Towards an Affective Video Recommendation System

Yancarlos Diaz
Rochester Institute
of Technology
yxd3549@rit.edu

Cecilia O. Alm
Rochester Institute
of Technology
coagla@rit.edu

Ifeoma Nwogu
Rochester Institute
of Technology
ion@cs.rit.edu

Reynold Bailey
Rochester Institute
of Technology
rjb@cs.rit.edu

***Abstract*—Video streaming services are prominent in people's lives and there is a need for improved video recommendation systems that adapt to their users in a personalized way. This project uses affective computing and non-invasive sensing to address this issue. Our objective is to develop an approach that uses the viewer's emotional reactions as the basis for recommending new content. To achieve this goal, we must first understand how viewers react to videos. We conducted a study where subjects' facial expressions and skin-estimated pulse were monitored while watching videos. Results showed that our approach can estimate dominant emotions 70% of the time. We found no correlation between the number of emotional reactions people have and how they rate the videos they watch. The pulse estimation is reliable to measure important changes in pulse, however it can still be improved.**

***Index Terms*—Affective computing; video recommendation; photoplethysmography**

I. Introduction

The use of video streaming services is popular. Such services rely heavily on recommendation systems to suggest new content. These suggestions are generated by algorithms based on the user's viewing trends as well as tags that come with every video. However, recommendation systems may fail to properly adapt to the user, in particular in real-time.

This study applies affective computing to address this issue. Affective computing is computing that relates to, analyzes, and interprets human emotions [10]. Affective data may include human facial expressions, speech prosody, body posture, and physiological data such as pulse and galvanic skin response.

We believe that video recommendation systems can benefit from analyzing user-extracted affective data. However, before addressing measurement-based recommendation, we must first explore the prospects and limitations of measured human reactions in such contexts. The purpose of this study is to better understand the nature and temporal characteristics of the affective data we can obtain from users as they view videos with emotional content. We also want to know whether we can collect data from users in a way that is physically non-invasive and does not require wearable devices. For that reason, we are limiting the use of sensors to simple webcams.

This paper is organized as follows: A discussion about previous work done by other researchers that relate to our goal and methods is followed by a detailed description of how we designed and ran our experiment. Next, we provide an overview of results. Closing remarks discuss future work and applications.

II. Related Work

Many research efforts have improved the way developers implement video recommendation systems. Most current recommendation systems use the users' viewing trends as an input for the algorithm. These systems need to avoid what some researchers call the cold start problem. In essence, the cold start problem is the initial lack of data to describe the users' viewing trends. To solve this issue, some studies have examined systems that use genre correlations as their basis for recommendation [3]. Others have used metadata for the same purpose [14]. Our affective, measurement-based approach represents a different point of view.

A. Affective Computing in Video Recommendation

Only rarely have affective data been used as the main source for recommendation. Zhao et al. were among the first to pioneer facial expressions as the basis for a recommendation system, but their system relied heavily on exaggerated, less natural facial expressions [16]. Furthermore, facial data alone is not enough to capture the complexity of human emotion [2]. Rajenderan expanded on the use of facial expressions by introducing photoplethysmography to calculate the heart rate of users while they watch a video [12]. However, his work looked at facial expression data and heart rate data separately. We aim to study these measures together.

To better understand human emotions in response to videos, studies have monitored the viewers' experiences with the help of various sensing devices. For instance, Fleureau et al. used galvanic skin response, heart rate, and electromyograms to compute physiological signals while their subjects watched a video [6]. As another example, Wollmer et al. looked at video reviews and developed a machine learning algorithm to automatically label the video review as positive or negative based on the reviewers' facial expressions and words used in the video [15]. Both studies obtained promising results, but their data collection does not fit the intended purpose. A video recommendation system should not require invasive or wearable devices nor verbal response of users.

Instead, this study proposes a noninvasive way of obtaining data that only requires the use of webcams. We extract two types of data from the video stream of standard webcams. Like Zhao et al. we analyze facial expressions to obtain information about the emotion of the viewer [16]. Furthermore, we also estimate the viewer's heart rate from the video through the use of photoplethysmography.

