The Effect of Personalized Music in Advertisement on Emotions and Memory Performance

Anna Horwath, Ernests Lavrinovics, Christian Neurohr,
Lilla Júlia Tóth, Marcus Alexander Dyrholm, Nasir Mohamoud

Abstract

The advertising market is large and competitive with companies attempting to develop campaigns that will increase their product sales as much as possible, thus requiring efficient and well thought-through advertisement content. It is known that content, that affects valence and arousal of a person, can improve their memory retention of an experience and music triggers emotional responses in people. Music is normally carefully selected to match the theme of the advertisement, which means that it is not interchangeable and provides the same experience for every viewer. The advertisement market does not produce material of where visual content is delivered to a user with overlaid music that is personalized to the individual instead of the advertisement. It is also not clear how such a combination would impact the changes on an individual's emotional state and the reception of the material. To explore the given context, an experiment was set up with a group of test subjects that are active users of Spotify streaming services. An experimental framework was developed that leverages test subject Spotify listening history with Spotify's provided songs audio features such as valence and arousal. The song features were mapped against the reported valence and arousal of each individual test subject at the time of conducting the test to determine the best matching song that was then overlaid on the advertisement and presented to the subjects. Preliminary results showed no statistical significance between test subjects experiencing the advertisement with and without music that was personalized to them. Flaws such as the low sample size of test subjects, potentially mismatching advertisement for the target audience and sub-par data gathering were noted. Based on this it cannot be conclusively said that combining an advertisement with music personalized to an individual has no effect on the experience.

1 Introduction

The goals pursued with a promotional video are varied and differ from company to company. However, many want their product to stand out, to be perceived positively by potential customers, and to be manifested in their minds. This should eventually lead to the purchase of the product. When asked how effective the promotion of a product is, previous scientific research has often failed to provide clear and unbiased results. This is partly due to the fact that emotional processes that take place in a person's brain during consumption have not been taken into account. But emotions have been shown to have a major impact on individuals' decisions and should therefore be considered when assessing the success of an advertisement [1, 2]. In addition, studies have shown a positive correlation between emotional events and memory, which is why the identification and interpretation of emotions in the marketing sector has also received more attention in recent years [1, 3, 4].

The effectiveness of emotional response has been proven for recollection of advertisements. If an advertisement has a higher effect on arousal and pleasure, viewers are, by a large margin, more susceptible to remembering brand names and products [2]. According to our knowledge, however, the effect of personalized music on the perception of product commercials has not yet been investigated, although it has already been proven that enjoyable as well as personalized music leads to increased activity in brain regions associated with emotion [5, 6]. Therefore, the question arises, whether personalized music causes a higher emotional reaction in the viewer compared to predetermined music and whether this is correlated to the perception of the promotional video and the product itself. Furthermore, it will be investigated whether the extent of emotion during consumption has an influence on short-term memory and recall of details.

2 Related Work

2.1 Advertising and Arousal-Valence

Behavioral sciences and the advertising industry share a common interest in the research topic of emotions. Yet emotions are challenging to measure although arguably the emotional spectrum stems from a mixture of *Valence* and *Arousal* [7]. It's crucial to investigate what other researchers have used for measuring emotions in advertising and how arousal-valence is involved in the advertisement as well as memory retention of the content.

Bolls and Lang [8] explored how valence and arousal has an effect on memory retention in radio advertisements. They tested the effect by using heart rate as a physiological measure of arousal. The participants listened to 24 60 second advertisements that had either a positive or negative valence. Bolls collected data on heart rate and memory retention of participants. The listeners' data on the heart rate indicates that negative information receives more attention than positive ones as well as when participants experienced higher arousal they would have better memory retention [8].

Hadinejad, A. [9] conducted an experiment where tourists from Iran and Australia were invited to a laboratory to watch a tourism advertisement. The data was collected by using a video camera to record the facial expressions of the tourists and using the software FaceReader™ to produce the data. The outcome of the experiment indicated that the tourists had overall low arousal and positive emotions while watching the video.

2.2 Memory, Valence and Arousal

In a study by Gomes et al. [4] the researchers investigated how valenced words, i.e. positive or negative words, influenced recollection. Several participants were asked to recall a list of words separated into three categories of valence: neutral, negative and positive. The study found that the participants had a significantly higher recollective ability for positive and negative valenced words compared to words with neutral valence. The same was found for when arousal and valence were measured in combination. With the combination of positive valence and arousal, participants had a much higher recollective ability than only with valence manipulation [4].

It is not only long term memory that is improved by arousal and non-neutral valence. Events of neutral valence can also be remembered more clearly if they are connected with some form of emotion and arousal [10]. This fact is important as it can be argued that advertisements fall within the category of a neutral event. However, long term memory is not the only one that should be targeted by advertisements. The visuospatial working memory is the component of memory that stores visual and spatial information in the short term [11]. Constanzi et al. investigated what effect arousal and valence had on visuospatial working memory [12]. In this experiment, participants had to relocate rectangles that were overlapped with either emotional or non-emotional pictures. They found that if the subject has a high arousal combined with neutral valence, the errors in the relocation task diminished significantly for both short and long term memory tests [12].

2.3 Facial Expression Analysis Techniques

People share universal emotions expressed through facial expressions, regardless of their origin or ethnicity. Ekman et al. [13] therefore divided emotions into six different states, including fear, joy, sadness, aggression, disgust, and astonishment. In advertising, facial expression analysis techniques can be used to detect these emotions and measure their strength in relation to the stimuli. Different approaches exist for measuring emotions in facial expressions, which are explained in the following sections.

Human Live Observation: According to studies, human performance in recognizing emotions in facial expressions of database images and videos generally ranges from 60% to 80% and usually does not exceed 90%. Compared to other emotional facial expressions, "happy" was most frequently rated as correct by human observers. It was substantially more difficult to recognise non-happy facial expressions and especially the recognition of fear [14].

Facial Electromyography (fEMG): In this technique, electrodes are placed on the surface of the skin to measure the electrical impulses of the facial muscles, which are amplified by the fEMG. A major drawback of fEMG systems is their limitation to use in an experimental context. Therefore, fEMG is not suitable for analyzing facial expressions of individuals in their natural and social environments [15].

