Utrecht University

Recognition of Emotion in Facial Expressions: the Comparison of FaceReader to fEMG and Self-report.

Master Thesis

Linde Booijink

4185641

Supervisor: Marloes Eidhof

Master Thesis

Faculty of Social Science, Psychology

University of Utrecht

14-07-2016

Abstract

In the current research, the differences between FaceReader, facial EMG and self-report in detecting facial expressions were examined. Additionally, the accuracy of the three measurement methods based on their distinctiveness was taken into account. Prior research suggested that FaceReader has an accuracy of 89%. However, further findings were limited. On the other side, fEMG proved itself as a sensitive and reliable tool in multiple studies. Therefore, it is of interest whether FaceReader, due to its advantages against fEMG, could perform on the same level as fEMG. Expected was that FaceReader and fEMG would show similar results on facial expression scores. This is studied by exposing 26 undergraduate participants to three emotion inducing film clips (sadness, disgust and fear) and one neutral film clip. Each film clip was presented five times. FaceReader and fEMG were performed at the same time. The results show that FaceReader, fEMG and VAS did not show any similarity on their facial emotion expression outcome. Moreover, findings suggest that the intended facial emotions were not expressed by the participants. It seems FaceReader is not capable of distinguishing between the emotion expression of sadness, disgust and fear. However, FaceReader and fEMG could observe a decrease in the facial expression outcome across time. This suggests that FaceReader is a potential measurement method in detecting changes over time of negative valence. Further research must demonstrate whether FaceReader could be implemented in clinical practise as a useful tool for measuring facial emotions.

Keywords: facial emotion expression, FaceReader, fEMG.

For centuries, scientists have studied the recognition of emotion in facial expression. This line of research started with Charles Darwin's *The Expression of Emotions in Man and Animals,* first published in 1872 (Darwin, 1998). Darwin gathered evidence that, applying to all animals, some emotions have a universal facial expression. Additionally he proposed principles explaining why specific expressions occur for specific emotions. To amplify Darwins' study, Ekman and Friesen (1971) carried out a cross-cultural research on facial expressions. They told adults and children from New Guinea a story, showed them s a set of three faces with different expressions and asked them to select which emotional expression fitted the story. The results showed that there is a universal agreement on emotion perception, all coming down to six basic expressions: happiness, anger, sadness, disgust, surprise, and fear. These primary emotions carry action potentials with important functions, organizing our behaviour to help us survive (Brosch, Pourtois, & Sander, 2010).

Expressing emotions and recognising emotions in other people are key components of human interactions, especially through facial expressions. In that way, we provide information that we can use to make social situations more predictable and easier to manage (Elfenbein, der Foo, White, Tan & Aik, 2007). However, not all individuals have the ability to show context-appropriate affective responses and express their feelings to others (Fanti, Kyranides & Panayiotou, 2017). Failing to recognize emotions not only undermines someone's productive potential, but can have negative interpersonal consequences as well Moreover, impaired emotional awareness and disturbances in emotion recognition are associated with many psychological disorders. That is why most psychotherapy approaches focus on awareness and acceptance of emotions. Research suggests that accurate recognizing of facial emotions is critical to psychological health and social adjustment (Sloan & Kring, 2007).

Not only is emotion recognition through facial expression important for clients in mental health care, but also for psychotherapists and psychotherapy researchers. Therapeutic empathy requires that psychotherapists are able to recognize both the quality and intensity of a client's emotional experience. Additionally, psychotherapy researchers must be capable of identifying the emotions that psychotherapy samples express to determine the role that facial expression plays in psychotherapeutic improvement (Greenberg, Rice, & Elliott, 1993). Understanding the role of emotions can assist all mental health practitioners in being more in tune with clients' emotions and further assist in the helping process (Egan, 2010).

Given the importance of recognition of emotion through facial expressions in daily life and in psychotherapy, it is imperative that psychometrically strong and useful measures of facial emotions are available. Indeed, there are numerous methods that are being applied for measurements of emotions including the traditional conscious (non)verbal approach and those involving cognitive, physiological and behavioural expressions (Churchill & Behan, 2010). An example of the traditional approach is a self-report questionnaire usually scaling a number of criteria, like the Visual Analog Scale (VAS; Miller & Ferris, 1993). The VAS is a type of single-item measure in which the patient indicates his or her mood or feelings on a line or scale, in which the anchors are for example 'very happy and 'very unhappy'.