B. *Facial Analysis*

Analyzing facial expressions is a convenient and noninvasive way to gain insight into affective states. Darwin proposed a link between facial expressions and emotions [4]. Ekman et al. developed the Emotion Facial Action Coding System (EMFACS) to explicitly link specific facial expressions to emotions [5]. We base our facial analysis on seven fundamental emotions: happiness, sadness, surprise, fear, anger, disgust, and contempt.

C. *Photoplethysmography*

During the cardiac cycle, changes in the volume of the blood vessels in the face modify the path length of the incident light. We can use the changes in the amount of reflected light to calculate the viewer's heart rate [11].

In one study, subjects sat in front of a camera used to calculate their pulse with this technique, wearing an FDA-approved finger pulse sensor [11]. The authors found that the mean error between the estimated pulse and the pulse recorded by the finger pulse sensor was of .95 beats per minute with a standard deviation of .83 beats per minutes.

We further build on the work of other researches who have assessed the reliability of this technique on different areas of the body, with results suggesting that forehead and the area in between the eyes are the best areas to estimate pulse with this technique [1], [9].

III. Experimental Design

Figure 1 shows a photograph of the experiment setup. We collected data from 30 subjects (13 females, 17 males) between the ages of 18 and 35. The data from two participants had to be removed due to problems with their data and a misunderstanding of the instructions. Every subject filled out a survey about their watching and rating trends. They then watched five videos while their behavior was being monitored by two Logitech pro 9000 webcams. These webcams recorded at 20 frames per second at a resolution of 720p.

We used the stream from one webcam for facial expression analysis and the other stream to calculate the subject's pulse throughout the video. We processed the subjects' videos using an Affectiva module for facial expression analysis in iMotions [8]. We calibrated the threshold for emotion detection to account for common false positives.

For the pulse measurement, we adapted a version of a program that uses the stream of a camera to calculate someone's pulse by looking at the forehead [7], [13], seen in Figure 2.

At the end of every video, subjects answered the following four questions about the video:

1) *Have you watched this video before?*

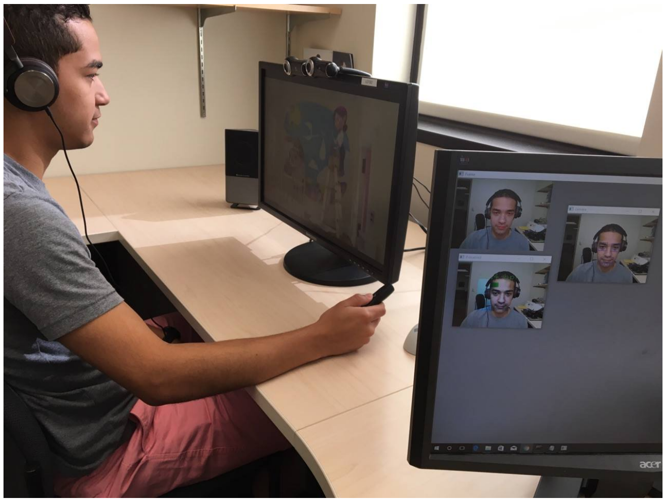


Fig. 1: Overview of experiment setup. Subjects sat in front of a computer screen and watched five videos while two Logitech webcams recorded their faces. Half the subjects also used a clicker to signal an emotional reaction.

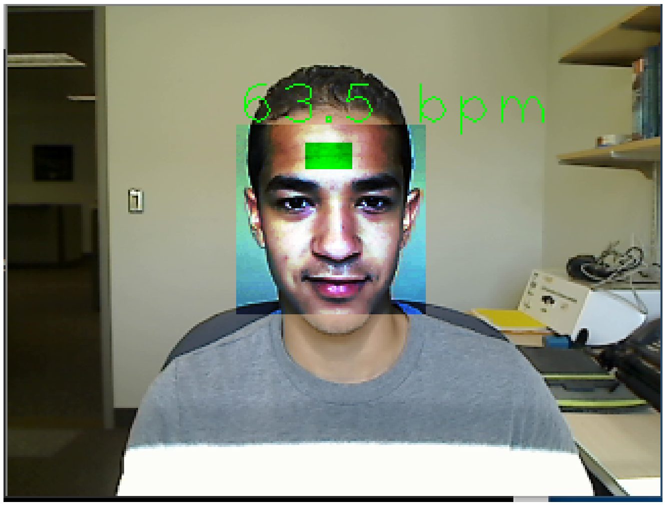


Fig. 2: Pulse estimation with photoplethysmography. We first detect the face then choose a central area just above the eyebrows to estimate pulse in real-time.

