Remote Photoplethysmography on Publicly Available Videos of Soccer

Managers in Neutral and Active Emotional States

Bart Haak

0899809

Supervisor: Daniël Lakens

Eindhoven University of Technology

Abstract

Remote photoplethysmography (rPPG) is a relatively new method to extract heart rate with. It uses colour differences of the face to determine a person's heart rate. To this date, most studies about this technique were set up in labs. The current study, however, used rPPG from FaceReader on publicly available YouTube videos of four well-known soccer managers. The general performance of rPPG was analysed on these videos. Furthermore, it was examined whether rPPG is able to detect higher heart rates during active emotion (happy, sad or angry) compared to neutral emotion. This effect was expected based on previous research. For both emotional states, 100 videos of 30 seconds were collected of either an interview or press conference. The raw rPPG output suffered from inaccuracy, but the output was more realistic after manual filtering of heartbeat peaks. Nonetheless, only 160 videos were good enough for analysis and variances of the heartbeat intervals were unexpectedly large. Furthermore, no significant effect of active versus neutral emotion was detected, even though mean active heart rates were higher. Overall, rPPG performed too poorly to accept the validity of the results. Face detection was sufficient, but registration of heartbeats still had its issues. Conclusively, rPPG needs more development before public videos can be used for reliable heart rate analysis.

Introduction

In today's world videos are everywhere. There are thousands of movies, television series and television shows. Not to mention 300 hours of new video content which are uploaded to YouTube every minute ("YouTube Company Statistics," 2016). All these video data is mostly consumed rather than utilized as a source for data analysis. In the current research, this is not the case, as information gets extracted from public videos. More specifically, a new technique called remote photoplethysmography (rPPG) derives heart rate (HR) from public figures appearing in YouTube videos.

Remote Photoplethysmography

Heart rate measures. There are different techniques to measure physiological parameters. The electrocardiogram (ECG) measures the cardiac pulse through gel the person wears or straps attached to the chest. Other techniques require a person to wear clips with sensors to earlobes or fingertips. A more convenient way to measure one's pulse is through photoplethysmography (PPG). As blood absorbs more light than surrounding tissue, the reflectance of light depends on blood volume variation (Wim Verkruysse, Svaasand, & Nelson, 2008). PPG utilizes this property to measure cardiac signals. The method is relatively cheap, non-invasive and easy to use. Nonetheless, the technique requires physical contact with the skin. PPG was already introduced in 1937 (Hertzman & Spealman), but more recently, research focuses on the development of a new type of PPG, called remote PPG.

The technique. Remote photoplethysmography is similar to PPG, except that no contact with the skin is necessary. The idea of contactless PPG was examined in 2005 (Wieringa, Mastik, & Van der Steen). A CMOS monochrome camera shot video images of the participant's skin located at the left-inner arm. It retrieves well-correlated signals with the true heart rate but suffers from noise and is not proven to work with diversity in skin type. A few years later standard digital cameras were used to extract respiration and pulse rates by only using ambient light (Takano & Ohta, 2007; Verkruysse et al., 2008). However, these studies lack convincing evidence. Poh, McDuff and Picard (2010) show rPPG its real potential through successful results based on physiological and mathematical models instead of the previous manual and heuristic interpretations. Moreover, Poh et al. (2010) implement rPPG such that it can handle slight head movements of a person.

Techniques suggested back then are largely still in use today. The study of Poh et al. (2010) describes the complete method to extract pulse signals. Many studies follow a similar approach, like Y. Hsu, Lin and W. Hsu (2014), Kublanov and Purtov (2015) as well as Rouast, Adam, Dorner and Lux (2016). First, a face tracker detects a face. Subsequently, this face is analysed by software which locates the regions of interests (ROIs) on the faces. ROIs consist out of the cheeks, the forehead or both. Colour information at ROIs gets separated into the three RGB (red, green and blue) channels. Post-process, these wavelength signals are transformed with a Blind Source Separation (BSS) technique to uncover the underlining cardiac signal.

The advantages. Remote PPG has some significant advantages over normal PPG, besides that no physical contact is necessary. Verkruysse and Bodlaender (2010) mention that rPPG can be used with little to no cooperation of the subject. Which implies that people can focus on their task without any disturbance. Secondly, it can be applied to existing video footage or it could be integrated into cameras (Verkruysse et al., 2010), since a video with ambient light and moderate technical specifications is enough to measure a heart rate signal. The current study exploits these advantages.

Used software

The present study utilizes the method proposed by Tasli, Gudi and Den Uyl (2014b), which uses a close resemblance of the above-mentioned technique. A software program called FaceReader (Den Uyl & Van Kuilenburg, 2005) detects faces and tracks ROIs through facial landmarks produced by the active appearance model (AAM) technique (Edwards, Taylor, & Cootes, 1998). After correct face detection, landmark recognition and colour acquisition of the ROIs, the next step processes the colour data obtained. Finally, a noise-free signal is generated to determine the actual heart rate. Results from FaceReader and Poh et al. (2010), both in preferred recording conditions, are comparable. The average error of resting heart rates is 2.9 beats per minute (bpm) and 2.29 bpm respectively. In the study of Poh et al. (2010) performance decreases in videos with moving faces, with an average error of 4.63 bpm.

From the technical description follows that colour differences should be clearly visible in order to obtain successful results. Unsurprisingly, bad lighting is the main limitation of FaceReader (Tasli, Gudi, & Den Uyl, 2014a). Furthermore, rPPG is in general possibly less accurate for darker skin (Wang, Stuijk, & De Haan, 2014). Besides optimal lighting and skin tone for good detection of colour differences, other settings are recommended for optimal performance. Based on extensive internal testing by the creators of FaceReader (VicarVision), the resolution should be 1280x720 pixels with a frame rate 30 frames per second (fps). The distance of the camera to a person should preferably be $0.5 - 1.0$ meters. Finally, the video should ideally have a frontal view (less than a 40° angle between face and camera) of the person without sudden head movements. Nonetheless, according to VicarVision, not all these video settings are required for a reliable signal. For instance, the minimum resolution can be 640x480 pixels with 15 fps and faces can be detected when they just make up for 8% of the video image size. Furthermore, FaceReader performs relatively well with head movements, since these movements are largely compensated for due to sophisticated facial landmark recognition (Tasli et al., 2014b). Hence, in videos with movement, the heart rate has presumably a better error than 4.63 bpm.

A limitation that stays present is the 8.53 seconds FaceReader needs to initialize any heart rate output. Even though settings allow to adjust this window size to 4.27 seconds, this smaller window size provides less accurate results according to VicarVision.