An example of the physiological and behavioural expressions for the measurement of emotions, is facial electromyography (fEMG; Dimberg, 1982). Facial EMG allows for the detection of subtle movements of facial muscles, and provide a good description of the time course of the movements (Fridlund & Izard, 1983). Since the start of fEMG it has been revised and tested several times. Dimberg (1990) for example, undertook a series of studies in an attempt to systematically explore whether the fEMG response is a general component of the emotional reaction. In all studies the method and the laboratory situation were similar; participants were exposed to different projected slides on a screen. Surface electrodes were attached over the corrugator and zygomatic muscle regions (Dimberg, 1990). Dimberg demonstrated that the fEMG technique is a sensitive tool for measuring emotional reactions. This is in line with other literature (Fridlund & Cacioppo, 1986; Hess, 2009; Tassinary, Hess, & Carcoba, 2012). However, there are limitations regarding the use of fEMG as well. First, a limitation with respect to the applicability of this technique is the extensive preparation that is required. Second, the connection with the electrodes to the recording equipment may impede with natural and spontaneous behaviour and is difficult to use in naturalistic settings (van Boxtel, 2010). Third, the equipment are very expensive to purchase and therefore not available in various settings.

FaceReader is a system for fully automatic facial expression analysis, that may overcome these limitations (Noldus Information Technology, Wageningen, The Netherlands, 2014). This software locates a person's face, and then reconstructs the face three-dimensionally. Robustness and reliability were tested in different studies, including the study by Den Uyl and van Kuilenburg (2005). Den Uyl and Kuilenburg compared the judgments of trained observers with the FaceReader outcome, showing that FaceReader can recognize facial expressions by distinguishing the six basic emotions and a neutral expression with an accuracy of 89%. Moreover, Lewinski, Fransen and Tan (2014) found that FaceReader measures of facial expressions of happiness correlated sufficiently high with participants' self-reports of their happiness. FaceReader has been applied for conducting research in various fields, for example in consumer behaviour. De Wijk, Kooijman, Verhoeven, Holthuysen, and De Graaf (2012) analysed the facial expressions that participating consumers showed when tasting or smelling liked or disliked food with FaceReader.

The most important advantages of FaceReader against fEMG are, its accessibility, wide reach and flexibility. Because a computer and a webcam are the only required tools, FaceReader is accessible to a much broader target group than fEMG. Additionally, FaceReader has huge potential to have positive effects on health outcomes in resource-poor settings. According to Donner (2008) in these low income countries, travel for health is for most people expensive, time-consuming, exhausting and physically challenging. FaceReader has the advantage of reaching to populations where care was previously slight. There are not only advantages of FaceReader against fEMG, but also against VAS. First, VAS is less reliable because it is a subjective measurement. Second, the VAS is difficult to accomplish for people who are not capable of indicating their own emotions (Fanti et al., 2017). Third, VAS is more time consuming.

However, studies examining the validity and applicability of FR are still scarce. One recent study that is conducted by Fanti et al. (2017) examined facial reactions of individuals with varying levels of callous-unemotional traits and impulsive aggression to violent and comedy films. These facial reactions were measured within two experiments, where FaceReader technology was used in one and fEMG in the other. These outcomes were combined, but still consisted of separate data retrieved from different participant groups. Thus, not yet has the accuracy of FaceReader directly been compared to other facial expression measurements.

Therefore, in the present study FaceReader, fEMG and VAS were performed at the same time, as a multi-method approach, to examine their differences in detecting facial expressions. In this experiment, young women were exposed to three emotion inducing film clips (sadness, disgust and fear) and one neutral film clip. Each film clip was presented five times to purchase different response strengths. In a pilot study, it was found that woman showed more facial expressions then men. That is why in the present study only women were included. Moreover, it was decided to focus solely on the three negative emotions because those are most important when suffering from a psychological disorder.

Since both measurements have previous shown to be effective in detecting emotions, it was hypothesized that FaceReader and fEMG have similar outcomes. VAS was used as an additional comparison tool. The accuracy of the three measurement methods based on their distinctiveness was taken into account in this study as well. FaceReader is potentially the most efficient method of the three, therefore it is worthwhile to investigate whether FaceReader is capable of detecting emotions as accurate as fEMG.