2) *What did you feel watching this video?*
3) *On a scale from 1 to 5, how would you rate this video (1 being the worst rating and 5 being the best rating)?*
4) *Would you want to watch similar videos?*

For the second question, subjects were given a list of the seven basic emotions. They were allowed to choose multiple emotions per video.

Half of the subjects had a clicker. They were instructed to use the clicker whenever they had an emotional reaction to the content in the videos. We wanted to test whether their click would match with an increase in heart rate or a change in their facial expression. We also wanted to examine possible differences between subjects with a clicker and subjects without a

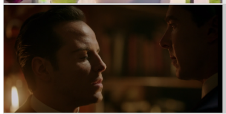
Video	Video ID	Source Film	Description	Length	Expected Viewer Affective Reaction
	1	Interstellar	A father watches his children grow up and eventually give up the hope to see him again	4 minutes, 15 seconds	Sadness
	2	Up	We see a couple get married, lose a child, grow old, and be separated by death	4 minutes, 20 seconds	Happiness and Sadness
	3	Sherlock – The Abominable Bride	A suspenseful video where two armed men build tension until one of them shoots himself	5 minutes, 30 seconds	Surprise
	4	The Watsons Go to Birmingham	Two boys go into a restaurant and are denied food because of their race	2 minutes	Anger and Disgust
	5	Gabriel Iglesias – Hot and Fluffy	A standup routine featuring a comedian known for his surprising sound effects	5 minutes, 10 seconds	Happiness

Fig. 3: Overview of the characteristics of the videos we showed in the experiment. All subjects viewed these five videos.

clicker. We leave this comparison for future work.

Figure 3 shows an overview of the videos the subjects watched. We chose videos that we thought would evoke different emotions. Video 1 was a scene from *Interstellar* expected to evoke sad emotions. Video 2 was the starting scene from the Pixar movie Up, meant to evoke a mix of emotions. The third video was a scene from the show *Sherlock* where the nemesis Moriarty unexpectedly uses a weapon on himself but survives the shock, surprising the viewer. Video 4 was a scene from the film *The Watsons Go to Birmingham* causing anger or disgust. Video 5 was a scene from *Hot and Fluffy*, a stand-up comedy intended to make the viewer laugh. We randomized the order in which the subjects watched these videos.

IV. Results and Discussion

In this section we look at the results obtained from the facial expression analysis software, the webcam pulse estimation, and the surveys subjects filled out about the videos. First, we consider how the emotions the software computed compare to the emotions people reported feeling for every video. We then study the relationship between the number of times people used the clicker and how they rate the videos. Finally, we discuss how we can observe and learn from all data as a whole.

A. Emotion Analysis Results

Figure 8 shows an overview of the results we obtained from the facial expression analysis of video observers and the surveys at the end of every video. Most subjects reported feeling the emotions that the videos were expected to evoke. In a majority of cases, the most reported emotions also matched the emotions that iMotions computed. However, it is important to note that subjects could report multiple emotions whereas the analysis software computes one dominant emotion per face video.

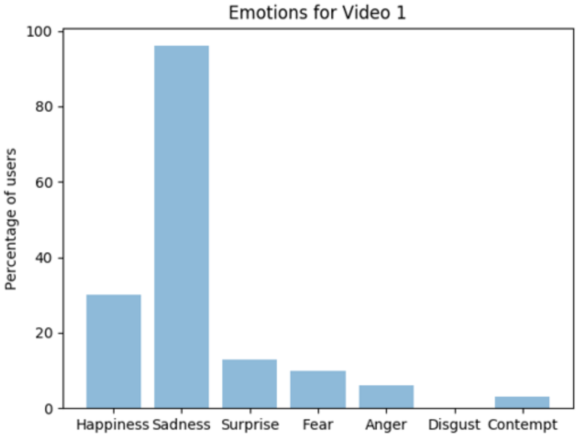


Fig. 4: Self-reported emotions for video 1 (*Interstellar*). As expected, sadness was the most prominent emotion for this video.

