composite sketch and also make it possible to produce a sketch without the assistance of a sketch artist. If we could enable ordinary people to create sketches without outside assistance, we could eliminate the message and comprehension errors, see Figure 2.

Although this may appear to be an easy solution for reducing composite sketch error, there is a reason we use sketch artists. As a whole, witnesses are generally poor artists. Very few will be as skillful as a sketch artist. In order to solve this problem, it is necessary to give witnesses a tool which will afford them great skill in relating the idea of physical appearance to a third party. It is not necessary that the witnesses actually draw the suspect, only that they are able to accurately provide others with the knowledge of the target's appearance. We propose that a digital medium will be more effective for our witnesses since it is more likely that they will have basic computer skills than that they will be skillful artists.

This senior thesis project focused on affording users the ability to create a composite of a target image (similar to a suspect) by decreasing the amount of time and effort required and increasing the user's ability to produce a composite highly similar to the target. Throughout this two term thesis, the focus of study shifted twice. The original project focused on the mass production of avatars according to user provided characteristics. The original project is described in the Appendix. Eventually, the project was narrowed to the generation of a single user's idea of the most attractive possible avatar. Finally the project was again narrowed to creating a framework for an algorithm to assist user's in real-time as they design an avatar. We will discuss our final project in the main section of this paper. In the appendix, we provide a general description of the second project and a very brief description of the original project.

2 Real-Time User Assistance With Parameterized Avatar Face Generation

After designing the experiment designed in the previous section, we concluded that the scope of the project was too wide. The current project stemmed from the preceding study and narrowed the focus to the design of an avatar after a target image. The difference between the sketch artist scenario and designing a most attractive avatar is actually very small, but there is one important difference. Designing a most attractive avatar is dependent upon being able to design an avatar according what a person believes himself to be most attracted to, while designing an avatar after a target is designing an avatar after a concrete image a person has in his or her mind. In other words, we narrowed our focus by removing the component where participants attempted to create an image based on what they believed themselves to be attracted to and instead had all participants use standardized target images.

We believe this was a good decision because in the previous study, we would have investigated males' physical preferences in long and short term mating contexts using a system which would not have been previously examined or tested. After the completion of the current study, it may either be reasonable to perform the previously described experiment, or there may still be a need for further research.

2.1 Intent

In the current study, we intent to investigate how users of three-dimensional avatar design software go about designing an avatar when they have a specific target in mind. We intend to investigate this process by designing a framework for an algorithm to assist users in real-time as they modify a parameterized avatar face.

2.2 Related Work

2.2.1 Mentally Processing a Human Face

Before an individual can even begin to design a human face, the idea of how the target face appears must be processed. This will include first recognizing that it is a face and often times being able to recognize whose face it is. This human element is discussed first because it must be considered for the design of a system to enable users to first process a target face and then translate that face to a new composite.

There are two main strategies for mentally processing a human face.[8][12] The first strategy is configural, or analytical, processing. This is the act of perceiving relationships between the physical structures of any stimulus, such as a face, in order to distinguish between them. There are three forms of configural processing: first-order, second-order and holistic processing. First-order processing is the ability to detect physical facial structures. Second-order processing is the ability to detect spacing between physical facial structures. Holistic processing is the ability to piece all of the structures together into a gestalt, or the shape of an entity's complete form.

The second strategy is featural processing, which is very similar to first-order processing. The main difference in using first-order processing is that the individual can identify more about a face. In configural processing, the individual will first detect the facial structures, detect the spacing between the structures and then piece the information together in order to make judgments about the face. Featural processing however, only looks at the individual facial structures and does not need to piece them together in order to judge the face as a whole[12]. The important idea to take away from this is that there are multiple strategies human beings employ in order to take a visual stimulus and identify it as a human face in order to eventually store the image in memory.

By adulthood, human beings are considered experts at recognizing human faces. Although humans are adept at identifying a human face, there are situations which can lead to difficulty. In a study by Boutet and Chaudhuri [1], the authors found that participants were able to recognize facial features of upright faces, but had great difficulty identifying upside-down images (see Figure 3). These findings can most likely be attributed to the fact that humans evolved in an environment where faces were normally viewed in the upright position, making inverted faces a relatively novel phenomenon.

Figure 3: Mental Processing[1]

A study by Maurer et al.[8] demonstrates this point effectively. The authors compared how participants were able to distinguish between pictures of faces with differences in spacing between features and faces with differences in the actual shape of those features. In Figure 4, the top row contains pictures of faces with differences in spacing and the bottom row contains faces with differences in structural shape. The authors found that participants were able to distinguish between these faces when they were presented upright as seen in Figure 4. However, when turned upside-down, participants had difficulty distinguishing between faces with spacing differences (the top row), but not shape differences (the bottom row). Bearing in mind how humans mentally process facial stimuli, next we will examine past systems for composite creation by inexperienced users.

Figure 4: Differences in Mental Processing[8]

2.2.2 Two-Dimensional Approaches

Past research with two-dimensional images produced methodologies for modifying images of photographed people and assisting the user in performing this task. Although past implementations have certain drawbacks, the current project built upon past methodologies, extending functionality of the past approaches while attempting to limit or remove previous drawbacks.

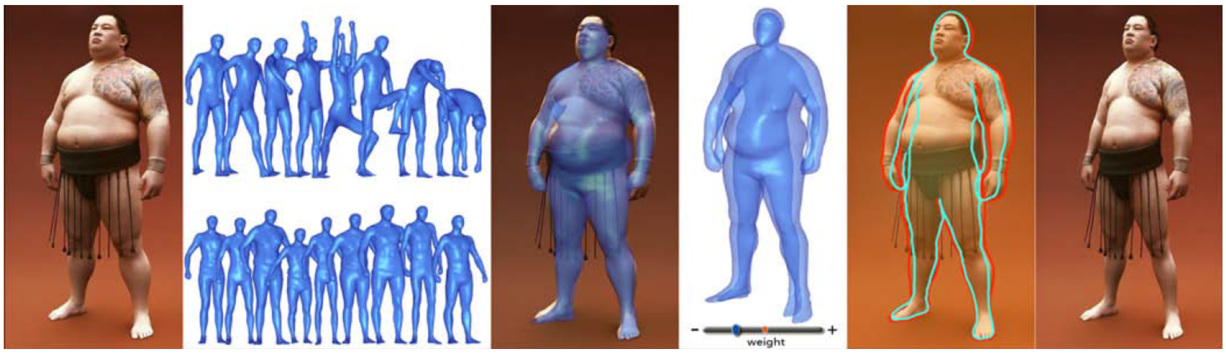


Figure 5: Two Dimensional Methodology[13]

Zhou et al.[13] used a three dimensional posable model matched with a two dimensional image in order to modify the image according to user specifications. In this system (see Figure 5), the user started with an image of a person and then matched a three dimensional posable model to the input image. The image and model were then locked together so that any future changes to the posable model would be reflected in final image as well. The researchers also created a parameterized user interface, allowing users to modify predetermined features such as waist girth, breast girth and leg length by moving various associated sliding bars (see Figure 6). In the sequence of images in Figure 5, the second image shows how a posable model can be deformed to match any pose that a human being might take. In the first image, we see an example input image, which is then matched with a posable model in the third image. In the fourth image, the posable model is resized. As explained previously, because the three dimensional posable model and the input image are locked together, the changes to the posable model are reflected in the input image as well. These changes can be seen in the last two images in the sequence, in which the sumo wrestler appears considerably smaller.

Although the Zhou et al.[13] study did provide a photo-realistic method for a parameterized image editor for human bodies, it was not without several weaknesses: pixelation, preservation of the original image's texture, and an inability to re-render a modified image. The first weakness stemmed from the method (see Figure 5) through which the images were edited. The researchers enabled the software only to stretch the input image, possibly creating a pixelated final image. Therefore, the extent to which an image could be stretched or modified by the software was due to the density of pixels in the input image, not the range of

Figure 6: Parametric Image Editor[13]

humanly normal physical features. This process was further limited because the images were only stretched in locations corresponding to a single physical feature, as seen in Figure 6. This could exaggerate the drawback to this approach because an observer of the output image may be able to see that some areas are more densely populated with pixels than other areas.

The second drawback to this approach was that the image texture (what an observer actually sees) is preserved even through modifications made to the image. Whatever the picture was originally of, a sumo wrestler[13], for example, the picture would still be of a sumo wrestler when the modification process was finished. However, it is important to note that this defies what should be expected to happen if one was to modify an image. For example, in Figure 5, although the sumo wrestler is shown to be made thinner, he still has the exact same image texture of a very large sumo wrestler. The folds of skin are still present and the texture remains unchanged, even though the physical appearance has been changed. If you imagine that the wrestler was modified again to become smaller, it would be possible to make him thin enough to be considered within the range of what is humanly possible, and yet the image texture would still reveal a sumo wrestler with large folds of skin. The image texture and outline of the two-dimensional sumo wrestler would be contradicting. The texture would seem to indicate an overweight individual and yet the shape would reveal a very thin individual.

This was closely related with the inability to re-render a modified image, the third drawback to this approach[13]. Although the user was enabled to stretch or constrict different areas of the person in the input image to create different shapes, the user was unable to change the underlying image texture according the modifications made. One possibility to make this happen would have been if the image could have been re-rendered to correct the image texture according to the newly modified person's shape. If this study had employed a three-dimensional model of the individual instead of a two-dimensional photograph coupled with the three-dimensional posable model, this could have been fixed. Because the user could re-rendered the image after making modifications to the hypothetical input three-dimensional model, the image texture and generated new and more accurate shadows for the modified physical shape of the model.

While the previous example used a three-dimensional posable model in order to manipulate a two-dimensional image, another proposed solution was to manipulate control points which correspond to various physical structures of a model (see Figure 7). When any single control point was moved, the physical structure it was linked to moved as well. Additionally, because the series of control points were linked together, the areas between the linked control points were modified as well. if an area between a group of three control points was enlarged, then the image area between those control points was also enlarged to fit the new boundary.

Frowd, Hancock and Carson created such a system using series of control points mapped to different facial structures. They enabled a user to create a target face through modifications to the control points, which in turn changed the size and location of facial features in the final output image. This study also suffered from pixelation, preservation of the original image's texture, and an inability to re-render a modified image just as in the Zhou et al. study[13]. If the control points were modified to an extreme, it was possible for the image to become noticeably pixelated. There was also no attempt in this study to manipulate the image texture, making it impossible, for example, to change something such as skin color. As with the Zhou et al. study, it was not possible for the authors to re-render any images they produced to more correct shadows for the modified facial structures. However, it did present a new methodology for manipulating a two dimensional image. Instead of using a three dimensional posable model, the authors used a vertex based control point system. This is interesting because it may prove useful for a three dimensional approach to modifying avatar faces. Since three-dimensional models are vertex based, the two-dimensional vertex based control point implementation in the Frowd, Hancock and Carson study could be employed with only slight modifications.