Application of rPPG

Public videos. Remote PPG is well suited for existing public videos, according to advantages of rPPG (Verkruysse et al., 2010). Furthermore, FaceReader its above-mentioned video requirements are easily met. Up to this date, however, rPPG on public videos is a scarcely utilized application. Only brief analyses on existing public videos, consisting of well-known people, are conducted (Tasli et al., 2014a; Verkruysse et al., 2010). The current study takes this opportunity to more elaborately examine rPPG on public videos. As a proof of concept 15 YouTube videos are collected for brief analysis. The videos, consisting of iconic television moments that are widely known across the world, are analysed with FaceReader. Examples are shown in Figures $1 - 4$, but all results and links to the videos can be found in Appendix A and Appendix D respectively. Thirteen videos result in enough output to determine HR with. Which proofs the concept that rPPG can be used on public YouTube videos.

Figure 1. Heart rate during Bill Clinton's infamous lie about his sexual relations with Lewinsky.

Figure 2. Heart rate of Günter Schabowski announcing the fall of the Berlin Wall.

Figure 3. Heart rate of David Cameron during his resignation speech after the Brexit result.

Figure 4. Heart rate of Michael Jackson during the allegedly most watched interview of all time ("The Michael Jackson Interview: Oprah Reflects," 2009) with Oprah.

Emotional differences. Heart rate measurement techniques can be used for multiple purposes. In this study, the public videos are used to detect differences in emotion. Ekman, Levenson and Friesen (1983) show that reliving a moment of anger and fear let the HR increase with 8 bpm on average, followed by sadness with almost a 7 bpm increase. Happiness only results in a 2.6 bpm increase, while surprise and disgust even cause lower HR changes. In replication studies with minor adaptions (Boiten, 1996; Levenson & Ekman, 2002) and a metaanalysis on the relation of emotion and physiological signals (Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2000) similar patterns are presented. And not only quick emotions but emotional states over a longer period of time (moods) are possibly related to HR as well. Nevertheless, results point in slightly different directions. Schubert (1977) investigated pulse and moods according to the Profile of Mood States (POMS). Five moods were measured: fatigue, anxiety, confusion, depression and aggression. Fatigue and confusion show significant negative correlations with HR, but the other mood states have no convincing relation. Pollock, Cho, Reker and Volavka (1979) conducted similar research, but have positive correlations on fatigue and confusion. Moreover, their results show significant positive correlations on depression, aggression and anxiety. Shapiro, Jamner, Goldstein and Delfino (2001) measured moods differently. Their study shows that HR increases with more aggression and anxiety but decreases when tired. They only used women in the study, who may have stronger emotions than men (Barrett, Robin, Pietromonaco, & Kristen, 1998). In the end, the exact effect of moods on HR is difficult to define. If the mood is an active state, however, HR is likely to increase according to the PA-NA model of positive (e.g. interested, excited) and negative (e.g. sad, frustrated) affect (Ilies, Dimotakis, & Watson, 2010).

Studies about HR and emotional states are mostly conducted in a controlled laboratory environment since PPG or other typically used HR measures need preparation. However, according to Sebe, Cohen, Gevers and Huang (2005), a close resemblance of a natural situation is desired to measure emotions, which can be achieved by measurements in real-world environments rather than in a lab setting. It can be deduced that rPPG is a great tool to measure emotions with since it can be used without the subject knowing of its existence in real-life environments. Hence, rPPG can be useful to extract HR differences between neutral and active emotions in videos in real-life settings.

Purpose of this study

In this study, the potential of rPPG is explored by exploiting the advantages and possibilities of rPPG. Public videos are collected and emotions are measured with rPPG in the non-lab environments. More specifically, it is questioned whether, compared to neutral HR, the cardiac pulse increases in active emotional states of anger, sadness or happiness in soccer managers. These managers appear in publicly available videos of interviews and press conferences. Professional soccer managers are relatively suitable for rPPG analysis since they give many interviews and press conferences and sit or stand mostly still in front of good quality cameras. What's more, the match results influence emotional states after a match (Crisp, Heuston, Farr, & Turner, 2007; Hanin, 2007). According to results of HR increase during active emotional states (Cacioppo et al., 2000; Ekman et al., 1983; Ilies et al., 2010) and the reasonable error rates of rPPG (Tasli et al., 2014b), it is expected to obtain accurate heart rates and to see an effect of active emotional states on HR. Active emotion may even lead to higher HR increases then seen from Ekman et al. (1983) because videos are recorded in non-lab settings (Sebe et al., 2005). Nevertheless, the extent to which interviews and press conferences are similar to a real-world environment as described by Sebe et al. (2005) is questionable.

In the end, this study indicates the possibilities of rPPG on videos that are public domain. Since previous research investigated more artificial environments, the quality of the rPPG signal of the public videos is extensively described. Subsequently, HR increase during active emotion is tested to discover rPPG performance on a more complex task. Eventually, it is concluded whether rPPG gives a proper signal on public videos and whether the algorithm is able to detect emotional activation.

Method

Design

The current study researched HR in soccer managers in different emotional states. The emotional state was the examined independent variable with values 'neutral emotion' and 'active emotion', of which the latter was subdivided into three states of 'anger', 'sadness' and 'happiness'. Other independent variables that were not included in the statistical analysis include video resolution, frame rate, face proportion with respect to the video image, face angle with respect to the camera, face movement, lighting conditions, type of background, participant's age and participant's skin tone. Heart rate was the dependent variable measured in beats per minute (bpm). This study used multiple videos per manager for both emotional states, so HR variations between emotions within individuals were measured. Furthermore, HR between participants was expected to differ. Therefore, the structure of the data is hierarchical.

Participants

Videos of four publicly known male soccer managers were collected (Figure 5). For each manager, 50 videos were collected (25 videos in neutral and active state each). In some cases, videos of the same person were recorded in different years (mean variance of age within a person was $.62 \pm 0.33$ years old), so age differed within participants as well as between them. Hence, for each participant the average age was calculated, which led to a calculation of the mean age of all participants (53.8 \pm 7.4 years old). All four managers were considered to be white-coloured. Still, some slight variations in skin tone were present.

Participants were nationally or internationally considered to be well known. Moreover, they appeared on publicly available video material. During the recording, participants knew that they were broadcasted on national television, but did not know about their participation in the present study.

Figure 5. Participants of research. With, from left to right, Guardiola and Koeman at the top and Mourinho and Van Gaal at the bottom.