As described before, the understanding of emotions is crucial in daily life and in therapy and therefore beneficial for both therapists and clients. A therapist could for example use FaceReader when providing e-health therapy to support the interpretation of facial emotions of clients. Moreover, the ability of therapists to better understand clients can result in more informed judgments regarding threats to oneself and others (Hurley, 2011). With regard to clients FaceReader could be helpful when they participate in emotion therapy aiming at improving their emotion-recognition skills. By studying the accuracy of the new technology FaceReader, a helpful contribution to the scientific debate about facial emotion recognition will be made.

Method

Participants

A total of 33 individuals participated in this experiment; all of them were young women. From those participants, seven were excluded from further analysis due to errors of FaceReader or fEMG. Five of these seven excluded participants fell out of the recording frame, causing detecting problems for FaceReader. For one of the excluded participant FaceReader failed to record all the data and two excluded participants' fEMG data was not recorded properly. Other exclusion criteria were: background knowledge, facial anaesthetics, medication and plastic surgery, because this would impede spontaneous or natural facial expression. The remaining 26 subjects had a mean age of 23 years (*SD* = 2.53). The participants were undergraduate students from the Utrecht University and received either course credits or 4 euros for participation. The participants had been recruited through advertisement at the campus, most of them had a Dutch nationality (*N*=18). Participants experienced all emotion inducing film clips (fear, neutral, sadness and disgust), in a randomized order. The participants completed the experiment in English.

Materials

The Visual Analogue Scale (VAS; Miller & Ferris, 1993) was used to measure self-reported emotion. The emotion VAS is a continuous scale comprised of a horizontal line, 10 centimetres in length, anchored by 2 descriptors, one for each emotion extreme. In the present study the VAS was used to measure the amount of self-reported happiness (left = not happy at all; right = extremely happy), sadness (left = not sad at all; right = extremely sad), anger (left = not angry at all; right = extremely angry), anxiety (left = not scared at all; right = extremely scared), disgust (left = not disgusted at all; right = extremely disgusted) and surprise (left = not surprised at all; right = extremely surprised). Participants marked on this horizontal line at the computer screen, the point that they felt represented their perception of their emotional state.

One questionnaire for demographic data was used and included 11 questions in total. Five questions about gender, age, current occupation, nationality and native language. Four questions about facial anaesthetic, medication and plastic surgery and three questions about the background knowledge of the content of this study. As described before, the four final questions were used as exclusion criteria.

Three emotion-inducing film clips and one neutral film clip were used that have been tested and validated in previous studies, and proved to elicit discrete emotions. The sadness clip was a scene from the movie 'The Champ' and showed a small boy crying over his dead hero. The disgust clip was a scene from the movie 'Pink Flamingos' (Gross & Levenson, 1995) and showed a weirdly dressed woman eating dog poop. The neutral clip was a scene from the movie 'The Lover' and showed a girl getting into a car. The fear clip was a scene from the movie 'Copycat' and showed a woman that was looking for a perpetrator (Schaefer, Nils, Sanchez & Philippot, 2010). All the video clips were shortened to 29 seconds.

Facial Electromyography (fEMG; Dimberg, 1982) was used to measure the muscles and the emotion of the face. Six shielded electrodes were connected to different facial muscles of the participant. Two below the lip on the *Depressor Anguli Oris* that measured sadness, two next to the nose on the *Levator Labbi Superioris* that measured disgust and two at the forehead on the *Medial Frontalis* that measured fear and sadness (van Boxtel, 2010). An unshielded ground electrode was attached behind the participant's left ear. Facial EMG activity was recorded bipolarly using sintered Ag/AgCI electrodes. The fEMG signals were logged with MindWare software (EMG 3.0.21; Mindware Technologies, Gahanne, OH). The signals were rectified and smoothed with a 20 Hz lowpass filte with a time interval of 100 ms. With the software BioLab the data was acquired from MindWare (BioLab 60-0107-3.1; Guerrero, Bataller, Soria & Magdalena, 2007).