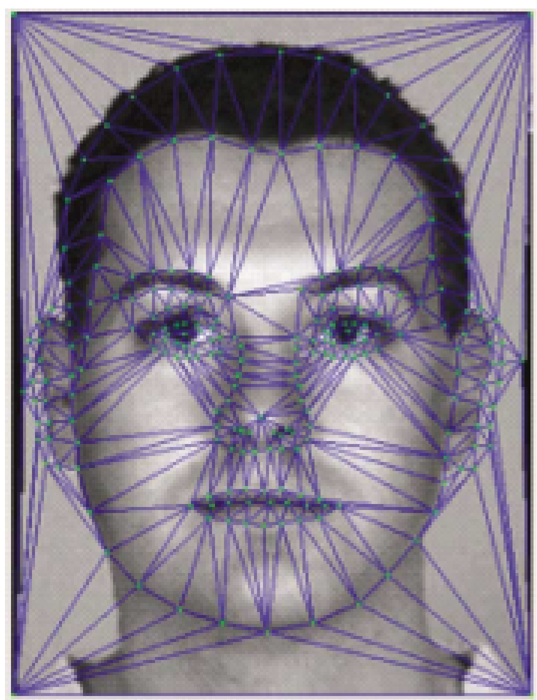


Figure 7: Facial Control Points[6]

In examining various methodologies for manipulating physical appearance, it is important to emphasize the importance of being able to manipulate individual facial features while allowing the user to maintain the context of the entire face. The Frowd, Hancock and Carson managed this taking a holistic approach for editing facial features. Once they mapped the control points to a photograph of a face, the image was edited as a whole, not based upon individual facial features. For example, when the authors moved one control point in the face, it affected how all directly adjacent areas on the face were displayed. Additionally, because all of the control points were modified together and then presented to the participants in their study, facial features were never displayed to participants out of context of the face they were trying to design. This is congruent with past research by Tanaka and Sengco who found that a single facial feature is best recognized in the context of the actual face trying one is trying to create. Based on this finding, if a user is close to creating an accurate composite, the user should be expected to become increasingly proficient in modifying the composite. Although this means that when initially beginning to create the composite from a default starting point, the current state of the composite may be very different from the end goal. However, it should still be considered more relevant than if the user was only given a variety of noses to choose from, absent a human face associated with each nose.

With this in mind, if a user was able to jointly edit facial features that are known to vary together to some degree, he or she should be expected to create a target face quicker and more accurately. This is because the user would be not only able to edit multiple features simultaneously, but the user will also be better able to recognize those facial features as they would be in the context of the target face. Past approaches (for example, by sketch artists) in which the user edits single facial features also allow for the features in question to be in the context of the target face, but only towards the end of the user's experience, when the facial composite would be most accurate. Tanaka and Sengco's study would favor an approach that edits multiple features together, as this will allow the user to edit a more complete, or in-context, face.

2.2.3 Three-Dimensional Approaches

MakeHuman is an open source software package intended for the production of fast, anatomically correct and realistic avatars. This software package served as the base for this project. The authors created a plug-in for MakeHuman which extends functionality by assisting users in making modifications to an avatar. MakeHuman was selected because it had a pre-existing environment which afforded users the previously discussed positive features for creating a composite of a human being.

Upon opening the MakeHuman software package, the user is presented with various menus for editing the different parts of the human body. For this study, the face menu will be most important. The base avatar, presented in the center of the screen, is a very tan, gender ambiguous model with hair clipped tightly to the skin. This is the model all users will start with when preparing to make an avatar.

To create the avatar a user sees, MakeHuman employs thousands of targets which which act as a prototype for how a physical trait should appear. Each target consists of a list of x, y, z coordinates. Each of these targets corresponds to various body conditions and physical features. For example, there are targets for an elderly female's nose chin, a young female's chin, a young male's large eye or a young female's small eye. Any modification a user makes to a parameter does not directly effect the avatar's topology, or physical form. Instead, when a modification is made, target files are accessed and the degree to which a parameter is modified will determine how close the vertices of the avatar will be moved towards the locations prescribed in a target file.

When editing the body, a user can edit parameters including gender, age, muscle tone, weight and height. These parameters are controlled via a sliding bar mechanism similar to the Zhou et al. study (Figure 6). When editing the face, similar parameters are used, except these parameters are facial structure specific, rather than global. That is, parameters such as nose width, nose size, eye diameter, or eye height, do not have effects on the surrounding facial features. The difference between the approach used for modifying facial features and body features is different in MakeHuman. This makes sense because the parameters used for the body are far less pronounced in the face, particularly muscle tone, weight and height. However, it is unfortunate because structure specific parameters should be expected to be less effective than a parameter which is able to modify multiple structures at the same time (assuming these modifications are all beneficial).

Unlike the global body parameters, the facial parameters are additive. That is to say that the effect of a modification to any parameter will effect the avatar jointly, although independent of, subsequent modifications to parameters of nearby structures. By contrast, a non-additive parameter, such as those used for bodily modification in MakeHuman, have dependent and joint effects upon the avatar. In practice, this means that when a user modifies the width of the avatar's nose, the eyes are not effected. This can become problematic because the width of the nose may encroach upon where the eyes are set, creating an unnatural appearance. A similar event could occur with any adjacent physical facial structure.

2.3 Our Approach

The current study expanded upon past two-dimensional and three dimensional implementations for editing the human body. Specifically, we focused on facial modification and design. We designed an algorithm which enabled experienced and inexperienced users to design and produce an avatar's face faster and with greater accuracy than users who did not use it.

This was accomplished by designing a framework for an algorithm which assisted users in the making modifications to the avatar's face. Whenever a user made a modification, the algorithm checked to determine if the result was within the realm of what is considered humanly normal and possible. The algorithm accomplished this by checking a previously determined list of parameters which could potentially interact with each other. If such an interaction was possible, the algorithm then checked to see if the modification made would interfere with realism of the avatar as determined by inappropriate proximity of two facial structures, meaning that two structures were too close together, or even intersecting. And also by checking to see if any structures were at opposing extremes of a spectrum. For example, it would probably be unnatural for an individual to have an extremely wide mouth if the lips were also extremely thin.

In cases where the algorithm found such events, the algorithm modified parameters slightly in order to keep within the range of what is considered humanly normal. The algorithm gave preference to the most recent modifications and assisted the user by only modifying the less recently changed parameters. This was because the most recent modifications were more likely to still be desired by the user and changing them may upset the user's goal.

The framework we created is scalable in that it will be able to hold future modifications to the current algorithm. Although the algorithm we designed does assist the user in real-time, there is still room for improvement. The important point to stress, however, is that the framework we set up in the MakeHuman system is robust and provides a strong platform for future work to build from. This will be further detailed in the closing discussion.

Beyond algorithmic changes to MakeHuman, we also felt it was important to make a few minor changes to the user interface. In the original version of MakeHuman, some of the sliders controlled features which could be best observed from a side view. The original creators implemented such that for these particular sliders, the view port would switch to a side view in order to give the user the best vantage point. Although this was a great feature for the original purpose of MakeHuman (avatar design), we did not feel that it was beneficial for our purposes. We chose to modify the interface of MH such that while modifying any slider, the view port was always locked to a frontal view.

2.4 HYPOTHESIS

If asked to recreate a target avatar, then users will design the avatar faster and more accurately with our parametric non-additive facial generator than than without it's assistance. We will measure the speed of creation through both the number of modifications the user makes before completing the task and also the amount of time they spent editing a composite. Specifically, we predict that task completion times will be lower when participants are assisted by our algorithm than when they are not assisted. Users assisted by our algorithm will make less modifications to the composite than users unassisted by our algorithm. Participant ratings of similarity to the target will be higher when participants are presented with our modified system than the original unmodified version. Participants will also report greater ease of use with our system than the original.

2.5 Method: Creation Phase

2.5.1 Participants

In the first phase of the experiment, 16 (9 females, 7 males) of participants were self-selected by responding to flyers posted in the campus center at Union College. All participants opted to participate for a cash reward of 10 dollars.

2.5.2 Design

Participants experienced one of two possible treatments. In the control condition, eight participants were not assisted by our algorithm during the experiment. In the experimental condition, 8 participants were assisted by our algorithm. Each participant completed three consecutive trials, in which they were asked to design a composite avatar, before completing our study. The experimenters tracked speed as measured by both time until completion and number of modifications made. The experimenters tracked participants' belief of success by recording the participant's self-reported similarity scores for their composites to the target images.

The procedure of this experiment was based off of a similar study by Hasel and Wells (2007) focused on witness accuracy in the courtroom and is described in detail in the following sections.

2.5.3 Materials

Participants used the MakeHuman system to recreate target images. As previously described, MakeHuman affords users the ability to modify a three-dimensional avatar by dragging sliders which pertain to physical characteristics in the avatar's face. Figure 8 depicts the MakeHuman user interface our participants were exposed to.

The sliders on the top left side of figure 8 can be dragged to the right or left in order to increase or decrease a physical trait. Currently, all of the sliders are in their left-most position. On the right-hand side, in the "category" panel, the user can select a set of traits to modify. Currently, the interface is set to modify the avatar's head shape. The set of visible sliders on the left is dependent upon the category chosen. For example, if the user selected "Right Eye," then a new set of sliders would appear in place of the "Head Shape" sliders.

As previously explained, we chose to slightly modify the interface of the original MakeHuman such that the view port was always locked to a frontal view when a user was in the process of modifying any feature. Although it was locked in the frontal position during a modification, between modifications, users still had the ability to pan around their composite. We chose to do this because of the style of target images we presented to our participants. Since all targets were frontal face images, we felt that during a modification,

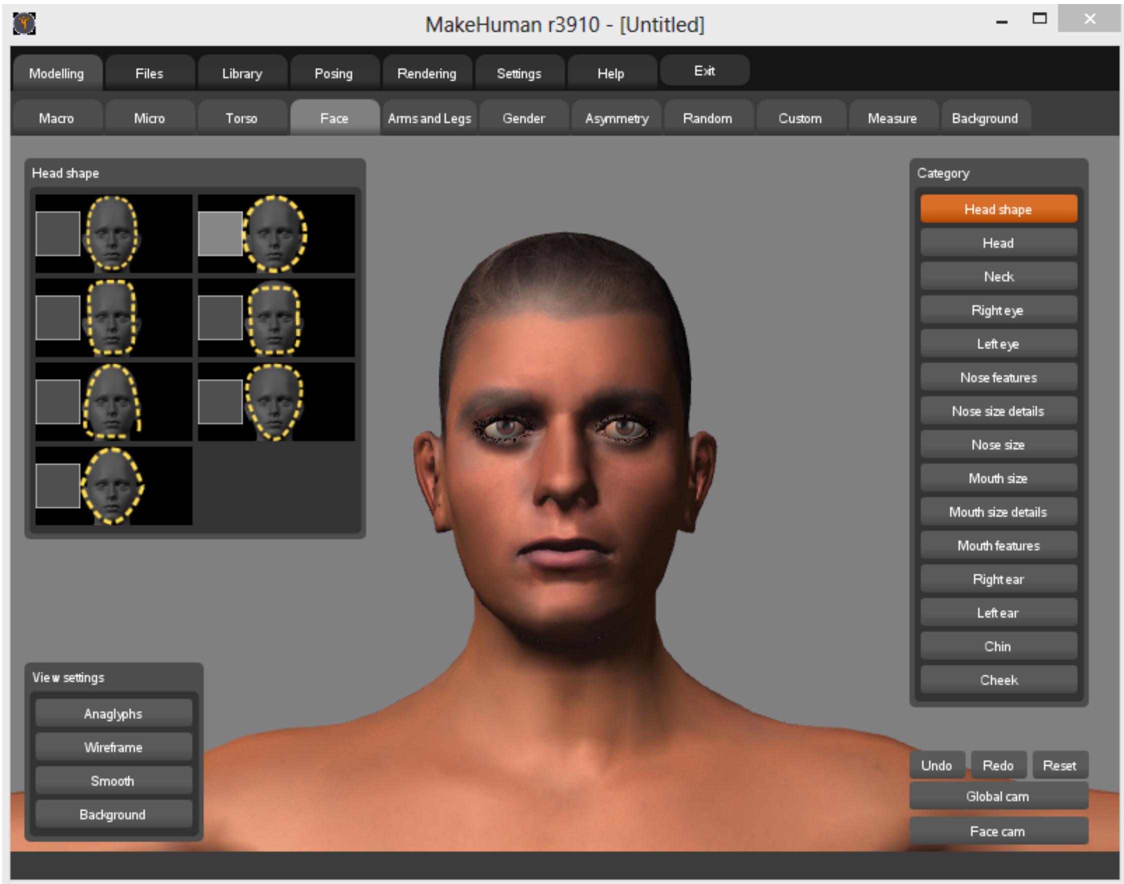


Figure 8: MakeHuman User Interface

the view port should also be locked in the frontal position.

