Facial composites:

a comparison between a traditional and new technique for conducting facial composite techniques used in the field of investigation

Master thesis

Sinitta Haridat

s1534858

Faculty of Behavioural Sciences, University of Twente

First supervisor: Dr. Elze G. Ufkes

Second supervisor: Dr. Ir. Peter W. de Vries

Specialization: Psychology of Conflict, Risk & Safety

Date: 14-03-2016

Author's note

I first would like to thank Bas Boïng for making it possible to conduct this research with the Dutch police and Gerard de Groot for creating the feature-based composites. Second a sincere thanks to Mirjam Hardeman and her colleagues for their assistance during the experiment and conducting the interview. I also would like to thank René Lewis for his critical remarks regarding the Reverse Correlation Image Classification technique. A final thanks goes out to my supervisors Dr. Elze G. Ufkes and Dr. Ir. Peter W. de Vries, who not only helped me set up this research but also guided me in the making of this thesis.

Abstract

The current way of composing composites for eyewitness purposes is not in line with the way how humans perceive and retrieve human faces from memory. Humans perceive and remember faces holistically; yet current facial composite systems require a feature-based approach. The aim of this research was to test to what extent the Reverse Correlation Image Classification (RC) technique could be used in compositing composites for eyewitness purposes compared to existing feature-based facial composite systems. In study 1 participants had to compose a composite using either the RC technique or a feature-based system. In study 2 participants rated the composites based on their resemblance to the original picture. Overall, the feature-based system composites had a higher mean resemblance rating compared to the composites produced by the Reverse Correlation technique. Even though the results are not fully in favour of the Reverse Correlation Image Classification technique, this technique does hold a potential for composing composites in the field of investigation.

Facial composites:
a comparison between a traditional and new technique for conducting facial composite techniques used in the field of investigation

On March 1974, a nine-year-old boy was dragged to a baseball field and raped. The boy told the police what he thought the victim looked like. A relative of the boy thought that this description fitted Jimmy Bain. Therefore Bain's photograph was included in a photo line-up and the victim positively identified Bain as the perpetrator. Even though Bain had a supporting alibi he was arrested. Semen was found, but the forensic team could not make a 100% positive match and stated that it could have been Bain's semen based on a weak resemblance of his blood group. Jimmy Bain was convicted of rape, kidnapping, and burglary and sentenced to life in prison. It was not until 2001 that certain cases could be reopened for DNA testing. The court denied Bain's case five times before the Innocence Project came to help. After serving 35 years in prison, Jimmy Bain was excluded as the source of the DNA (Innocence Project, 2015).

The case above does not stand on its own. According to The Innocence Project (2015) eyewitness misidentification is the greatest contributing factor to wrongful convictions. 75% of the convictions overturned through DNA testing in the United States of America, were due to eyewitness misidentification. Since eyewitnesses hold an important role in the criminal justice system and in identifying the perpetrator (Brewer & Wells, 2011), one might expect that this area of the work field should have as little mistakes as possible, because a wrongful conviction takes the life of an innocent person with the real perpetrator still on the loose.

A witness is normally required to describe the events of a crime and the people concerned through a cognitive interview (Frowd et al., 2005). The witness has to provide a detailed description of the appearance of the face, including individual facial features such as the nose, eyes, hair and mouth. Frowd (2012) however states that remembering, and therefore

describing, individual features are not natural tasks for humans. Humans process faces as whole images (i.e. the face in its entirety), which means that face recognition is a holistic process rather than a recall process (Tanaka & Farah, 1993, 2003). Holistic processing can be defined as a psychological process where faces are processed in terms of their multidimensional similarity to other known faces instead of processing of individual facial features (e.g., eyes, mouth, nose) (Wells & Hasel, 2007). The composite systems used by the police however are mostly feature-based. Systems that require retrieval of memory for isolated facial features probably do not work well, since they do not hold in account the holistic way of processing faces (Wells & Hasel, 2007) , which means that the composed images do not resemble the target to the extent it should.

A new approach for creating facial images is a technique that is not yet used in the field of eyewitnesses, but may have potential to do so. The Reverse Correlation Image Classification (RC) technique is a technique whereby a participant is shown a base face, but instead of choosing features, noise is projected onto the base face to change the face-features (Dotsch & Todorov, 2012). Participants are provided with two faces at different trials, both with different noise patterns on the base face and have to choose which image most resembles the target. This technique has been successfully used in modelling famous persons (Tom Cruise vs. John Travolta), gender categories (male vs. female), and emotional expressions (trustworthy vs. untrustworthy and dominant vs. submissive) (Dotsch & Todorov, 2012; Mangini & Biederman, 2004). As mentioned earlier this technique has not been used in the field of eyewitnesses, but this technique might be a new approach in compositing target faces. The technique enables researchers to generate images that reflect participants' internal representations of faces (Dotsch & Todorov, 2012). To the authors' knowledge the Reverse Correlation Image Classification technique has not been subjected to a comparison of composites with a feature based composite system that is already in use. The research

question of this study is therefore; to what extent can the Reverse Correlation Image Classification (RC) technique be used in compositing face images for eyewitness purposes compared to existing feature-based facial composite systems?

a. Do facial composites based on whole face systems provide higher resemblance composites than feature-based systems?
b. Are facial composites based on whole face systems more often correctly identified in photo line-up than composites made by a feature-based system?