Materials and setting

Materials. An alpha version of FaceReader 8 was used for analysis. The rPPG algorithm of the software obtained HR signals from the participants. The participants themselves were recorded by professional cameras. Nonetheless, more information of the cameras was unknown. All the recordings were collected on YouTube. The YouTube videos were downloaded with an online tool [\(https://www.onlinevideoconverter.com/video](https://www.onlinevideoconverter.com/video-converter)[converter\)](https://www.onlinevideoconverter.com/video-converter). A pilot video analysis was performed to correlate HR from a raw video with HR of the same video uploaded to YouTube and subsequently downloaded from YouTube with the above-specified converter tool. Best correlations were found when the YouTube video was downloaded in MP4 format $(r = 0.98)$. The number of peaks in the raw video (101 peaks) differed from the downloaded video (96 peaks). Videos were downloaded in AVI format as well, but HR was less correlated $(r = .96)$ and the peak quantity differed to a larger extent (109) peaks).

Setting. The setting differed slightly across videos. All videos were either an interview $(n = 40)$ or press conference $(n = 160)$. In an interview, the participant was standing, while he sat down in a press conference. Almost all videos had a static wall with company logos in the background (Figure 5), which is typical for these interviews and conferences. Eight videos differed in this aspect. In six of these, the background was static but different from the typical sponsored wall. Two videos were recorded on the pitch, which resulted in a slightly more dynamic background (Figure 6). Overall, lighting conditions were consistent within videos and in all videos the face was well lit. Still, the light conditions differed to some extent across videos (Figure 6).

Figure 6. From left to right: an interview with a pitch as background, a press conference with a highly lit face and small face-to-screen proportion, an interview with a normally lit face.

Procedure

Video selection. Managers had to appear in the two emotional conditions. Nonetheless, there was limited information on the actual feeling of participants. Therefore, assumptions were made on those feelings based on the context of the video. From results found by Crisp et al. (2007), happiness was assumed to increase after important wins and anger or sadness after important losses. Besides these general assumptions, the emotional distinctions were also based on the verbal content, vocal emotion classifications (Murray & Arnott, 1993) as well as visual indicators (Ekman & Friesen, 2003). In some cases, videos unrelated to the soccer matches were used, as long as they had clear verbal or non-verbal emotional indicators. Generally, neutral states were found from videos unrelated to results of a match where no emotional indicators were apparent. Finally, to ensure the independence of HR per video, videos were not allowed to be recorded on the same date.

Within the interview or press conference itself, 30 seconds of video was used to measure HR on. Similar to the research of Ekman et al. (1983) where participants relived an emotional state for 30 seconds. Which 30 seconds of the complete video to use for further analysis was determined according to two criteria. First, the displayed content needed to be approximately suitable for the algorithm. Therefore no blockages of the face were allowed to be present during the entirety of the 30 seconds. Furthermore, the angle between face and camera needed to be approximately less than 40° for the majority of the video. Secondly, if the video fragment fitted the first criterion, the fragment with the strongest apparent emotion was taken. In case of the neutral emotional state, the fragment with the least apparent emotion was chosen.

The participants in the selected videos stayed steady but moved their head on some occasions. Therefore, the angle of the face direction with respect to the camera differed across videos and within the video itself. Additionally, the face-to-screen proportion differed for each video (Figure 6), even though the face was detected in all videos. It was tried to minimize the differences. Nevertheless, the videos were recorded in real-world environments, so within a video as well as between videos, the deviations which appeared were expected.

Videos had a resolution of $1280x720$ pixels and $23 - 30$ fps ($M = 25.11$ fps, $SD = .74$ fps), of which most (96%) were 25 fps. From test videos, it was concluded that the HR outcome did actually differ between 25 fps and 30 fps videos (approximately 2 bpm difference in mean HR). Nonetheless, with YouTube as a database, it was not possible to collect enough 30 fps videos in time. It was assumed that using the slight frame rate inconsistencies did not influence outcomes of the research significantly.

Sample size determination. The necessary amount of participants was estimated according to an A priori power analysis with the data analytical software G*power and simulation data. The expected mean difference of HR in active and HR in neutral emotional conditions ($\mu_{diff} = 5.67$ bpm), along with its standard deviation ($\sigma = 2.78$ bpm) were calculated from the simulation data. The data was based on differences of active emotional states compared to neutral states from results provided by Ekman et al. (1983). The generated simulation data consisted of cases where happy ($\mu_{\text{diff}} = 2.6$ bpm, $\sigma_{\text{diff}} = 1.0$ bpm), sad ($\mu_{\text{diff}} = 7$ bpm, $\sigma_{diff} = 1.5$ bpm) and angry ($\mu_{diff} = 8.0$ bpm, $\sigma_{diff} = 1.8$ bpm) equally contributed. It was highly expected that an active state raised HR compared to a neutral state, so only a one-tailed test was considered. Finally, a significance level and power were indicated ($\alpha = 0.05$, $1 - \beta =$ 0.90) to enable sample size indications. From this information a sample size of $n = 4$ was calculated to be necessary, thus heart rates of four different managers were used.

The 25 videos per participant in the same emotional state were collected to expand HR measurements for each condition. Additionally, the use of multiple videos contributed to balance out the possible signal noise. The number of videos per condition was estimated according to a simple sample size calculation. First, the standard deviation of rPPG was estimated. The used videos were assumed to contain some movement, but the movement was expected to be better tracked with FaceReader than in the study of Poh et al. (2010). Consequently, the presumed standard deviation was set to 3.8 bpm, which is the rounded up mean of 4.6 bpm (Poh et al., 2010) and 2.9 bpm (Tasli et al., 2014b). The underlying heart rate within the neutral condition was assumed to be the same. Similarly, the underlying HR within the active condition was presumably the same as well. Besides, it was assumed that both emotional conditions were distributed normally. A sample size calculation for normal distributions was performed to calculate the necessary video quantity. From the deduced standard deviation (σ = 3.8), the z-score (z = 1.96) and a margin of error (E = 1.5) the number of videos was calculated to be equal to 24.7. As a consequence, the video quantity for the neutral and the active state was set to 25 videos each. Table 1 shows the summary of collected videos. Appendix D gives links to all collected video material.

Table 1

Number of videos per participant neutral and active states

Note. $A =$ angry, $S =$ sad and $H =$ happy.

Measurements and Statistical analysis

The measurements were conducted by the rPPG algorithm of FaceReader. The software converted all videos to 30 fps and to a resolution of 1024x1024 pixels. Furthermore, the window size was set to 8.53 seconds. Similar to the videos shown in the proof of concept (Figures $1 - 4$), the forehead of the participants was the region of interest. These settings were chosen according to recommendations of VicarVision, based on internal testing results.

From the FaceReader output, data was described to deduce general rPPG performance. Subsequently, the effect between the neutral and active state was determined by a comparison of the neutral HR mean and the active HR mean whilst accounting for the HR differences between the four managers themselves. More specifically, hierarchical linear modelling was

performed on the videos. The corresponding *t-*value and *p-*value were reported. As an alpha value, the standard of $\alpha = .05$ was used.