Participants were exposed to three consecutive target pictures of human faces from www.facity.com. Facity provides a free image database of human faces that have no copyright or royalty related clauses. We selected our three images from this database. Figures 9, 10, and 11 depict the three target images used in our study. We made a few difficult choices before deciding upon these three targets.

Originally, we planned to use two male and two female targets. However, since we estimated participants would require 15 minutes to complete a composite, we decided to limit the number of targets to three. We believed that an hour might be too long for participants to be using our system and fatigue may invalidate our results beyond this point. Using only three targets would have made it difficult to design a strong study

Figure 9: Target One [3]

around male and female target images. Each participant would have received two targets of one gender but only one of the other. Reducing the number further to two would not have solved the problem either. Although we intended to record data for all targets, we expected that the first image might be used as a "practice" image, in which participants did poorly as they familiarized themselves with the MakeHuman system.

Furthermore, possible male targets tended to have facial hair, which obscured the bone structure participants were attempting to recreate. Since our participants' task was to replicate facial structure and not appearance, this was in serious conflict with our goal. These two factors led to our decision to limit our study to three female targets.

At the time of this research project, the current version of MakeHuman does not support different skin

Figure 10: Target Two [3]

tones or hair. The original female targets we chose were Caucasian and of a light skin color. Because the skin color in MakeHuman was locked as a default setting to be darker, our participants would have been forced to attempt to create a composite which would have appeared fundamentally different than the target. Although the participant might have been successful in recreating the facial structure of the target, they could not replicate the skin tone. This was determined to potentially be very distracting to our future participants. We chose to select darker skinned female targets from Facity.

One advantage of using pictures from the Facity database [3] was the format of the images. All photographs are cropped halfway up the forehead, include the full width of the face and down to the base of the neck. This was important for our study because it allowed for some uniformity between target images. This made it much easier to make a claim about progressive differences from the participants experience with the first target to their experience with the second and then the third.

Finally, participants were also given a questionnaire with four questions. The first three questions were

Figure 11: Target Three [3]

the similarity assessments between each composite they produced and its associated target on the seven point scale. The final question was a rating of their ease of use with the software package used in their condition. This scale was also out of seven points.

2.5.4 Procedure

Participants were asked to create a total of 3 composites. The targets were randomly chosen from the three target pairs described in the materials section above. The participants were randomly assigned to use either the parametric facial avatar generator or MakeHuman's facial avatar generator. Participants were allowed as much time as necessary to recreate each target image. The number of modifications made by participants was logged by our software. The amount of time required was logged after each completed avatar. Before beginning to design the composites, participants were given the following instruction in both hard copy and spoken orally by the experimenter:

> You will be asked to create three composites using facial design software. For each target picture, we ask that you try to complete your design both quickly and accurately. When you complete one design, please notify the experimenter and and he will present you with the next target image. After each of the three designs, you will be asked to rate your creations for similarity to the original target. Additionally, we will ask you to indicate the ease or difficulty of your experience with the facial design software at the end of your participation.

The participants were given the seven point scale for similarity and another for realism upon completion of both composites.

2.6 Method: Judgment Phase

2.6.1 Participants

In the second phase of the experiment, 13 (3 females, 10 males) participants were self-selected by responding to flyers posted in the campus center and dormitories at Union College. All participants opted to participate for a cash reward of 4 dollars.

2.6.2 Design

Participants rated the similarity of composites made in the creation phase to the target images. These participants were asked to rate the composite images in order to determine how successful the participants in part 1 actually were in creating similar composites to the original target.

2.6.3 Materials

Participants were presented with a 48 page slide show on a computer. Each slide contained the seven point similarity rating scale, a composite image, the composite's target image and a randomized set of letters for later identification. Figure 12 depicts an example of a slide from the slide show presented to our participants.

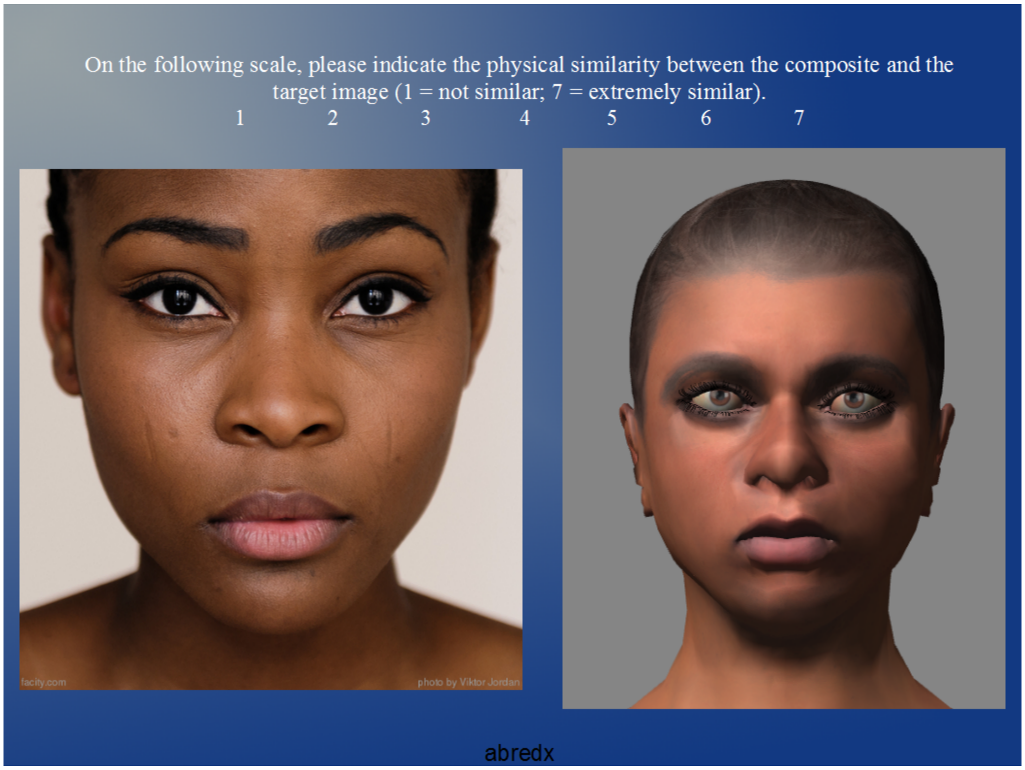


Figure 12: Example Slide

It should be noted that we did randomly change the order of the slides in the slide show to prevent any ordering effects. However, we always made sure that the sequence of composite/target pairs remained the same. The sequence was a loop of the first, second and third composite target pairs. In other words, the order of when any given pair was presented may have changed, but the order of targets displayed was always looped in the same order to all participants. Additionally, in the sequence of composites of target one, target two and target three, the authors of each composite may not have been the same. We only ensured that the sequence of targets remained constant.

2.6.4 Procedure

We drew a new sample of participants for the Judgment Phase. These participants were presented with the slide show previously described. Participants were asked to go through each slide, in order, and complete the similarity scale assessment between the two images presented on each slide. Each slide (see Figure 12) contained a target avatar and a composite avatar. The target avatars were the same targets used by participants in the the creation phase. The composite avatars also came from the creation phase. These were the composite images the participants created.

The researchers had the participants rate the similarity between the created avatars and the targets in order to determine how successful each participant in the creation phase had been in creating a composite. In total, the participants made 48 similarity assessments between a composite and target image, three for each of the 16 participants from the previous phase.

During the design of the judgment phase, the experimenters made multiple revisions of the procedure before settling on the previously described process. Originally, we considered asking participants to also give realism assessments in addition to similarity assessments. We were curious to see if there was a difference between similarity and realism. Although we knew that there would certainly be a correlation between realistic composites (those that looked human) and composites that were similar to the intended target, we were not sure of what differences there might be between the two. However, we ultimately decided not to include this measure in the current study, because we did not feel that the current system was not yet strong enough to warrant a second assessment of success. We felt that a participants assessment of similarity would be partially based upon whether or not the composite was simply human looking. Because of this, we felt that including the extra realism assessment would be redundant and a waste of our participants' time until a more effective and refined system is created.

The experimenters also considered having participants give similarity assessments between composites, not just between composites and the intended target. This assessment would have allowed for the experimenters to determine the variability between participants' composites. This information could have been useful in the interpretation of our data. If it happened to be that all the composites were identical and yet they were not particularly similar to the targets, we would have been able to deduce that our software package needed to allow for a either greater degree of variability or fine-tuning in the modifications users were able to make. In other words, the degree of variability could have indicated the degree to which our system could create different results.

We did not choose to have our participants make these extra assessments, however, because it would have required a large amount of their time. For just one of the three targets, participants would have had to make comparisons between all 16 of the composites created. This would have amounted to 136 separate comparisons. In order to get multiple ratings, we would have either needed to get a larger number of participants or greatly increase the number of comparisons our participants would make. We felt it may have been detrimental to our data to increase the number of comparisons participants made because they might "burn out" towards the end and start to produce unreliable data. Ultimately, we also decided that the amount of extra information that this addition to our study would contribute would likely be worth less than the cost in time to our participants.

2.7 RESULTS

In the Creation Phase, we were most concerned with three different measurements. We were interested in the amount of effort it took a participant to create a composite (measured in both time and the number of modifications made), how similar the participant believed his or her composite was to the intended target and also the relative ease of use expressed by the participant upon completion of our study.

First we will investigate the amount of time participants used to create a composite (See Figure 13). For both the first, $t(13.74) = -0.567$, $p = 0.58$, and the second target, $t(7.89) = 2.17$, $p = 0.062$, we were unable to find a significant difference in the amount of time participants required to complete a composite. However, when participants were creating a composite for the third target, participants used significantly less time than did the participants in the control condition, $t(7.7) = 2.54$, $p = 0.03552$. After analyzing the data, we decided to combine the data from the second and third targets. We found that participants in the experimental condition were significantly in creating composites for the final two targets than participants in the control condition, $t(17.5) = 3.41$, $p = 0.004$.