Memory

The memory process can be divided into three stages: encoding (making new memories), storage (consolidation of information) and retrieval (remembering what was previously encoded). Even though our cognitive abilities are incredible, the human memory does not function as a tape recorder and is therefore highly subjective to bias. These errors can begin in the encoding stage. As one can imagine, for accurately remembering a specific event, it is important to pay attention to what is happening. The classic experiment by Simon and Chabris (1999) showed that when participants were preoccupied counting bouncing basketballs, they failed to see a man in a gorilla suit. In the field of police practice, the weapon focus effect (Loftus et al. 1987; Steblay, 1992) is a well-known phenomenon where a weapon attracts attention, which leaves the witness with not enough time to encode to culprits face. These two examples show that humans can pay attention, but because they are preoccupied they might miss other ques. Besides errors in the encoding phase, errors can also occur in the storage phase. These errors can be estimator variables, these variables are factors that are not under the control of the justice system but negatively affect the reliability of eyewitness memory. (Gary L. Wells, 1978). Retention interval is one of these variables which often are used in the research domain of police practise. Retention interval refers to the time between the initial

event and the retrieval of the event.

Retrieving faces. The retrieval of information can happen in two ways, either by free recall or by recognition. By recalling information, one retrieves previous learned information from the memory, like filling in an open questions exam (e.g. who wrote the play 'The Tempest'?). Recognition is more like a multiple choice exam, were you only need to identify the information previously learned. This is the same with retrieving faces.

Humans are able to store a great number of individual faces (Werner, Kühnel, & Markowitsch, 2013). Babies can recognize their mothers' faces within a few days after their birth (Le Grand, Mondloch, Maurer, & Brent, 2004). There is however something special about face perception, as Farah, Wilson, Drain, and Tanaka (1998) argue. Humans use different brain areas for face recognition and other types of object recognition. O'Toole (2004) states that our sensitivity for the small differences between facial characteristics is higher than for any other object category.

Humans do not perceive faces as different features that make a face, but as one whole object. Tanaka and Farah (1993) found that the parts of a face are better perceived if presented in the context of a whole face than in the context of a scrambled face and that the recognition of whole faces was better than recognition of isolated parts. This means that faces are generally processed, stored, and retrieved at a holistic level rather than at the level of individual facial features (Tanaka & Farah, 2003). The favoring of face recognition (and hence holistic processing) over feature-based face recall might be explained using an evolutionary approach. From a survival point of view it is likely that survival is in favour of those who could readily recognize faces as to make rapid judgments of a friend versus foe (Wells & Hryciw, 1984).

Holistic representations facilitate recognition whereas feature-based representations identify the culprit (if present) in a photo line-up (recognition) (Wells & Hryciw, 1984).

However when there is a lack of physical evidence such as DNA or CCTV footage, eyewitnesses are asked to provide a description of the culprit and select features either provided by a sketch artist or by using facial composite system (recall). Through a cognitive interview an eyewitness is asked to provide a detailed description of the event and culprit (Frowd et al., 2005). An eyewitness can also be asked to create a facial composite either by a sketch artist or by a facial composite system. Wells & Hasel (2007) state that there is a mismatch between the task demands in composing a facial composite and the way that faces are usually perceived and remembered. Most composite systems demand a feature-based approach, but since humans encode and store faces as a whole, it is hard to retrieve the different features of a face.

Facial composite systems

There is a variety of facial composite systems, ranging from non-computerised feature-based systems to computerised whole face based systems (see table 1). The first facial composite system is in fact not an actual system but a sketch artist (Wells & Hasel, 2007). Sketch artists follow a feature-by-feature approach to compose a sketched image of a perpetrator (Frowd et al., 2012). One can imagine that the likeness of such a sketch depends not only of the eyewitness' description, but also on the drawing skills of the artist, which might vary from person to person.