For example, for video 1 as illustrated in Figure 4, 96% of the subjects reported feeling sad and 30% reported feeling happy. Figure 6 shows the results of the facial analysis. Facial expression analysis revealed sadness as the dominant emotion for 45% of the subjects and happiness for 14% of the subjects. Thus for 45% of the subjects, sadness was the most often displayed emotion. It is possible that these subjects also felt happy at some point in the video; however, only the emotion shown for a majority of the video is reported. For this reason we are not able to directly compare the percentages. Arguably,

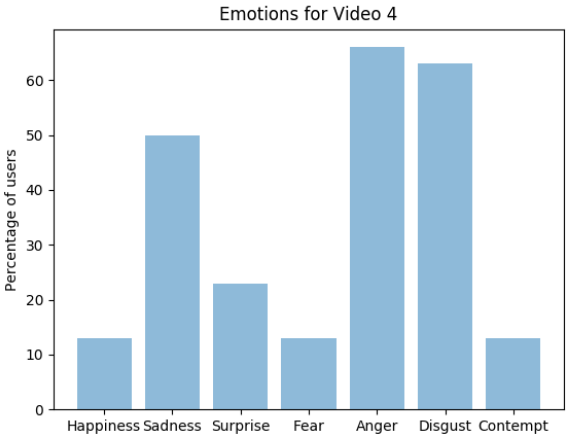


Fig. 5: Self-reported emotions for video 4 (*The Watsons Go to Birmingham*). Most subjects reported feeling angry, disgusted, or sad, illustrating the complexity of negative emotions.

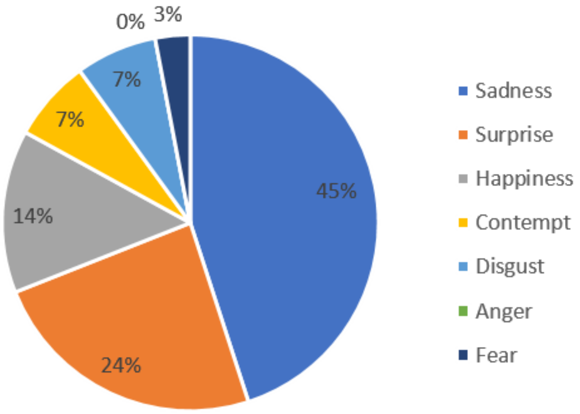


Fig. 6: Computed emotions for video 1. The top emotions (sadness, surprise, and happiness) also correspond to the most self-reported emotions in Figure 4.

a better way to assess the reliability of this system is to look at the top emotions in each category and see if they match.

With the exception of video 4, the software computed the emotions we expected from participants (shown in bold in Figure 8). Figure 5 and Figure 7 show the reported and computed emotions for video 4, respectively. It appears that, for our set of videos, video 4 was the exception because of the nature of the expected emotion. Even though viewers reported feeling angry, the face analysis did not register this, which may reflect that anger is less prominently detected via changes in the face than modalities such as speech prosody.

B. *Clicks vs. Rating*

Many recommendation systems base their recommendations on user ratings. However, our pre-experiment survey revealed that people rarely rate videos. Figure 9 shows the answers to the question *How often do you rate the videos you watch?* 50% of subjects reported rarely rating their videos and 30% went as far as saying that they never rate their videos.

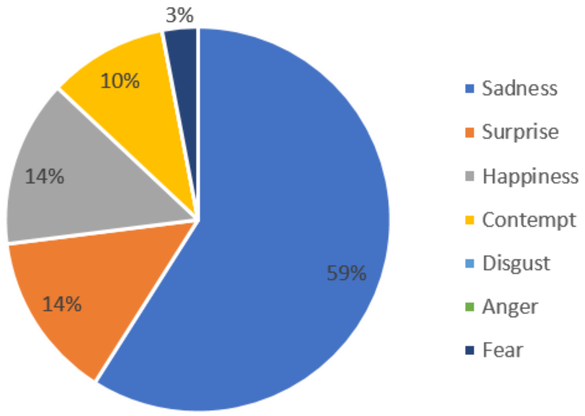


Fig. 7: Computed emotions for video 4. This video had the most mismatches between computed emotions and reported emotions (see Figure 5).