We performed the same statistical tests to determine the difference in required effort as defined by the number of modifications performed by our participants (See Figure 14). For the first target, we did not find a significant difference between participants in the experimental and control condition, $t(9.5) = 0.624$, $p = 0.547$. We found the same result for the second target, $t(7.78) = 1.99$, $p = 0.083$, as well as for the third,

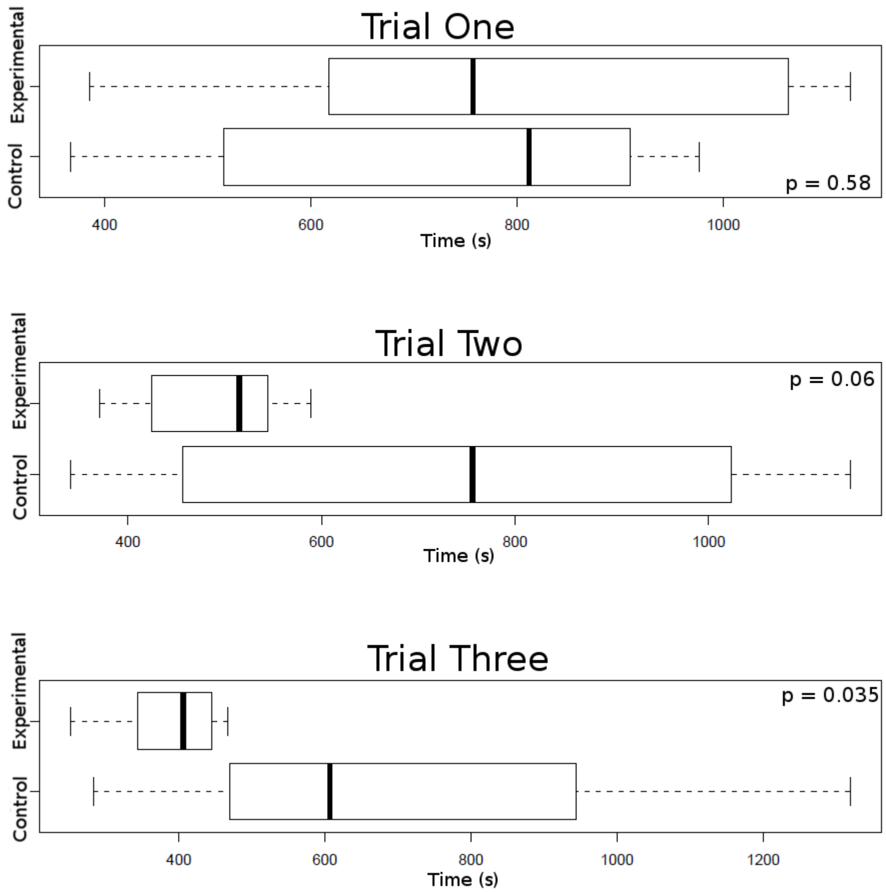


Figure 13: Creation Phase: Time to Completion

t(9.02) = 2.03, p = 0.072. It is worth pointing out, however, that there may have been a significant result found if we were able to have increased our number of participants for the second and third targets. This also led us to combine the data from our second and third targets. We found that there was a significant difference between the control and experimental participants in the combined and that experimental participants used significantly less modifications to complete their composites, t(17.9) = 2.88, p = 0.009.

To evaluate the perceived similarity between a creation phase participant's composite and the target, we performed similar tests using the self-reported scores on the seven-point similarity scale described earlier (See Figure 15). We did not find a significant difference between reported scores between control and experimental

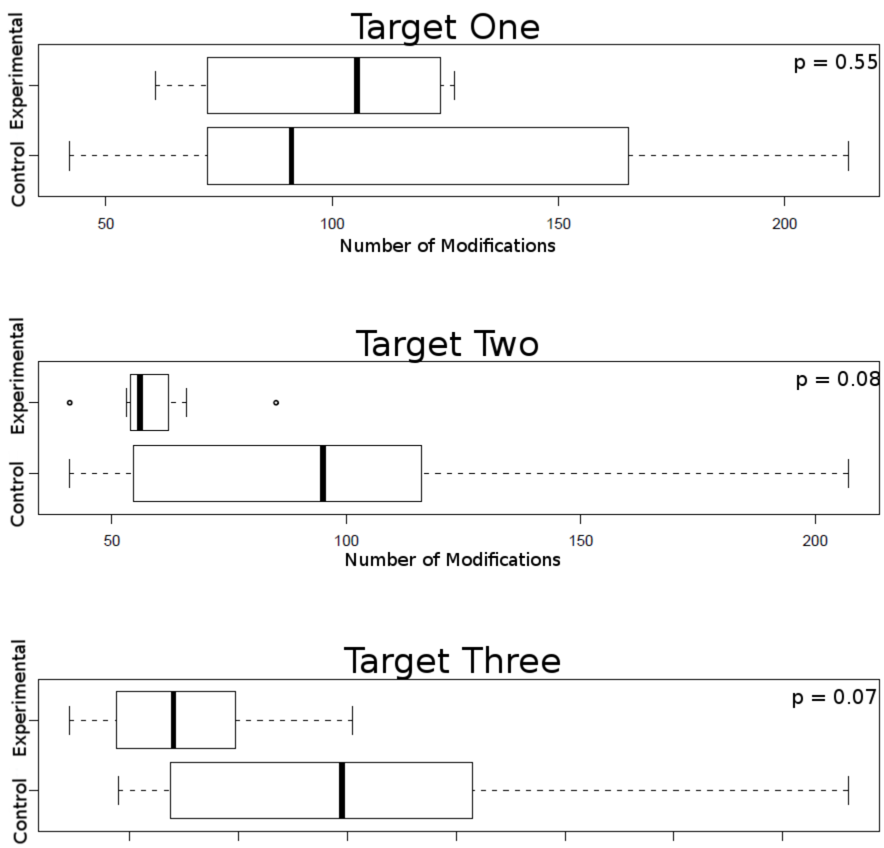


Figure 14: Creation Phase: Number of Modifications Made

participants for the first target, t(10.7) = 1.55, p = 0.15. We did not find a significant difference between participants scores for the second target, t(13.95) = 0.46, p = 0.654. Similarly, we did not find any significant difference between participants' self-reported similarity scores for the third target either, t(13.8) = 1.26, p = 0.227.

Finally, we were also unable to find a significant difference between the experimental and control participants' self-reported ease of use with their respective software packages, t(12.29) = -1.11, p = 0.288 (See Figure 16).

In the Judgment Phase of our study, we were only concerned with evaluating the similarity of the com-

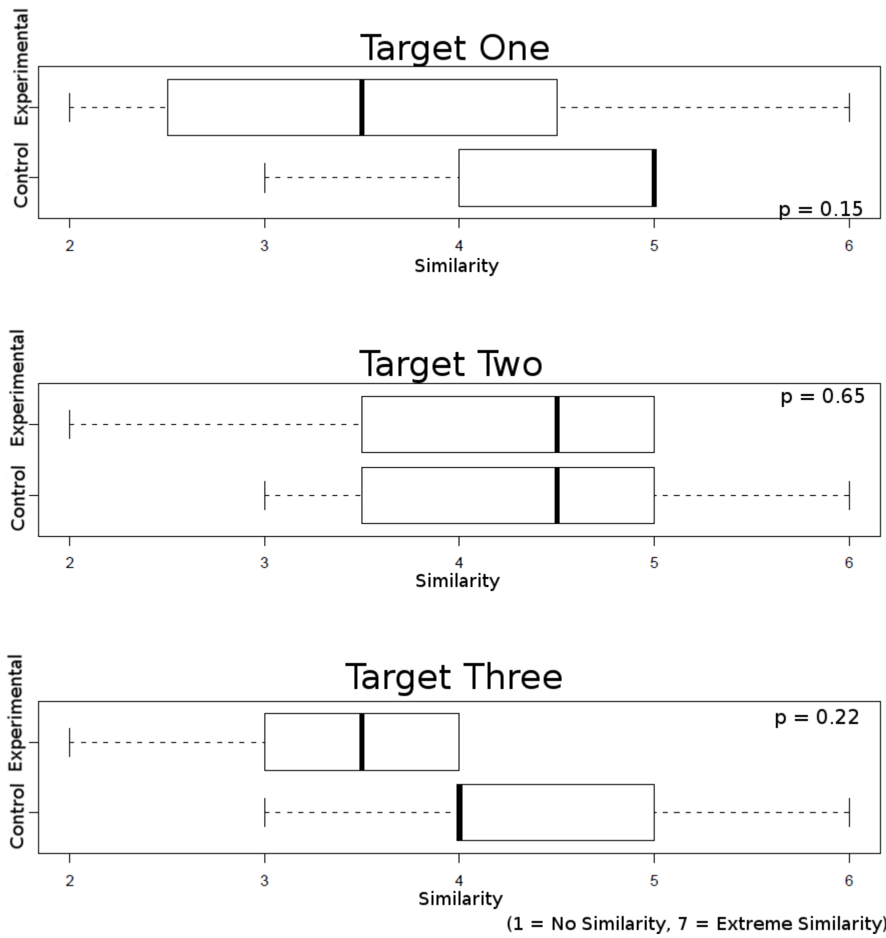


Figure 15: Creation Phase: Self-Score of Similarity

posites created by participants in the creation phase with their intended target images. This was done with the same similarity scale as described earlier. First, we analyzed the data from ratings of the experimental and control composites as a whole. We were unable to find any statistically significant differences between the two groups, t(612.5) = -0.31, p = 0.75. Next we tested for differences between the composites as made for each target based upon their being authored by a participant in the control or experimental group. For composites of the first target, we did not find any significant differences in similarity ratings, t(203.5) = 0.522, p = 0.602. For composites of the second target, again we did not find any significant differences, t(205.1) = 0.13, p = 0.895. Finally, for the set of composites of the third target, we did not find any signif-

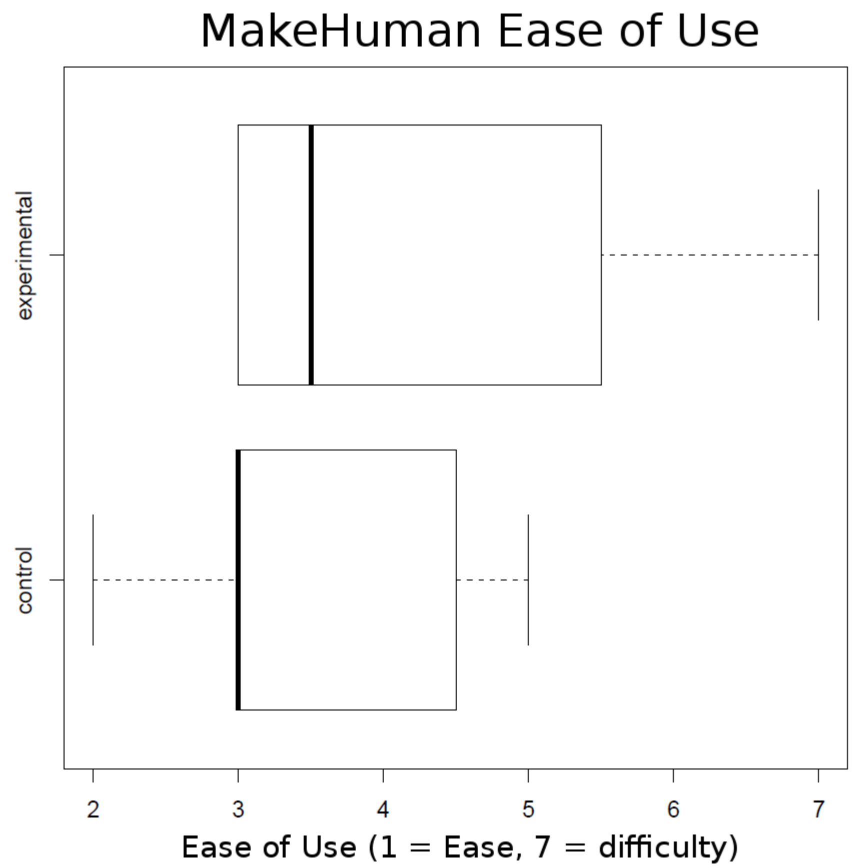


Figure 16: Creation Phase: Ease of Use

icant differences, t(199.0) = -1.21, p = 0.224. We decided to combine the data from the second and third targets as we previously did for the data from the Creation Phase. Again, we were unable to find any significant differences in the similarity of composites produced by participants in the control and experimental conditions, t(406.5) = -0.72, p = 0.5 as shown in Figure 17.