The Noldus FaceReader 4.0 software was used to code facial expressions (Noldus Information Technology, Wageningen, The Netherlands). A Sony SRG300H webcam was used to film the procedure, which was installed on top of the computer screen facing the participants. Some important requirements for FaceReader to produce reliable results is to make sure the participants' face is well lighted, well situated in front of the camera with minimal head rotations. Therefore the wall of the experiment room was covered with white papers and two big lamps were pointed at the participants face. The recordings were processed by Media Recorder (Noldus Information Technology, Wageningen, The Netherlands), saved as AVI files and analysed frame by frame by using FaceReader 4.0 software. The six emotions and neutral expression were scaled from 0 (not present at all) to 1 (maximum intensity of the fitted model).

Procedure

Participants were first brought to the experiment room and explained the study protocol. After giving their written consent, the facial electrodes were attached an participants were asked to make some facial movements to make sure the electrodes were well attached. The subjects were told that the purpose of the experiment was to measure physiological reactions to facial stimuli. To reduce the subjects' attention to their facial muscles, they were told a cover story that the electrodes were measuring internal activities (Dimberg, 1982). Also, participants were instructed to sit straight and keep looking at the screen to minimize FaceReader errors. On a computer screen, participants watched the four film clips five times each, in a randomized order. Each clip lasted 29 seconds with a 10 seconds black screen between the display of each clip. While the participants watched the film clips, one of the experimenters sat in the corner of the room to maximize the emotion response effect due to social adjustments. After the final film clip, the participants had to fill in the VAS on the computer screen, using the mouse to drag the scale lines. On these scale lines, participants were asked to indicate how they felt at the moment when they first saw the video clip (described as 'woman and the dog', 'young boy', 'girl in the car' and 'woman in bathroom') and at the moment they saw the video clip for the last time. During this task the experimenters left the room. Finally, participants completed the demographical information questionnaire and received their compensation.

Data preparation and analysis

The design of the current study was a within-subjects design. Prior to the statistical analyses, the data were prepared. First step of the preparation was to average the FaceReader data. This data consisted of facial expressions, that were measured each millisecond. The data of each movie was averaged for each participant individually. In order to compare the measurement outcomes of FaceReader, fEMG and VAS scores, the data of all three methods was standardized. This resulted in z-scores. Moreover, the data was analysed for outliers.

To check whether the manipulation worked, Helmert contrasts were conducted for each film clip (sadness, disgust, neutral, fear) and each measurement. This way it was checked whether FaceReader, fEMG or VAS indicated that one emotion was expressed stronger than the average of the others.

To answer the research question, the three measurements methods were compared to each other. Additionally was tested whether the measurement methods could pick up the participant facial expression response differences during the first seen film clip and last seen film clip. Therefore three 2 (time: first and fifth) x 3 (measurements: fEMG, FaceReader and VAS) repeated measures analysis of variance (ANOVA) were carried out for sadness, fear and disgust. The factors time and measurements were both used as within-subject factors.

Results

Two outliers were found for the sad film clips (VAS, first time showed), 11 for the disgust film clip (FaceReader first time showed; FaceReader last time showed; VAS first time showed), 8 outliers for the neutral film clip (FaceReader, first time showed, scared and disgust emotion), and 9 for the fear film clip (FaceReader, first and fifth time showed; VAS fifth time showed). Due to the small sample size it was decided to winsorize the outliers. However, winsorizing the outliers did not change the results compared to keeping all data in the analysis.

Manipulation

The Helmert contrasts showed that detected by FaceReader and during the sadness video, the emotion sadness was significantly more expressed than the two other emotions [F (24,2) = 109.64, $p < .001$). During the disgust video, the emotion sadness was significantly more expressed than the two other emotions [F (24,2) = 50.92, p = < .001]. Additionally, during the fear video, the emotion sadness was also more expressed than the two other emotions [F (24,2) = 135.73, $p < .001$]. During the neutral video, the emotion sadness was significantly more expressed that the two other emotions [F (24,2) = 231.81, p = < .001]. The

Helmert contrasts showed that detected by fEMG and during the sadness video, the emotion sadness was more expressed than the two other emotions. However, this result was not significant [F (24,2) = 2.65, $p = .091$]. During the disgust video, the emotion sadness was significantly more expressed than the two other emotions [F (24,2) = 5.63, $p = .010$]. During the fear video, the emotion sadness was more expressed than the two other emotions.