Table 1

Overview facial composite systems

	Non-computerised	Computerised
Feature-based	Sketch artist	E-FIT
	Photo-fit	PRO-fit
	Identikit	Identikit 2000
	Identikit II	FACES
		Mac-a-Mug
Whole-face based		Evo-FIT

Other examples of non-computerised feature-based systems are Photo-fit, used in the UK and Identikit in the USA. Both systems use collections of facial features (e.g., noses, eyes, hair styles) that can be combined to create a face (Wells & Hasel, 2007). Originally the Identikit contained line drawings and used facial features printed on view foil (special kind of paper). Now the Identikit II uses photographic elements, and is similar to Photo-fit. A critical point for both systems is that eyewitnesses select facial features apart instead of using a whole face construction.

More modern computerized versions but still feature-based systems are E-FIT and PRO-fit in the UK, and FACES and Identikit 2000 in the USA (Frowd et al., 2005). These systems include more possible facial features and have more realistic visual results. E-FIT is one of the leading composite systems in the UK (Frowd et al., 2005), and uses photographed features to produce more realistic-looking composites (Frowd, Hancock & Carson, 2004). A system that is quite similar to E-FIT is Mac-a-Mug Pro. PRO-fit is a composite system used in the UK, and contains a large number of facial features. This system has separate databases for different features from different races and has an artwork package for enhancing the appearance of facial features, so one can really composite an image as wanted. The features selected are switched in and out of the presented face, which also is applicable to E-FIT. A system that is found to be quite similar to PRO-fit is the Identikit 2000. This is the most popular facial composite system in the USA, even more so than Mac-A-Mug Pro (McQuiston-Surrett, Topp, & Malpass, 2006). Besides Identikit and Mac-a-Mug Pro, there is another facial composite system popular in the US: FACES. FACES has quite a few limitations compared to the earlier mentioned composite systems (Frowd, McQuiston-Surret, Anandaciva, Ireland & Hancock, 2007).

According to Wells & Hasel (2007) no composite system could have enough facial-

feature choices to represent the physiognomic variability of the human face. It could therefore be that a face which already poses all these face features would perform better when used for a composite task. A computerised facial composite system that is whole face based is EvoFIT. Eyewitnesses first select face shapes and are then presented with a set of faces and have to choose which of these faces resembles the target at best. The chosen faces are then made into another set of faces which ultimately results in the composite face (Frowd et al., 2005). This system takes into account that faces are processed holistically (Frowd et al., 2005), but the programme does have a selection of noses, face shapes, eyes et cetera which one should choose from. It is therefore the question to what extent EvoFIT takes into account this holistically processing.

Reverse Correlation Image Classification technique

The Reverse Correlation technique has its roots in the domain of auditory perception during the seventies (Dotsch & Todorov, 2012). Later the technique was adapted for research on vision (Ahumada, 1996, 2002; Beard & Ahumada, 1998; Solomon, 2002 in Dotsch & Todorov, 2012) and neurophysiology. Just recently this technique has been used in the field of social perception. Mangini and Biederman (2004) described the Reverse Correlation Image Classification technique. This technique starts with a base face that creates different stimuli (or faces) by adding or subtracting. By using different noise patterns the base face changes. The noise placed on the base face image makes that the face looks different with each different random noise pattern (Dotsch & Todorov, 2012). A participant then decides to what extent the presented stimulus resembles the target. The faces that are found to resemble the target at best are merged into one so an image of the target can be created.

The Reverse Correlation Image Classification has been successfully applied in a study Mangini and Biederman (2004) conducted. In previous research only group stereotypes were used, but Mangini and Biederman (2004) used individual faces and asked participants after

the composite was created if they could recognise the intended person. Participants had to classify famous persons (Tom Cruise/John Travolta), gender categories (female/male), and emotional expressions (happy/unhappy). Dotsch and Todorov (2012) also used the Reverse Correlation Image Classification technique. Based on the noise pattern, the participant had to choose which of the stimuli resembles a trustworthy, untrustworthy, dominant or submissive face. By averaging the chosen images, a classification image was conducted which should represent the intended face.

It is said that groups make better judgements on average than experts and most individuals do, some people might overestimate while others might underestimate. This is called the *wisdom of the crowd effect* (Surowiecki, 2004). Hasel and Wells (2007) found that morphing composites from independent participants produced a composite that was of better likeness than the individual ones. This might also be the case when combining individual composites using the Reverse Correlation Image Classification technique. As stated by Bruce, Ness, Hancock, Newman, and Rarity (2002) four heads are better than one. By morphing (i.e. combining) the selected images, the likeness of the composites should be higher in comparison to the individual composites.

According to Dotsch and Todorov (2015) the technique "enables researchers to generate images that reflect participants' internal representations of faces, without making any assumption about what those representations might look like". This is because the Reverse Correlation Image Classification technique is data-driven and therefor has no assumptions about what a face or face-feature should look like. Besides the input of a base face, no other information is provided by either a researcher, police men, or by the programme itself. It is this point of difference compared to other composite systems that makes that the Reverse Correlation Image Classification technique holds potential for being a composite system used in the field of investigation.