Automatic Facial Expression Recognition (FER): The automatic recognition and evaluation of facial expressions in static images and videos with regard to the analysis of sentiments is made possible by Emotion Recognition or Affective Computing (AC) [16]. A typical FER system flow consists of three steps as shown in appendix A. To achieve an accurate result, the subject is often first separated from the background [17]. In the face detection stage, input images or sequences are used to discover a face region. Following face positioning, discriminatory information is extracted. Finally, facial expression recognition is performed.

The implementation of Deep Learning (DL) methods, such as Convolutional Neural Networks (CNN), has contributed significantly to improving facial expression recognition results. Existing studies on face recognition with CNN achieve high accuracy of over 90% [18] [19] [20]. A common approach is to solve the tasks of background removal and facial expression recognition in a single CNN network. Other approaches include splitting the tasks between two separate convolutional neural networks to reduce the complexity of the system [17]. Another form of CNN is the Multi-Task Convolutional Neural Network (MCTNN), which is a method for simultaneous face recognition and alignment based on a neural network with deep convolution. Compared to the traditional method, MTCNN has a better performance, and can localize faces more precisely [21]. It has been shown to outperform state-of-the-art methods on a number of benchmarks [22].

Popular software that uses Deep Learning include FaceReader and iMotions' AFFDEX and FACET modules. But despite the popularity of these systems, compared to human observers, only FACET seems to perform better in matching emotions to facial expressions [14].

3 Methods

To collect experimental data in which researchers can observe behavior in a controlled environment, two primary test designs are generally used. In within-subject design, subjects are exposed to two experimental conditions. For example, results can be obtained on how behavior changes after subjects are exposed to certain stimuli. However, it was decided against this procedure, as there could be disadvantages due to the research question and the circumstances. For example, subjects could perceive more details in two runs and this would falsify the results of the memory test. Therefore, the between-subjects design was chosen. Here, the test subjects are individually exposed to only one experimental condition. The participants were divided into two groups and exposed to two different scenarios. Further details on the procedure of the experiment are given in section 6.2. In this test setup, the commercial with and without personalized music is the independent variable. And the measured dependent variables are the scores of the short term memory test and the results of the emotion responses in the self-report and the facial emotion recognition technique [23]. These measurements are explained in detail in the following section.

4 Measures

4.1 Self-reported emotions

To capture the emotions, the Self-Assessment Manikin (SAM) method was used. This is, as the name suggests, a method in which the test participants can self-report their emotions. SAM is a reliable method of recording the emotional impressions of a person. It is a non-verbal, graphic depiction with the three major affective dimensions distinguished: Valence, Arousal and Dominance [24]. The test participants are presented with illustrations of five various emotional states of each emotional dimension. In Appendix D under question 3, the emotional levels are shown from unpleasant to pleasant. This is to represent the Valence dimension. In question 4 the figure represents Arousal. Here the figures range from calm(left) to very excited with eyes wide open(right). The manikin in question 5 represents dominance. The dominance dimension represents changes in control with changes in the size of the figure. The small figure on the left means that the person does not feel in control of the situation and the large figure on the right, represents having full control over the situation [25]. Participants can also select positions in between pictures to more accurately represent their emotional state, making the scale range from 1 to 9.

The figures shown are kept neutral, which eliminates the problems with non-verbal measures that are based on human photographs. Another positive feature of the SAM is that there is no language barrier. In addition, the scale can be answered/checked in less than 15 seconds, so there is little time between stimuli and measurement [24].

4.2 Emotion recognition through FER

In order to measure emotions directly via facial expressions of the test subjects, FER by Justin Shenk was used. It uses an emotion score to determine if a certain emotion is present at a certain point of time. Videos of the test participants while they were watching the product commercial are used as input. The face recognition itself is carried out by a MTCNN. The facial expressions are automatically analyzed and the output is a list with the 6 basic emotions as defined by Ekman and a score for each of these emotions, which can range from 0 to 1, where 0 indicates that the emotion is most likely not present, while 1 indicates that the corresponding emotion is most likely present. The list also includes a value for a neutral facial expression to cover this case as well [26] [13]. In addition, the facial expressions were observed by the test conductor.

4.3 Memory Retention

To test the effect of emotions on short-term memory and recall of specific information, 10 cued recall questions were asked about the content of the product commercial. To avoid biasing the results, open questions were asked. Closed questions that give possible answers may lead participants to select the correct answer by guessing without actually remembering the detail or event shown in the video.

5 Experimental Software

In order to test the hypothesis and to gather results, an experimental software prototype was produced that performs automated questionnaire analysis and determination of a song that matches the best to the test subject's reported mood.

The experimental software design takes the pre-test questionnaire of the control group as input in a CSV file format. Afterwards all questions regarding the mood are parsed and the data is linearly scaled from 0 to 1. Heavy reliance is made on Spotify's online API that allows to query the service for songs, expose the audio content as a downloadable link as well as provides coefficients of already pre-made analysis of emotion and artistic related features of the song [27]. The preview URL exposes downloadable audio content that is encoded as a 30 seconds long MP3 96kb/s.

The project primarily is interested in two features for each song that are described as coefficients from 0 to 1 - *valence* and *energy*. *Valence* describes how happy and cheerful a song is, the higher the coefficient - the happier the song[27]. Similarly with *energy*, it describes the activity and intensity levels of a song, higher coefficient denotes higher energy. These two features are used to define the circumplex model of emotion [7], while energy is not quite arousal it still serves as a surrogate. As suggested by Mauro et al. [28], Spotify open API's audio features alone may not be an ideal representation of a song's true emotional value. Despite this, as a first iteration of the framework, full reliance was made on Spotify's openly available estimations.

For each of the 50 songs that are provided by the test subject, a numerical distance is computed of the song's X_i *Valence* and *Energy* parameters against the subject's mood, the best match is selected by taking a song with the smallest numerical distance. Afterwards the audio track's RMS value is normalized to $-24dB_{FS}$ +/- 2dB to prevent loudness biases, and is then overlaid with the advertisement video to be presented to the test subjects.

6 Evaluation

6.1 Participants

There was a total number of 8 people (n=8) participating in the study. There were 4 people in the control group, and 4 people in the experimental group. The sample consisted of participants from a variety of backgrounds, with ages ranging from 14 to 29 (M=25, SD=5.976). 50% of participants were male and 50% were female. Participants were drawn from a pool consisting of the acquaintances of the research group by the non-probabilistic method of convenience sampling. The mean age of the experimental group was 21,75 (SD=7.365), while for the control group it was 28.25 (SD=0.957). Each group had the same distribution of sexes. Participants in the experimental group were all active users of Spotify services.