This lack of user input confirms the need for a way of estimating how people experience a video without directly asking them. Our hypothesis was that the number of emotional reactions people had to certain videos would correlate to how they rate videos. We asked half the subjects to click whenever they had an emotional reaction in order to test this hypothesis.

The highest correlation between clicks and ratings was 0.5, illustrated in Figure 10, which is not enough to say that the number of clicks is related to how people rate videos. The rest of the videos had an even lower correlation, like the one illustrated in Figure 11. This could reflect that introspection of emotional reactions in real-time is less natural. People's ratings can also be based on video quality, acting, and other biases.

C. *Affective Graphs*

While the clicks represent one way to gain self-estimated reference data of when people experienced emotional reactions to content, a recommendation system should not require the user to signal their reactions (just as it should not require the user's explicit rating). Instead, we believe that pulse estimations could be used to analyze users' reactions. Thus, we examined links between changes in facial expression and a click event or a change in pulse.

To that end, we visualize pulse, click events, and facial expressions jointly in what we term an *affective graph*. Figure 12 shows two sample graphs for two users, in which clicks (gray dots) often match a drastic change in pulse (blue line), which supports our hypothesis that a change in pulse would appear near an emotional reaction. However, it is hard to tell whether

Video	Average Rating	Average Clicks and Standard Deviation	Most Reported Emotions	Most Computed Emotions
	4.3 (0.9)	4.7 (2.8)	**Sadness (96%)** Happiness (30%) Surprise (13%)	**Sadness (45%)** Surprise (24%) Happiness (14%)
	4.4 (1.3)	5.1 (3.2)	**Sadness (90%)** **Happiness (66%)** Surprise (13%)	**Happiness (35%)** **Sadness (24%)** Surprise (24%)
	4.1 (1.3)	4.9 (2.1)	**Surprise (73%)** Fear (43%) Disgust (43%) Happiness (23%)	Happiness (31%) **Surprise (28%)** Sadness (21%) Contempt (10%)
	3.7 (1.4)	3 (2.8)	**Anger (66%)** Disgust (63%) Sadness (50%) Surprise (23%)	Sadness (59%) Happiness (14%) Surprise (14%) Contempt (10%)
	3.8 (1.5)	6.9 (4.6)	**Happiness (96%)** Surprise (36%) Disgust (10%)	**Happiness (66%)** Surprise (17%) Contempt (10%)

Fig. 8: Overview of results. Self-reported emotion often matches the dominant emotion inferred from face capture.

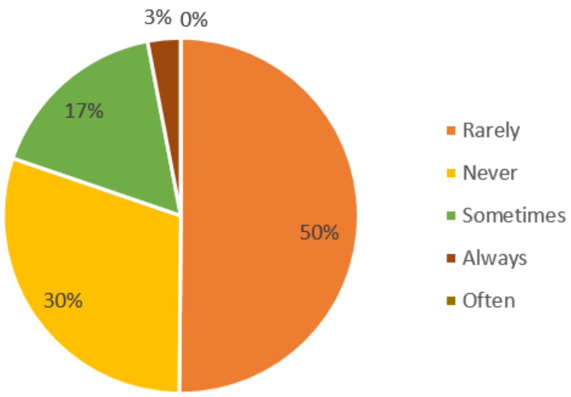


Fig. 9: Answers to the question *How often do you rate the videos you watch?*

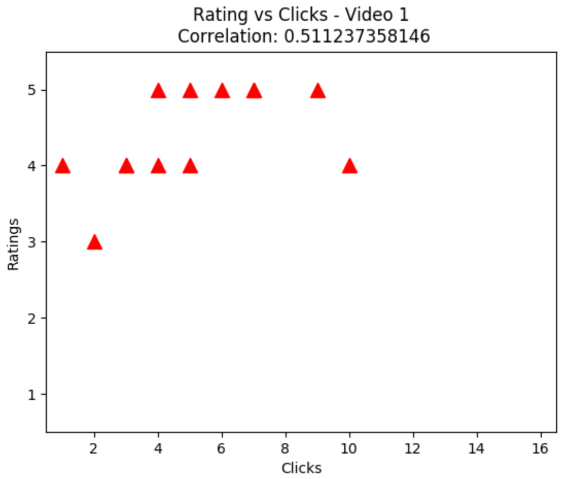


Fig. 10: Rating vs. clicks for video 1. This video had the highest correlation between the number of clicks and how people rated the video.

a change in emotional facial expressions (indicated by intervals rendered in different color by different emotions, e.g., blue is sadness and green is happiness) correspond to click events or a drastic change in pulse.