Composing the face

Humans perform badly when it comes to composing an composite, even when they know the target (Davies & Valentine, 2013). The human memory is highly subjective to bias, Wells and Hryciw (1984) state that feature-based processing harms later recognition of the whole face. If the recognition of the whole face is harmed, it could be that the eyewitness chooses the wrong target when shown a photo line-up. The description earlier provided might not be the real target, but eyewitnesses tend to commit to the composite they provided (McQuiston-Surrett et al., 2006). As noted in the opening section, distorted memories from eyewitnesses account for 75% of wrongful convictions. This is a serious matter that needs further research.

Hypotheses

Even though the few studies have been able to show that whole-face systems compose better composites than feature-based systems (Davies & Valentine, 2006), based on the idea that faces are processed holistically, whole faced systems should, based on theory, perform better at creating composites of faces. It is therefore expected that the composite composed by the reverse correlation technique will have a better resemblance of the target

Hypothesis 1: Composites composed by the Reverse Correlation Image Classification technique have higher resemblance rates than composite composed by a feature-based system

Eyewitnesses are asked to compose a face composite when there is a lack of physical evidence. This composite is important for the identification of the culprit. For this reason the composite must be useable and actually represent the target. The Reverse Correlation Image Classification technique holds into account the holistic way of processing faces and it expected to provide higher resemblance rates regarding the composites than a feature-based system. It is therefore expected that the produced composites made by the RC technique will have a higher true hit rate in a photo line-up.

Hypothesis 2: Composites composed by the Reverse Correlation Image Classification technique will have a higher true hit rate in a photo line-up than composites composed by a feature-based system

It is suggested that morphing individual composites create a higher likeness than the mean of the composites individually (Hasel & Wells, 2007). It is therefore expected that the morphs than can be created for the Reverse Correlation Image Classification technique, have a higher resemblance rating than the composites using the RC technique individually.

Hypothesis 3: The mean resemblance ratings for the morphs will be higher than the mean resemblance ratings for the individual composites

Method

This study consisted of two parts. In the first part participants were asked to compose a composite either by using the reverse correlation technique of by using the feature-based system Compo-fit which is currently used by the Dutch Police for creating composites. Participants produced composites of a target perpetrator (either Stefan Raab; target 1 or Till Schweiger; target 2). The choice was made to choose two targets, in case a participant recognised one of the targets. No effects based on target were expected. In the second part of this study the composites from part one will be rated for their resemblance of the target. Each method in this study will be discussed separately.

Part 1: Creating the composites using either the Reverse Correlation Image Classification technique or Compo-fit

Participants

Participants were 21 volunteers (nine men and 12 women) collected by social media who received a travel allowance. A total of 13 participants were living in Enschede and

surroundings in the East of the Netherlands, six participants were located in the Western part of the Netherlands, leaving one participant in the South and one up North. The age ranged from 18 to 63 (M = 27.62, SD = 11.87). Thirteen of the participants were students from which nine attended a university, six attended a university of applied sciences. The other participants were either working full time (six in total) or are unable to work or find a job. All participants carry the Dutch nationality. Seventeen participants were randomly assigned to the composition technique and target. Due to the fact that this study wanted to research the extent to which the RC technique can be used as a composition system, and only two participants were assigned, despite the randomisation, to compose a composite using the RC technique, four additional participants were asked to the composite target 1.

In total, seven composites for target 1 were created using Compo-fit, one participant created a composite for target 2 using Compo-fit, since one participant created a composite of target 1 that was not useable. For the composites using the Reverse Correlation Image Classification technique both targets had six composites each.

Procedure

Participants received a link to an online survey, for which they were asked to open this link 48 hours before their appointment for composing the composites. All but two participants followed this instruction. The reason why the participants had a retention interval of 48 hours, was to create a more real-life experience, since eyewitnesses are not asked to create a composite hours after the violation has taken place, but it is more likely that a few days pass.

Participants first read a case stating that a violation had been taking place in a local park and that they had to imagine being at the park and witnessing this violation. After reading the case, the participants were told that would see the perpetrator for two minutes. Two days later the participants were invited to come to a police station in Amsterdam, where they composed the composites. Participants were brought to a room, where they first had to

sign the informed consent and after that would undergo a cognitive interview conducted by a detective. After the interview the participants started composing the composite. After composing the composite the participants filled out a questionnaire about their demographics and their experience. The participants who used the Reverse Correlation Image Classification technique would first read the instructions about how to work with the programme, and that they could take a break whenever they wanted due to the long procedure the technique has, it took about 50 minutes to complete the task. After the instructions the participants first completed a practice trial with 30 pairs of faces. Participants selected one of two faces that most resembled the target they saw two days before the experiment. When they finished the practice trials and had no further questions, the task was started including 770 pairs of faces. The participants in the feature-based system group, where first introduced to the sketch artist. Following a standardised questionnaire participants were asked to describe the target. The participants we not able to watch the making of the actual composite, they saw the composite when it was finished based on their description. When the composite was shown they could make some alterations in for example the size of the face. Due to the fact that the programme is quite outdated, facial hair was added by pencil. After finishing the composite the participants were asked to rate the composite on a 1-10 scale based on the likeness of the composite and the image in their head. The procedure of the police is that when a composite is rated lower than a six, it won't be published. It is therefore that only one composite was produced in the feature-based target 2, because the participant rated the composite with a 5.