6.2 Procedure

The testing was completed remotely. Participants were contacted online through video chat. After establishing connection and ensuring good video and audio quality, participants were given a brief introduction to the study and the agenda of the testing. No details about the emotions and memory aspect of the study were revealed not to introduce bias to the behaviour of the participants. Participants were only told that the research is concerning advertisements and their effects. After the introduction participants were asked to fill out a consent form. The consent form can be seen in appendix B. After consent was given, the recording of the experiment was started.

Participants from the experimental group were asked to fill out an additional preliminary questionnaire. This questionnaire inquired about how positive and energetic they felt at the time of filling out the survey. Participants could rate each question on a scale of 0(not at all) to 10(extremely). The questionnaire also required participants to supply their top 50 songs played from Spotify of the last six months. They were instructed to use an online tool for acquiring the statistics [29]. This questionnaire can be seen in appendix C.

Then both the experimental and the control group received the same questionnaire. This main questionnaire consisted of two sections - pre-ad and post-ad. The pre-ad gathered demographic information and baseline data on emotional state of participants before watching the advertisement. Also during the pre-ad section, participants were asked to look into the camera and just have a normal resting facial expression for 10 seconds. This was done to gather baseline data for the facial expression recognition.

After the pre-ad questionnaire was complete, users were sent the advertisement video. Control group participants received the advertisement with the original audio, while the experimental group participants received advertisements with one of the 50 songs that matched the best to their self-reported happiness and energy scores. After watching the ads, the post-ad section of the questionnaire was administered. Here the knowledge and perception of the brand was probed, as well as the effect of the music on their perceptions. These questions were followed by the SAM to once again record emotional state post-advertisement and then 10 memory retention questions. The main questionnaire can be seen in Appendix D.

7 Results

All of the questions used in the evaluation of this project with the exception of the memory retention question yield ordinal type data. The resulting answers from the memory retention task are transformed into scores based on how many correct answers participants gave, therefore these results are interval level data. Due to the scales of measurements used and the small sample size of each group in the experiment, non-parametric methods will be used to evaluate the results.

None of the participants had prior knowledge of the brand. To begin with, the pre and post advertisement Self-Assessment Manikin scores were compared within each group, to see if there were any changes. Descriptive statistics for both groups can be seen under section 1 of Appendix E. The Wilcoxon signed-rank test was used to compare the repeated measure of the SAM. There was no statistically significant difference for the control group in Valence ($t=1.5$, $P=1.0$), Arousal ($t=1.5$, $P=1.0$) or Dominance ($t=0$, $P=0.15$) between the pre and post-advertisement results. There was also no statistically significant difference between the Valence ($t=0$, $P=0.317$), Arousal ($t=2$, $P=0.564$) and Dominance ($t=1.5$, $P=1$) of the experimental group's before and after advertisement measures.

Moving on, the differences in the SAM between the two independent groups was compared with the Mann-Whitney U test. Once again, no statistically significant difference was detected in the change of Valence ($t=10$, $P=0.62$), Arousal ($t=6.5$, $P=0.76$) or Dominance ($T=11$, $P=0.429$) between the two groups. As stated in the Methods section, additionally to the self-reported measure of the SAM, Facial Expression Recognition was used to gather behavioral data on the emotional state of participants. The recordings of participants watching the advertisement were analysed with the FER neural network. The raw data from the analysis was then formatted to return the top emotion per frame. Based on this data it was calculated how many emotions and for how long they were present, see Appendix F. No statistically significant differences were found in the within-group comparisons of baseline to reaction data carried out with the Wilcoxon signed-rank test. The same is true for the between-group comparisons of the control and experimental groups' reactions done with the Mann-Whitney U test. Descriptive statistics of the facial expression data can be found in section 6.1 of Appendix E. Results of the within-group comparisons are located in section 6.2 , while the between-group statistics are in 6.3 of Appendix E.

Results of the memory retention task were also compared via the Mann Whitney U test. Yet again, no statistically significant difference was found between the two groups' scores ($t=7.5$, $P=1$). Additional data was also gathered on the participants perception of the advertisement and the product, the music used in the commercial, the perceived impact of music on their perception of the ad and how likely they were to recommend the product to their friends. None of the additional data comparisons revealed a statistically significant difference. The results can be seen in section 4 of Appendix E.

8 Discussion

As the results of the experiment showed no significant differences in any of the measures, it is pertinent to discuss the possible causes of this. Firstly the product shown in the commercial was a predominantly targeted at men. Given the small sample size and that the distribution of participant sexes were equal, it could be argued that the female portion of the participants would have had a more neutral reaction to the advertisement. Combined with the male portion not being large enough could be the cause for the lack of detectable differences. The small sample size in general could also explain the lack of measured difference in emotion and recall. Therefore we recommend that for future work, the samples size should be significantly larger. Also for future work it could be a good idea to use many different types of advertisements tested on the audiences it was intended for.

Another cause of discontinuity between the related research and this implementation, in terms of arousal, could be caused by the less accurate measurements employed in this experiment. Using only behavioral data obtained via FER alone could be the cause of the lack of differences. Furthermore, the FER framework was not calibrated on a per-subject basis and qualitative analysis of results showed that misclassifications with *neutral* and *sad* emotions were frequent, see AppendixA. For future implementations and explorations within advertisements and personalised music, a multi-modal measuring technique for arousal should be employed for gathering psychophysiological data [30].

Additionally, concerning the cued recall portion of the evaluation, the sequences in the advertisement was quite rapid. The rapid succession could have overloaded the visuospatial working memory of the participant, causing the lack of commitment to long/short-term memory. Another measurement could also be included in the evaluation of any future work on this subject, such as free recall. This could give participants the ability to give a more detailed recollection of the content they had viewed, possibly resulting in a more accurate measurement of memory retention.

The lack of significant differences in the self-reported music rating, and perceived impact of music between the two groups could be because musical fit might be a more important factor for advertisements, or the methodology used for matching a song with subject's emotions is sub-par to accurately make a reliable mapping. Lastly, the accuracy of the neural network was lower than anticipated which created a lower consistency in the FER results. To rectify this issue, the last layers of the neural network should be retrained with the participants baseline as training data.

9 Conclusion

Based on the data gathered throughout this study and the evaluation thereof it can be concluded that personalised music did not affect memory retention of participants when compared to the control group. Furthermore, personalised music did not elicit statistically significant arousal in the viewers of the advertisement. Also no significant difference was detected in emotional response data obtained through facial expression machine learning analysis. That is not to say conclusively that a difference does not exist, however further studies are required with larger sample sizes and refined methods.