However, this result was not significant [*F* (24,2)=2.14, *p* = .139]. During the neutral video, the emotion sadness was more expressed that the two other emotions. However, this was not significant [*F* (24,2) = 3, *p* = .069].

The Helmert contrasts showed that detected by VAS and during the sadness, the emotion sadness was significantly more experienced the two other emotions [*F* (24,2)=24.95, *p* < .001]. During the disgust video, the emotion disgust was significantly more experienced than the two other emotions [*F* (24,2) = 158.48, *p* < .001]. During the fear video, the emotion fear was significantly more experienced that the two other emotions [*F* (24,2) = 23.85, *p* < .001]. During the neutral video, the emotion fear was significantly more experienced that the two other emotions [*F* (24,2) = 4.46, *p* = <.023]. According to the participants (VAS), the intended emotions were experienced. Meaning that the manipulation has worked. However, mere experience of an emotion does not necessarily imply that it also is expressed.

Research question

Sadness. Findings from the first repeated measures ANOVA suggested that participants showed no decrease in facial expression of sadness across time in the sadness film for all three measurement methods [*F* (1,25) = 1.63, *p* = .061]. See Table 1 for the means and standard deviations. Moreover, a main effect of measurement was found [*F* (2,50) = 35,89, *p* = < .001]. This main effect imply that the measurements significantly differ from each other on the expression of sadness in the sadness film. Additionally, a significant Time by Measurement-type was found [*F* (2,24) = 3.61, *p* = .043]. This interaction effect imply that the measurements significantly differ on the average progress score of the emotion sadness in the sadness film. The significant main effects of measurement for FaceReader and fEMG were further tested with paired t-tests. This in order to check whether the same emotions at Time point 1 (first time showed) significantly differ between FaceReader and fEMG for each film clip. Findings from the paired t-tests showed that FaceReader and fEMG significantly differ on the expressions of sadness and disgust in the sadness film clip. These results can be seen in Table 2.

Table 1

The Means and Standard Deviations for Sadness, Disgust, Fear Film Clips at Two Time Points for FaceReader, fEMG and VAS.

	FaceReader				fEMG				VAS			
	Time 1		Time 5		Time 1		Time 5		Time 1		Time 5	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Sadness	1.215	.598	1.492	.555	.3106	1.285	-.072	.747	1.221	.731	.832	.745
Disgust	-.725	.174	-.077	.118	1.36	2.037	-.131	.742	1.94	.577	1.482	.852
Fear (vid)	-.588	.135	-.568	.143	-.188	.822	.051	.923	.994	.761	-.469	.399

Table 2

Results of the Paired T-Tests: t and df on the Differences Between FaceReader and Facial EMG as well as the Means and Standard Deviations for FaceReader and fEMG for the Three Emotions (Sad, Disgust, Fear) in the Four Film Clips (Sad, Disgust, Neutral, Fear).

		95% CI for Mean Difference				FaceReader		fEMG	
		Lower	Upper	*t*	*df*	*M*	*SD*	*M*	*SD*
Sad clip	Sad	.925	1.726	6.815**	25	1.215	.598	-.111	.847
	Disgust	-.796	-.111	-2.725*	25	-.752	.222	-.299	.795
	Fear	-.602	.184	-1.094	25	-.486	.369	-.277	.852
Disgust clip	Sad	-1.482	-.116	-2.408*	25	.971	.846	1.77	1.699
	Disgust	-2.895	-1.318	-5.502*	25	-.727	.171	1.379	1.998
	Fear	-1.289	.061	-1.873	25	.475	.388	.139	1.591
Fear clip	Sad	.65	1.383	5.716**	25	1.125	.455	.366	.843
	Disgust	-.917	-.169	-2.996*	25	-.756	.194	-.213	.847
	Fear	-.734	-.065	-2.461*	25	-.588	.135	-.188	.822
Neutral clip	Sad	.446	1.071	4.993**	25	1.234	.575	.222	.941
	Disgust	-1.144	-.387	-4.162*	25	-.757	.123	.008	.926
	Fear	-.682	.034	-1.863	25	-.626	.135	-.302	.867

* = $p < .05$, ** = $p < .001$

Fear. Findings from the second repeated measures ANOVA suggested that participants showed a decrease in the facial expression of fear across time in the fear film for all three measurement methods [F (1,25) = 21.05, p= < .001]. Moreover, a main effect of measurement was found [F (2,50) =15.69, p = <.001]. This main effect imply that the measurement methods significantly differ from each other on the expression of fear in the fear film clip. Additionally, a significant Time by Measurement-type was found [F (2,50) = 54.13, p = <.001]. This interaction effect imply that the measurements significantly differ on the average progress score of the fear emotion in the fearful film. Findings from the paired t-tests showed that FaceReader and fEMG significantly differ on the expressions of sadness, disgust and fear in the fear film clip.