2.8 DISCUSSION

Our results indicate that the current system does not produce significantly more similar composites than our control version of MakeHuman, we believe that with future improvements, however, we will find significant

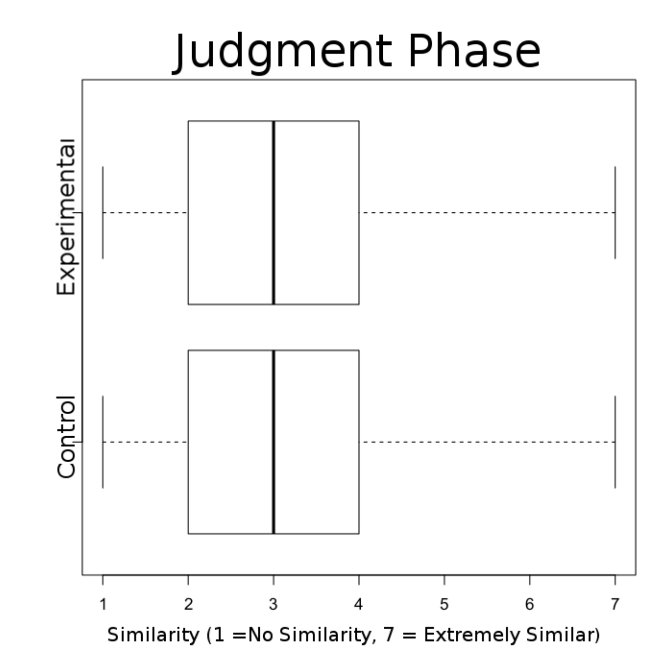


Figure 17: Judgment Phase: Similarity Scores

advantages with our system. It is interesting that we found our system was significantly faster, but not more similar. The increased speed with our system is most likely due to the fact that it allows participants to modify multiple features simultaneously. Although this clearly increased speed, we believe it may have had mixed effects on similarity. That is, in some ways it was beneficial, but in others, it may have actually been detrimental the the similarity of the composite to the target.

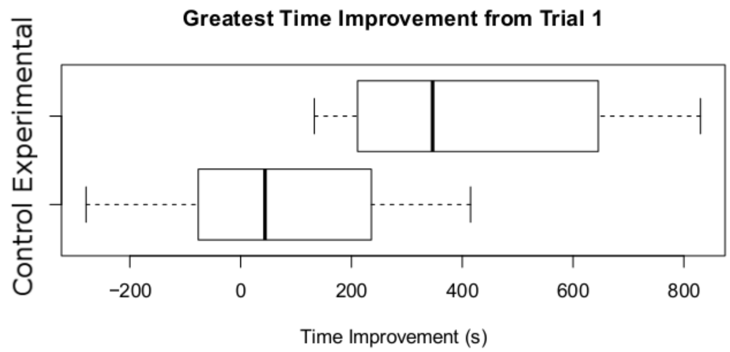
Before investigating our results further, we would like to point out the possibility that because our implementation enabled users to complete composites faster it may also mean that with future improvements to the algorithm, our system may produce more similar composites as well. We believe that this is indeed the case. In this discussion, we will explain why we believe the current system still warrants future research, even though we did not find a significant improvement in similarity under the current conditions. The reason our system was not able to produce more similar composites was due to disadvantages experienced by our experimental participants (mostly with the undo feature) and advantages experienced by our control participants (mostly due to side effects of our implemented algorithm involving forced symmetry).

Let us first touch upon the purpose and motivation for this research once again. Since we are trying to design a system that will allow a user to quickly and accurately design a composite of a human being after some target image that they have in their mind, we should discuss what our actual goals are and how this project helps to realize them.

Many of the choices we made in our research project were due to the fact that that this study was, as far as we know, the only one thus far to attempt to design a system for composite creation using digital avatars. Because it was the first study, we chose to limit our focus to keep our study very simple such that extraneous factors (which may be interesting) would not dilute the data we gathered. One important flaw we would like to acknowledge that we hope will be corrected in the future was the focus of our data collection. Although we believe it was important and best that we designed our experiment the way we did, in the future, the focus should not rest upon the creation of composites for separate targets. We used three separate targets in our study in order to gather more data (as compared to if we only had each participant create one composite).

In our three-target design, participants in the creation phase were asked to design composites consecutively for multiple targets. In a real-life application of our software, this would not be the case (unless there were three suspects). Because of this, the focus of our study should have been on the first composite made by each participant. In a real setting, we do not want users to be required to first sit down and make practice composites before providing the police with a final product. However, we expected that there would be a significant learning curve with our software because the interface can still benefit from improvements. The more intuitive we can make the interface and the more control we can give the user over our software, the more effective this software will become.

If one looks at the data chronologically in terms of the order in which the participants created the composites one will find that participants became increasingly faster with our version of MakeHuman (See Figure 2.8). Although this data is compelling, it is unfortunately not supportive of our ultimate goal in this project. While the data may indicate that participants are better able to learn and improve with our system than the original, this was not our goal. Our goal was to enable users to create composites for a target they have in their mind. In a real-life setting, we do not want users to need to get all of this practice in order to eventually be able to produce an accurate composite. We want a witness to a crime to sit down immediately after the event and without any previous training, quickly produce a composite for the police to utilize.



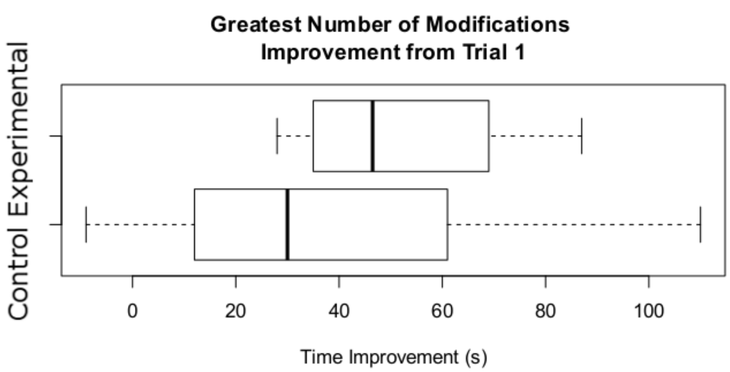


Figure 18: Improvement Across Targets

To better interpret our data, it is important to recognize how our system was assisting users. For many features, our algorithm was enforcing facial symmetry. We should note here that our system did not actually enforce "symmetry," as much as it enforced "feature value symmetry." That is to say, our base model was not created through a mirror image. There are very minute differences between the right and left sides of our base model. However, our system used symmetry by making the same relative changes to the left side of the body as it did to the right side of the body. For example, if the right eye was changed, this same change was applied to the left eye (and vice-versa) such that the eyes would always be nearly mirror images. However, since the base model was not perfectly symmetrical to begin with, even with enforced feature value symmetry, the model is unlikely to actually be perfectly symmetrical at any one point in time. We believe our algorithm was most beneficial in the early stages of a composite's design because it would allow the user to get the most obvious characteristics set in the face. However, we know human faces are not perfectly symmetrical. Although our system did not force symmetry, it did not allow the user to specify dissymmetry either. The user was unable to make small changes to one side of the face that would not be present upon the other side.

We believe this may be partially why our participants were able to produce composites faster in the experimental condition. For features in which symmetry was enforced, our experimental participants only had to perform half of the work to create the same result as participants in the control condition performed. However, this also sheds light on a weakness in our experiment. A consequence of the enforced symmetry was that participants in the control condition may have actually had an advantage in creating a more similar composite. Although they may have required more time, participants in the control condition were not limited by the enforced symmetry rules in the experimental condition. Without enforced symmetry, these participants were able to create composites that did not have perfect facial symmetry. In this regard, they may have been able to produce a more realistic and similar composite.

The images in Figures 19 and 20 depict examples of the differences in how images can differ from the target image. Although neither of these images are highly accurate, they do help to illuminate the pros and cons of our system. From Figure 19, we can clearly see that participants in the control condition are able to create images which look very strange. They are not given symmetry enforcement or assisted with shaping the different facial features. Figure 20 displays an image created with the experimental version of MakeHuman. It is interesting that the images created with our new system tend to have similar mistakes. We believe a lot of the errors made by our participants in the experimental condition stem from either the algorithm forcing them to make modifications they do not want to make or that they can not create composites without symmetry. Figure 21 depicts an avatar that was considered to be of higher similarity in our study.

Although the control participants may have had this advantage, they did not have the advantage of the algorithm assisting them in real-time. Enforcing symmetry was only one of the ways in which it assisted

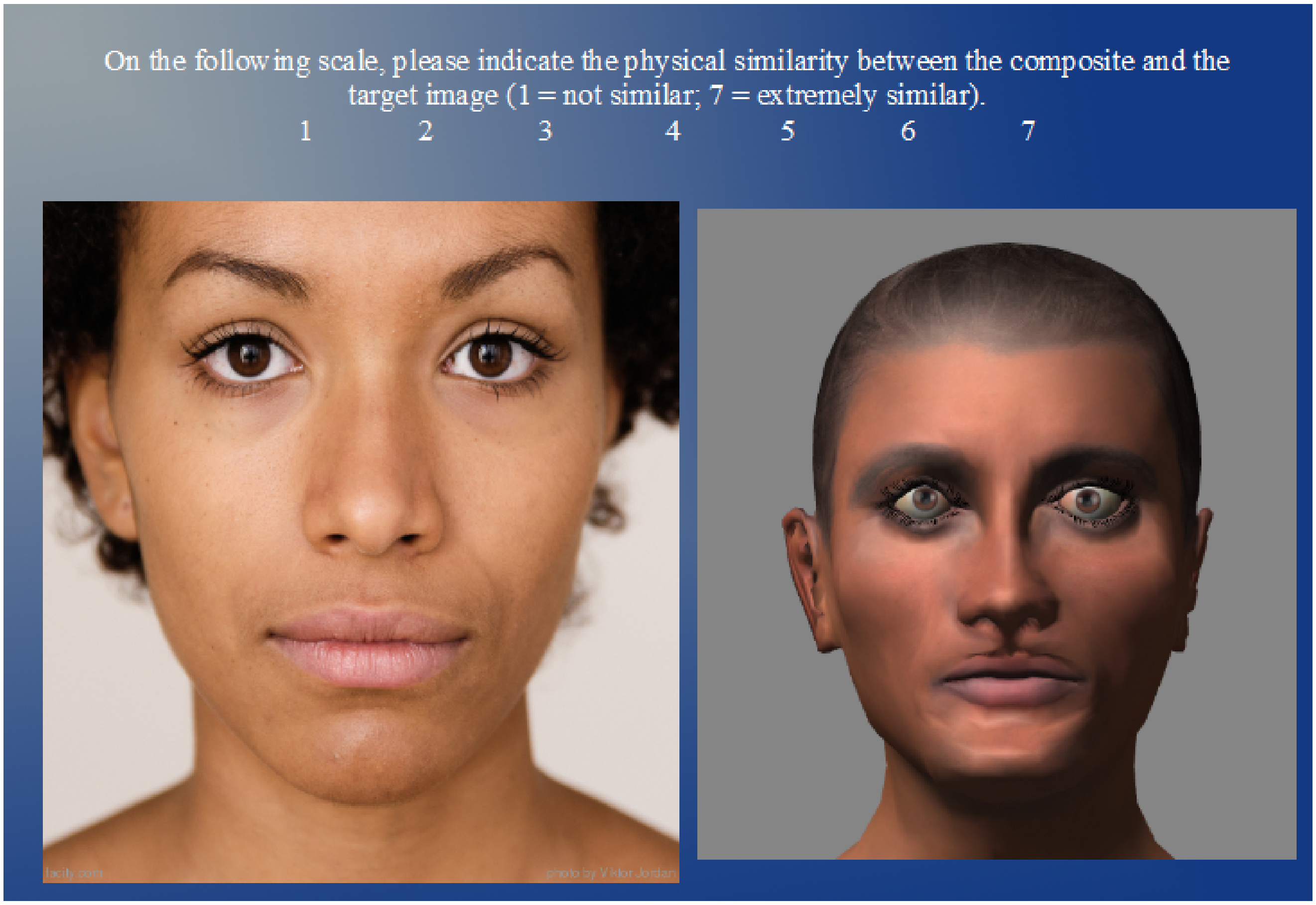


Figure 19: Example of Low-Similarity Control Composite

our experimental participants. Although our main focus was in the creation of a framework to support an algorithm, it is worthwhile to examine the initial algorithm we introduced in this study. The algorithm we implemented was intended to restrict the facial features to remain within what we considered to be "humanly possible." This is an important point to understand, because the bounds of what we considered "humanly possible" were arbitrarily set by us. We set these bounds by making modifications to the avatar in MakeHuman and determining at which point the avatar no longer looked "natural." Once the avatar reached an unnatural state, we noted the degree of modifications we had made and adjusted the algorithm to not allow the current combinations of physical features. We enabled our algorithm to avoid such combinations by

Figure 20: Example of Low-Similarity Experimental Composite

setting general percentage based rules for relationships between features. Specifically, we created a "sweet spot," a range in which our algorithm would make no modifications (except for symmetry enforcement) beyond what the user chose. However, once the user made a change that moved the state of the avatar outside of this sweet spot, our algorithm made additional changes to the necessary features to bring the avatar back inside what we considered to be humanly possible. We implemented the algorithm such that it would always allow the user to make any change he or she wanted to make. In other words, we assumed that any change the user made was intentional and correct. Our algorithm never modified the change most recently made by the user, but was free to modify any other features as necessary.