10 Acknowledgements

For our experiment, we were kindly provided with a product commercial produced by Steve Giralt and RITE Media Group. Therefore, we would like to take this opportunity to thank them for their support. We would also like to thank all test participants for their active participation in the experiment.

References

[1] F. Otamendi and Dolores Lucia Sutil Martín. The emotional effectiveness of advertisement. *Frontiers in Psychology*, 11:2088, 2020.

[2] Stephen J Newell, Kenneth V Henderson, and Bob T Wu. The effects of pleasure and arousal on recall of advertisements during the super bowl. *Psychology & Marketing*, 18(11):1135–1153, 2001.

[3] Linda Levine and David Pizarro. Emotion and memory research: A grumpy overview. *Social Cognition - SOC COGNITION*, 22:530–554, 10 2004.

[4] Carlos F A Gomes, Charles J Brainerd, and Lilian M Stein. Effects of emotional valence and arousal on recollective and nonrecollective recall. 2012.

[5] Hussain-Abdulah Arjmand, Jesper Hohagen, Bryan Paton, and Nikki S. Rickard. Emotional responses to music: Shifts in frontal brain asymmetry mark periods of musical change. *Frontiers in Psychology*, 8:2044, 2017.

[6] Joris H. Janssen, Egon L. van den Broek, and Joyce H. D. M. Westerink. Personalized affective music player. In *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*. IEEE, September 2009.

[7] James A Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980.

[8] Paul D. Bolls and Annie Lang. I saw it on the radio: The allocation of attention to high-imagery radio advertisements. *Media Psychology*, 5(1):33–55, February 2003.

[9] Noel Scott A nna Kralj Arghavan Hadinejad, Brent D. Moyle. Emotional responses to tourism advertisements: the application of facereader™. *article*, 44:131–135, 2019.

[10] A. K. Anderson, P. E. Wais, and J. D. E. Gabrieli. Emotion enhances remembrance of neutral events past. *Proceedings of the National Academy of Sciences*, 103(5):1599–1604, January 2006.

[11] Alan Baddeley. Working memory. *Science*, 255(5044):556–559, 1992.

[12] Marco Costanzi, Beatrice Cianfanelli, Daniele Saraulli, Stefano Lasaponara, Fabrizio Doricchi, Vincenzo Cestari, and Clelia Rossi-Arnaud. The effect of emotional valence and arousal on visuo-spatial working memory: incidental emotional learning and memory for object-location. *Frontiers in psychology*, 10:2587, 2019.

[13] Paul Ekman and W V Friesen. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17 2:124–9, 1971.

[14] Sabrina Stöckli, Michael Schulte-Mecklenbeck, Stefan Borer, and Andrea Samson. Facial expression analysis with affdex and facet: A validation study. *Behavior Research Methods*, 50, 12 2017.

[15] Karsten Wolf. Measuring facial expression of emotion. *Dialogues in Clinical Neuroscience*, 17:457–462, 12 2015.

[16] Martin Cooney, Sepideh Pashami, Anita Sant'Anna, Yuantao Fan, and Slawomir Nowaczyk. Pitfalls of affective computing: How can the automatic visual communication of emotions lead to harm, and what can be done to mitigate such risks. Republic and Canton of Geneva, CHE, 2018. International World Wide Web Conferences Steering Committee.

[17] N. Mehendale. Facial emotion recognition using convolutional neural networks (FERC). *SN Appl. Sci. 2*, 2020.

[18] Yucong Shen and Kuanyuan Chen. Face detection with deep learning. 2018.

[19] Xiangyun Zhao, Xiaodan Liang, Luoqi Liu, Teng Li, Yugang Han, Nuno Vasconcelos, and Shuicheng Yan. Peak-piloted deep network for facial expression recognition, 2017.

[20] Wenting Liu, Li Zhou, and Jie Chen. Face recognition based on lightweight convolutional neural networks. *Information*, 12(5), 2021.

[21] Hongchang Ku and Wei Dong. Face recognition based on mtcnn and convolutional neural network. *Frontiers in Signal Processing*, 4, 01 2020.

[22] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016.

[23] Janina AM Lehmann and Tina Seufert. The influence of background music on learning in the light of different theoretical perspectives and the role of working memory capacity. *Frontiers in psychology*, 8:1902, 2017.

[24] J. D. Morris. Observations: Sam: the self-assessment manikin; an efficient cross-cultural measurement of emotional response. *Journal of advertising research*, 6:63–68, 1995.

[25] Margaret M Bradley and Peter J Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994.

[26] Justin Shenk, Aaron CG, Octavio Arriaga, and Owlwasrowk. justinshenk/fer: Zenodo, 09 2021.

[27] Spotify AB. Documentation — spotify for developers, 2022.

[28] Renato Panda, Hugo Redinho, Carolina Gonçalves, Ricardo Malheiro, and Rui Pedro Paiva. How does the spotify api compare to the music emotion recognition state-of-the-art? 07 2021.

[29] Spotify listening stats: Your top artists, songs & genres - chosic. https://www.chosic.com/spotify-listening-stats/. (Accessed on 01/10/2022).

[30] Nicu Sebe, Ira Cohen, Theo Gevers, and Thomas S Huang. Multimodal approaches for emotion recognition: a survey. In *Internet Imaging VI*, volume 5670, pages 56–67. International Society for Optics and Photonics, 2005.

[31] Renuka Deshmukh, Shilpa Paygude, and Vandana Jagtap. Facial emotion recognition system through machine learning approach. 05 2017.

Appendices

Appendices

A Flow of an FER system

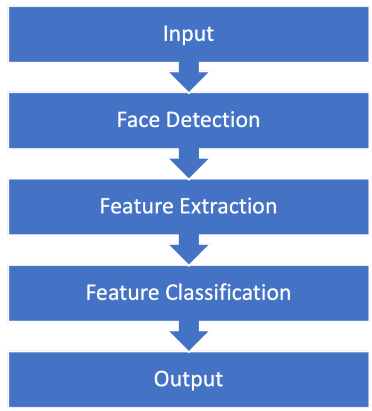


Figure 1: FER System Flow. In accordance with [31]

B Consent Form

Consent Form

We are a group of Medialogy Master students at the University of Aalborg conductiong a research on advertisements.