Disgust. Findings from the third repeated measures ANOVA suggested that participants showed a decrease in facial expression of disgust across time in the disgust film for all three measurement methods [F (2,24) = 19.6, p = < .001]. Moreover, a main effect of measurement was found [F (2,48)= 58.72, p = < .001]. This main effect imply that the measurement methods significantly differ from each other on the expression of disgust in the disgust film clip. Additionally, a significant Time by Measurement-type was found [F (2,48) = 12.68, p = < .001]. This interaction effect imply that the measurements significantly differ on the average progress score of the disgust emotion in the disgusted film. Findings from the paired t-tests showed that FaceReader and fEMG significantly differ on the expressions of sadness and disgust in the disgust film clip.

Neutral. Findings from the paired t-test showed that FaceReader FaceReader and fEMG significantly differ on the expressions of sadness and disgust in the neutral film clip.

Discussion

The current study employed a multi-method design to examine differences between three measurements methods, FaceReader, fEMG and VAS, in detecting facial expressions. Moreover, the accuracy of FaceReader, fEMG and VAS was examined based on their distinctiveness. This was done by measuring participants' emotional expression responses to emotion inducing film clips. Hypothesized was that FaceReader and fEMG would show similar outcomes on the facial expression scores, comparable with the VAS outcome.

An important first finding was that the emotion expression responses detected by FaceReader, fEMG and VAS individually, significantly differed from each other. Moreover, FaceReader and fEMG significantly differ in their facial expression outcome of sadness and disgust in all four film clips (sadness, disgust, fear and neutral). Additionally, FaceReader and fEMG did significantly differ in their facial expression outcome of fear in the fear film clip, but did not in the other three film clips (sadness, disgust and neutral). The findings are remarkable since prior work shows that both measurements, fEMG and FaceReader, are respectively sensitive and accurate tools for measuring facial emotions (Dimberg, 1990; Den Uyl & van Kuilenburg, 2005). Facial EMG, specifically, has established itself as a reliable and sensitive emotion detection tool (Fridlund & Cacioppo, 1986; Hess, 2009; Tassinary et al., 2012).

Another finding of this study was that FaceReader and fEMG could observe a decrease in the facial expression outcome of the three emotions across time. This finding provides us more information about FaceReaders' capabilities. In the study of Fanti et al. (2017), where no decreases of facial expressions across time were expected, the participants indeed showed similar expressions across time. Both results are in favour of FaceReaders' ability to detect facial expressions differences over time.

However, as well as the two other measurement methods, it seems FaceReader is not capable of distinguishing between the emotion expressions sadness, disgust and fear. In this study, Helmert contrast analysis suggested that solely the sadness film clip significantly evoked more sadness expression on the faces of the participants, compared with the remaining two emotions detected by FaceReader. However, the other film clips also evoked significantly more sadness expression on the faces of the participants, compared with the remaining two emotions detected by FaceReader. For fEMG, the same results were found. To summarize, no significant effects were found between the film clip and its intended expression of emotion. Participants were specifically more likely to express emotions of sadness in response to the film clips.

Previously discussed findings suggest that the intended facial emotions were not expressed by the participants. However, participants reported on the VAS that they did experience the intended emotions. This means that the participants self-report is inconsistent with the outcomes of FaceReader and fEMG. As indicated before, the connection with the electrodes to the recording equipment may have impeded with natural and spontaneous behaviour of the participants (van Boxtel, 2010). Another possibility for this contradiction is that the wires of the electrodes may have negatively affected the FaceReader program's ability to assess specific facial responses. Furthermore, the self-report may be biased by such factors as social desirability concerns, because it was quite obvious to participants which emotions they 'should have' felt when watching each movie (Churchill & Behan, 2010).