V. Conclusion

The presented results increased our understanding of how viewers react to videos. We gained insights from using physically noninvasive measurements, which are relevant for our long-term goal of their use in an affective video recommendation system.

However, our contributions are not limited to recommendation systems. Our approach and the affective graph visualization may also be meaningful in the film-making industry for evaluating effectiveness of certain scenes. This work could also be applicable in online learning environments that lack face-to-face interactions of traditional learning contexts.

The limitations of our study include certain aspects of the experiment design. An experimenter was present when subjects were watching the videos, which might have influenced their emotional reactions. We also only focused on one video per emotion, as opposed to multiple videos per emotion. Furthermore, although the facial analysis software detects clear facial expressions, it also regularly fails to differentiate a resting face from contempt.

In the future, we would like to make our data collection

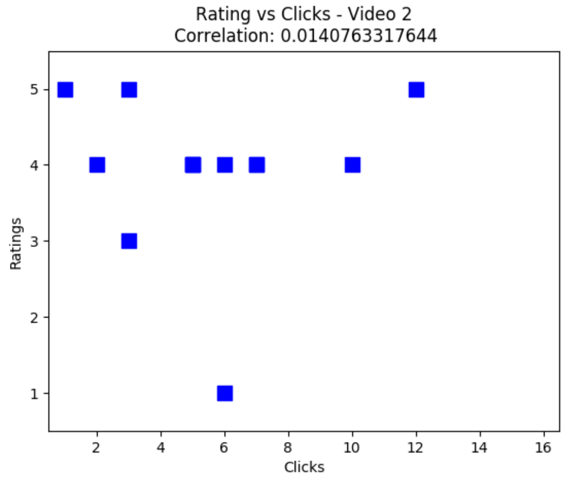


Fig. 11: Rating vs. clicks for video 2 did not produce a strong correlation between number of clicks and ratings.

methods increasingly intuitive, scalable, and transportable across settings. To reach such goals, we intend to explore other ways of estimating pulse in a noninvasive way. We also plan to use machine learning to estimate video ratings based on the measurement data. Finally, we would like to study the quantitative relationships between ratings and pulse variation as well as facial expressions.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Award Nos. IIS-1559889 and HRD-120248. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] L. A. Aarts, V. Jeanne, J. P. Cleary, C. Lieber, J. S. Nelson, S. B. Oetomo, and W. Verkruysse. Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit - a pilot study. *Early Human Development*, 89(12):943–948, 2013. doi: 10.1016/j.earlhumdev.2013.09.016

[2] Y. Baveye, C. Chamaret, E. Dellandrea, and L. Chen. Affective video content analysis: A multidisciplinary insight. *IEEE Transactions on Affective Computing*, 2017. doi: 10.1109/taffc.2017.2661284

[3] S.-M. Choi, S.-K. Ko, and Y.-S. Han. A movie recommendation algorithm based on genre correlations. *Expert Systems with Applications*, 39(9):8079–8085, 2012. doi: 10.1016/j.eswa.2012.01.132

[4] C. Darwin. *The Expression of the Emotions in Man and Animals*. John Murray, 1872.

[5] P. Ekman, W. V. Freisen, and S. Ancoli. Facial signs of emotional experience. *Journal of Personality and Social Psychology*, 39(6):1125–1134, 1980. doi: 10.1037/h0077722

[6] J. Fleureau, P. Guillotel, and Q. Huynh-Thu. Physiological-based affect event detector for entertainment video applications. *IEEE Trans. on Affective Computing*, 3(3):379–385, 2012. doi: 10.1109/t-affc.2012.2

[7] T. Hearn. Github repository: webcam-pulse-detector, 2013.