In this testing procedure, you will watch an advertisement and fill out some questions. The testing should last no more than 10 minutes.

The testing procedure will be video recorded.

Thank you for your time and participation in the test of this project.
Now, please fill out the details below.

*Required

1. I confirm that I have heard and understood the aforementioned information about the testing. I have had the opportunity to consider the information, ask questions, and have had these answered satisfactorily. *

   Tick all that apply.

   ☐ Yes

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason. *

   Tick all that apply.

   ☐ Yes

3. I understand that all personal information will be anonymized and treated confidentially. *

   Tick all that apply.

   ☐ Yes

4. I hereby consent to the video recording of the testing. *

   Tick all that apply.

   ☐ Yes

C Experimental Group Preliminary Questionnaire

Survey

*Required

1. What is your name? *

2. How positive do you feel today? *

Mark only one oval.

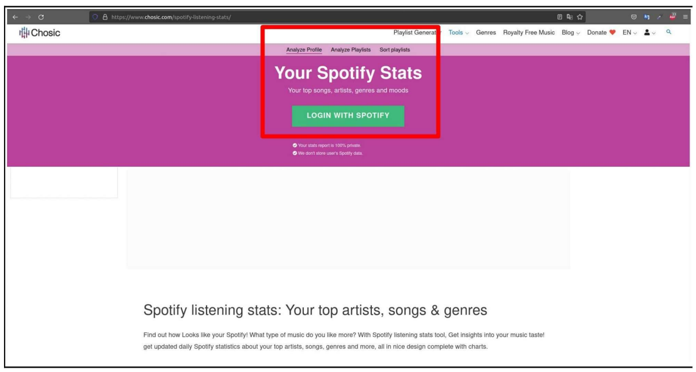
	0	1	2	3	4	5	6	7	8	9	10	
Not at all												Extremely

3. How energetic do you feel today? *

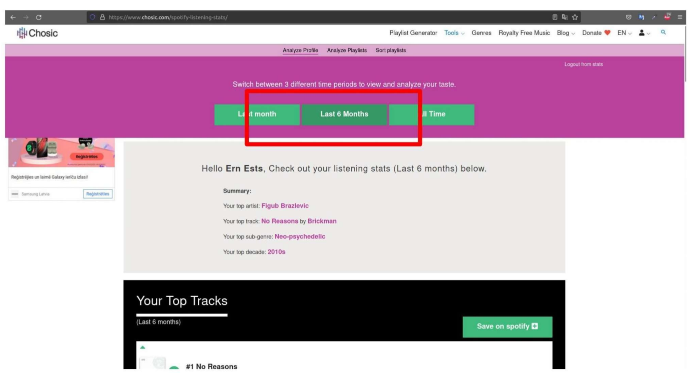
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all												Extremely

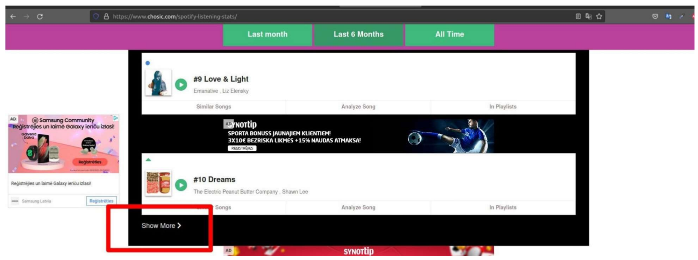
Please open the following website and log in with your spotify account: https://www.chosic.com/spotify-listening-stats/



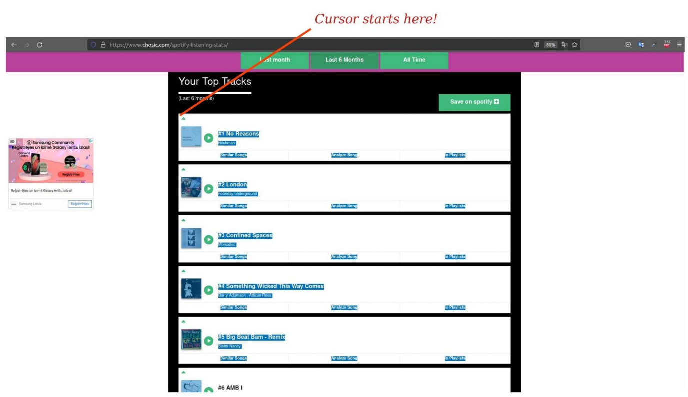
Then click on your last 6 months statistics



Scroll down end press "Show more"



Place your cursor on the top-left corner of the first, white box. Press and hold left mouse button and scrollall the way down to the last song to highlight all of yours songs.



Copy and paste the list of songs at the end of this questionnaire where is says "Paste tracklist here". If everything went correct, the text should be pasted and automatically formatted in a similar way as below.

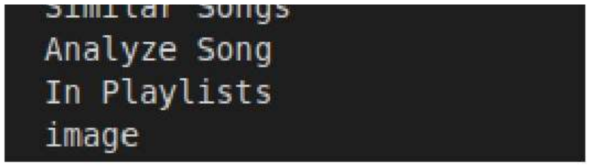
```
#1 No Reasons

Brickman
Similar Songs
Analyze Song
In Playlists
image
#2 London

noonday underground
Similar Songs
Analyze Song
In Playlists
image
#3 Shalala

Moses Gunn Collective
Similar Songs
Analyze Song
In Playlists
image
#4 From A Distance

STL
Similar Songs
```



4. Paste tracklist here *

This content is neither created nor endorsed by Google.

Google Forms

D Main Questionnaire

Questionnaire

Please fill out the information below.

*Required

1. Age *

2. Gender *

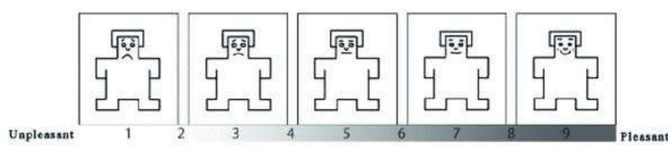
Mark only one oval.

- Male
- Female
- Prefer not to say

Part 1

Please rate how do you feel right now on the scales below.

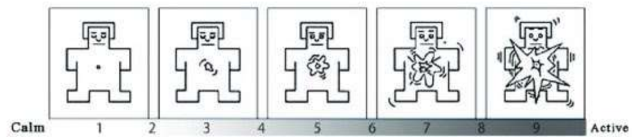
3. Please rate how do you feel. *



Mark only one oval.