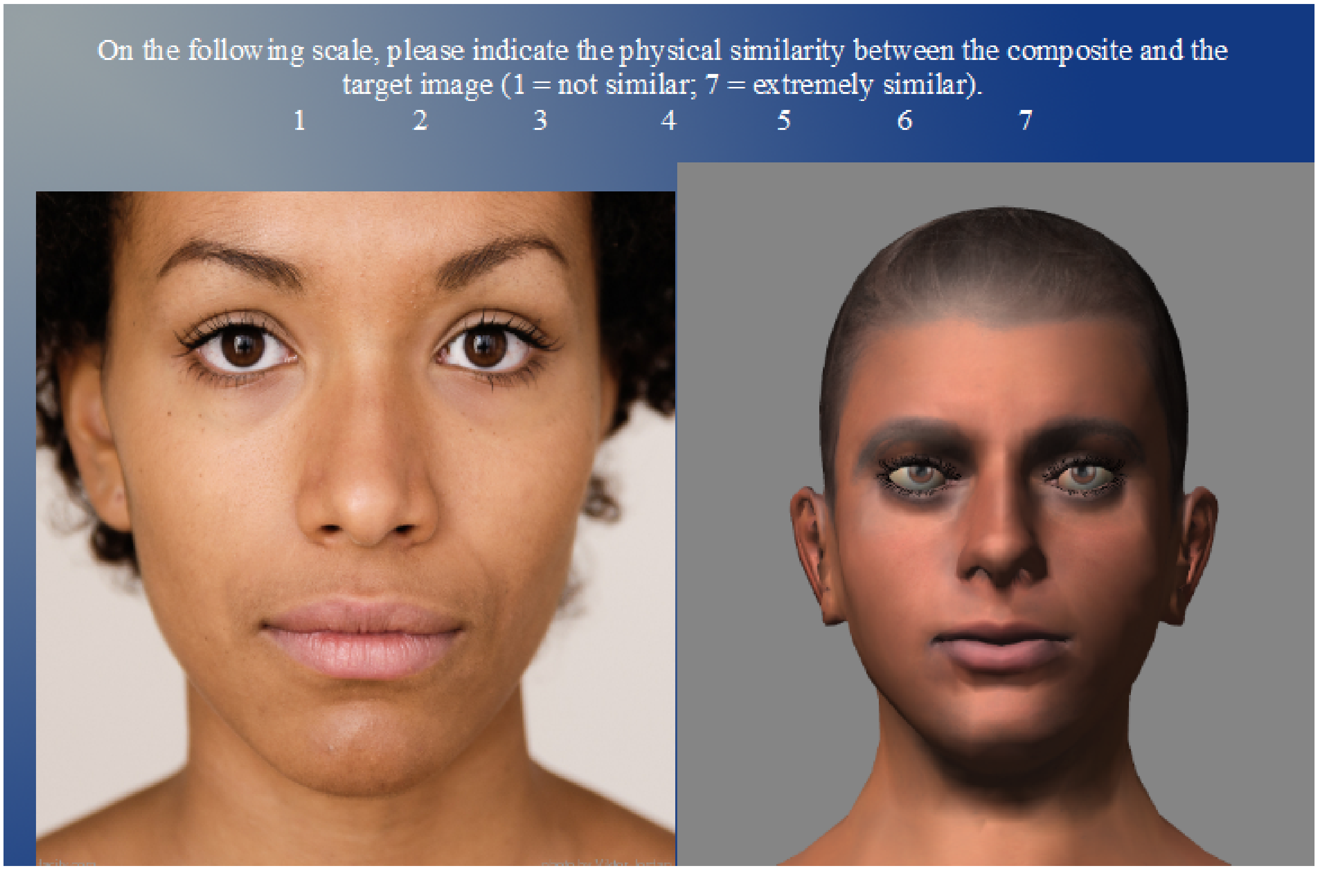


Figure 21: Example of High-Similarity Experimental Composite

As we previously discussed, our system locked the view port to a frontal view while a modification was taking place. Although this decision was justified for our system (it mimicked the frontal view of our target images), we believe this was not the most realistic scenario for our participants. Outside of the laboratory, humans do not witness a crime (or simply meet new people) through a photograph. In real life, humans use their own eyes to see an actual person. We did not choose to design our study this way however because using a living target would require that we had three individuals ready and present for all of our participants in the creation phase. Since each participant required approximately 45 minutes to complete our study, we did not believe it was worth the time of the hypothetical targets (or the cost in our funding). Another problem with this approach would be that all of our participants would not get the same exact stimulus from which to create their composite. Yes, all participants might see the same people, but if one of our targets became sick, changed his appearance, or was forced to drop out from our study, we would have had a very serious design problem. A second alternative would have been to utilize a pre-recorded video of our participants performing a task (or committing a mock-crime). This was a strong possibility which we considered. It would have greatly increased the realism of our study. This design choice would have enabled us to make strong claims about how our system is actually used in real life. This gets at one weakness of our study. Our design has a low degree of realism and our experiment clearly takes place in a lab setting. However, we believe this was a good decision.

Because our study was a very early investigation into the use of three-dimensional parameterized avatars for the purpose of designing human physical appearance, it was not previously known how successful we would be. We did not want to complicate our study further with possible differences that could arise outside of the laboratory. We chose to maintain a high degree of control in our study at the expense of realism of the scenario. We chose to use a still frame image as opposed to a video because it does not change. For the first test of our system, we believed it was important to have as much control as possible over the experiment. Using a still image allowed us to give participants the equivalent of perfect memory. As described in our introduction (see Figure 1 and 2) a large part of the error in the composite creation process comes from the memory of the individual attempting to recall the physical appearance of the target. We chose to completely eliminate this source of error. In our study, we wanted the only error in similarity to come from the participant's ability to create a composite.

In the most life-like scenario, our participants would have been presented with the target stimulus and then asked to create a composite after we removed the stimulus from their view. Obviously this would have made it difficult to limit the amount of error stemming from participants' memory. However, this does not fully explain why we did not use a video which the participants could choose to replay over and over while they designed a composite. We felt that this could potentially cause the same problem because the participant would need to constantly replay the video in order to find the right vantage point to best see the features currently being modified. We felt this could create a similar problem to what would occur if we simply took the target stimulus away during composite creation. Even though the participants could replay the video, there would be times when they would be forced to recall the targets appearance from memory. Because any one view point does not allow the participant to see both the right and the left side of a target's head, the use of memory would still be (sometimes) required. To avoid this, we felt it was best to simply lock the view in place and choose target images which would compliment the frontal vantage point used in our interface.

Regarding our target, another difficult choice we faced we faced was in the participant's ethnicity as compared to the target's. When we designed our experiment, we originally planned on using Caucasian targets since we expected most of our participants to be Caucasian. However, the original MakeHuman software package does not yet implement skin tone. The default skin tone that we were forced to use was darker than most of our participants and that of our Caucasian targets. We felt that this could be a problem in the use of our system. Even though skin tone was not the focus of our study, our participants were being asked to create composites of the targets we presented. Although the only tools we gave them allowed them to manipulate facial structure (not skin tone), we felt that it would be natural for our participants to be distracted by the discrepancy between target and composite skin tones. To avoid this problem, we chose to only pick targets in the Facity [3] database which were of darker skin.

One negative consequence of making this decision, however, was that now we were using dark skinned targets and most of our participants were Caucasian. Although we believe this was better than the alternative of requiring participants to make dark skinned composites of light skinned targets, we still believe it is a weakness in our study. In future implementations of this study, it would most likely be beneficial to choose new targets that will match the ethnicity of the participants. This will require, however, that the MakeHuman system be updated such that the skin tone of the base model can be manipulated to match the targets provided.

The fact that there were both beneficial and negative aspects to the system we implemented makes it difficult to deduce how effective our system truly is. Although we are able to say with certainty that our system allows for faster use, we were unable to make positive claims about creating more similar composites with our system. However, we can infer from our data and participants' anecdotal evaluations after completion of our study where we might improve.

Many of our participants expressed frustration with the way our version of MakeHuman allows users to undo modifications. Currently, the undo does not take the modifications made by the algorithm into account when undoing past actions. This was very frustrating for participants in the experimental condition because if they wanted to undo the changes made by the algorithm, they would have to do it manually. We do not believe this to be a flaw in our system, however, as we hope that this issue will be addressed in future work. We believe that an entirely new system for handling different types of undo will be required for the types of algorithms that are likely to be used in the future.

Some of our participants also expressed that the algorithm we employed would, in some cases, hinder them from completing the task. For example, when a user moved the right eye towards the outside of the face, the face would eventually become wider such that it could still contain the eyes. Because our system was making the head wider, it would also slightly increase head height in order to maintain the ratio of head height to width. However, this was not always what the user wanted. In some cases, it surely was, but in others, the user may have only wanted the head to increase in width. We believe that our proposed new system for handling undo will solve this problem in the future.

2.9 Future Work

In the future, we believe there is important work to be done in order to improve our system, both in the framework and the algorithm employed, and also changes in the methodology used to evaluate our system in order to increase our knowledge in this field of research. While we found in our study that participants were enabled to produce composites faster, but not necessarily more similar to the target, it may be the case that further improvements to our system will illuminate how the current system is in actuality a better system for creating composites. In future studies, we may find that if it were not for the drawbacks previously described in the discussion, this system could be both faster and more accurate.

We believe the most profound improvement for our system will come from the development for a new undo system. We chose not to modify the current undo feature in our system such that it would undo both the user's and algorithm's changes because we felt this would not be adequate for our purposes. Instead, we believe that a new system should be developed that will be tailored specifically for this project.

When a user wishes to undo a past modification, there are multiple facets to this task. First, the user will want to undo the last change they made. Secondly, they will want to undo the changes the algorithm made when trying to assist them. However, this is only the most basic case. The user may find that the change they made was correct, but the changes made by the algorithm had a negative effect. Or, the user might discover that only a few of the changes made by the algorithm were unwanted. The best system for this problem will be one that will allow the user to pick and choose what changes are helpful and which modifications should be thrown out. Ideally, this could be realized with two new frames added to the user interface. In one frame, a list of past modifications made by the user could be displayed. If the user clicked on any of these past modifications, the second frame would display a list containing the change the user made and all of the relevant changes the algorithm made for that change as well. Once the user is viewing this list, the user could theoretically then remove the unwanted modifications, independent of whether they were made by the user or the algorithm.