Materials

The photographs used as the target face belong to two German celebrities, namely Stefan Raab and Till Schweiger (Figure 1). Their photographs were found online from Google Image Search. The participants in part one of this study had to be unfamiliar to the target, because when an eyewitness knows the target, there is not a reason to compose a composite.

Figure 1. **Shown stimuli Stefan Raab (left), Till Schweiger (right)**

The questionnaire used during the Cognitive Interview was created by one of the involved detectives, the questions mainly focused on the violation and the target.

The Reverse Correlation Image Classification was programmed using the programme Inquisit, making it possible to alter welcome texts, instructions and the base face with different kind of projected noise. The base face used for the Reverse Correlation Image Classification is a standardised Caucasian male face (Figure 2) collected from the emotion lab from the Karolinska Institute.

Figure 2. **Base face from the Karolinska Institute. On the left the used base face, on the right an example of the base face with projected noise**

Each participant completed 30 practice trials followed by 770 trials, a total of 800 trials. Each trial consisted of two face, or stimuli, being a or b. All the stimuli contained the same face but at each trial the base face was projected with a different noise pattern. Participants had to choose which of the two faces most closely resembled the target, they could choose from 1: Clearly A, 2: Probably A, 3: Probably B, 4: Clearly B. Based on the data-files twelve composites were produced, for each target a morph containing all six composites was created (figure 3).

***Figure 3*. Morphs, target 1 (left) and target 2 (right)**

One of the composite systems the Dutch police use is Compo-fit, which is also the system that is used in this study. The system is quite outdated, which makes that is partially computerised and non-computerised. The system contains different faces, eyes, mouths et cetera that can be altered in size. Yet special characteristics (e.g. a beard) have to be drawn by hand. This system has not yet been subjected to research to the authors' knowledge. A selection of composites that were created in this study is shown in figure 4.

***Figure 4*. Example of composites using Compo-fit. Target 1 (left) and target 2 (right)**

Part 2 rating the composites

Participants and design

A total of 100 participants completed the survey; 30 uncompleted surveys were excluded from any analysis. The remaining 100 participants consist of 39 men and 61 women. Participants were recruited by social media and research exchange groups belonging to University Twente. The age ranged from 16 to 55 (M = 26.08, SD = 7.34). The vast majority of participants (n = 96) are Dutch, with four participants being German and most of the participants (64%) having their residence in Enschede and surroundings. Most participants were students (n = 52) or working full-time (n = 39). Each participant filled out the same survey.

Procedure

Participants were first informed about the research objective and after that were asked to sign the informed consent. The first task on this study involved a photo line-up. A photo line-up was created using the photographs used in part was of this study and adding five other photographs of other male Caucasian celebrities, for both targets a photo line-up was created

(Figure 5 & Figure 6). Participants were randomly shown eleven photo line-ups with different composites. The composites used in this task were the eight composites produced by the feature-based system and the two morphs produced by the Reverse Correlation Image Classification technique.

The next task the participants had was to rate the composite based on the similarity to the photograph the participants in part one saw before composing the composite. This was done for 22 composites; eight feature-based composites, 12 reverse correlation composites and two morphs created by combining the singular composites using the RC technique. All participants saw the 22 photo line-up randomly and had to answer the question "to what extent does this picture resemble the perpetrator?" this was answered by a 7 point Likert scale (1; not at all to 7; entirely). After completing this task participants were asked about their demographics.

Figure 6. **Photo line-up for target 1.**

***Figure 6*. Photo line-up for target 2.**

Materials

The materials used in this part of the study where the 22 composites produced in part one of this study. The 22 composites consists of six composites of target 1 plus the morph of target 1 and six composites of target 2 plus the morph of target 2 using the Reverse Correlation Image Classification technique, seven composites of target 1 using Compo-fit and one composite of target 2.

For the photo line-up an additional of 10 photographs of Caucasian males who looked to some extent like the targets, were collected on google image search.

Results

To test *hypothesis 1: Composites composed by the Reverse Correlation Image Classification technique have higher resemblance rates than composites composed by a feature-based system* a repeated-measures ANOVA was conducted, with the mean resemblances scores of

the composites produced by the feature-based system and by the Reverse Correlation Image Classification technique, this was done for both targets, so the means of in total four groups were analysed.