Data selecting methods

Per video, the rPPG algorithm outputted a CSV file with three variables showing the video time stamps, their corresponding heartbeat peak values (either 0 or 1) and live calculated heart rates. From this output, there were several possible methods to calculate the heart rate with. These methods were judged on accuracy performances.

Raw output methods. There were three simple methods using the raw data output. The first method calculated heart rate from the number of peaks and the video duration. The accuracy judgment of this first method was solely based on the resulting heart rate. Average HR of 49 – 104 bpm were considered realistic, as the boundary values were three standard deviations from a previously found mean heart rate (Umetani, Singer, McCraty, & Atkinson, 1998).

The second method used the average interval time between heartbeats (RR interval) to calculate HR with. Performance on the mean RR interval as well as the standard deviation of RR intervals (SDNN) within a video was analysed. From the HR range found by Umetani et al. (1998) the range of reasonable average RR intervals was set to .58 – 1.22 seconds. Other research reported comparable ranges (Bigger et al., 1995; Stein, Kleiger, & Rottman, 1997) even though values were about 10 ms less for both boundaries. The range of Umetani et al. (1998) was used in order to be consistent with other ranges defined in the current study. Furthermore, based on 5-minute SDNN measurements conducted in previous research (Umetani et al., 1998), the SDNN values of $3 - 99$ ms were concluded to be realistic. Similar findings were presented in other research (Abhishekh et al., 2013; Bigger et al., 1995; Fei, Copie, Malik, & Camm, 1996). Nevertheless, SDNN is usually measured with a measurement time of at least 5 minutes and preferably 24 hours (Malik, 1996). Therefore, the SDNN values obtained from the 30-second videos were expected to slightly differ from the ranges fetched from the literature.

For the third method, the average of the live calculated heart rates and its standard deviation were looked into. Similar to the first method, the average was expected to fall within the range of $49 - 104$ bpm.

Filtered output methods. Based on expected ranges, the raw data failed to produce realistic results (Appendix B). Consequently, more sophisticated methods were needed. These methods consisted of several manual filtering methods based on RR intervals or the live calculated HR. Another method used transformed RR intervals by Kubios (Tarvainen, Niskanen, Lipponen, Ranta-aho, & Karjalainen, 2014) on two different threshold levels. Distinctions between different filtered methods were noticeable but less prominent than the difference to the raw data performances (Appendix B). Kubios was not chosen for further analysis since manual filtering allowed for more control in interval selection. Moreover, there was no profound reason to assume that Kubios resulted in more accurate results.

Used filter. The filter in this research made use of RR intervals since HR calculated from the RR intervals was more accurate than the live calculated HR according to VicarVision. For the filter, unreasonable RR intervals needed to be removed from the data. With a combination of realistic mean RR intervals and SDNN values (Umetani et al., 1998) a broad realistic RR interval range of .48 – 1.32 seconds was established. Graphs with all RR intervals as well as RR intervals that fell within this range were plotted (Appendix C). After analysis of these graphs, a filter on the realistic RR intervals was applied. In 165 of the 200 videos, there was at least one isolated realistic interval (i.e. no realistic interval directly pre- or succeeding it). The validity of the isolated intervals was questionable since values of surrounding RR intervals were untrustworthy. Conclusively, all RR intervals that were used

for HR calculation needed to be within the realistic interval range and needed to have at least one directly pre- or succeeding interval which was realistic as well. Of course, a sequence of at least three, four or five realistic intervals instead of just two might have led to better results. Nonetheless, sequences of at least two realistic intervals were appropriate for the current research, since any higher threshold would have caused significant data loss.

Results

Data cleaning

After applying the manual filter, some videos still resulted in unusable output. In general, SDNN values were considerably large ($M = 157$ ms, $SD = 44$ ms), considering the realistic range of 3 – 99 ms. Nevertheless, no videos were classified as an outlier at first since none were three standard deviations from the SDNN mean. The same applies for the mean RR interval and mean heart rate. The latter two were expected since extreme RR intervals were already filtered out.

Furthermore, the quantity and distribution of the filtered intervals within all videos were analysed. From these two indicators, it was concluded that in several videos the intervals were insufficient to calculate heart rate with. Firstly, if the filtered interval quantity was less than five, the video was disqualified for further analysis because the heart rate calculation would be based on sparse amounts of data. Secondly, a combination of the quantity and its distribution was examined. Generally, if there were less than 10 filtered intervals and the largest interval sequence was less than 4, the rPPG output of the video was insufficient to calculate HR with. This rule is not proven to be true but was applied because videos with few peaks and small sequences of realistic intervals were considered to have too unstable rPPG data for HR calculation. Nonetheless, in one case (video 104) this rule was ignored since the filtered RR intervals were in close proximity to each other and there was a relatively low

SDNN among them. In contrast, some other videos were left out even though the general rule did not apply. In the latter case, the rPPG output of the video was considered not to be valid based on quantity, distribution, SDNN values or a combination of the three. In total 40 out of 200 videos were classified as unusable. There possibly were better rules to disqualify certain videos, but the current outlier handling method took care of most unstable and unrealistic rPPG values while enough videos remained present for analysis.

Conclusively, 160 of the 200 videos were actually used in further analysis. Table 2 specifies the reasons why the other 40 videos got removed. Table 3 shows the video quantity per condition after video removal. For the remainder of this paper, outliers are excluded unless stated otherwise.

Table 2

Reasons why certain videos were not used on further analysis

Table 3

Number of videos per participant in neutral and active states after outlier removal

Note. $A =$ angry, $S =$ sad and $H =$ happy.

Data description

Measurement Time. The timestamp of the first measurement subtracted from the final measurement was defined as the initial measurement duration. Remarkably, FaceReader did not measure during the first 833 ms and final 866 ms in any collected video. Furthermore, within the boundaries of the first and final measurements, several data points were missing. If no data was lost, time intervals between consecutive measurement points were equal to 33 ms. Nevertheless, if the video was 25 fps, this time interval was transformed to 67 ms once every 200 ms because it needed to be converted to 30 fps. Hence, if the time between the measurements was longer than 67 ms then data was considered missing. The corresponding missing time was calculated if missing data occurred. This missing time was subtracted from the initial measurement duration to accomplish a final measurement time. On average $27.2 \pm$ 2.2 seconds out of the 30 seconds were actively used to measure the heart rate with. Figure 7 shows the boxplot distribution.