This implementation would allow the user to create dissymmetry in a composite by making a change and then removing the symmetry enforcing modifications made by the algorithm. Similarly, the user could also remove only the "unnecessary" modifications made by the algorithm as well. However, we do not believe it will be necessary for our future users to rely upon the undo mechanism in order to produce similar composites. It is only a tool through which to finely tune the composite toward the intended goal. Once this new system is ready, we envision a new methodology for best use of our system. Initially, we believe our algorithm was very helpful to our participants. The enforced symmetry saves a lot time for users as they attempt to get the facial features as close as possible to the target. However, we believe there will come a point when enforcing symmetry is no longer beneficial. The first solution that a user could benefit from is to utilize the undo mechanism to undo any unwanted changes. However, presumably the user will eventually come to a point when the composite no longer needs symmetry to be enforced for any facial features. This will occur once the only changes to be made to the composite are those which involve creating differences between the two sides of the face.

At this point in the design process, we envision a new feature will help to improve our system. We believe users should be enabled to modify the algorithm assisting them. Once the user does not want facial symmetry to be enforced, the user could simply turn off symmetry enforcement, allowing him or her to make fine grain detail changes on either side of the face without it effecting the opposite side. Implementing this feature would allow the algorithm to be more flexible based upon the user's current goals. As we previously discussed, we believe that a large factor in why participants using the experimental version of MakeHuman did not produce significantly more similar to the composites was because they were not able to produce dissymmetry in their composites. We believe that making this change will allow users to exert more control over the assistance given by the algorithm which will in turn enable them to produce more similar composites.

We also encourage improvements to be made to our implemented algorithm. As we previously explained, our algorithm is far from complete. In the future, we believe an interesting study would be to keep the logic of the algorithm the same but instead to vary the bounds on which the algorithm acts. In other words, we believe the range in which the algorithm considers a composite to be within what is "humanly possible" should be altered. We recommend that one treatment in such a future study should involve bounds determined by human anatomy. Previously we explained that we determined the bounds used in our system arbitrarily. This may be a large weakness and a possible source of error in our system. It may be the case the we chose a poor range for what should be humanly possible. Instead, if this range could be set according to what ranges of human facial features are actually considered possible (within reason), we believe great improvements in similarity might be discovered. This could be a very important improvement because even if our system had the most effective algorithm possible working to assist users, if it was making decisions based on faulty knowledge of human anatomy, it would not produce as similar a result as a user working with an algorithm that had a more accurate database of what is possible in human anatomy.

As new features are added to improve the MakeHuman system, we encourage that these features will also be integrated such that the user may choose to turn them off if needed. We believe this is an important feature for a user to have because, in truth, human physical form can be very unique. Although there may be a general rule or pattern which human faces will follow, human beings are not all the same and there are outliers who appear very different from the average person. In order to enable our system to handle these cases, it is essential that our users are able to disable features which would normally prevent them from creating such different features. However, in general, it is our belief that this will not always be necessary.

Returning again to our user in the design process, after they have reached the point at which they must turn off facial symmetry in order to create differences on either side of the face (along with any other newly implemented future features), we believe that this may be the point at which they might choose to disable help from the algorithm entirely. This will give them free reign in order to make any remaining changes as necessary to their composite.

Although allowing users to disable algorithmic features may be beneficial, it is important to note possible drawbacks to this approach as well. In many cases, once a feature of the algorithm is turned off, it can not be turned back on unless the user also chooses to undo all changes up to the point when they actually turned off the algorithm's assistance. This is because features such as symmetry enforcement will not be able to work properly once they are re-enabled. If there are changes present which are not symmetrical, the algorithm will be forced to make a decision about whether the change on the left or right side is correct. This will simply be a 50 percent chance for the algorithm to choose what the user intended. This will likely be problematic for other algorithmic features as well. For example, one job the algorithm performed was to ensure that the eyes did not exceed a certain percentage width of the face. If the algorithm was disabled and the user made such a change, it would be allowed without any additional modifications being made. Now, assume more similar changes were made in other areas of the face. If the user were to now re-enable the entire algorithm, the algorithm would make a large number of modifications attempting to bring the composite back into the range of what it considers humanly possible.

One solution we initially considered to this problem would be that the algorithm only be allowed to act upon immediate changes. So the algorithm would not make a large mass of changes in this case. However, this would still not be a good solution because it would only delay the unwanted changes. When the user went to make a modification that had any effect upon a feature which was not in sync with what the algorithm would have expected according to the rules it enforces, it would then go ahead to make the possibly unwanted changes. Also, simply using the new selective undo feature will not fix this problem because the algorithm will be making decisions based upon choices it will not expect. This will cause the algorithm to make many choices which will not be helpful to the user, as the algorithm will be attempting to undo the changes the user has made. Because of this, we believe the best solution is to simply enforce that the algorithm can not be re-enabled by the user unless he or she first undoes all changes up to the point where the algorithm was disabled.

Future work on this project is not only limited to improvements upon our system, but also on the methodology we employed. Although we believe our methodology was well grounded and justified, we were unable to investigate certain topics in this area of research because there was not a yet a foundation from which to build from. As far as we know, there have not been any previous studies in this area of research which used a parameterized avatar generation/modification in order to investigate the design of human facial appearance. However, now that we have completed this first study, we hope that future studies will expand upon our work.

One area which could provide potential future research surrounds the targets we used in our study. In our study, our targets were chosen from the Facity [3] database. This database provided only frontal face images of human beings. As we previously explained, this was why we chose to limit the view in our user interface to a frontal view during modification. In future studies, this could be changed such that participants could be presented with a video in which to observe the target individual. Since a video will allow the participant to get a panoramic view of the target, it would be justified for the interface to be returned to its previous state which allowed the user to observe certain traits from the profile view when advantageous.

Also, an interesting new methodology for determining success would be to remove the target stimulus from our participants' view so that they would be forced to retrieve the idea of the target's appearance from memory. It may very well be the case that some implementations of composite design software are more effective in assisting users as they retrieve this information than others. If this were the case, these systems would be seen to produce more similar results in experiments with greater realism but they would fail to perform as well in experiments like ours, in which we help participants by simulating perfect recall by allowing them to view the photograph for the entire creation phase.

A final interesting point to explore in the future will be to only have participants complete a composite for a single target. As we previously explained, we believe that it is more realistic to have participants only create one composite because that is what would most likely happen in the real world. The only time a participant would need to create multiple composites would be if he or she saw multiple suspects committing a crime together. This change would have made our results less significant in the current study, but we believe that in the future, if the changes we have discussed are made, significant results may still be found.

3 Appendix: Mass Generation of Avatars According to User Specifications

3.1 Intent

In our original research project, we wished to examine how a system could large numbers of avatars according to a user's specifications. This study would have been the same as our current study except on a much larger scale. Instead of attempting to create a single avatar that looks like a target, the participant would have created a range of avatars (all slightly different from one another). We originally planned to use the software package Blender for this research project (See Figure 22).

Eventually, we moved on from this project to adopt a project with deeper roots in psychology, as described in the following section.

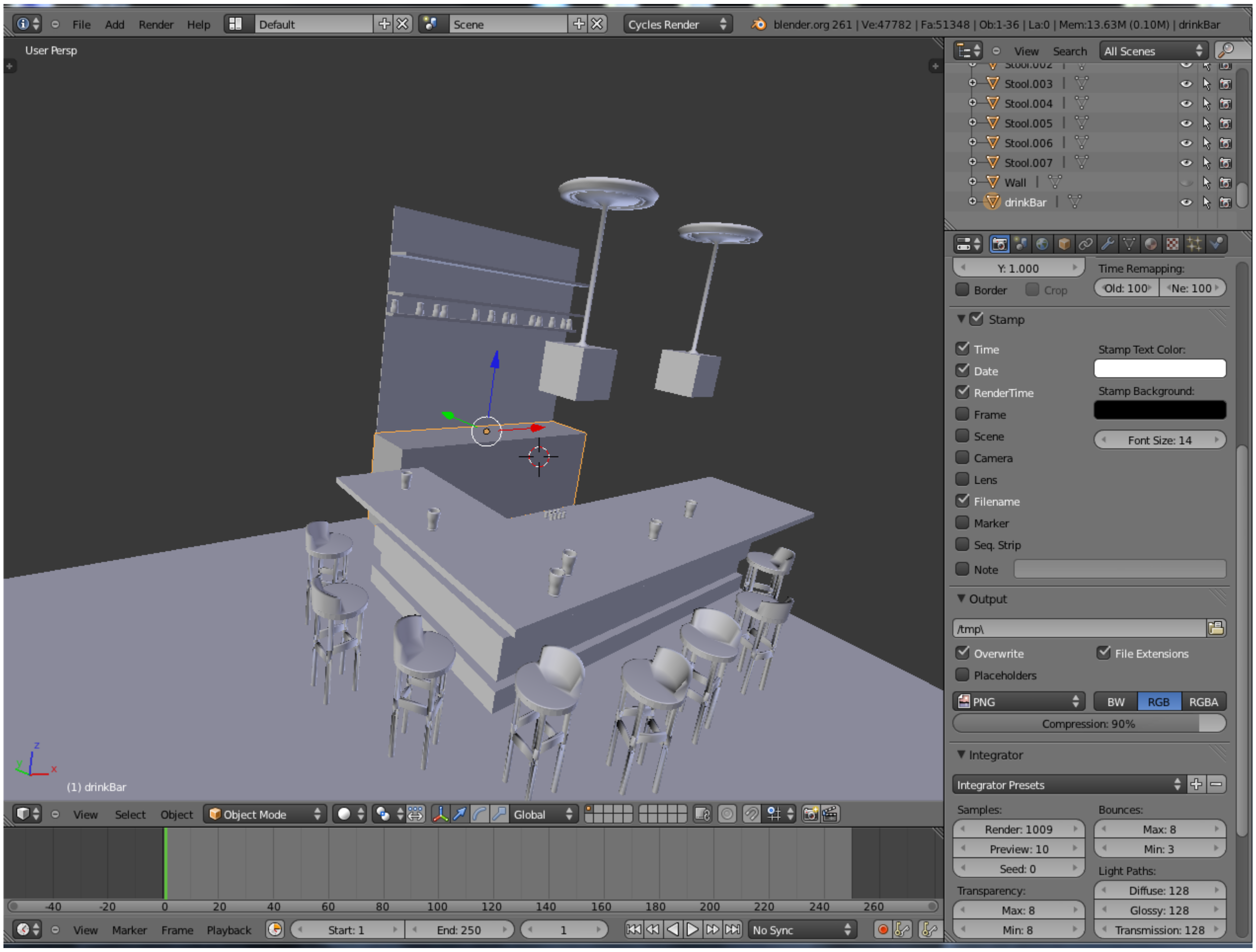


Figure 22: Blender User Interface

4 Appendix: User Creation of a Most Physically Attractive Avatar

4.1 Intent

We wished to investigate patterns of attraction in heterosexual males. We sought to investigate differences in what males physically find attractive in both long and short term mating contexts.

4.2 Related Work

4.2.1 Evolutionary Psychology

We approached this project from an Evolutionary Psychology view point. That is to say, we believed that the differences we might find and the results from past work in this area can be explained with the theory of evolution. Evolutionary theory claims that throughout human history, physical and psychological structures have been created through slow adaptations derived from genetic mutations which have ended up being adaptive to the organisms possessing them [5]. Over the millions of years of evolution resulting in the existence of the human species, slowly accumulated changes and adaptations have created what we now know as the human body and mind. Dawkins [5] details the example of the creation of the human eye as an example of this process. He explains how the first organisms did not have eyes, but slowly, over time, organisms were selected to survive by the environment as they came to inherit mutated genes which allowed for the basic approximations of a foundation for an eye. Over time, more complex features were added until the human eye, as we know it, now exists.