There was a significant main effect of the type of composite system used on ratings of the composites, F (1, 99) = 8.76, p = .004. Participants rated the composites produced by feature-based system higher in resemblance (M = 3.34, SD = 0.96), than the composites produced by the Reverse Correlation Image Classification technique (M = 2.94, SD = 0.89). Hypothesis 1 is therefore rejected.

There was also a significant unexpected main effect of the target used on ratings of the composites, F (1, 99) = 78.20, p <.001.

Analysis revealed significant interaction effect of used composite technique and target, F (1, 99) = 105.92, p < .001. This indicates that the target used had different effects on participants' ratings depending on which type of composite system was used. Due to this significant interaction a simple main effect analysis was conducted that revealed that in the feature-based group resemblance ratings for target 1 were higher (M = 3.39, SD = 0.96), than resemblance ratings using the reverse correlation classification image technique. (M = 2.01, SD = 0.82). For target 2 the resemblance ratings using the reverse correlation technique were higher (M = 3.79, SD = 1.24), than resemblance ratings for composites produced by the feature-based programme (M = 3.05, SD = 1.71).

Hypothesis 2

To test *hypothesis 2: Composites composed by the Reverse Correlation Image Classification technique will have a higher true hit rate in a photo line-up than composites composed by a feature-based system* four chi-square tests of goodness fit were performed, one for each group to determine whether observed sample frequencies differ significantly from the expected change level. The sample size includes 100 participants, by dividing these 100 participants

over six possible targets, it was expected that based on a chance level 16.7%, participants selected the correct target. For target 1 using the feature-based system seven composites were rated, for this analysis the composite with the highest resemblance rate was used. The chi-square showed that the percentage of participants that the selected the correct target was not equally distributed χ^2 (5, N=100) = 41.96, p <.001, 37% of the participants selected the correct target. Analysis of target 2 using the feature-based system showed that 29% of the participants selected the correct target χ^2 (5, N=100) = 99.32, p <.001. Using the Reverse Correlation Image Classification technique the distribution for target 1 again was not equally distributed χ^2 (5, N=100) = 58.76, p <.001, 15% of the participants selected the correct target. Analysis of target 2 showed a significant preference of target χ^2 (5, n=100) = 62.6, p <.001, 44% of the participants selected the correct target.

Hypothesis 3

To test *hypothesis 3: The mean resemblance ratings for the morphs will be higher than the mean resemblance ratings for the individual composites*, the means of the morphs from each target were compared to the mean of the individual composites combined. A paired sample t-test was conducted to test whether the morphs were rated higher on resemblance compared to the mean individual composites. Comparison of the individual composites resemblance ratings for target 1 (M = 2.01, SD = 0.82), and the morph for target 1 (M = 1.93, SD = 1.06) revealed no significant differences between resemblance scores t (99) = 0.91, p = .365. The mean resemblance ratings for the individual composites for target 2 (M = 3.79, SD = 1.24) were rated lower than the ratings for morph created for target 2 (M = 4.40, SD = 1.64). This difference, -0.61, BCa 95% CI [-0.85, -0.37], was significant t (99) = -5.11, p < .001.

Discussion

The current way of composing composites for eyewitness purposes is not in line with the way how humans perceive and retrieve human faces from memory. The aim of this research was to determine to what extent the Reverse Correlation Image Classification (RC) technique could be used in compositing composites for eyewitness purposes compared to existing feature-based facial composite systems. This study shows that Reverse Correlation Image Classification (RC) technique does hold potential to be used as a composite system. There was an unexpected difference in results between the two targets when the RC technique was used; target 2 had the best overall results, even compared to the results using the feature-based programme. It might even be typical to find mixed results when comparing different composite systems, as previous research also shows mixed results (Frowd et al. 2005; Frowd et al. 2007). Depending on the different kind of measures, one programme might outdo the other one. Note however that the previous mentioned research used different measures to compare the composites produced than the measures used in this study. It is difficult to compare measures across studies, it is even more difficult to describe some ‘‘absolute’’ level of performance, as stated by Wells and Hasel (2007).

As to determine if a more holistic approach to composing composites favours a traditional feature-based approach it might be said using the results in this study that both approaches are compatible. Research regarding facial composite systems up till this point has not been able to proof that, composite systems that take into account the holistic way of perceiving and storing faces, produce higher in likeness composites than traditional systems (Frowd et al., 2005). Wells and Hryciw (1984) did show that when participants were asked to re-create a face that was made using Identi-kit the resemblance rating on a 7-points Likert scale was only a 2.0.This implies that even when participants had the option to create the exact same face, they were unable to select the different kind of features.