Figure 7. Boxplot of active measurement time. $(n = 160)$

RR-intervals. Several aspects of the RR intervals were explored. First, the number of RR intervals within a video. Due to the application of the manual filter, multiple intervals that appeared in the raw data were disregarded. Secondly, the mean RR interval and SDNN value were gathered per video. Finally, the corresponding heart rate was determined from the mean RR interval. Table 4 shows the description of the above-mentioned statistics. Additionally, the values from raw data are described as a means to compare the filtered method with.

Table 4

Description of data for raw and filtered data

Note. Statistics of the raw data are based on the second raw output method. For the raw as well as the filtered data $n = 160$.

Heart rates. The heart rate was most important for this research and hence described in more detail. Figures $8 - 12$ show plots of measured heart rates calculated from the manual filter for each video separately. The graphs give an indication of the heart rate variations within a video. They show heart rates calculated from each filtered RR interval in grey dots, with a line connecting the data points. The horizontal red line indicates the mean HR based on these intervals. The mean HR was the only relevant value for statistical testing.

Video 9

Video 15

Video 25

Video 30

Video 6

Video 11

Video 17

ाः

Video 23

Video 27

Video 33

Video 22

Video 41

N MÃ I

Figure 8. Heart rate plots for videos with ID $1 - 43$, excluding outliers. Time $(0 - 30 s)$ along the x-axis and heart rate $(40 - 125$ bpm) along the y-axis.

Video 7

Figure 9. Heart rate plots for videos with ID $44 - 82$, excluding outliers. Time $(0 - 30 s)$ along the x-axis and heart rate $(40 - 125$ bpm) along the y-axis.

Video 89

Video 93

Video 105

Video 109

Video 114

Video 119

λM ¥

バス

Video 87

Video 91

Video 99

Video 103

Video 107

Video 112

Video 116

t

Video 100

Video 104

Video 108

Video 113

Video 117

Video 88

Video 92

ΔŻ

Video 102

Video 106

Video 111

Video 115

Video 121

Figure 10. Heart rate plots for videos with ID 83 – 121, excluding outliers. Time $(0 - 30 s)$ along the x-axis and heart rate $(40 - 125$ bpm) along the y-axis.

¥

Figure 11. Heart rate plots for videos with ID $122 - 161$, excluding outliers. Time $(0 - 30 s)$ along the x-axis and heart rate $(40 - 125$ bpm) along the y-axis.

جوالا

Video 165

Video 175

Video 171

Video 177

Video 187

Video 191

Video 196

Video 173

Video 167

Video 178

Video 179

Video 185

1

Video 189

Video 193

Video 168

Video 174

Video 181

Video 184

Video 194

Video 200 ς,

٨r

ÅΛ

Figure 12. Heart rate plots for videos with ID $162 - 200$, excluding outliers. Time $(0 - 30 s)$ along the x-axis and heart rate $(40 - 125$ bpm) along the y-axis.

Statistical testing

The 81 videos of the neutral emotional state were compared to the 79 active state videos. The hierarchical linear model used for the statistical analysis resulted in slightly insignificant results, $t(155) = 1.86$, $p = .065$. With an approximately normal distribution of residuals. Nonetheless, data showed higher heart rates for active states in all four managers. From all data, the active state had a 1.72 bpm higher average heart rate than the neutral state. Table 5 shows the heart rates per manager in the neutral and active emotional state. Furthermore, most of the differences in HR were within a subject caused by the emotional state, since the differences between managers only accounted for 7.5 % of the variance.

Table 5

Mean heart rate in each condition

Note. Standard deviation indicated after the \pm sign.

Discussion

Explanation descriptive data

In the current study, the HR output from the rPPG algorithm performed poorly compared to previous research (Poh et al., 2010; Tasli et al., 2014b). In contrast to the prior studies, current videos were recorded in non-lab environments. So in general, the collected videos contain more movement, a worse face to camera angle and larger differences in

lighting conditions than lab videos. This section discusses specific reasons for the relatively poor performance.

Measurement time. Overall, the measurement time was sufficiently large. But Figure 7 shows some low measurement time values as well. There appear to be two causes of the low values. Firstly, missing data occurred if the face could not be found. A probable cause is a too severe angle between face and camera (Figure 13) since all videos with face detection errors lack a sufficient face angle. On the contrary, sometimes angles are severe but the face can still be detected (Figure 14). Therefore, it is deduced that face angle influences face detection, but that other factors affect this detection algorithm as well. Which specific factors contribute cannot be determined from the current study. However, movements do not seem to significantly affect the face detection in the collection of videos, as long as the face angle remains sufficiently small. Although, more extensive research should be done to confirm this theory. Furthermore, new studies can research angles in combination with other factors such as lighting conditions or contrast with the background to infer the minimally required settings in natural public videos.

Figure 13. Severe angles of face to camera, where no face was detected

Figure 14. Severe angles of face to camera, where the face was detected

Secondly, in four videos (1, 56, 83, 167) the final approximate eight seconds (similar to the window size) are measured by the software but not outputted to the CSV file. In all four cases, the final measurement point is missing due to the absence of face detection. For instance, if the face is not detected at $29 - 30$ seconds, the CSV file shows data up until 22 instead of 29 seconds. Most likely, this is a software bug which can be updated.

RR intervals. The interval quantity, as shown in Table 4, significantly decreases for the filtered data. The content of videos together with the peak graphs (Appendix C) are inspected in order to explain the large number of removed intervals. Despite no missing measurements, substantially long RR intervals appear in unfiltered video data. Nonetheless, even in extreme cases, no apparent cause can be found. For instance, video 108 has a large interval of $19 - 25$ seconds, but during this time the face is steady and frontal to the camera (Figure 15). Moreover, some instances which contain sudden head movements or a relatively large face to camera angle still result in seemingly valid peak output (e.g. within video 30, 78 or 190). Figures 15 and 16 illustrate the inconsistency. More elaborate research needs to be conducted to explain flawed peak detection in apparently suitable video content.

Figure 15. Screenshot during a 5 second period with no heartbeat, but with face detection.

Figure 16. Screenshot during a 5 second period with movement and realistic heartbeats.

Secondly, almost all SDNN values are unrealistically large, even for the filtered data. Since SDNN is usually measured for at least 5 minutes (Malik, 1996), SDNN of the 30 second videos may be more susceptible to noise and consequently larger. Furthermore, the output from videos with a large SDNN value shows that in some cases interval times between sequences differ. For example, a sequence of intervals at the start of the video have approximately similar durations. Nevertheless, the average RR interval in this first sequence differs considerably from another sequence at the end of the video. This effect appears in some of the video heart rate plots from figures $8 - 12$ (e.g. video 20, 99 or 191). Still, after video content analysis no cause for the large RR interval differences between sequences is noticed. Moreover, there is a substantial amount of videos where the interval time within a sequence was inconsistent as well (e.g. video 40, 41 or 193), for which no cause with respect to video content could be found either. Therefore, causes for large SDNN values remain unclear and need to be researched in future studies as well.