On the one hand this discrepancy could be the cause of limitations of the three measurement methods. On the other hand the film clips could not have caused emotional expression. The clips may be too outdated and new clips need to be used for a modern audience. While these film clips have reliably elicited emotion in the past (Gross & Levenson, 1995; Schaefer et al., 2010), emotional connection to the characters may require a longer film clip. Moreover, the scary film clip included one single scary moment leading to one emotional peak. When averaging the total score of the expressed emotions within each film clip, this could have caused a distorted outcome. Lewinski et al. (2014) took for each participant the average score of the top 10% peak values for all facial expressions. They chose this approach in order to analyse the most prominent facial expressions and took into account the frequency of their occurrence during exposure. With this approach the outcomes of the present study could be more distinct.

In order to explain the contradictory finding that the expression responses detected by FaceReader, fEMG and VAS significantly differed from each other, it is possible that the sample size and the power were too small to find statistically significant effects. Some of the effect sizes were considerably big even though no significant effect was found. Thus, it could be that findings from earlier studies are true and could have been confirmed with this study if the sample size was bigger.

Positive aspects of this study should also be noticed. In this study, FaceReader and fEMG were directly compared to each other, limiting the in-group differences. Comparing fEMG and FaceReader directly has not been done in prior work. Research of Fanti et al. (2017) did combine the FaceReader and fEMG data, however they were retrieved from separate participant groups. Another advantage is that the stimuli used in the current study were more realistic compared to static emotional pictures that were used in study of Den Uyl & Kuilenborg (2010), creating a more naturalistic setting.

As described earlier, there are several limitations regarding the use of fEMG, such as its long preparation time, its impediment with natural responses and its high expense (van Boxtel, 2010). FaceReader outperforms fEMG in these regards and has some great advantages with concerning its accessibility and wide reach. Therefore, it would be valuable if FaceReader and fEMG showed matching measurement outcomes. Prior research findings of FaceReader were limited but very promising, because high correlations between FaceReader measures and participants' self-reports were found (Lewinski et al., 2014).

It is clear that according to this current study, it is too early to introduce FaceReader into clinical practice. However, based on current findings, FaceReader has some prospects. Namely, it can detect differences in muscle activation over time of negative valence. This could suggest that FaceReader can distinguish between negative and possibly positive emotions, but is not ready to distinguish between sadness, disgust and fear. Further research is needed to give an insight in this matter and to overcome the limitations of the current study. This study was designed to test only the similarities between FaceReader, fEMG and VAS and to give an insight into their accuracy, but further experiments could aim toward: (a) improvements of the program FaceReader (b) cross-validating FaceReader against other measures (e.g. PrEmo or AdSAM®); (c) testing different kind of emotion eliciting stimuli (e.g. longer and newer clips); (d) create more naturalistic and comfortable settings (e.g. at home); (e) a bigger sample size.

A general downside of measuring facial expressions is that the human face does not only show affective responses, but also various behavioural activities that are not related to emotional expressions like mental fatigue, performance motivation and speech (van Boxtel, 2010). This can result in biased internal consistency estimates. Moreover, even in experimental studies it is difficult to control for these disturbing influences, let alone during practical applications. This can be regarded as an overall limitation within facial emotion detection and therefore hard to control.

To conclude, this was one of the first studies in which FaceReader was directly compared to fEMG and moreover, a first examination of FaceReader as a useful measurement into clinical practice. This way the current study contributed to the emotion detecting technique discussion and gives an insight into the remaining questions to address in future research.

References

Brosch, T., Pourtois, G., & Sander, D. (2010). The perception and categorisation of emotional stimuli: A review. Cognition and Emotion, 24(3), 377-400.

Churchill, A., & Behan, J. (2010). Comparison of methods used to study consumer emotions associated with fragrance. Food quality and preference, 21(8), 1108-1113. doi: 10.1016/j.foodqual.2010.07.006

Darwin, C. (1998). The Expression of the Emotions in Man and Animals, 3rd edit.

Den Uyl, M. J., & Van Kuilenburg, H. (2005). The FaceReader: Online facial expression recognition. In L. P. J. J. Noldus, F. Grieco, L. W. S. Loijens, & P. H. Zimmerman (Eds.), Proceedings of Measuring Behavior (pp. 589–590). Noldus Information Technology, Wageningen B.V., The Netherlands.