1	2	3	4	5	6	7	8	9

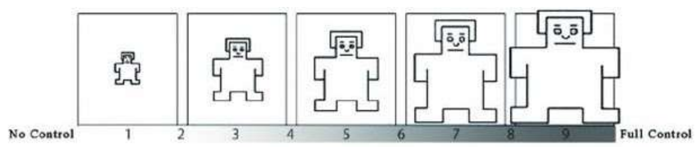
4. Please rate how do you feel. *



Mark only one oval.

1	2	3	4	5	6	7	8	9
◯	◯	◯	◯	◯	◯	◯	◯	◯

5. Please rate how do you feel. *



Mark only one oval.

1	2	3	4	5	6	7	8	9
◯	◯	◯	◯	◯	◯	◯	◯	◯

Part 1 over

Please let the test conductor know that you have done the first part of the quesionnaire before proceeding, so they can show you an advertisement.

Part 2

Please fill out the questions below after watching the advertisement.

6. Did you know the brand before watching this commercial? *

Mark only one oval.

- Yes
- No

7. How would you rate the advertisement in general? I think the commercial that I just watched is ... *

Mark only one oval.

	1	2	3	4	5	
Very bad						Excellent

8. The music in the commercial was ... *

Mark only one oval.

	1	2	3	4	5	
Very bad						Excellent

9. The choice of music has significantly improved my perception of the commercial. *

Mark only one oval.

	1	2	3	4	5	
Stongly disagree						Strongly agree

10. After seeing the commercial, I am convinced that it is a high quality product. *

Mark only one oval.

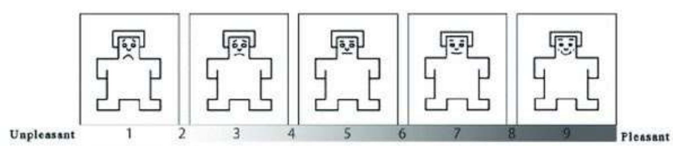
	1	2	3	4	5	
Stongly disagree	◯	◯	◯	◯	◯	Strongly agree

11. After seeing the commercial, I would recommend the product to a friend. *

Mark only one oval.

	1	2	3	4	5	
Stongly disagree	◯	◯	◯	◯	◯	Strongly agree

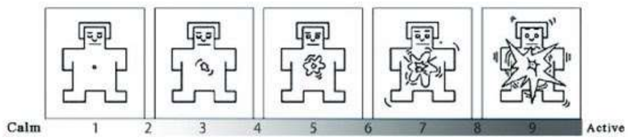
12. Please rate how do you feel. *



Mark only one oval.

1	2	3	4	5	6	7	8	9
◯	◯	◯	◯	◯	◯	◯	◯	◯

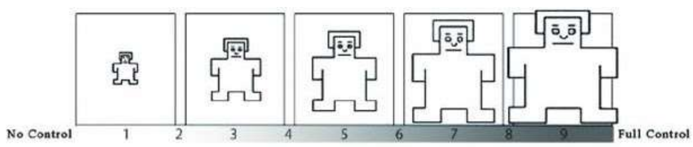
13. Please rate how do you feel. *



Mark only one oval.

1	2	3	4	5	6	7	8	9
◯	◯	◯	◯	◯	◯	◯	◯	◯

14. Please rate how do you feel. *



Mark only one oval.

1	2	3	4	5	6	7	8	9
◯	◯	◯	◯	◯	◯	◯	◯	◯

15. What was the very first object seen in the commercial? *

____

16. What kind of shoes were featured in the commercial? *

____

17. What color was the outside of the shoes? *

____

18. What color was the inside of the shoes? *

19. What color were the shoelaces? *

20. What colour was the packaging of the product? *

21. What color was the pencil in the commerical? *

22. What was the brand name of the product? *

23. How many scissors were in the commerical? *

24. How many hammers were in the commerical? *

Thank you for your participation!

E Results of the Statistical Evaluation

Untitled

January 10, 2022

[179]:
```
import pandas as pd
from scipy import stats
control=pd.read_csv('Control.csv')
pd.set_option("display.max_columns", None)
```

1 Descriptive Statistics

1.1 Control Group

[180]:
```
control.describe()
```

[180]:

	Age	Valence	Arousal	Dominance	Advertisement	Music
count	4.00000	4.000000	4.000000	4.000000	4.000000	4.000000
mean	21.75000	6.750000	4.250000	7.000000	3.750000	3.000000
std	7.36546	0.957427	2.061553	1.414214	0.957427	0.816497
min	14.00000	6.000000	2.000000	5.000000	3.000000	2.000000
25%	16.25000	6.000000	2.750000	6.500000	3.000000	2.750000
50%	22.00000	6.500000	4.500000	7.500000	3.500000	3.000000
75%	27.50000	7.250000	6.000000	8.000000	4.250000	3.250000
max	29.00000	8.000000	6.000000	8.000000	5.000000	4.000000

	Music-choice	Quality	Recommendation	Valence-post	Arousal-post
count	4.0	4.000000	4.000000	4.00	4.000000
mean	2.5	3.500000	2.000000	6.75	4.250000
std	1.0	1.290994	1.154701	1.50	2.629956
min	1.0	2.000000	1.000000	5.00	2.000000
25%	2.5	2.750000	1.000000	5.75	2.000000
50%	3.0	3.500000	2.000000	7.00	4.000000
75%	3.0	4.250000	3.000000	8.00	6.250000
max	3.0	5.000000	3.000000	8.00	7.000000

	Dominance-post	Memory
count	4.000000	4.000000
mean	6.500000	6.500000
std	1.914854	2.081666
min	4.000000	4.000000
25%	5.500000	5.500000

1

```
50%          7.000000  6.500000
75%          8.000000  7.500000
max          8.000000  9.000000
```

1.2 Experimental Group

[181]:
```python
experimental=pd.read_csv('Experimental.csv')
experimental.describe()
```