Explanation testing results

From the statistical testing results cannot be concluded that an active emotional state of happiness, anger or sadness leads to a higher HR than a neutral state. This is inconsistent with findings of previous research (Boiten, 1996; Cacioppo et al., 2000; Ekman et al., 1983; Ilies et al., 2010). Besides, the current research found 1.72 bpm difference between the active and neutral state, even though it was assumed to be around 5.67 bpm (Ekman et al., 1983). Nonetheless, the current study differs substantially from previous research, so results are not completely unexpected. What's more, in spite of the poor performance as described in the data description section, the final *p-*value of .065 seems decently low. Possibly, there is some underlying truth in rPPG measurements which is largely overshadowed by the generated noise.

Limitations

The current study has a couple of substantial limitations. First of all, no ground truth measurements could be measured. It was assumed that the video quantity would outbalance the inconsistencies from rPPG data. Nonetheless, 40 videos were left out so the result is less accurate than planned upfront. Besides, the determined number of videos followed from, possibly poor, assumptions. Not all sample size assumptions were met since the HR of Koeman in the neutral state was not normally distributed. Furthermore, the rPPG output was less accurate than the results where the assumptions were based on (Poh et al., 2010; Tasli et al., 2014b). Hence, even though the heart rates have realistic values (Umetani et al., 1998), there is no reason to conclude that the found heart rates are the true heart rates. Future research can try to combine ground truth measures in natural environments to ensure accurate HR output.

Another limitation is the lack of knowledge about the video recordings. First, the emotion might be inaccurately estimated, because there was often no confirmation of the emotional state by the manager himself. The manager may appear neutral from the outside but experiences emotions in reality. Secondly, there is not much background information on the manager his activities before the interview. He might have performed some intense physical activity that increases heart rate. Furthermore, the interview might in some cases be a stressful event for a manager, which temporarily increases HR (Kudielka, Schommer, Hellhammer, & Kirschbaum, 2004). Additionally, resting HR could differ within the same person. Factors like age and time of day differ between videos and are known to influence one's resting HR (Ewing, Neilson, Shapiro, Stewart, & Reid, 1991). But factors like prolonged stress (Shapiro et al., 2001; Vrijkotte, van Doornen, & de Geus, 2000) effect the resting HR as well. Moreover, managers were interviewed in either standing (interview) or sitting (press conference) positions which possibly corresponds with a different resting HR (Robinson,

Epstein, Beiser, & Braunwald, 1966). Besides, press conferences tend to take much longer (about 15 minutes). Hence, the used video moments were on average collected after more time passed compared to the interviews (which take about 3 minutes). This means that HR during press conferences had more time to settle down from possible HR increasing activities done upfront. Nonetheless, no large difference appears between the HR during press conferences (64.21 \pm 6.09 bpm) or interviews (64.18 \pm 5.55 bpm).

A large part of the results is based on the manual filtering method. The method is derived from realistic values (Umetani et al., 1998) and the outputted data. Nonetheless, there are possibly more appropriate ways to deal with inaccurate data measurements, especially if the video attributes (video length, movement of a person, etc.) differ from the current research. Therefore, future research can focus on better general filtering methods. Filters like Kubios (Tarvainen et al., 2014) or other HRV handling methods (Singh & Bharti, 2015) possibly help to find a preferred way to deal with raw post measurement rPPG data.

An obvious limitation in the current research is the possible loss of important data during the multiple conversions of the video. First, the raw video gets uploaded to an online platform. Then the video is downloaded by a YouTube-user. Possible disturbances of the video signal during this process are completely unknown, and therefore some signal related issues may unknowingly have been present in the collected videos. Subsequently, the video is uploaded by this YouTube-user, at a later date downloaded through the converter tool and finally loaded into FaceReader. A brief pilot study was conducted to determine accuracy during these phases. As mentioned previously, the output from converted MP4 and converted AVI videos differed compared to the raw video. Future research can examine video formats and reliable conversion methods to achieve little data loss, in case raw videos are unavailable. At the end of the process, FaceReader itself does some conversion as well. Videos of 25 fps were converted to 30 fps. This yields that every 40 ms one frame occurs in the actual video

but the output shows intervals of 33 ms with five times a second an interval of 67 ms. This possibly affects HR calculations to some extent. Furthermore, as noted earlier, this study assumes that 25 fps or 30 fps videos are not significantly different. Still, differences between 25 fps and 30 fps pilot videos differed with approximately 2 bpm. Even though internal VicarVision testing states that at least 15 fps is necessary, more research can be conducted to examine if the minimum threshold holds in less trivial environments like the ones investigated in the current study.

The used measurement settings are a possible limitation as well. Settings like window size and region of interest were set according to recommendations from VicarVision. Nevertheless, not much research studied the performance with different time windows and ROIs of the FaceReader software. Hence, future research can determine preferred settings in several natural or artificial video conditions.

A final limitation is the software program used in this study. The version used for the video analysis was in the alpha stage during the research. The above-mentioned frame rate limitation can relatively easily be improved by using 40 ms intervals for 25 fps videos. Furthermore, the previously called out bug which decreases measurement time is expected to be solved in updated versions. Moreover, other improvements on the rPPG algorithm and FaceReader software are expected in the final version.

Implications

The current study has elaborately explored the rPPG algorithm on public video data in relatively natural settings. This is a new step in rPPG research as previous research used rPPG on videos recorded in lab environments. New issues arise in this unfamiliar setting consisting of interviews and press conferences. The algorithm occasionally suffers from missing measurement points due to errors in face detection. But more importantly, a large part of RR

interval measurements are inaccurate, which leads to large SDNN values and possibly wrong heart rates. To improve the rPPG algorithm, these problems need to be studied and understood before rPPG can be applied as HR measurement tool in natural environments. Moreover, more elaborate post-processing methods need to be developed which filter or transform rPPG output such that heart rate can be accurately calculated. Still, the positive HR difference between active and neutral emotion suggests that, in spite of the insignificant result and the limitations, there is a modest start to accomplish more complex rPPG tasks like the distinction of emotional state in public videos.

Furthermore, if more complex tasks like the one examined in the current study are successfully completed, other aspects of rPPG can be researched. For instance, the performance in less optimal video conditions. It can be questioned whether rPPG can be used as a tool built into laptops, phones or security cameras. Besides, even if this would be possible from a technical perspective, the ethical aspect needs to be kept in mind. The ease at which rPPG can be used makes that many people are potentially analysed without their knowledge. This raises a new ethical issue (Sebe et al., 2005) that can be researched more in detail as well.