[8] iMotions. iMotions Biometric Research Platform 6.0, 2016.

[9] M. V. Kopeliovich and M. V. Petrushan. Optimal facial areas for webcam-based photoplethysmography. *Pattern Recognition and Image Analysis*, 26(1):150–154, 2016. doi: 10.1134/s1054661816010120

[10] R. W. Picard. *Affective Computing*. MIT Press, 2000.

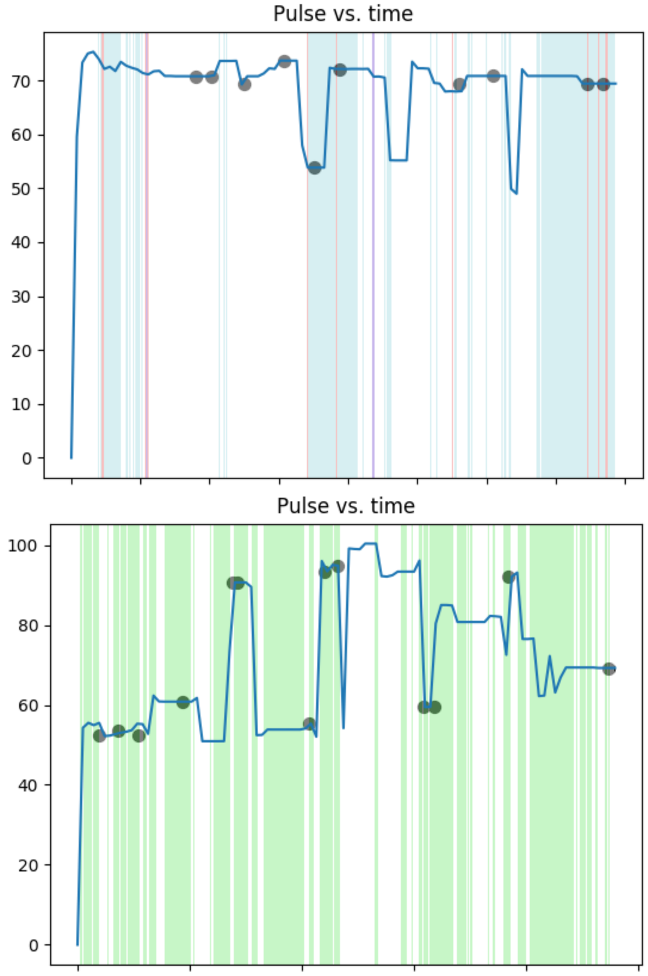


Fig. 12: Affective graphs from two different subjects watching video 1 (*Interstellar*) and video 5 (*Hot and Fluffy*), respectively. Pulse is y-axis and time is x-axis. The blue curve represents the pulse throughout the span of the video. The gray dots are click events and the background colors stand for the emotions computed at that point in time. Blue, the most dominant color in the top graph, stands for inferred sadness (expected emotion for this video) per face analysis. The green in the graph below stands for inferred happiness.

[11] M.-Z. Poh, D. J. Mcduff, and R. W. Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Transactions on Biomedical Engineering*, 58(1):7–11, 2011. doi: 10.1109/tbme.2010.2086456

[12] A. Rajenderan. An affective movie recommendation system. Master's thesis, Rochester Institute of Technology, 2014.

[13] A. Rajenderan. Github repository: webcam-pulse-detector, 2014.

[14] M. Soares and P. Viana. Tuning metadata for better movie content-based recommendation systems. *Multimedia Tools and Applications*, 74(17):7015–7036, Mar 2014. doi: 10.1007/s11042-014-1950-1

[15] M. Wollmer, F. Weninger, T. Knaup, B. Schuller, C. Sun, K. Sagae, and L.-P. Morency. YouTube movie reviews: Sentiment analysis in an audio-visual context. *IEEE Intelligent Systems*, 28(3):46–53, 2013. doi: 10.1109/mis.2013.34

[16] S. Zhao, H. Yao, X. Sun, P. Xu, X. Liu, and R. Ji. Video indexing and recommendation based on affective analysis of viewers. *Proceedings of the 19th ACM international conference on Multimedia - MM 11*, 2011. doi: 10.1145/2072298.2072043