Dimberg, U. (1982). Facial reactions to facial expressions. Psychop/zysiology, 19, 643-647. doi: 10.1111/j.1469-8986.1982.tb02516.x

Dimberg, U. (1990). Facial electromyography and emotional reactions. Psychophysiology. doi: 10.1111/j.1469-8986.1990.tb01962.x

Donner J. (2008): Research Approaches to Mobile Use in the Developing World: A Review of the Literature. Information Society. 2008, 24: 140-159. doi: 10.1080/01972240802019970.

Egan, G. (2010). The skilled helper: A problem-management and opportunity-development approach to helping (9th ed.). Belmont, CA: Brooks/Cole.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. Journal of Personality and Social Psychology, 17(2), 124-129. doi: 10.1037/h0030377

Elfenbein, H. A., Der Foo, M., White, J., Tan, H. H., & Aik, V. C. (2007). Reading your counterpart: The benefit of emotion recognition accuracy for effectiveness in negotiation. Journal of Nonverbal Behavior, 31(4), 205-223. doi: 10.1007/s10919-007-0033-7

Fridlund, A. J., *&* Cacioppo, J. T. *(*1986*).* Guidelines for human electromyographic research*.* Psychophysiology., 23, 567–589. doi*:* 10.1111/j.1469-8986.1986.tb00676.x

Fanti, K. A., Kyranides, M. N., & Panayiotou, G. (2017). Facial reactions to violent and comedy films: Association with callous–unemotional traits and impulsive aggression. Cognition and emotion, 31(2), 209-224. Doi: 10.1080/02699931.2015.1090958

Greenberg, L. S., Rice, L., & Elliott, R. (1993). Process-experiential therapy: Facilitating emotional change. New York: Guilford.

Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. Cognition & emotion, 9(1), 87-108. doi: 10.1080/02699939508408966.

Guerrero, J. F., Bataller, M., Soria, E., & Magdalena, R. (2007). Biolab: An educational tool for signal processing training in biomedical engineering. IEEE Transactions on Education, 50(1), 34-40. doi: 10.1109/TE.2006.886463

Hess, U. *(*2009*).* Facial EMG. *In* E. Harmon-Jones *&* J. S. Beer *(Eds.),* Methods in social neuroscience *(pp.* 70–91*).* New York, NY*: Guilford Press.*

Hurley, R. F. (2011). The decision to trust: How leaders create high-trust organizations. John Wiley & Sons.

Lewinski, P., Fransen, M. L., & Tan, E. S. H. (2014). Predicting advertising effectiveness by facial expressions in response to amusing persuasive stimuli. Journal of Neuroscience, Psychology, and Economics, 7(1), 1-14. doi: 10.1037/npe0000012

Miller, M. D., & Ferris, D. G. (1993). Measurement of subjective phenomena in primary care research: the Visual Analogue Scale. Family Practice Research Journal.

Van Boxtel, A. (2010). Facial EMG as a tool for inferring affective states. In Proceedings of measuring behavior (pp. 104-108). Wageningen: Noldus Information Technology.

Tassinary, L. G., Hess, U., *&* Carcoba, L. M. *(*2012*).* Peripheral physiological measures of psychological constructs. APA handbook of research methods in psychology, Vol 1: Foundations, planning, measures, and psychometrics. *(pp.* 461–488*):* Washington, DC*: American Psychological Association.*

Schaefer, A., Nils, F., Sanchez, X., & Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. Cognition and Emotion, 24(7), 1153-1172. doi: 10.1080/02699930903274322

Sloan, D. M., & Kring, A. M. (2007). Measuring changes in emotion during psychotherapy: Conceptual and methodological issues. Clinical Psychology: Science and Practice, 14(4), 307-322. doi: 10.1111/j.1468-2850.2007.00092.x

de Wijk, R. A., Kooijman, V., Verhoeven, R. H., Holthuysen, N. T., & de Graaf, C. (2012). Autonomic nervous system responses on and facial expressions to the sight, smell, and taste of liked and disliked foods. Food Quality and Preference, 26(2), 196-203. doi: 10.1016/j.foodqual.2012.04.015