In conclusion, the current research shows that rPPG fails to convincingly measure heart rate in non-trivial conditions. Before rPPG can be used on natural video occurrences, the algorithm needs to improve in heartbeat registration. Nonetheless, the potential of rPPG is not diminished since statistical test results cautiously hint in a positive direction.

References

Abhishekh, H. A., Nisarga, P., Kisan, R., Meghana, A., Chandran, S., Trichur Raju, & Sathyaprabha, T. N. (2013). Influence of age and gender on autonomic regulation of heart. *Journal of Clinical Monitoring and Computing*, *27*(3), 259–264.

https://doi.org/10.1007/s10877-012-9424-3

- Barrett, L. F., Robin, L., Pietromonaco, P. R., & Kristen, M. (1998). Are Women the "More Emotional" Sex ? Evidence From Emotional Experiences in Social Context. *Cognition & Emotion*, *12*(4), 555–578. https://doi.org/10.1080/026999398379565
- Bigger, J. T., Fleiss, J. L., Steinman, R. C., Rolnitzky, L. M., Schneider, W. J., & Stein, P. K. (1995). RR Variability in Healthy, Middle-Aged Persons Compared With Patients With Chronic Coronary Heart Disease or Recent Acute Myocardial Infarction. *Circulation*, *91*(1), 1936–1943.
- Boiten, F. (1996). Autonomic response patterns during voluntary facial action. *Psychophysiology*, *33*, 123–131.
- Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2000). *The Psychophysiology of Emotion*. *Handbook of emotions* (Vol. 2).
- Crisp, R. J., Heuston, S., Farr, M. J., & Turner, R. N. (2007). Seeing Red or Feeling Blue: Differentiated Intergroup Emotions and Ingroup Identification in Soccer Fans. *Group Processes & Intergroup Relations*, *10*(1), 9–26.

https://doi.org/10.1177/1368430207071337

- Den Uyl, M. J., & Van Kuilenburg, H. (2005). The FaceReader: Online facial expression recognition. *Proceedings of Measuring Behavior*, *30*, 589–590.
- Edwards, G. J., Taylor, C. J., & Cootes, T. F. (1998). Interpreting face images using active appearance models. *Automatic Face and Gesture Recognition*, 300–305. https://doi.org/10.1109/AFGR.1998.670965
- Ekman, P., & Friesen, W. V. (2003). *Unmasking the face: A guide to recognizing emotions from facial clues*.
- Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic Nervous System Activity in Emotion. *Science*, *221*(4616), 1208–1210.
- Ewing, D. J., Neilson, J. M., Shapiro, C. M., Stewart, J. A., & Reid, W. (1991). Twenty four hour heart rate variability: effects of posture, sleep, and time of day in healthy controls and comparison with bedside tests of autonomic function in diabetic patients. *British Heart Journal*, *65*(5), 239–244. https://doi.org/10.1136/hrt.65.5.239
- Fei, L., Copie, X., Malik, M., & Camm, A. J. (1996). Short-and long-term assessment of heart rate variability for risk stratification after acute myocardial infarction. *The American Journal of Cardiology*, *77*(9), 681–684.
- Hanin, Y. L. (2007). Emotions in Sport : Current Issues and Perspectives. *Handbook of Sport Psychology*, *3*(3158), 22–41.
- Hertzman, A. B., & Spealman, C. R. (1937). Observations on the finger volume pulse recorded photoelectrically. *American Journal of Psysiology*, *119*, 334–335.
- Hsu, Y., Lin, Y. L., & Hsu, W. (2014). Learning-based heart rate detection from remote photoplethysmography features. *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4433–4437. https://doi.org/10.1109/ICASSP.2014.6854440
- Ilies, R., Dimotakis, N., & Watson, D. (2010). Mood, blood pressure, and heart rate at work: An experience-sampling study. *Journal of Occupational Health Psychology*, *15*(2), 120– 130. https://doi.org/10.1037/a0018350
- Kublanov, V. S., & Purtov, K. S. (2015). Researching the possibilities of remote photoplethysmography application to analysis of time-frequency changes of human heart rate variability. *International Conference on Biomedical Engineering and Computational Technologies (SIBIRCON)*, 87–92. https://doi.org/10.1109/SIBIRCON.2015.7361857
- Kudielka, B. M., Schommer, N. C., Hellhammer, D. H., & Kirschbaum, C. (2004). Acute HPA axis responses, heart rate, and mood changes to psychosocial stress (TSST) in humans at different times of day. *Psychoneuroendocrinology*, *29*(8), 983–992. https://doi.org/10.1016/j.psyneuen.2003.08.009
- Levenson, R. W., & Ekman, P. (2002). Difficulty does not account for emotion-specific heart rate changes in the directed facial action task. *Psychophysiology*, *39*, 397–405. https://doi.org/10.1017.S0048577201393150
- Malik, M. (1996). Heart rate variability standards of measurement, physiological interpretation and clinical use. Task force of the european society of cardiology and the North American society of pacing and electrophysiology. *Circulation*, *93*(5), 1043–1065.
- Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America*, *93*(2), 1097–1108. https://doi.org/10.1121/1.405558
- Poh, M.-Z., McDuff, D. J., & Picard, R. W. (2010). Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics Express*, *18*(10), 10762–10774. https://doi.org/10.1364/OE.18.010762
- Pollock, V. B. A., Cho, D. W., Reker, D. R. N., & Volavka, J. M. D. (1979). Profile of Mood States: the factors and their physiological correlates. *The Journal of Nervous and Mental Disease*, *167*(10).
- Robinson, B. B. F., Epstein, S. E., Beiser, G. D., & Braunwald, E. (1966). Control of Heart Rate by the Autonomic Nervous System. *Circulation Research*, *XIX*(2), 400–411.
- Rouast, P. V, Adam, M. T. P., Dorner, V., & Lux, E. (2016). Remote Photoplethysmography : Evaluation of Contactless Heart Rate Measurement in an Information Systems Setting.
- Schubert, D. S. P. (1977). Alertness and clear thinking as characteristics of high naturally occurring autonomic nervous system arousal. *Journal of General Psychology*, *97*, 179– 183. https://doi.org/10.1080/00221309.1977.9920836
- Sebe, N., Cohen, I., Gevers, T., & Huang, T. S. (2005). Multimodal Approaches for Emotion Recognition: A Survey. *Proc. SPIE*, *5670*, 56–67. https://doi.org/10.1117/12.600746

Shapiro, D., Jamner, L. D., Goldstein, I. B., & Delfino, R. J. (2001). Striking a chord: Moods ,

blood pressure, and heart rate in everyday life. *Psychophysiology*, *38*, 197–204.