[181]:
```
             Age   Valence  Arousal  Dominance  Advertisement     Music  \
count   4.000000  4.000000     4.00        4.0            4.0  4.000000
mean   28.250000  6.750000     4.75        5.5            4.0  4.000000
std     0.957427  1.707825     0.50        1.0            0.0  0.816497
min    27.000000  5.000000     4.00        4.0            4.0  3.000000
25%    27.750000  5.750000     4.75        5.5            4.0  3.750000
50%    28.500000  6.500000     5.00        6.0            4.0  4.000000
75%    29.000000  7.500000     5.00        6.0            4.0  4.250000
max    29.000000  9.000000     5.00        6.0            4.0  5.000000

       Music-choice   Quality  Recommendation  Valence-post  Arousal-post  \
count      4.000000  4.000000        4.000000      4.000000      4.000000
mean       3.000000  4.000000        3.000000      7.250000      4.500000
std        1.825742  0.816497        0.816497      1.258306      1.290994
min        1.000000  3.000000        2.000000      6.000000      3.000000
25%        1.750000  3.750000        2.750000      6.750000      3.750000
50%        3.000000  4.000000        3.000000      7.000000      4.500000
75%        4.250000  4.250000        3.250000      7.500000      5.250000
max        5.000000  5.000000        4.000000      9.000000      6.000000

       Dominance-post     Memory
count        4.000000   4.000000
mean         5.500000   6.750000
std          1.290994   2.362908
min          4.000000   5.000000
25%          4.750000   5.000000
50%          5.500000   6.000000
75%          6.250000   7.750000
max          7.000000  10.000000
```

2

2 Within-group Emotional State Differences

2.1 Control Group

2.1.1 Pre and Post-Ad Valence Difference

[217]:
```
stat, p = stats.wilcoxon(control['Valence'],control['Valence-post'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=1.500, p=1.000
fail to reject H0
```

2.1.2 Pre and Post-Ad Arousal Difference

[183]:
```
stat, p = stats.wilcoxon(control['Arousal'],control['Arousal-post'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=1.500, p=1.000
fail to reject H0
```

2.1.3 Pre and Post-Ad Dominance Difference

[184]:
```
stat, p = stats.wilcoxon(control['Dominance'],control['Dominance-post'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=0.000, p=0.157
fail to reject H0
```

2.2 Experimental Group

2.2.1 Pre and Post-Ad Valence Difference

[185]:
```
stat, p = stats.wilcoxon(experimental['Valence'],experimental['Valence-post'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
```

```
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=0.000, p=0.317
fail to reject H0
```

2.2.2 Pre and Post-Ad Arousal Difference

[186]:
```
stat, p = stats.wilcoxon(experimental['Arousal'],experimental['Arousal-post'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=2.000, p=0.564
fail to reject H0
```

2.2.3 Pre and Post-Ad Dominance Difference

[187]:
```
stat, p = stats.
↪wilcoxon(experimental['Dominance'],experimental['Dominance-post'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=1.500, p=1.000
fail to reject H0
```

3 Between-group Emotional State Differences

3.1 Valence Difference

[188]:
```
CVD=control['Valence']-control['Valence-post']
EVD=experimental['Valence']-experimental['Valence-post']
stat, p = stats.mannwhitneyu(CVD,EVD)
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

4

```
Statistics=10.000, p=0.620
fail to reject H0
```

3.2 Arousal Difference

[189]:
```
CAD=control['Arousal']-control['Arousal-post']
EAD=experimental['Arousal']-experimental['Arousal-post']
stat, p = stats.mannwhitneyu(CAD,EAD)
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=6.500, p=0.760
fail to reject H0
```

3.3 Dominance Difference

[190]:
```
CDD=control['Dominance']-control['Dominance-post']
EDD=experimental['Dominance']-experimental['Dominance-post']
stat, p = stats.mannwhitneyu(CDD,EDD)
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=11.000, p=0.429
fail to reject H0
```

4 Between-group Comparisons of Additional Measures

4.1 Advertisement Rating Difference

[191]:
```
stat, p = stats.
 ↪mannwhitneyu(control['Advertisement'],experimental['Advertisement'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=6.000, p=0.617
fail to reject H0
```

5

4.2 Music Rating Difference

[192]:
```
stat, p = stats.mannwhitneyu(control['Music'],experimental['Music'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=3.000, p=0.172
fail to reject H0
```

4.3 Perceived Impact of Music Difference

[193]:
```
stat, p = stats.
 ↪mannwhitneyu(control['Music-choice'],experimental['Music-choice'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=6.500, p=0.766
fail to reject H0
```

4.4 Quality Perception Difference

[194]:
```
stat, p = stats.mannwhitneyu(control['Quality'],experimental['Quality'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=6.000, p=0.653
fail to reject H0
```

4.5 Likelihood of Recommendation Difference

[195]:
```
stat, p = stats.
 ↪mannwhitneyu(control['Recommendation'],experimental['Recommendation'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
```

```
    print('reject h0')
```

```
Statistics=4.000, p=0.278
fail to reject H0
```

5 Between-group Comparison of Memory Retention Scores

[196]:
```
stat, p = stats.mannwhitneyu(control['Memory'],experimental['Memory'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=7.500, p=1.000
fail to reject H0
```

6 Facial Expression differences

6.1 Descriptive Statistics

6.1.1 Control Group Baseline

[197]:
```
CR=pd.read_csv('control-emotion.csv')
CB=pd.read_csv('control-baseline.csv')

ER=pd.read_csv('experimental-emotions.csv')
EB=pd.read_csv('experimental-baseline.csv')