Likeness

The first hypothesis expected higher resemblance rating of the composites for the reverse correlation image classification. Results in general show that the feature-based system provides higher likeness composite than the Reverse Correlation Image Classification technique. However, this is only true for target 1, the overall mean rating of target 2 shows that the RC technique provides higher likeness composites. The main issue seems to be the low rating of target 1 using the Reverse Correlation Image Classification technique. Composites were showed in a random order, so the problem could not just depend on viewing order, altering ones expectations based on the previous composite. The low rating of target 1 using the RC technique could be due to the procedure, since four out of six were in fact not created at the police station but in a study room. It could be that the surroundings have affected the participants in the making of the composite. A last explanation of the low ratings of target 1 could be due to the fact that the RC technique has no function to create a beard or facial hair. One participant in this study said that he was looking for a face with facial hair, this could imply that facial hair is an important factor for face recognition. There is however little literature found about the importance of facial hair in face recognition. De Marsico (2014) notes that facial hair may affect the reliability of recognition negatively, because information about the mouth is now partial missing.

Photo line-up

Analysis of hypothesis two showed that the percentage of participants that the selected the correct target was not equally distributed and that again the composite of target 2 using the Reverse Correlation Image Classification technique had the highest true hit rate compared to the other composites. Yet, the composite of target 1 using the Reverse Correlation Image Classification technique had the lowest true hit rate, which even lies below the chance level. Even though this means that hypothesis 2 has to be rejected, the overall results are consistent

and the results found are significant different from the chance level. This implies that participants did in fact see a difference between the faces in the photo line-up, which made them choose the correct target.

Morphing

For hypothesis 3, the mean resemblance rating of the individual composites of target 1 compared to the mean resemblance rating of the morph of target 1 had no significant difference. The mean resemblance rating of the individual composites of target 2 compared to the mean resemblance rating of the morph of target 2 did have a significant difference. This is in line with the results found in the study of Hasel and Wells (2007). Morphs were rated as more similar to the target face than were the mean ratings of the individual composites.

Strengths and limitations of the present study

This study was designed to compare a feature-based composite system and the Reverse Correlation Image Classification technique; this was done in a realistic setting. Participants conducted the experiment at a real police station and were interviewed by experienced detectives. The feature-based composites were created by a sketch artist whose composites are used in real investigation cases. Because of this real life setting we were able to get input from those who work in the field of investigation.

A cooperating detective in this study observed that some of the participants provided different descriptions in the cognitive interview conducted by a detective and the questionnaire the sketch artist used. It would have been interesting to conduct a photo line-up with the participants in part 1 of this study, because in the RC group the participants were only interviewed once, but in the feature-based group twice. Research shows that when eyewitnesses are asked to describe an event, the describing might interfere with the actual memory (Frowd & Fields, 2011). The Reverse Correlation Image Classification technique

does not require an additional interview for composing the composite, so this technique could be part of a solution for the verbal overshadowing effect.

A limitations most of the participants using the Reverse Correlation Image Classification technique mentioned was that the task itself is really boring, too long and that after a while it was hard to focus of the difference between the two classification images. One might assume that in a real-life case the stakes are higher and therefor the willingness to produce a composite that can be used is higher, but still 770 pairs might be too many. A solution could be that instead of running of premeditated file with faces, a selection is made based on the choice the eyewitness made. In research conducted by Dotsch and Todorov (2012) using the RC technique, participants saw 300 pairs of faces, whereas in the study of (Mangini and Biederman (2004)), participants were subjected to 390 trials. The two previous mentioned studies focus mostly on social perception and it might be that for recognizing social perceptions a lesser amount of pairs is needed to produce a composite that can be used. It would be interesting to research the results obtained from this study to a deeper level. Since our data files contain the data of 770 pairs of faces, it could be researched if the resemblance ratings are affected if we were to take 440 pairs of faces, for example.

Furthermore, analysis of hypothesis 1 revealed an interaction effect between composite system and target. Participants in part 2 of this study rated all the composites, irrespectively of the system used (RC or Compo-fit) or shown target, it might therefore be interesting how the ratings of the different composite systems would have been if not rated by the same participants.

Recommendations

For future research it is recommended to work on the usability of the Reverse Correlation Image Classification technique for eyewitness purposes. This can be done by researching the amount of necessary pairs of faces in order to create a composite that can be used in the field

of investigation. Furthermore, a few participants mentioned the fact that they had to use a mouse for selecting the images, but they preferred using the keyboard. It is therefore recommended to also incorporate a possibility to use the keyboard.

For the purpose of this study it was not a limitation that the data-files created using the RC technique were not immediately converted into actual composites, for the use in practise this is important since the eyewitness has to determine if the composite looks the way the eyewitness intended it to be.