- Singh, B., & Bharti, N. (2015). Software tools for heart rate variability analysis. *International Journal of Recent Scientific Research*, *6*(4), 3501–3506.
- Stein, P. K., Kleiger, R. E., & Rottman, J. N. (1997). Differing Effects of Age on Heart Rate Variability in Men and Women. *The American Journal of Cardiology*, *80*(3), 302–305.
- Takano, C., & Ohta, Y. (2007). Heart rate measurement based on a time-lapse image. *Medical Engineering and Physics*, *29*(8), 853–857. https://doi.org/10.1016/j.medengphy.2006.09.006
- Tarvainen, M. P., Niskanen, J. P., Lipponen, J. A., Ranta-aho, P. O., & Karjalainen, P. A. (2014). Kubios HRV - Heart rate variability analysis software. *Computer Methods and Programs in Biomedicine*, *113*(1), 210–220. https://doi.org/10.1016/j.cmpb.2013.07.024
- Tasli, H. E., Gudi, A., & Den Uyl, M. J. (2014a). Integrating Remote PPG in Facial Expression Analysis Framework. *Proceedings of the 16th International Conference on Multimodal Interaction*, 74–75. https://doi.org/10.1145/2663204.2669626
- Tasli, H. E., Gudi, A., & Den Uyl, M. J. (2014b). Remote PPG based vital sign measurement using adaptive facial regions. *Image Processing (ICIP)*, 1410–1414.
- The Michael Jackson Interview: Oprah Reflects. (2009). Retrieved from http://www.oprah.com/entertainment/oprah-reflects-on-her-interview-with-michaeljackson/1
- Umetani, K., Singer, D. H., McCraty, R., & Atkinson, M. (1998). Twenty-four hour time domain heart rate variability and heart rate: Relations to age and gender over nine decades. *Journal of the American College of Cardiology*, *31*(3), 593–601. https://doi.org/10.1016/S0735-1097(97)00554-8
- Verkruysse, W., & Bodlaender, M. P. (2010). *A novel biometric signature: multi-site, remote (> 100 m) photo-plethysmography using ambient light*. *Philips Research*.
- Verkruysse, W., Svaasand, L. O., & Nelson, J. S. (2008). Remote plethysmographic imaging using ambient light . *Optics Express*, *16*(26), 21434–21445.
- Vrijkotte, T. G. M., van Doornen, L. J. P., & de Geus, E. J. C. (2000). Effects of Work Stress on Ambulatory Blood Pressure, Heart Rate, and Heart Rate Variability. *Hypertension*, *35*(4), 880–886. https://doi.org/10.1161/01.HYP.35.4.880
- Wang, W., Stuijk, S., & De Haan, G. (2014). Exploiting Spatial-redundancy of Image Sensor for Motion Robust rPPG. *IEEE Transactions on Biomedical Engineering*, *62*(2), 415– 425. https://doi.org/10.1109/TBME.2014.2356291
- Wieringa, F. P., Mastik, F., & Van der Steen, A. F. W. (2005). Contactless Multiple Wavelength Photoplethysmographic Imaging: A First Step Toward "SpO2 Camera" Technology. *Annals of Biomedical Engineering*, *33*(8), 1034–1041. https://doi.org/10.1007/s10439-005-5763-2
- YouTube Company Statistics. (2016). Retrieved from https://www.statisticbrain.com/youtube-statistics/

Appendix A

Live heart rate on iconic videos

Video 6

Video 9

Video 12 Infmous interview where Diana mentions her three person relationship

Appendix B

Data description of several filtering methods

Boxplots of the SDNN values, RR intervals and heart rates of the raw data on all videos (including outliers) as well as several filtering methods on all videos.

- **RP** (Raw Peaks) = The HR calculated from the number of peaks occurring within the time interval from the appearance of the first peak to the final peak.
- **RL** (Raw Live) = The HR calculated from the live reported HR by FaceReader
- **RI** (Raw Interval) = The SDNN, RR interval and HR calculated from the time between intervals
- **FL** (Filtered Live) = The HR calculated from the live reported HR by FaceReader disregarding measurements smaller than 49 bpm or larger than 104 bpm (based on results found by Umetani et al. (1998))
- **FI** (Filtered Interval) = The SDNN, RR interval and HR calculated from the time between peaks disregarding intervals smaller than .48 s or larger than 1.32 s (based on results found by Umetani et al. (1998))
- **FI2** (Filtered Interval 2) = The SDNN, RR interval and HR calculated from the time between peaks disregarding intervals smaller than .41 s or larger than 1.23 s (based on results found by Bigger et al. (1995))
- **(Filtered Interval 3) = The SDNN, RR interval and HR calculated from the time** between peaks disregarding intervals smaller than .48 s or larger than 1.32 s (based on results found by Umetani et al. (1998)) and intervals where the preceding and succeeding intervals are outside this range. Therefore only sequences of at least two intervals within the range are taken into account
- KW (Kubios Weak) = The SDNN, RR interval and HR calculated from the transformed RR intervals between peaks by Kubios with the very low threshold

- **KS** (Kubios Strong) = The SDNN, RR interval and HR calculated from the transformed RR intervals between peaks by Kubios with the strong threshold

Appendix C

Peak plots all videos (including outliers)

These graphs give an indication where peaks are detected and where peaks appear to be missing. Time is shown on the x-axis $(0 - 30 s)$ and the peak value (0 for no peak and 1 for a peak) on the y-axis. The latter indicates whether there was a heartbeat (value $= 1$) or not (value $= 0$). The big green dots indicate the peaks which have a time difference with the preceding peak of .48 – 1.32 seconds (based on Umetani et al. (1998)). Only if there were at least two consecutive time intervals within this range the intervals were used for analysis. In the graphs this means that red (small) dots and green dots surrounded by two red dots were disregarded for the applied filter method.

Appendix D

Links of all video material

List of links to all videos with the corresponding time stamps per video. The 30-

second videos are uploaded to the Open Science Framework as well.

Guardiola

Koeman

Mourinho

Van Gaal

Iconic videos

Appendix E

Declaration of Scientific Conduct

Declaration concerning the TU/e Code of Scientific Conduct for the bachelor's final project

I have read the TU/e Code of Scientific Conducti.

I hereby declare that my bachelor's final project has been carried out in accordance with the rules of the TU/e Code of Scientific Conduct

See: http://www.tue.nl/en/university/about-the-university/integrity/scientific-integrity/
The Netherlands Code of Conduct for Academic Practice of the VSNU can be found here also. More information about scientific integrity is published on the websites of TU/e and VSNU

July 2 2015