CB.describe()
```

[197]:

	Neutral	Happy	Angry	Fear	Sad
count	4.000000	4.0	4.000000	4.0	4.000000
mean	9.940048	0.0	0.108333	0.0	0.008333
std	0.288977	0.0	0.216667	0.0	0.016667
min	9.533333	0.0	0.000000	0.0	0.000000
25%	9.883333	0.0	0.000000	0.0	0.000000
50%	10.005005	0.0	0.000000	0.0	0.000000
75%	10.061719	0.0	0.108333	0.0	0.008333
max	10.216847	0.0	0.433333	0.0	0.033333

6.1.2 Control Group Reaction

[198]:
```
CR.describe()
```

[198]:

	Neutral	Happy	Angry	Fear	Sad
count	4.000000	4.00	4.00	4.000000	4.000000
mean	23.188690	0.61	0.01	0.020425	5.942960
std	11.107904	1.22	0.02	0.023595	11.560859
min	6.716667	0.00	0.00	0.000000	0.000000
25%	21.479167	0.00	0.00	0.000000	0.156380
50%	27.837198	0.00	0.00	0.020000	0.244254
75%	29.546721	0.61	0.01	0.040425	6.030833
max	30.363697	2.44	0.04	0.041701	23.283333

6.1.3 Experimental Group Baseline

[199]:
```
EB.describe()
```

[199]:

	Neutral	Happy	Angry	Fear	Sad
count	4.000000	4.0	4.0	4.0	4.000000
mean	7.166667	0.0	0.0	0.0	2.658333
std	3.669998	0.0	0.0	0.0	3.819286
min	1.900000	0.0	0.0	0.0	0.000000
25%	6.075000	0.0	0.0	0.0	0.000000
50%	8.383333	0.0	0.0	0.0	1.266667
75%	9.475000	0.0	0.0	0.0	3.925000
max	10.000000	0.0	0.0	0.0	8.100000

6.1.4 Experimental Group Reaction

[200]:
```
ER.describe()
```

[200]:

	Neutral	Happy	Angry	Fear	Sad
count	4.000000	4.000	4.000000	4.000000	4.000000
mean	20.962500	2.775	0.004167	0.208333	6.075000
std	8.615558	5.550	0.008333	0.416667	8.563807
min	11.233333	0.000	0.000000	0.000000	0.000000
25%	15.183333	0.000	0.000000	0.000000	1.875000
50%	21.316667	0.000	0.000000	0.000000	2.766667
75%	27.095833	2.775	0.004167	0.208333	6.966667
max	29.983333	11.100	0.016667	0.833333	18.766667

6.2 Within-Group Facial Expression Differences

6.2.1 Control Group

6.2.2 Neutral

[201]:
```
stat, p = stats.wilcoxon(CB['Neutral'],CR['Neutral'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
```

```
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=1.000, p=0.250
fail to reject H0
```

6.2.3 Happy

[216]:
```
stat, p = stats.wilcoxon(CB['Happy'],CR['Happy'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=0.000, p=0.317
fail to reject H0
```

6.2.4 Angry

[203]:
```
stat, p = stats.wilcoxon(CB['Angry'],CR['Angry'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=1.000, p=0.655
fail to reject H0
```

6.2.5 Fear

[204]:
```
stat, p = stats.wilcoxon(CB['Fear'],CR['Fear'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=0.000, p=0.180
fail to reject H0
```

6.2.6 Sad

[205]:
```
stat, p = stats.wilcoxon(CB['Sad'],CR['Sad'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=0.000, p=0.109
fail to reject H0
```

6.2.7 Experimental Group

6.2.8 Neutral

[206]:
```
stat, p = stats.wilcoxon(EB['Neutral'],ER['Neutral'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=0.000, p=0.125
fail to reject H0
```

6.2.9 Happy

[207]:
```
stat, p = stats.wilcoxon(EB['Happy'],ER['Happy'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=0.000, p=0.317
fail to reject H0
```

6.2.10 Angry

[208]:
```
stat, p = stats.wilcoxon(EB['Angry'],ER['Angry'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=0.000, p=0.317
fail to reject H0
```

6.2.11 Fear

[209]:
```
stat, p = stats.wilcoxon(EB['Fear'],ER['Fear'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=0.000, p=0.317
fail to reject H0
```

6.2.12 Sad

[210]:
```
stat, p = stats.wilcoxon(EB['Sad'],ER['Sad'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=0.000, p=0.109
fail to reject H0
```

6.3 Between-Group Facial Expression Differences

6.3.1 Neutral

[211]:
```
stat, p = stats.mannwhitneyu(CR['Neutral'],ER['Neutral'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=10.000, p=0.686
fail to reject H0
```

6.3.2 Happy

[212]:
```
stat, p = stats.mannwhitneyu(CR['Happy'],ER['Happy'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=7.500, p=1.000
fail to reject H0
```

6.3.3 Angry

[213]:
```
stat, p = stats.mannwhitneyu(CR['Angry'],ER['Angry'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=8.500, p=1.000
fail to reject H0
```

6.3.4 Fear

[214]:
```
stat, p = stats.mannwhitneyu(CR['Fear'],ER['Fear'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
else:
    print('reject h0')
```

```
Statistics=9.000, p=0.869
fail to reject H0
```

6.3.5 Sad

[215]:
```
stat, p = stats.mannwhitneyu(CR['Sad'],ER['Sad'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
alpha = 0.05
if p > alpha:
    print('fail to reject H0')
```

```
else:
    print('reject h0')
```

```
Statistics=6.500, p=0.772
fail to reject H0
```

F Results of Emotion Detection

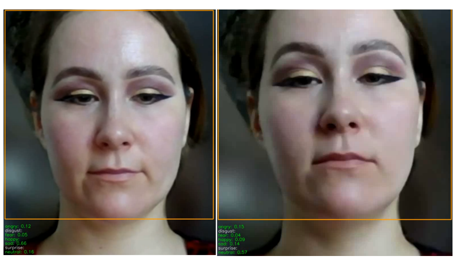
Table 1: Overview of Number of Frames with Detected Emotions During Baseline and Presentation of the Advertisement

Participant / FPS	Neutral	Happy	Angry	Fear	Sad
1 Baseline (29.97)	300				
1 Reaction (29.97)	910				
2 Baseline (25)	250				
2 Reaction (25)	660	61	1	1	7
3 Baseline (60)	572		26		2
3 Reaction (60)	403				1397
4 Baseline (23.98)	245				
4 Reaction (23.98)	702			1	5
5 Baseline (60)	600				
5 Reaction (60)	1799		1		
6 Baseline (10)	93				
6 Reaction (10)	165	111			25
7 Baseline (30)	224				76
7 Reaction (30)	784			25	91
8 Baseline (30)	57				243
8 Reaction (30)	337				563

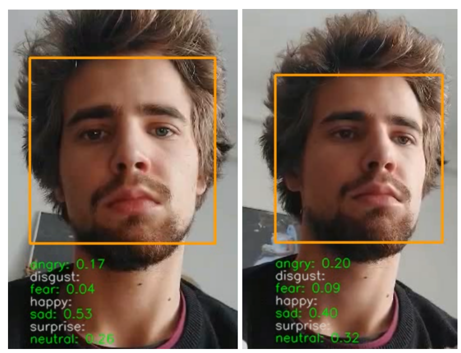
Test subjects 1-4 are from the control group, 5-8 are from the experimental group. The table depicts the number of frames for which a given emotion had the highest prediction coefficient. Varied number of FPS was dependant on test subjects webcam hardware.

G Qualitative Overview of the Emotion Estimation Framework

Figure 2



(a) Subject 3. Baseline Neutral (Left) - Reaction Sad (Right)



(b) Subject 8. Baseline Sad (Left) - Reaction Sad (Right)