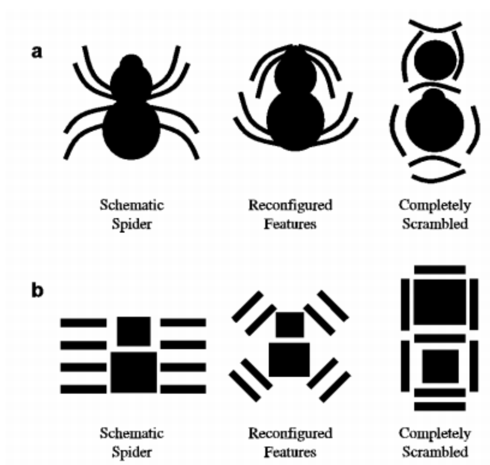


Figure 23: Spider Silhouettes[10]

The same has been true for psychological features as well. Over time, organisms who exhibited behaviors which led to survival tended to live longer and pass on their genetic information to later generations. It is not the actual psychological structures which are selected, but the overt behaviors which result from having such structures which natural selection is able to act upon. For example, researchers found that young infants spend more time looking at silhouettes of large spiders than other sillouetes comprised of the same shapes, but in a different arrangement, as displayed in Figure 23 [10]. The researchers theorized that this was because it is adaptive for infants to be wary of possibly deadly creatures and so more wary infants tended to survive throughout human evolution. Eventually, this resulted in human infants tending to be wary of spider silhouettes. The important distinction to make here is that it is not the act of being wary that was selected, but infants who were wary. Since these infants tended to survive more than others, they were also more likely to pass on their genetic material. While having the ability to think is adaptive, natural selection only acts upon the choices the organism makes as a result of being able to think, such as deciding avoid danger or not to go near the poisonous spider [11]. In other words, it is not enough to be able to think, as this does not make one fit for survival, instead, it is that the choices the individual makes as a result of thinking are productive and positive. Using the theory of evolution as applied in Evolutionary Psychology, we can investigate differences in physical attraction by males in short and long term mating contexts.

4.2.2 Short and Long Term Mating Strategies

Given that the most fit individuals will generally survive and reproduce, it makes sense that mating strategies should be important for reproduction. Furthermore, there may be different mating strategies for different situations [4].

Mating strategies can manifest themselves in various ways. It is important to recognize that these strategies are formed for specific types of scenarios and are most effective only for the circumstances for which it was most likely developed. For example, in figure 24, the advantages both sexes face in using a long or short-term mating strategy are explained. As we are focusing mainly on males, we will investigate further into some of the choices males must make.

Before an individual can make decisions about the number of partners to seek, one must first decide upon criteria for mate selection. The more stringent the rules are, the more difficult it will be for that individual to select a mate. One indicator used (by both men and women) has been the amount of time an individual has known the potential mate. Although the potential mate has most likely not changed a great deal over the period of time the individual is becoming acquainted with the new prospect, the individual will be likely

type of mating	men's reproductive challenges	women's reproductive challenges
short-term	• partner number • identifying women who are sexually accessible • minimizing cost, risk and commitment • identifying women who are fertile	• immediate resource extraction • evaluating short-term mates as possible long-term mates • attaining men with high-quality genes • cultivating potential backup mates
long-term	• paternity confidence • assessing a woman's reproductive value • commitment • identifying women with good parenting skills • attaining women with high-quality genes	• identifying men who are able and willing to invest • physical protection from aggressive men • identifying men who will commit • identifying men with good parenting skills • attaining men with high-quality genes

Figure 24: Short and Long-Term Mating Strategies[2]

to obtain more details allowing him or her to make a better evaluation of the potential mate's value. Figure 25 shows the differences between males and females for how long it takes one to become familiar enough with a person to decide to mate. Males and females both become more likely to consider having intercourse with a partner as they spend more time with a partner (although after one year, a male's likeliness to engage in intercourse does decrease slightly). However, males are initially much more likely than females to consider mating a possibility. This supports the notion that male and female mating strategies are determined by their mating goals.

If we accept that passing on one's genetic information is a goal of reproduction (although a subconscious one) these sex differences can easily be interpreted. The main difference between male and female reproductive goals stems from their physical differences. Females have an implicit cost in the reproductive process that males do not share. When a female becomes pregnant, she must bear a child for approximately nine months and will not be able to bear a child for a short time afterwards. Furthermore, bearing a child can bring a physical cost as well. Not only is it physically draining, but there may be complications which end in the death of the mother. Because of these added costs, it is in a woman's best interest to only bear the child of a highly valuable mate. Conversely, it is in a male's best interest to produce as many children as possible, since he does not have the implicit cost of being unable to produce a new child after conception occurs.

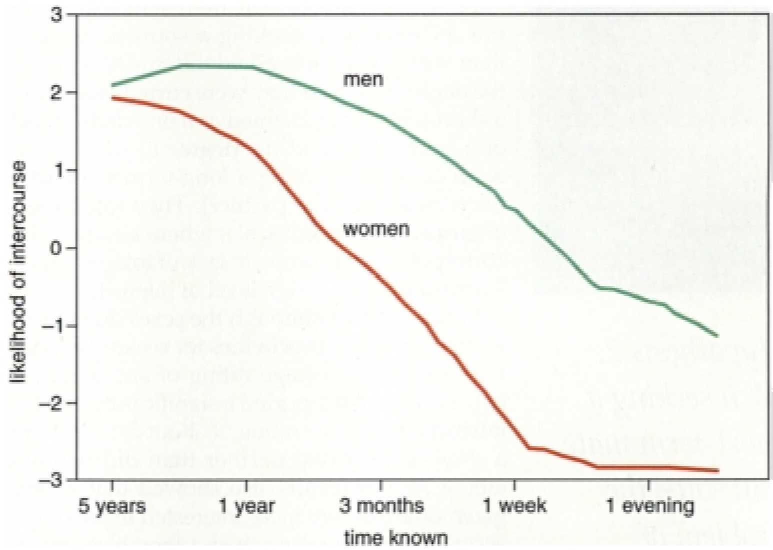


Figure 25: Time to Consider Mating By Gender[2]

Females spend more time evaluating a mate before copulation because they experience a greater risk than males do. If a male were to impregnate his partner, he could choose to leave and not help the female take care of the child. Although males do not share in this risk, they must still fulfill their own reproductive goals in order to successfully pass on their genetic material. Accepting that it is a male's reproductive goal to produce as many viable children as possible, we can better understand the short and long-term mating strategies for males [2]. In the short-term strategy, a male will attempt to mate with multiple partners. The weaknesses of this strategy are that the female may or may not be fertile at the time of copulation and also that the female's strategy is likely to be in contradiction to the male's. In the long-term strategy, the male will be more likely to have the opportunity to mate with the female when she is fertile than a short-term mate would. However, a male using a long-term mating strategy will also be more concerned with losing his mate to another male and also that it is more costly in a long-term mating strategy if one's female mate is unable to bear children.

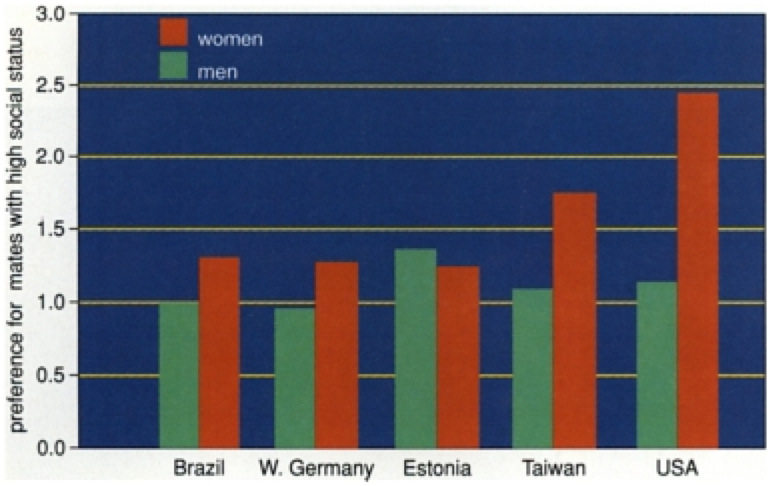


Figure 26: Preference for Mate with Social Status By Gender[?]

4.2.3 Mate Value

Attempting to determine the value of a mate can be a difficult task. One might argue that a good indicator could be social status. Figure 26 depicts the difference in importance of social status to both genders for different countries[2]. We can see that social status is used more by females than by males. This is likely due to females being more concerned with a male's ability to provide for her offspring than a male would be. For males, a better indication of a female mate's value (in terms of producing viable offspring) is determined by both her reproductive value and her current fertility.

4.2.4 More than Just a Pretty Face

In one experiment [4], half of the participants were told to evaluate their interest for a short-term mate, while the other half were told to make an evaluation for a long term mate. The image of the potential mate which was presented to the participants was occluded such that the actual female was not visible. However, the participants were given a choice. They could either remove the face occlusion or the body occlusion before making their decision. The participants making short-term decisions tended to choose to remove the body box more often than participants making long-term mating decisions. The authors interpreted this finding to indicate that males seeking a short-term mate tend to use the body as a cue to gauge their own interest while males seeking a long-term mate will use the face as a cue. It is important to note that participants were randomly selected to either the short or long-term conditions, so it is not the case that the experimenters studied participants who were previously known to seek short or long-term relationships. Instead, when participants were asked to indicate their interest in a short or long-term mate, they behaved differently[4].

4.3 Our Approach

We planned to approach this problem in a very similar fashion to our experiment design in our study. We planned to have participants create avatars using a software package we provided. The difference in this study, however, was that our participants would not be given target images. Instead, we planned to ask them to design their ideal physical partner for a romantic relationship. Our original plan, stemming from the previous project, was to use the Blender software package (See Figure 22), but we ultimately decided it would be much easier to do this using MakeHuman.

5 ACKNOWLEDGEMENTS

I would like to thank my advisers, Aaron Cass and Carin Perilloux, at Union College for their guidance in this project. I would also like to thank Union College for awarding a student research grant.

References

[1] Isabelle Boutet and Avi Chaudhuri. Multistability of overlapped face stimuli is dependent upon orientation. *Perception*, 30(6):743 – 753, 2001.

[2] David M. Buss. *The evolution of desire: Strategies of human mating.* Basic Books, 1994.

[3] Hannes Caspar. www.facity.com, 2010.

[4] Jaime C. Confer, Carin Perilloux, and David M. Buss. More than just a pretty face: Men's priority shifts toward bodily attractiveness in short-term versus long-term mating contexts. *Evolution and Human Behavior*, 31(5):348 – 353, 2010.

[5] Richard Dawkins. *The blind watchmaker: Why the evidence of evolution reveals a universe without design.* WW Norton & Company, 1986.

[6] C.D. Frowd, P.J.B. Hancock, and D. Carson. Evofit: A holistic, evolutionary facial imaging technique for creating composites. *ACM Transactions on Applied Perception (TAP)*, 1(1):19–39, 2004.

[7] Felicity Jenkins and Graham Davies. Contamination of facial memory through exposure to misleading composite pictures. *Journal of Applied Psychology*, 70(1):164 – 176, 1985.

[8] D. Maurer, R.L. Grand, and C.J. Mondloch. The many faces of configural processing. *Trends in cognitive sciences*, 6(6):255–260, 2002.

[9] David H Rakison and Jaime Derringer. Do infants possess an evolved spider-detection mechanism? *Cognition*, 107(1):381–393, 2008.

[10] Lawrence S. Sugiyama. *Physical Attractiveness in Adaptationist Perspective.*, pages 292 – 343. John Wiley Sons Inc, 2005.

[11] G. Van Belle, P. De Graef, K. Verfaillie, T. Busigny, and B. Rossion. Whole not hole: Expert face recognition requires holistic perception. *Neuropsychologia*, 48(9):2620, 2010.

[12] S. Zhou, H. Fu, L. Liu, D. Cohen-Or, and X. Han. Parametric reshaping of human bodies in images. *ACM Transactions on Graphics (TOG)*, 29(4):126, 2010.