Conclusion

This research shows that the Reverse Correlation Image Classification technique is compatible with composite systems already in use. Participants were able to compose compositions that to some extent represented the target. The composites created using the RC technique are not ready to be used for investigation though; the likeness with the target is not high enough. Besides the before mentioned, the task might be too cognitive demanding for victims of serious crimes. Before the RC technique can be implied as a composite system, it is wise to check the Dutch law and regulations for creating composites, as the current legislation is based upon feature-based systems. Even though humans have difficulties composing composites, maybe when a holistic based system, such as the Reverse Correlation Image Classification technique, is implied clearance rates using composite might improve.

Literature

Brewer, N., & Wells, G. L. (2011). Eyewitness Identification. *Current Directions in Psychological Science, 20*(1), 24-27. doi: 10.1177/0963721410389169

Bruce, V., Ness, H., Hancock, P. J. B., Newman, C., & Rarity, J. (2002). Four heads are better than one: Combining face composites yields improvements in face likeness. *Journal of Applied Psychology, 87*(5), 894-902. doi: 10.1037/0021-9010.87.5.894

Davies, G. M., & Valentine, T. (2013). Facial composites: Forensic utility and psychological research. *Handbook of eyewitness psychology: Memory for people*, 59-83.

De Marsico, M. (2014). *Face Recognition in Adverse Conditions*. Hershey, Pennsylvania IGI Global.

Dotsch, R., & Todorov, A. (2012). Reverse Correlating Social Face Perception. *Social Psychological and Personality Science, 3*(5), 562-571. doi: 10.1177/1948550611430272

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review, 105*(3), 482-498. doi: 10.1037/0033-295X.105.3.482

Frowd, C. D., Carson, D., Ness, H., Richardson, J., Morrison, L., McLanaghan, S., & Hancock, P. (2005). A forensically valid comparison of facial composite systems. *Psychology, Crime & Law, 11*(1), 33-52. doi: 10.1080/1068316031000163431 3

Frowd, C. D., & Fields, S. (2011). Verbalization effects in facial composite production. *Psychology, Crime & Law, 17*(8), 731-744. doi: 10.1080/10683161003623264

Frowd, C. D., McQuiston-Surrett, D., Anandaciva, S., Ireland, C. G., & Hancock, P. J. B. (2007). An evaluation of US systems for facial composite production. *Ergonomics, 50*(12), 1987-1998. doi: 10.1080/00140130701523611

Frowd, C. D., Skelton, F., Atherton, C., Pitchford, M., Hepton, G., Holden, L., . . . Hancock, P. J. B. (2012). Recovering faces from memory: The distracting influence of external facial features. *Journal of Experimental Psychology: Applied, 18*(2), 224-238. doi: 10.1037/a0027393

Hasel, L. E., & Wells, G. L. (2007). Catching the bad guy: morphing composite faces helps. *Law and Human Behavior, 31*(2), 193-207. doi: 10.1007/s10979-006-9007-2

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2004). Impairment in holistic face processing following early visual deprivation. *Psychological Science, 15*(11), 762-768.

Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: estimating the information employed for face classifications. *Cognitive Science, 28*(2), 209-226. doi: 10.1207/s15516709cog2802_4

McQuiston-Surrett, D., Topp, L. D., & Malpass, R. S. (2006). Use of facial composite systems in US law enforcement agencies. *Psychology, Crime & Law, 12*(5), 505-517. doi: 10.1080/10683160500254904

O'Toole, A. J. (2004). Psychological and Neural Perspectives on Human Face Recognition. 349- 370.

The Innocent Project, (2015). Eyewitness Misidentification. Retrieved on May, 2015, from http://www.innocenceproject.org/causes-wrongful-conviction/eyewitness-misidentification

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, socities and nations.* New York: Random House.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology Section A, 46*(2), 225-245. doi: 10.1080/14640749308401045

Tanaka, J. W., & Farah, M. J. (2003). The holistic representation of faces. *Perception of faces, objects, and scenes: Analytic and holistic processes*, 53-74.

Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology, 36*(12), 1546-1557. doi: 10.1037/0022-3514.36.12.1546

Wells, G. L., & Hasel, L. E. (2007). Facial Composite Production by Eyewitnesses. *Current Directions in Psychological Science, 16*(1), 6-10. doi: 10.1111/j.1467-8721.2007.00465.x

Wells, G. L., & Hryciw, B. (1984). MEMORY FOR FACES - ENCODING AND RETRIEVAL OPERATIONS. *Memory & Cognition, 12*(4), 338-344. doi: 10.3758/bf03198293

Werner, N.-S., Kühnel, S., & Markowitsch, H. J. (2013). The Neuroscience of Face Processing and Identification in Eyewitnesses and Offenders. *Frontiers in Behavioral Neuroscience, 7*, 189. doi: 10.3389/fnbeh.2013.00189